Econometric Theory, **25**, 2009, 1143–1179. Printed in the United States of America.
doi:10.1017/S0266466608090464

BIAS REDUCTION AND LIKELIHOOD-BASED ALMOST EXACTLY SIZED HYPOTHESIS TESTING IN PREDICTIVE REGRESSIONS USING THE RESTRICTED LIKELIHOOD

WILLA W. CHEN
Texas A & M University

ROHIT S. DEO
New York University

Difficulties with inference in predictive regressions are generally attributed to strong persistence in the predictor series. We show that the major source of the problem is actually the nuisance intercept parameter, and we propose basing inference on the restricted likelihood, which is free of such nuisance location parameters and also possesses small curvature, making it suitable for inference. The bias of the restricted maximum likelihood (REML) estimates is shown to be approximately 50% less than that of the ordinary least squares (OLS) estimates near the unit root, without loss of efficiency. The error in the chi-square approximation to the distribution of the REML-based likelihood ratio test (RLRT) for no predictability is shown to be $(\frac{3}{4}-\rho^2)n^{-1}(G_3(\cdot)-G_1(\cdot))+O(n^{-2})$, where $|\rho|<1$ is the correlation of the innovation series and $G_s(\cdot)$ is the cumulative distribution function (c.d.f.) of a χ^2_s random variable. This very small error, free of the autoregressive (AR) parameter, suggests that the RLRT for predictability has very good size properties even when the regressor has strong persistence. The Bartlett-corrected RLRT achieves an $O(n^{-2})$ error. Power under local alternatives is obtained, and extensions to more general univariate regressors and vector AR(1) regressors, where OLS may no longer be asymptotically efficient, are provided. In simulations the RLRT maintains size well, is robust to nonnormal errors, and has uniformly higher power than the Jansson and Moreira (2006, *Econometrica* 74, 681–714) test with gains that can be substantial. The Campbell and Yogo (2006, *Journal of Financial Econometrics* 81, 27–60) Bonferroni Q test is found to have size distortions and can be significantly oversized.

Part of this research was completed when Deo was at the University of Texas–Austin. We thank four referees and the editor for comments that led to a significantly improved paper. We also thank Michael Jansson and Marcelo Moreira for providing their code and Yi Wang for his assistance with programming. Chen's research was supported by NSF grant DMS-0605132. Address correspondence to Rohit S. Deo, 8-57 KMC, New York University, West 4th Street New York, NY 10012, USA; e-mail: rdeo@stern.nyu.edu.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

1. INTRODUCTION

A question of interest in financial econometrics is whether future values of one series $\{Y_t\}$ can be predicted from lagged values of another series $\{X_t\}$. The hypothesis of no predictability is commonly tested under the assumption that the two series $\{Y_t\}_{t=1}^n$ and $\{X_t\}_{t=0}^n$ obey the following model:

$$Y_t = \eta + \beta X_{t-1} + u_t, \tag{1}$$

$$X_t = \mu + \alpha X_{t-1} + v_t, \tag{2}$$

where $|\alpha| < 1$, $u_t = \phi v_t + e_t$, $(e_t, v_t) \sim N\left(\mathbf{0}, \text{diag}(\sigma_e^2, \sigma_v^2)\right)$ are an independent and identically distributed (i.i.d.) series and $X_0 \sim N\left(\mu(1-\alpha)^{-1}, \sigma_v^2(1-\alpha^2)^{-1}\right)$. Interest generally centers on the case where the regressor series $\{X_t\}$ possesses a strong degree of autocorrelation with the autoregressive (AR) parameter α lying close to the unit root. It is well known (Stambaugh, 1999) that the standard ordinary least squares (OLS) estimate of β is biased when the errors $\{u_t, v_t\}$ are contemporaneously correlated, with the amount of bias increasing as α gets closer to unity. This bias results in the corresponding t-statistic being biased with poor size properties. However, Stambaugh (1999) provided a simple estimable expression for the bias in the OLS estimate of β that allows the researcher to compute a bias-corrected OLS estimate and also a t-statistic based on it. See, for example, Amihud and Hurvich (2004). There is currently, however, no known theoretical justification that such a t-statistic based on the bias-corrected OLS estimate will have improved size properties relative to the test based on the uncorrected OLS estimate. Indeed, Sprott and Viveros-Aguilera (1984) and Sprott (1990) point out that if inference is the goal of the researcher, computing pivotal t-statistics from bias-corrected point estimates is not necessarily guaranteed to improve finite-sample performance. Sprott and Viveros-Aguilera (1984, Sect. 6) provide an example where even the use of the exact bias correction does not result in accurate finite-sample inference because the resulting t-statistic can still be very far from normal. (One could perhaps use conservative bias corrections, such as in Lewellen, 2004, which can yield tests that are undersized but at the cost of significant power loss. See Lewellen, 2004, p. 232.) Instead, following Fisher (1973), Sprott argued in a series of fundamental papers (1973, 1975, 1980, 1990) that from an inferential point of view, issues such as the bias of point estimates may be irrelevant in small samples, and he stressed the importance of examining the likelihood.

At first glance, it may seem surprising that a likelihood ratio test (LRT) may provide well-behaved hypothesis tests in situations when the t-statistic does not because the two test statistics are closely related. More specifically, under standard regularity conditions we can expand the LRT, $LRT(\hat{\theta}, \theta)$, for a parameter θ in terms of its t-statistic, $t(\hat{\theta}, \theta)$, as

$$LRT(\hat{\theta}, \theta) = t^2(\hat{\theta}, \theta) + R_n(\hat{\theta}),$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

where the remainder term $R_n(\hat{\theta})$ converges to zero in probability. This expansion suggests that t-statistics with poor finite-sample size properties would correspond to LRTs that are also poorly behaved. However, this intuition is misleading because it ignores a crucial property of a likelihood that is not shared by the t-statistic, namely, invariance under 1-1 parameter transformations $g(\theta)$. As we argue next based on some of the key ideas in Sprott (1975, 1980), this property can prove invaluable for the LRT. Because the LRT for θ is identical to the LRT for $g(\theta)$ because of invariance, we can also expand the LRT in terms of the t-statistic for $g(\theta)$, yielding

$$LRT(\hat{\theta}, \theta) = LRT(g(\hat{\theta}), g(\theta)) = t^2(g(\hat{\theta}), g(\theta)) + R_n(g(\hat{\theta})). \quad \textbf{(3)}$$

Sprott argued that

1. there may exist some transformation $g(\theta)$ such that the remainder term $R_n(g(\hat{\theta}))$ in the preceding expansion is close to zero, thus making the likelihood ratio approximately quadratic in the t-statistic based on that transformed parameter $g(\theta)$;
2. the conditions that allow for this quadratic approximation to be adequate also improve the normal approximation to the distribution of the t-statistic $t(g(\hat{\theta}), g(\theta))$ in the transformed parameter.

As a result, it follows from these two observations and the expansion in (3) that the likelihood ratio will be well approximated by a chi-square variable in finite samples as long as some parametrization satisfying observation 1 exists, *even if one does not know what that parametrization is*. Sprott (1973, 1975, 1980) showed that such a parametrization would exist if the "curvature" of the log-likelihood,[1] as measured by a function of its higher order derivatives, was small. The use of such a likelihood would then result in a well-behaved LRT in finite samples. See also Efron (1975) and McCullagh and Cox (1986) for a geometrical approach to curvature and likelihood ratio based–hypothesis testing.

The approach we take in this paper is guided by the intuition given earlier. We find that in a univariate AR process of order one, the likelihood has very small curvature and hence yields tests with good finite-sample behavior when there is no intercept in the model. However, the inclusion of an intercept in the model causes the likelihood ratio to lose this property, thus pointing to the intercept as the source of the problem. This motivates the use of the restricted likelihood, which is free of the nuisance intercept parameter and hence able to imitate the likelihood of the no-intercept univariate model with its attendant small curvature. This suggests that the restricted likelihood will also be useful in the related predictive regression problem, which is a bivariate AR(1) with intercept. Indeed, we are able to obtain theoretical results that demonstrate that the LRT based on the restricted likelihood (RLRT) has good finite-sample performance for both estimation and inference in this context. A curvature-related approach to tackling the predictive regression problem for the model in (1) and (2) was also taken by Jansson and Moreira

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

(2006). We compare our results in more detail with those in Jansson and Moreira (2006), first on a theoretical basis in Section 3 and then through simulations in Section 5.

The layout of the paper is as follows. In Section 2, we provide motivation for considering the restricted likelihood by studying the related problem of inference in the univariate AR(1) model. We then obtain the restricted likelihood for the bivariate predictive regression model and provide results on the bias of the restricted maximum likelihood (REML) estimates. In Section 3, we state our result on the finite-sample behavior of the RLRT for β and compare its power under a sequence of Pitman alternatives to that of the restricted likelihood–based Wald and Rao score test. A comparison with the results in Jansson and Moreira (2006) is also provided. Section 4 provides results on extensions of the REML method to higher order AR processes for the regressor series, and also to multivariate AR(1) regressors. The finite-sample performance of the REML estimates and the RLRT is studied through simulations in Section 5 and is compared to the performance of the procedures developed by Jansson and Moreira (2006) and Campbell and Yogo (2006). All proofs are relegated to the Appendix at the end of the paper.

2. RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION

To understand why the restricted likelihood may yield well behaved LRT's in the predictive regression context, it is instructive to consider LRT's for α in the univariate AR(1) model given in (2). If $LRT_{\alpha,\mu}$ denotes the LRT for testing $H_0: \alpha = \alpha_0$ versus $H_1: \alpha \neq \alpha_0$ in that model for $|\alpha_0| < 1$, then Theorem 2 of van Giersbergen (2006) in conjunction[2] with the results of Hayakawa (1977, 1987), Cordeiro (1987), Barndorff-Nielsen and Hall (1988), and Chesher and Smith (1995) yields the following formal expansion:[3]

$$P(LRT_{\alpha,\mu} \leq x) - G_1(x) = 0.25(1+7\alpha_0)(1-\alpha_0)^{-1}n^{-1}(G_3(x)-G_1(x)) + O(n^{-2}), \tag{4}$$

where $G_s(x)$ is the cumulative distribution function (c.d.f.) of a χ^2_s random variable. On the other hand, if LRT_α denotes the LRT for testing $H_0: \alpha = \alpha_0$ versus $H_1: \alpha \neq \alpha_0$ in the univariate AR(1) model given in (2) *with* μ *known to be zero,* then it follows from Theorem 1 of van Giersbergen (2006) that we get the formal expansion

$$P(LRT_\alpha \leq x) - G_1(x) = -0.25n^{-1}(G_3(x) - G_1(x)) + O(n^{-2}). \tag{5}$$

It is obvious from (4) that $LRT_{\alpha,\mu}$ is very unstable when the AR parameter α is close to the unit root with the leading error term in the expansion going to infinity. In stark contrast, we see from (5) that LRT_α is very well behaved with the leading term in the expansion of its distribution being both very small and free of α. Figure 1 shows the empirical densities of LRT_α and $LRT_{\alpha,\mu}$ based on samples of

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

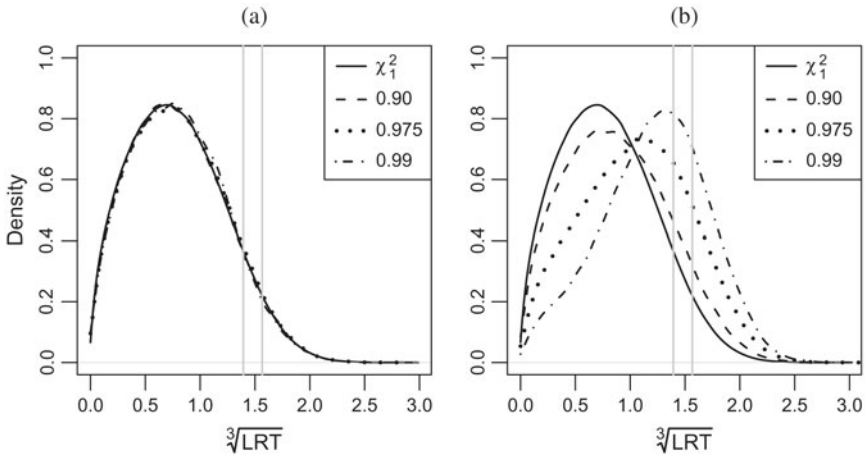


FIGURE 1. Empirical densities of cubic root transformed LRT statistics of AR(1) processes (a) without intercept and (b) with intercept. The vertical lines are 90th and 95th percentiles. Both plots are based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = 0.9, 0.975,$ and 0.99.

size $n = 100$ for various values of $\alpha = 0.9, 0.975$, and 0.99 plotted together with the limiting χ^2_1 density. (Because a χ^2_1 density is very right skewed, we plot the density of the cube root of the LRT to ensure a density that looks more symmetric to make the comparisons in the right tail clearer.) The empirical densities of LRT_α (the zero intercept case) are seen to be remarkably well approximated by the limiting χ^2_1 distribution, both when α is far from the unit root and when α is close to unity, whereas those of $LRT_{\alpha,\mu}$ (the intercept case) are far moved from that of the χ^2_1. Simulation results in Figure 1 of van Giersbergen (2006) further confirm the accuracy of the standard χ^2_1 approximation for LRT_α, even when α is close to unity. Not surprisingly, van Garderen (1999) has found that for the univariate AR(1) model in (2) *with μ known to be zero,* the Efron (1975) curvature (one of the standard measures of curvature of the likelihood) is very small,[4] being of the order $O(n^{-2})$ and converging to zero as $\alpha \to 1$ for every fixed n.

These results indicate that the culprit in the finite-sample failure of the LRT in the univariate AR(1) model is the unknown intercept μ. Because the bivariate prediction model in (1) and (2) is a vector AR(1) (with the first column of the coefficient matrix restricted to zero) with an intercept, one is led to suspect from the preceding discussion that the LRT for β will also be poorly behaved in finite samples because of the nuisance intercept vector. This is indeed the case, as seen in the simulation study presented in Table 1. Even when the AR(1) coefficient is 0.9, and thus far from the unit root, the usual LRT for β has inflated size, and the problem is exacerbated significantly as either the correlation between the innovations (u_t, v_t) or the AR coefficient increases. Thus, there is no advantage in

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

TABLE 1. Rejection rates of LRT for predictive regressions with 5% nominal rate: $\beta = 0$, $\sigma_v^2 = \sigma_u^2 = 1$

	ϕ	−0.5				−0.98			
α	n	50	100	200	400	50	100	200	400
0.90		0.0680	0.0620	0.0532	0.0528	0.1127	0.0838	0.0690	0.0604
0.95		0.0784	0.0696	0.0580	0.0546	0.1570	0.1149	0.0843	0.0696
0.99		0.0904	0.0882	0.0791	0.0639	0.3114	0.2392	0.1748	0.1260
0.995		0.0946	0.0936	0.0897	0.0713	0.3629	0.3095	0.2370	0.1700

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

using the usual LRT instead of the t-statistic in this case. However, the preceding discussion leads us to believe that the LRT for β might perhaps be well behaved if the intercept vector (η, μ) were known. Because the assumption that (η, μ) is known is extremely unrealistic, we are prompted to seek a likelihood that does not involve the location parameters and yet possesses small curvature properties similar to those of the model with known location parameters. The restricted likelihood turns out to be the one that has such properties, and we turn next to defining it and stating some of its properties.

The idea of restricted likelihood was originally proposed by Kalbfleisch and Sprott (1970) precisely as a means of eliminating the effect of nuisance location (more generally, regression coefficient) parameters when estimating the parameters of the model covariance structure in a linear model. More specifically, assume that we observe data on the vector $\mathbf{Z}$ that follows the linear model

$$\mathbf{Z} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{6}$$

where $\mathbf{W}$ and $\boldsymbol{\theta}$ are the design matrix and coefficient parameter vector, respectively, and the error vector $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\psi}))$ with $\boldsymbol{\psi}$ being the parameters that describe the variance-covariance matrix. Suppose that interest centers on the parameters $\boldsymbol{\psi}$ of the error covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\psi})$ and that the regression coefficients $\boldsymbol{\theta}$ are nuisance parameters. Kalbfleisch and Sprott (1970) defined the restricted likelihood to be the exact likelihood of the linearly transformed data $\mathbf{TZ}$, where $\mathbf{T}$ is any matrix of full row rank equal to $n - \text{rank}(\mathbf{W})$ such that $\mathbf{TW} = \mathbf{0}$. Thus, the likelihood of the transformation $\mathbf{TZ}$ does not depend on the nuisance regression coefficient parameters $\boldsymbol{\theta}$. The particular choice of the matrix $\mathbf{T}$ is irrelevant because the likelihood of $\mathbf{TZ}$ will change only by a multiplicative constant for different choices of $\mathbf{T}$ (Harville, 1974) and hence will have no effect on either estimation or testing of hypotheses. Harville (1974) showed that the restricted likelihood for the process (6), up to a multiplicative constant, is

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

given by

$$RL(\mathbf{Z}, \boldsymbol{\psi}) = |\mathbf{W}'\mathbf{W}|^{1/2}|\boldsymbol{\Sigma}(\boldsymbol{\psi})|^{-1/2}|\mathbf{W}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{W}|^{-1/2} \times \exp\left(-\frac{1}{2}\widetilde{\mathbf{Z}}(\boldsymbol{\psi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\widetilde{\mathbf{Z}}(\boldsymbol{\psi})\right), \tag{7}$$

where $\widetilde{\mathbf{Z}}(\boldsymbol{\psi}) = \mathbf{Z} - \mathbf{W}(\mathbf{W}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{Z}$. Harville (1977) showed that REML estimates of the parameters of the covariance structure do not suffer any loss of efficiency due to the linear transformation of the data. Harville (1974) also provided a Bayesian interpretation of the restricted likelihood, and Smyth and Verbyla (1996) showed that the restricted likelihood is also the exact conditional likelihood of the original data given the complete sufficient statistic for the regression coefficient parameters. Though the restricted likelihood has been studied primarily in the context of variance component models, there has also been some work on it in the context of time series models. See, for example, Tunnicliffe Wilson (1989) and Rahman and King (1997), among others. Francke and de Vos (2006) studied unit root tests for AR(1) models based on the restricted likelihood, and Chen and Deo (2006, 2007) showed that confidence intervals for the sum of the AR coefficients of univariate AR(p) processes based on the restricted likelihood have good coverage properties, even when the series is close to a unit root process. The REML estimates also are less biased than regular maximum likelihood estimates in nearly integrated univariate AR models with intercept (Cheang and Reinsel, 2000) and with trend (Kang, Shin, and Lee, 2003).

The restricted likelihood can also be interpreted as the exact likelihood of the maximal invariant (Lehmann and Romano, 2005, Sect. 6.2) that is invariant under transformations of the form $\mathbf{Z} \to \mathbf{Z} + \mathbf{W}\boldsymbol{\kappa}$, where $\boldsymbol{\kappa}$ is some vector. There are references in the econometrics literature that use the restricted likelihood in some version of the form given in (7) while explicitly stating that it is also the likelihood of the maximal invariant (see, e.g., Rahman and King, 1997; Francke and de Vos, 2006). The paper by Jansson and Moreira (2006) also uses the restricted likelihood for the model in (1) and (2), though the authors refer to it only as the likelihood of the maximal invariant. Because Jansson and Moreira (2006) make the simplifying assumption that $\mu = 0$ in (2), they are able to write the restricted likelihood of the process, which becomes the exact likelihood of $(y_t - y_{t-1}, x_t)$ without having to exploit the form in (7). There is another strand of the literature in econometrics (e.g., Dufour and King, 1991) that considers tests that are invariant under transformations of the form $\mathbf{Z} \to \mathbf{Z} + \mathbf{W}\boldsymbol{\kappa}$, but this literature focuses on point optimal tests and hence does not derive the likelihood of the maximal invariant of the form in (7). We now exploit the expression in (7) to obtain the restricted likelihood for an AR(1) with intercept, which helps develop some intuition for why the restricted likelihood can be of use in the bivariate predictive regression model.

When $\mathbf{X} = (X_0, \ldots, X_n)'$ follows the univariate AR(1) model in (2), we can express the vector $\mathbf{X}$ in the form (6), where now the design matrix $\mathbf{W}$ is a vector

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

of ones, the regression coefficient parameter $\boldsymbol{\theta}$ is given by $\mu(1-\alpha)$, and the error vector $\boldsymbol{\varepsilon}$ is a vector following a zero-mean AR(1) process. As a result, the restricted likelihood for this model is merely the exact likelihood of the first differences $\{X_t - X_{t-1}\}_{t=1}^{n}$, and from expression (7), the restricted log-likelihood of $\mathbf{X}$ is given by

$$L_R(\sigma_v^2, \alpha, \mathbf{X}) = -\left(\frac{n}{2}\right)\log\sigma_v^2 + \frac{1}{2}\log\left(\frac{1+\alpha}{(n-1)(1-\alpha)+2}\right) - \frac{1}{2\sigma_v^2}Q(\alpha), \qquad \textbf{(8)}$$

where

$$Q(\alpha) = \mathbf{X}'\boldsymbol{\Sigma}_X^{-1}\mathbf{X} - \frac{(\mathbf{X}'\boldsymbol{\Sigma}_X^{-1}\mathbf{1})^2}{\mathbf{1}'\boldsymbol{\Sigma}_X^{-1}\mathbf{1}} \qquad \textbf{(9)}$$

and $\boldsymbol{\Sigma}_X \equiv \mathrm{Var}(\mathbf{X})$. On the other hand, the regular likelihood of $\mathbf{X}$ for model (2) with known μ (set, without loss of generality, to zero) is

$$L(\sigma_v^2, \alpha, \mathbf{X}) = -\left(\frac{n+1}{2}\right)\log\sigma_v^2 + \frac{1}{2}\log(1-\alpha^2) - \frac{1}{2\sigma_v^2}\mathbf{X}'\boldsymbol{\Sigma}_X^{-1}\mathbf{X}. \qquad \textbf{(10)}$$

On comparing (8) and (10) and noting that the second term in $Q(\alpha)$ is $O(1)$ whereas $\mathbf{X}'\boldsymbol{\Sigma}_X^{-1}\mathbf{X}$ is $O(n)$, it immediately becomes apparent that the restricted likelihood in this case differs on a relative scale by only an order $O(n^{-1})$ from the likelihood of the AR(1) process with known intercept μ. As a matter of fact, Cheang and Reinsel (2000) show that the restricted likelihood for the AR(1) provides REML estimates of α whose bias is $-2\alpha n^{-1} + O(n^{-2})$, which is identical, up to order $O(n^{-1})$, to the bias of the maximum likelihood estimate when the intercept is known (Marriott and Pope, 1954). Because the bias of the maximum likelihood estimate of α when the intercept is not known is $-(1+3\alpha)n^{-1} + O(n^{-2})$, the REML estimate is able to achieve a bias reduction of approximately 50% when the AR parameter is close to the unit root. The ability of the restricted likelihood to imitate the regular likelihood of a zero-mean process goes even further. A little simple algebra shows that the higher order derivatives (with respect to the parameters α, σ_v^2) of the restricted likelihood in (8) are identical, up to $O(n^{-1})$, to those of the zero-mean regular likelihood in (10). As a result, we can use the calculations in van Garderen (1999) obtained for the zero-mean AR(1) model and establish that the Efron curvature properties of the restricted likelihood for the AR(1) are the same as those of the AR(1) model with known intercept, up to order $O(n^{-1})$. In addition, we also get the following theorem, which provides a formal expansion for the distribution of the RLRT in the model in (2) by arguments identical to those for Theorem 1 of van Giersbergen (2006), established for the zero-mean AR(1) model (see also note 2 for the expansion in (4)).

THEOREM 1. *Let* $\mathbf{X} = (X_0, \ldots, X_n)'$ *follow the univariate* AR*(1) model in (2) and let* $RLRT_\alpha$ *denote the RLRT based on the expression in (8) for testing* $H_0 : \alpha = \alpha_0$ *versus* $H_1 : \alpha \neq \alpha_0$, $|\alpha_0| < 1$. *Then,*

$$P(RLRT_\alpha \leq x) - G_1(x) = -0.25n^{-1}(G_3(x) - G_1(x)) + O(n^{-2}).$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Remark 1. It is worth noting that the result of Theorem 1 continues to hold if the initial value X_0 in the AR(1) model is assumed to follow $N(\mu, \sigma_v^2)$ instead of the stationary distribution. The reason for this is that the leading terms in the derivatives of the restricted likelihood are unaffected by the specification of the initial value.

Comparing the result in Theorem 1 with the expressions in (4) and (5), we see that the distribution of the RLRT for α in the model with intercept behaves like that of the regular LRT for the zero-mean model and should be well approximated by the χ_1^2 even when α is close to the unit root. This can also be seen in the empirical densities of the RLRT for α in an intercept model shown in Figure 2, which are plotted together with the limiting χ_1^2 density. (As in Figure 1, we plot the density of the cube root of the RLRT.) Furthermore, Table 2 shows that the simulation rejection percentages of the RLRT based on χ_1^2 critical values are close to the nominal levels. The preceding discussion shows that the restricted likelihood provides a great advantage for both hypothesis testing and estimation for the univariate AR(1) model through the elimination of the nuisance intercept parameter. The general problem of RLRT-based inference in univariate AR(p) models

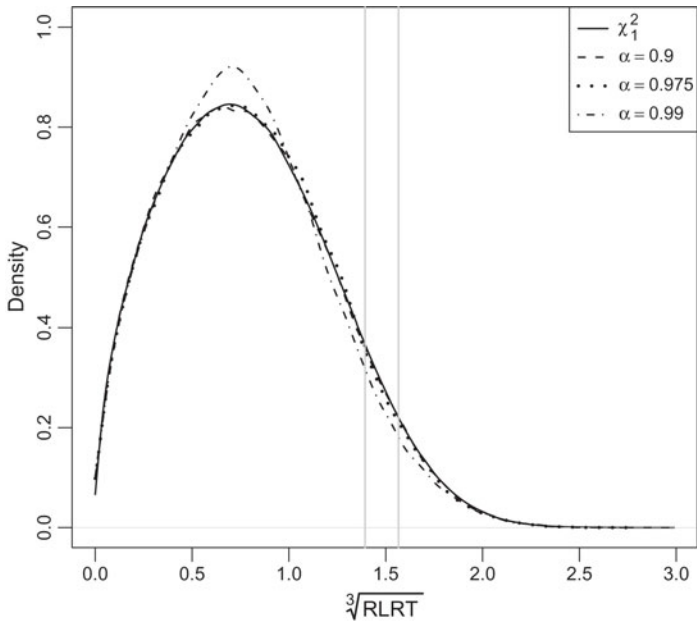


FIGURE 2. Empirical densities of cubic root transformed RLRT statistics of AR(1) processes with intercept. The vertical lines are 90th and 95th percentiles. The plot is based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = 0.9, 0.975$, and 0.99.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

TABLE 2. Rejection rates of RLRT for AR(1) processes

	Intercept			Intercept and trend		
α	10%	5%	1%	10%	5%	1%
0.90	0.1025	0.0504	0.0112	0.1026	0.0499	0.0106
0.95	0.1006	0.0505	0.0112	0.0903	0.0435	0.0091
0.99	0.0847	0.0404	0.0095	0.0927	0.0442	0.0088
0.995	0.0827	0.0403	0.0091	0.0927	0.0437	0.0089

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

with intercept/trend is studied in Chen and Deo (2007). In this paper, we focus our attention on the use of the restricted likelihood for carrying out inference on β in the bivariate predictive regression model in (1) and (2). We first obtain a tractable expression for the restricted likelihood for this model.

We start by noting that the vector $(\mathbf{Y}', \mathbf{X}')' = (Y_1, \ldots, Y_n, X_0, \ldots, X_n)'$ defined by (1) and (2) can be expressed in the form (6), where the design matrix $\mathbf{W}$ is now of the form

$$\mathbf{W} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n+1} \end{bmatrix}, \tag{11}$$

where $\mathbf{1}$ is a vector of ones and the regression coefficient vector is of the form $\boldsymbol{\theta} = (\eta + \beta\mu/(1-\alpha), \mu/(1-\alpha))'$. However, the covariance matrix of the resulting error vector $\boldsymbol{\varepsilon}$ in this representation takes on an awkward form. As a result, though in principle we could obtain the restricted likelihood of $(\mathbf{Y}', \mathbf{X}')'$ by using the form (7) given earlier, the resulting expression is not simple to study. Hence, we derive the restricted likelihood by appealing to its basic definition and exploiting the structure of the model. We start by noting that because the design matrix $\mathbf{W}$ now has the form (11), the restricted likelihood of $(\mathbf{Y}', \mathbf{X}')'$ is the exact likelihood of $(\mathbf{Y}'\mathbf{T}_1', \mathbf{X}'\mathbf{T}_2')'$, where $\mathbf{T}_1$ and $\mathbf{T}_2$ are full row rank matrices of dimension $(n-1) \times n$ and $n \times (n+1)$, respectively, satisfying $\mathbf{T}_1\mathbf{1} = \mathbf{0}$ and $\mathbf{T}_2\mathbf{1} = \mathbf{0}$. (In other words, the restricted likelihood in this context is the exact likelihood of $\{(Y_t - Y_{t-1})_{t=2}^n, (X_t - X_{t-1})_{t=1}^n\}$.) We next note that the exact likelihood of $(\mathbf{Y}'\mathbf{T}_1', \mathbf{X}'\mathbf{T}_2')'$ can be factorized as

$$L\big((\mathbf{Y}'\mathbf{T}_1', \mathbf{X}'\mathbf{T}_2')'\big) = L(\mathbf{T}_1\mathbf{Y} \mid \mathbf{T}_2\mathbf{X})L(\mathbf{T}_2\mathbf{X}).$$

By definition, the likelihood $L(\mathbf{T}_2\mathbf{X})$ is just the restricted likelihood of $\mathbf{X}$ given in (8). In the Appendix we obtain a simple expression for $L(\mathbf{T}_1\mathbf{Y} \mid \mathbf{T}_2\mathbf{X})$ by using the model structure, and in conjunction with (8) we thus obtain a simple expression for the restricted likelihood of the predictive regression model. This simple expression for the restricted likelihood that we state in Theorem 2 later in this

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

section allows both very easy calculation of the REML estimates and a useful expression for their finite-sample bias. We first define some quantities that will be useful to us in stating our results.

For the observed data $(\mathbf{Y}', \mathbf{X}')' = (Y_1, \ldots, Y_n, X_0, \ldots, X_n)'$ define $\mathbf{X}_1 = (X_1, \ldots, X_n)'$ and $\mathbf{X}_0 = (X_0, X_1, \ldots, X_{n-1})'$. Define the sample means $\bar{Y} = n^{-1}\mathbf{1}'\mathbf{Y}$, $\bar{X} = (n+1)^{-1}\mathbf{1}'\mathbf{X}$, $\bar{X}_1 = n^{-1}\mathbf{1}'\mathbf{X}_1$, and $\bar{X}_0 = n^{-1}\mathbf{1}'\mathbf{X}_0$ and the sample mean corrected data $\mathbf{Y}_c = \mathbf{Y} - \mathbf{1}\bar{Y}$, $\mathbf{X}_c = [\mathbf{X}_1 - \mathbf{1}\bar{X}_1,\ \mathbf{X}_0 - \mathbf{1}\bar{X}_0]$. Define

$$S(\phi, \beta, \alpha) = (\mathbf{Y}_c - \phi\mathbf{X}_{c,1} - (\beta - \phi\alpha)\mathbf{X}_{c,2})'(\mathbf{Y}_c - \phi\mathbf{X}_{c,1} - (\beta - \phi\alpha)\mathbf{X}_{c,2}) \quad \textbf{(12)}$$

and note that for computational purposes $Q(\alpha)$ given in (9) can be written as

$$Q(\alpha) = \sum_{t=0}^{n}(X_t - \bar{X})^2 + \alpha^2\sum_{t=1}^{n-1}(X_t - \bar{X})^2 - 2\alpha\sum_{t=0}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})$$
$$- \frac{(1-\alpha)\alpha^2}{(n-1)(1-\alpha)+2}[X_0 + X_n - 2\bar{X}]^2.$$

We now can state the following theorem.

THEOREM 2. *For the model given by (1) and (2), the REML log-likelihood up to an additive constant is given by*

$$L(\beta, \alpha, \phi, \sigma_v^2, \sigma_e^2) = -\left(\frac{n-1}{2}\right)\log\sigma_e^2 - \frac{1}{2\sigma_e^2}S(\phi, \beta, \alpha) \quad \textbf{(13)}$$
$$- \left(\frac{n}{2}\right)\log\sigma_v^2 + \frac{1}{2}\log\left(\frac{1+\alpha}{(n-1)(1-\alpha)+2}\right) - \frac{1}{2\sigma_v^2}Q(\alpha).$$

The REML estimates $\hat{\boldsymbol{\psi}} = (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\sigma}_v^2, \hat{\sigma}_e^2)$ *are given by*

$$\hat{\alpha} = \arg\min_{\alpha}\left\{n\log Q(\alpha) - \log\left(\frac{1+\alpha}{(n-1)(1-\alpha)+2}\right)\right\},$$

$$(\hat{\phi}, \hat{\beta})' = \begin{bmatrix} 1 & 0 \\ \hat{\alpha} & 1 \end{bmatrix}(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{Y}_c,$$

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi}, \hat{\beta}, \hat{\alpha})}{n-1},$$

and

$$\hat{\sigma}_v^2 = \frac{Q(\hat{\alpha})}{n}.$$

The bias in $(\hat{\alpha}, \hat{\beta})$ *is given by*

$$\mathrm{E}(\hat{\alpha} - \alpha) = -\frac{2\alpha}{n-1} + o(n^{-1}),$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

$$\mathrm{E}(\hat{\beta}-\beta)=\phi\mathrm{E}(\hat{\alpha}-\alpha)$$
$$=-\phi\left(\frac{2\alpha}{n-1}\right)+o(n^{-1}),$$

and

$$\mathrm{E}(\hat{\phi}-\phi)=0.$$

Remark 2. It is obvious that obtaining the REML estimates is computationally easy because all the estimates are obtained in succession after the optimization of a one-dimensional function that is almost quadratic.

Remark 3. Note that the restricted likelihood in (13) is well defined at the unit root $\alpha=1$, without having to assume that the initial value X_0 is fixed when $|\alpha|<1$.

It is interesting to compare the bias in $\hat{\beta}$ with the bias in the OLS estimate $\hat{\beta}_{OLS}$, given by (see Stambaugh, 1999)

$$\mathrm{E}(\hat{\beta}_{OLS}-\beta)=\phi\mathrm{E}(\hat{\alpha}_{OLS}-\alpha)=-\phi\left(\frac{1+3\alpha}{n-1}\right)+o\left(n^{-1}\right),$$

where $\hat{\alpha}_{OLS}$ is the OLS estimate of α. Thus, the bias in $\hat{\beta}$ depends upon the bias in $\hat{\alpha}$ in a manner identical to the way the bias in $\hat{\beta}_{OLS}$ depends on the bias in $\hat{\alpha}_{OLS}$. Consequently, the approximately 50% reduction in bias that $\hat{\alpha}$ achieves compared to $\hat{\alpha}_{OLS}$ close to the unit root is inherited by $\hat{\beta}$, relative to $\hat{\beta}_{OLS}$. The bias expression in Theorem 2 also suggests a bias-corrected version of the REML estimate of β that may be computed as

$$\hat{\beta}_c=\hat{\beta}+\hat{\phi}\left(\frac{2\hat{\alpha}_c}{n-1}\right),$$

where

$$\hat{\alpha}_c=\hat{\alpha}\left(\frac{n+1}{n-1}\right)$$

is the bias-corrected REML estimate of α. Because the bias correction term $\phi 2\alpha(n-1)^{-1}$ is smaller than $\phi(1+3\alpha)(n-1)^{-1}$, one would expect the bias-corrected REML estimate $\hat{\beta}_c$ to have both less bias and a smaller variance than its bias-corrected OLS counterpart,

$$\hat{\beta}_{OLS,c}=\hat{\beta}_{OLS}+\hat{\phi}\left(\frac{1+3\hat{\alpha}_c}{n-1}\right),$$

particularly because the parameter ϕ can take any value in $(-\infty,\infty)$.[5] This is indeed the case, as we see in the simulations reported in Section 5.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

The restricted likelihood given in (13) is derived for the situation where we assume that the initial value X_0 comes from the stationary distribution $N(\mu(1-\alpha)^{-1}, \sigma_v^2(1-\alpha^2)^{-1})$. A similar argument can be used to obtain the restricted likelihood for the model in (1) and (2) where the regressor series follows an asymptotically stationary process, given by $X_t = \mu + \widetilde{X}_t$, where $\widetilde{X}_t = \alpha \widetilde{X}_{t-1} + v_t$ for $t \geq 1$ and $\widetilde{X}_0 = v_0$. Under this assumption, the restricted likelihood is given by

$$\widetilde{L}(\beta, \alpha, \phi, \sigma_v^2, \sigma_e^2) = -\left(\frac{n-1}{2}\right)\log\sigma_e^2 - \frac{1}{2\sigma_e^2}S(\phi, \beta, \alpha) \tag{14}$$
$$-\left(\frac{n}{2}\right)\log\sigma_v^2 + \frac{1}{2}\log\left(\frac{1}{n(1-\alpha)^2+1}\right) - \frac{1}{2\sigma_v^2}\widetilde{Q}(\alpha),$$

where

$$\widetilde{Q}(\alpha) = (X_0 - \hat{\mu}(\alpha))^2 + \sum_{t=1}^{n}(X_t - \hat{\mu}(\alpha) - \alpha(X_{t-1} - \hat{\mu}(\alpha)))^2$$

and

$$\hat{\mu}(\alpha) = \frac{X_0 + (1-\alpha)\sum_{t=1}^{n}(X_t - \alpha X_{t-1})}{1 + n(1-\alpha)^2}.$$

The bias results of Theorem 2 continue to hold for the REML estimates under this initial value condition.

In the next section, we provide a theorem that shows that the REML-based LRT has very good finite-sample properties, in that its finite-sample distribution approaches the limiting one very quickly and is practically unaffected by nuisance parameters, while also maintaining power against local alternatives when compared to the Wald and score tests.

3. REML LIKELIHOOD RATIO TEST

One standard method for testing the composite hypothesis $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$ is the LRT, which compares the log-likelihood evaluated at the unrestricted estimates of the parameters to the log-likelihood evaluated at the parameter estimates obtained under the restriction that the null hypothesis $H_0 : \beta = 0$ is true. Using the quantities defined in (12) and just preceding it, it can be easily verified that under $H_0 : \beta = 0$ the restricted estimates $\hat{\boldsymbol{\psi}}_0 = \left(0, \hat{\alpha}_0, \hat{\phi}_0, \hat{\sigma}_{v0}^2, \hat{\sigma}_{e0}^2\right)$ are obtained as

$$\hat{\alpha}_0 = \arg\min_{\alpha}\ n\log Q(\alpha) - \log\left(\frac{1+\alpha}{(n-1)(1-\alpha)+2}\right) + (n-1)\log R(\alpha),$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

where

$$R(\alpha) = \mathbf{Y}_c' \left(\mathbf{I} - \mathbf{Z}_c(\alpha)(\mathbf{Z}_c'(\alpha)\mathbf{Z}_c(\alpha))^{-1}\mathbf{Z}_c'(\alpha) \right) \mathbf{Y}_c,$$

$$\mathbf{Z}_c(\alpha) = \mathbf{X}_1 - \mathbf{1}\bar{X}_1 - \alpha(\mathbf{X}_0 - \mathbf{1}\bar{X}_0),$$

$$\hat{\phi}_0 = (\mathbf{Z}_c'(\hat{\alpha}_0)\mathbf{Z}_c(\hat{\alpha}_0))^{-1}\mathbf{Z}_c'(\hat{\alpha}_0)\mathbf{Y}_c,$$

$$\hat{\sigma}_{v,0}^2 = \frac{Q(\hat{\alpha}_0)}{n}, \qquad \hat{\sigma}_{e,0}^2 = \frac{1}{n-1} R(\hat{\alpha}_0).$$

Just as in the unrestricted case, it is obvious that obtaining the restricted estimates requires only the optimization of a nearly quadratic one-dimensional function. The REML-based likelihood ratio test (RLRT) for testing $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$ is now given by

$$R_T = -2L(\hat{\boldsymbol{\psi}}_0) + 2L(\hat{\boldsymbol{\psi}}), \tag{15}$$

where $L(\cdot)$ is the REML log-likelihood presented in (13). Under $H_0 : \beta = 0$, the asymptotic distribution of R_T is χ_1^2, the chi-square distribution with one degree of freedom. The following theorem provides insight into the finite-sample behavior of the RLRT through a formal expansion of its distribution.

THEOREM 3. *Under $H_0 : \beta = 0$ in the model given by (1) and (2), we have*

$$P(R_T \leq x) = P(\chi_1^2 \leq x) + \left(\frac{3}{4} - \rho^2 \right) n^{-1} [P(\chi_3^2 \leq x) - P(\chi_1^2 \leq x)] + O(n^{-2}), \tag{16}$$

where $\rho = \mathrm{Corr}(u_t, v_t)$.

Theorem 3 in conjunction with the result of Barndorff-Nielsen and Hall (1988) yields the following corollary.

COROLLARY 1. *If the Bartlett-corrected RLRT is defined as*

$$R_{TB} = \left(1 + 2 \left(\frac{3}{4} - \hat{\rho}^2 \right) n^{-1} \right)^{-1} R_T,$$

where

$$\hat{\rho}^2 = \left(\hat{\phi}^2 \hat{\sigma}_v^2 + \hat{\sigma}_e^2 \right)^{-1} \hat{\phi}^2 \hat{\sigma}_v^2,$$

then

$$P(R_{TB} \leq x) = P\left(\chi_1^2 \leq x \right) + O\left(n^{-2} \right).$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Remark 4. Inspection of the proof of Theorem 3 shows that the expansion (16) depends entirely on the expected values of the higher order derivatives (with respect to the parameters) of the restricted log-likelihood. Hence, the results of Theorem 3 and Corollary 1 continue to hold for the restricted likelihood given in (14) when the initial condition X_0 is not from the stationary distribution.

The results of Theorem 3 obviously imply that the χ_1^2 approximation to R_T is very good and almost unaffected by the nuisance parameters. Most importantly, the leading term in the error is free of the AR parameter, which most affects the finite-sample performance of t-statistic-based tests, particularly when it is close to unity. Theorem 3 also suggests two very simple ways in which one could adjust the p-value when carrying out a test of $H_0 : \beta = 0$. One would be to use the first two terms on the right-hand side of (16), with ρ^2 replaced by $\hat{\rho}^2$. The other is to use the Bartlett-corrected statistic R_{TB}. However, (16) suggests that the original test R_T used in conjunction with the standard χ_1^2 distribution should be very well behaved and any improvements will be minimal at best. This belief is supported by the simulations that we provide in Section 5. Also, the correction factor $1 + 2\left(\frac{3}{4} - \hat{\rho}^2\right) n^{-1}$ in Corollary 1 is almost unity for any reasonable sample size, and hence the correction will be negligible. Hence, we do not pursue the use of the Bartlett-corrected test in Corollary 1, merely noting that it can achieve an $O(n^{-2})$ error rate. It is also worth noting that though the REML likelihood does not provide an unbiased estimate of β (indeed, the bias of $\hat{\beta}$ can be arbitrarily large because of the fact that ϕ is unbounded, as noted following Theorem 2), the REML likelihood yields a very well-behaved test for β, irrespective of how large ϕ is. This result serves to illustrate the point that it may be more desirable at times to carry out tests of hypotheses using appropriate likelihoods rather than using parameter point estimates and supports the idea that bias can be irrelevant in inference, as described in the Introduction.

It is worthwhile to compare our results with those provided in Jansson and Moreira (2006). We first note that the restricted likelihood approach not only provides a way of carrying out reliable inference but also yields point estimates with significantly reduced bias, in addition to an estimable bias correction, unlike the procedure in Jansson and Moreira (2006), which is only a hypothesis test. Jansson and Moreira assumed that the error covariance parameters $(\phi, \sigma_e, \sigma_v)$ are known, that the series $\{X_t\}$ has zero intercept (i.e., $\mu = 0$), and that the initial value X_0 is known to be zero. Under these assumptions, Jansson and Moreira obtained a test with certain finite-sample optimality properties. Such assumptions, however, would normally not be satisfied in practice. In their Theorem 6, Jansson and Moreira allow for an unknown μ taking potentially nonzero values. However, in this Theorem 6 they provide asymptotic results on the size and power properties of their test statistic only along a sequence of local-to-unity parameter values for the AR coefficient α, which is parametrized as $\alpha = 1 - cn^{-1}$ for some fixed $c \geq 0$. The limiting distribution for their test statistic $\hat{R}$, which is not "self-normalized," unlike, for example, the t-statistic for α in a univariate AR(1) process, will be

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

asymptotically degenerate even if the AR coefficient were approaching unity at a rate slower than n^{-1} (e.g., $\alpha = 1 - ck_n^{-1}$ for $k_n^{-1} + k_n n^{-1} \to 0$) based on the results of Phillips and Magdalinos (2007) and Giraitis and Phillips (2006). The k_n in the preceding framework may be arbitrarily "close" to n, such as $k_n = n/\log(\log n)$.

The result provided in Theorem 3 for the RLRT is for a fixed value α of the AR coefficient and is thus a pointwise result. Though we do not currently have any results on the behavior of the RLRT under a local-to-unity framework, our simulations reported subsequently show that the RLRT works very well even in such scenarios. At the boundary value $\alpha = 1$, the distribution of the RLRT will not be chi-square, and the chi-square approximation will fail. In practice however, the boundary value of $\alpha = 1$ is not relevant in most finance applications because the predictor series is stationary by construction. For example, Baker, Taliaferro, and Wurgler (2006) assert on page 1715 that "The predictor variables we consider are theoretically stationary by construction (although in any given small sample, of course, one might not be able to reject a unit root)." A similar case for stationarity of the predictor series is made on page 213 of Lewellen (2004). Furthermore, the dependent series y_t in most finance applications is a returns series, which is unquestionably stationary, and hence, as Lewellen (2004) states on page 213, "It also makes little sense to predict returns with a nonstationary variable." In light of this context in which the predictive regression model is most often used, we argue that the boundary case of the unit root, where the chi-square approximation to the RLRT will fail, is not as relevant as it is in univariate AR modeling of economic series.

The fact that the leading error term in (16) is minimized at $\rho = \pm\sqrt{3/4}$ and not at $\rho = 0$ seems somewhat puzzling in light of the observation that the t-statistic for β provides exact inference when $\rho = 0$. This puzzling result can be explained by comparing the quality of the chi-square approximation in Theorem 3, where all four parameters $(\alpha, \phi, \sigma_v^2, \sigma_e^2)$ are nuisance parameters, to the approximation in the case where the parameters $(\phi, \sigma_v^2, \sigma_e^2)$ are known and α is the only nuisance parameter. From equation (A.12) in the Appendix and the discussion following it in the proof of Theorem 3, it can be seen that if $(\phi, \sigma_v^2, \sigma_e^2)$ are known the relevant RLRT for testing $H_0 : \beta = 0$, denoted by $R_{T,1}$, would have a distribution that satisfies the formal expansion

$$P(R_{T,1} \le x) = P(\chi_1^2 \le x) - \rho^2 n^{-1}[P(\chi_3^2 \le x) - P(\chi_1^2 \le x)] + O(n^{-2}). \quad \textbf{(17)}$$

In this situation, the RLRT will be best approximated by a χ_1^2 variable when $\rho = 0$. As a matter of fact, some trivial calculations show that under the assumption that $(\phi, \sigma_v^2, \sigma_e^2)$ are known the RLRT when $\rho = 0$ is exactly a χ_1^2 variable. By comparing the result in (17) with that in Theorem 3, one sees that the quantity $3/4$ is a measure of the extent to which lack of knowledge of the innovation parameters (ϕ, τ_v, τ_e) affects the finite-sample distribution of R_T. Because $\sup_\rho \rho^2 > \sup_\rho |\rho^2 - 3/4|$ for $\rho \in (-1, 1)$, the chi-square approximation to the RLRT is better in a "minimax" sense over all possible values of ρ for the case when the innovation covariance parameters are unknown than when they are known.[6] Hence,

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

though the RLRT may not provide exact inference when $\rho = 0$, it is able to keep the error in check over the entire parameter space of ρ, whereas the t-statistic for β works perfectly at $\rho = 0$ but fails badly as ρ moves away from zero.

The result in Theorem 3 guarantees that the RLRT will yield a test that is almost of exact size in finite samples. However, one would also like to ensure that this is not achieved at the expense of loss of power. The obvious tests that are competitors to the RLRT are the Wald and Rao score tests based on the restricted likelihood. It is a well-known fact that just as these three tests share the identical limiting distribution under the null hypothesis, they also have the same power properties to first order. Hence, to distinguish between them one has to consider a sequence of local Pitman alternatives given by $H_a : \beta = \beta_0 + \xi n^{-1/2}$. The next theorem obtains the power function of the RLRT, Wald, and Rao score tests against such local alternatives.

THEOREM 4. *Let RLRT, W, and RS denote the LRT, Wald test, and Rao score test, respectively, of $H_0 : \beta = \beta_0$ based on the restricted likelihood in (13) of the model in (1) and (2). Assume that the true value of β is given by $\beta = \beta_0 + \xi n^{-1/2}$. Define*

$$\Delta = \frac{1}{1-\alpha^2}\frac{\sigma_v^2}{\sigma_u^2}\xi^2 + O(n^{-1}),$$

$$C_1 = \frac{-2\alpha\phi\sigma_v^4\xi^3}{(1-\alpha^2)^2\sigma_u^4},$$

and

$$C_2 = \frac{-3\alpha\phi\sigma_v^2\xi}{(1-\alpha^2)\sigma_u^2}.$$

Let $\bar{G}_{s,\Delta}(x)$ denote the survival function of a noncentral χ^2 random variable with s degrees of freedom and noncentrality parameter Δ. Then,

$$P(RLRT > x) = \bar{G}_{1,\Delta}(x) + \frac{C_1}{n^{1/2}}(\bar{G}_{3,\Delta}(x) - 0.5\bar{G}_{1,\Delta}(x) - 0.5\bar{G}_{5,\Delta}(x)) + O(n^{-1}),$$

$$P(W > x) = P(RLRT > x) + O(n^{-1}),$$

and

$$P(RS > x) = P(RLRT > x) + \frac{C_1}{n^{1/2}}(0.5\bar{G}_{5,\Delta}(x) - 0.5\bar{G}_{7,\Delta}(x)) + \frac{C_2}{n^{1/2}}(0.5\bar{G}_{1,\Delta}(x) - 0.5\bar{G}_{5,\Delta}(x)) + O(n^{-1}). \quad (18)$$

From Theorem 4 we see that the RLRT and the Wald test based on the restricted likelihood have identical power up to second order (i.e., up to $O(n^{-1})$) against local Pitman alternatives. Because $\bar{G}_{s,\Delta}(x) - \bar{G}_{l,\Delta}(x) < 0$ for all $x > 0$ when $l > s$, it follows from (18) that the RLRT will be guaranteed to be more powerful

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

than the Rao score test against local alternatives if $C_1 > 0$ and $C_2 > 0$. This will be the case if $\phi < 0$ and $\xi > 0$, which is exactly the part of the parameter space that is of relevance in empirical applications in finance and economics. It is also interesting to note that the noncentrality parameter Δ, which will be the main source of power, increases as α gets closer to the unit root.

In the next section we derive the REML likelihood under more general models for the regressor series and also for multiple regressors and discuss some efficiency and computational issues.

4. REML FOR MORE GENERAL REGRESSOR MODELS

It is easy to generalize the REML likelihood in two directions that are both of practical interest. One generalization is to the case where the predictor series is a multivariate AR(1) process. Applications of such models can be found, for example, in the paper by Amihud and Hurvich (2004), who considered dividend yield and earnings to price ratio as bivariate predictors of market returns. The other generalization is to the case where the univariate predictor follows a higher order AR process. We will state the REML log-likelihood for both these cases, starting with the multivariate AR(1) predictor model. The method of obtaining the log-likelihood is identical to that used in Theorem 2.

4.1. Multivariate Regressors

Assume that the data $(Y_1, \ldots, Y_n, \mathbf{X}_0', \ldots, \mathbf{X}_n')$ follow

$$Y_t = \eta + \boldsymbol{\beta}' \mathbf{X}_{t-1} + u_t, \tag{19}$$

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{X}_{t-1} + \mathbf{v}_t, \tag{20}$$

where $u_t = \boldsymbol{\phi}' \mathbf{v}_t + e_t$, $(e_t, \mathbf{v}_t')' \sim N(\mathbf{0}, \operatorname{diag}(\sigma_e^2, \boldsymbol{\Sigma}_v))$ is an i.i.d. series, and $\mathbf{A}$ is a $k \times k$ matrix with all eigenvalues less than unity in absolute value. Let $\boldsymbol{\Sigma}_v \equiv \operatorname{Var}(\mathbf{v}_t)$ and $\boldsymbol{\Sigma}_X \equiv \operatorname{Var}(\mathbf{X}_t)$, given by

$$\operatorname{vec}(\boldsymbol{\Sigma}_X) = \left(I_{k^2} - \mathbf{A} \otimes \mathbf{A}\right)^{-1} \operatorname{vec}(\boldsymbol{\Sigma}_v)$$

and define

$$\hat{\boldsymbol{\tau}} = \left[\boldsymbol{\Sigma}_X^{-1} + n(\mathbf{I} - \mathbf{A})' \boldsymbol{\Sigma}_v^{-1} (\mathbf{I} - \mathbf{A})\right]^{-1} \left[\boldsymbol{\Sigma}_X^{-1} \mathbf{X}_0 + (\mathbf{I} - \mathbf{A})' \boldsymbol{\Sigma}_v^{-1} (\mathbf{I} - \mathbf{A}) \sum_{t=1}^{n} \mathbf{X}_t\right].$$

LEMMA 1. *Then the REML log-likelihood up to an additive constant for the model in (19) and (20) is given by*

$$L_M = -\left(\frac{n-1}{2}\right) \log \sigma_e^2 - \frac{1}{2\sigma_e^2} S(\boldsymbol{\phi}, \boldsymbol{\beta}, \mathbf{A}) - \frac{1}{2} \log |\boldsymbol{\Sigma}_X| - \frac{n}{2} \log |\boldsymbol{\Sigma}_v| \tag{21}$$
$$-\frac{1}{2} \log \left| \boldsymbol{\Sigma}_X^{-1} + n (\mathbf{I} - \mathbf{A})' \boldsymbol{\Sigma}_v^{-1} (\mathbf{I} - \mathbf{A}) \right|$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

$$-\frac{1}{2}\left\{\left(\mathbf{X}_0-\hat{\boldsymbol{\tau}}\right)'\boldsymbol{\Sigma}_X^{-1}\left(\mathbf{X}_0-\hat{\boldsymbol{\tau}}\right)+\sum_{t=1}^{n}\left(\mathbf{X}_t-\hat{\boldsymbol{\tau}}-\mathbf{A}\left(\mathbf{X}_{t-1}-\hat{\boldsymbol{\tau}}\right)\right)'\right.$$

$$\left.\times\,\boldsymbol{\Sigma}_v^{-1}\left(\mathbf{X}_t-\hat{\boldsymbol{\tau}}-\mathbf{A}\left(\mathbf{X}_{t-1}-\hat{\boldsymbol{\tau}}\right)\right)\right\},$$

where

$$S(\boldsymbol{\phi},\boldsymbol{\beta},\mathbf{A})=\sum_{t=1}^{n}\left(Y_{t,c}-\boldsymbol{\phi}'\mathbf{X}_{t,c}-\left(\boldsymbol{\beta}'-\boldsymbol{\phi}'\mathbf{A}\right)\mathbf{X}_{t-1,c}\right)^2,$$

$$\mathbf{X}_{t,c}=\mathbf{X}_t-n^{-1}\textstyle\sum_{t=1}^{n}\mathbf{X}_t,\ \textit{and}\ \ \mathbf{X}_{t-1,c}=\mathbf{X}_{t-1}-n^{-1}\textstyle\sum_{t=1}^{n}\mathbf{X}_{t-1}.$$

To ease the computational burden during optimization, the likelihood can be defined in terms of the reparametrized set $(\boldsymbol{\Sigma}_v,\sigma_e^2,\mathbf{A},\boldsymbol{\phi},\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}=\boldsymbol{\beta}-\mathbf{A}\boldsymbol{\phi}$. This reparametrization allows us to concentrate $(\sigma_e^2,\boldsymbol{\phi},\boldsymbol{\gamma})$ out of the likelihood, thus reducing the dimensionality of the optimization problem. The likelihood can then be sequentially optimized, first over $(\boldsymbol{\Sigma}_v,\mathbf{A})$, with the REML estimates of $(\sigma_e^2,\boldsymbol{\phi},\boldsymbol{\gamma})$ being then obtained by OLS through the minimization of $S(\boldsymbol{\phi},\boldsymbol{\gamma})$. A further simplification of the preceding likelihood occurs if the coefficient matrix $\mathbf{A}$ is diagonal, given by $\mathbf{A}=\operatorname{diag}(\alpha_1,\ldots,\alpha_k)$ with $\max_i|\alpha_i|<1$, in which case one gets

$$\operatorname{Var}(\mathbf{X}_t)\equiv\boldsymbol{\Sigma}_X=\left(\left(\frac{\sigma_{v,ij}}{1-\alpha_i\alpha_j}\right)\right),$$

where $\boldsymbol{\Sigma}_v=((\sigma_{v,ij}))$. Amihud and Hurvich (2004) find evidence to support this model with a diagonal coefficient matrix in the empirical example that they consider. It should be noted that in the case where $\mathbf{A}$ can be assumed to be a diagonal matrix, the predictive regression model is no longer a seemingly unrelated regression (SUR) system and hence OLS will no longer be efficient. However, REML will clearly retain efficiency, no matter what the form of $\mathbf{A}$ is, thus giving it an advantage in terms of both asymptotic efficiency and power over any OLS-based procedure.

Because the dimension of the parameter space is very large in the vector case, it is not feasible to obtain a result such as Theorem 3 in the most general case. However, in the case where $\mathbf{A}$ is a diagonal matrix and where $(\sigma_e^2,\boldsymbol{\phi},\boldsymbol{\Sigma}_v)$ are assumed known with $\boldsymbol{\Sigma}_v$ diagonal, we are able to obtain the following result on the finite-sample behavior of the RLRT for testing $H_0:\boldsymbol{\beta}=\mathbf{0}$. The proof follows along lines similar to those for Theorem 3 and is omitted.

THEOREM 5. *In the model given by (19) and (20), assume that* $\mathbf{A}=\operatorname{diag}(\alpha_1,\ldots,\alpha_k)$, *with* $\max_i|\alpha_i|<1$, *and that* $(\sigma_e^2,\boldsymbol{\phi},\boldsymbol{\Sigma}_v)$ *are known with* $\boldsymbol{\Sigma}_v=\operatorname{diag}(\sigma_{v,11}^2,\ldots,\sigma_{v,kk}^2)$. *Let* R_M *denote the RLRT based on the restricted*

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

likelihood in (21) for testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$. *Then,*

$$P(R_M \le x) = P(\chi_k^2 \le x) - n^{-1}\left(\sum_{i=1}^{k} \frac{\phi_i^2 \sigma_{v,ii}^2 \sigma_e^{-2}}{1+\phi_i^2 \sigma_{v,ii}^2 \sigma_e^{-2}}\right) \times [P(\chi_{k+2}^2 \le x) - P(\chi_k^2 \le x)] + O(n^{-2}).$$

Because $0 < n^{-1}\sum_{i=1}^{k} \phi_i^2 \sigma_{v,ii}^2 \sigma_e^{-2}(1+\phi_i^2 \sigma_{v,ii}^2 \sigma_e^{-2})^{-1} < n^{-1}k$ trivially, the result shows that the χ^2 distribution once again provides a very good approximation to the RLRT in this situation. It is useful to note that Theorem 5 shows that the quality of the χ^2 approximation to the RLRT is affected only minimally by the dependence between u_t and $\mathbf{v}_t$, over which one has no control. However, once one X variable has been chosen, we can control which other X variables should be included in the model, and it is preferable to use a group of X variables that have low (ideally zero) correlation among themselves to avoid unnecessary multicollinearity. Hence, the assumption in the theorem that $\boldsymbol{\Sigma}_v$ is diagonal is not very unreasonable. Finally, from the discussion following equation (5) and following Theorem 3, one would expect the χ^2 approximation to continue to work well when $(\sigma_e^2, \boldsymbol{\phi}, \boldsymbol{\Sigma}_v)$ are unknown (with $\boldsymbol{\Sigma}_v$ diagonal), and, indeed, we find this to be the case in our simulations in Section 5. As a matter of fact, the simulations show that the RLRT behaves very well even when the cross-correlation in the $\mathbf{X}$ variables is as high as 0.9.

The restricted likelihood for the AR(p) regressor case can be derived in a manner analogous to that for the AR(1) case and is given next.

4.2. Higher Order Autoregressive Regressors

Let the observed data $(Y_1, \ldots, Y_n, X_{-p+1}, X_{-p+2}, \ldots, X_n)$ follow

$$Y_t = \eta + \beta X_{t-1} + u_t \tag{22}$$

and

$$X_t = \mu + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + v_t, \tag{23}$$

where $u_t = \phi v_t + e_t$ and $(e_t, v_t) \sim N\left(\mathbf{0}, \operatorname{diag}(\sigma_e^2, \sigma_v^2)\right)$ are an i.i.d. series. Furthermore, assume that all the roots of the polynomial $z^p - \sum_{s=1}^{p} z^{p-s}\alpha_s$ lie within the unit circle. Define

$$\mathbf{Y}_c = \mathbf{Y} - \mathbf{1}\bar{Y}, \qquad \mathbf{X}_c = \left[\mathbf{X}_1 - \mathbf{1}\bar{X}_1, \ldots, \mathbf{X}_{-p+1} - \mathbf{1}\bar{X}_{-p+1}\right],$$

where $\mathbf{X}_i = (X_i, X_{i+1}, \ldots, X_{n-1-i})$ and $\bar{X}_i = n^{-1}\mathbf{1}'\mathbf{X}_i$.

LEMMA 2. *The REML log-likelihood up to an additive constant for the model in (22) and (23) is given by*

$$L(\boldsymbol{\alpha}, \beta, \phi, \sigma_e^2, \sigma_v^2) = -\left(\frac{n-1}{2}\right)\log \sigma_e^2 - \frac{1}{2\sigma_e^2} S(\phi, \beta, \alpha_1, \ldots, \alpha_p) \tag{24}$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

$$-\left(\frac{n+p-1}{2}\right)\log\sigma_v^2-\frac{1}{2}\log|\mathbf{\Sigma}|-\frac{1}{2}\log\left|\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}\right|$$
$$-\frac{1}{2\sigma_v^2}(\mathbf{X}-\mathbf{1}\hat{\tau})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{1}\hat{\tau}),$$

where $\mathrm{Var}(\mathbf{X})=\sigma_v^{-2}\mathbf{\Sigma}$, $\hat{\tau}=(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{X}$, *and* $S(\phi,\beta,\alpha_1,\ldots,\alpha_p)=\mathbf{Z}_c'\mathbf{Z}_c$, *where*

$$\mathbf{Z}_c=\left(\mathbf{Y}_c-\phi\mathbf{X}_{c,1}-(\beta-\phi\alpha_1)\mathbf{X}_{c,2}+\sum_{i=2}^{p}\alpha_{c,i}\mathbf{X}_{c,i+1}\right).$$

Though at first glance the expression in (24) looks formidable, it is actually very easy to compute. The quantity $S(\phi,\beta,\alpha_1,\ldots,\alpha_p)$ is, of course, just a quadratic form. It is also well known that both the determinant $|\mathbf{\Sigma}|$ and bilinear/quadratic forms of the type $y'\mathbf{\Sigma}^{-1}x$ are very easy to compute for AR(p) models, because for such models $\mathbf{\Sigma}^{-1}$ can be easily expressed as $\mathbf{B}'\mathbf{B}$ for some lower triangular matrix $\mathbf{B}$. (See equation (A.6) in the Appendix.) After a reparametrization $\gamma=\beta-\phi\alpha_1$, the parameters $\left(\sigma_e^2,\sigma_v^2,\phi,\gamma\right)$ can be concentrated out of (24) and the concentrated log-likelihood optimized over the remaining parameters $\left(\alpha_2,\ldots,\alpha_p\right)$.

In the next section we report the results of our simulations.

5. SIMULATIONS

We first study the performance of four estimators of β: (i) the OLS estimate $\hat{\beta}_{OLS}$, (ii) the bias-corrected OLS estimate $\hat{\beta}_{OLS,c}$, (iii) the REML estimate $\hat{\beta}$, and (iv) the bias-corrected REML estimate $\hat{\beta}_c$. The bias-corrected estimators are defined following Theorem 2. The data were simulated from the model given by equations (1) and (2) with sample sizes $n=50,100,200$, and 400. For each sample size, the predictive regression slope coefficient β was set to zero whereas the AR coefficient α was set to $\alpha=1-cn^{-1}$ for $c=1,5$, and the initial value X_0 was drawn from its stationary distribution. The innovation variances were set as $\sigma_u^2=\sigma_v^2=1$ and $\phi=-0.98$. When $\sigma_u^2=\sigma_v^2=1$, the parameter ϕ reduces to the correlation between (u_t,v_t). This normalization of the innovation variances was also used in Campbell and Yogo (2006) and Jansson and Moreira (2006). The high negative value of ϕ was chosen to reflect the kind of values seen empirically (see Campbell and Yogo, 2006).

Tables 3 reports the simulation means and standard deviations of these four estimates based on 10,000 replications for each of the values of α. As predicted by the theory, the bias of $\hat{\beta}$ is uniformly less than that of $\hat{\beta}_{OLS}$ and approximately half its value when α is close to unity. At the same time, $\hat{\beta}$ does not suffer any loss of efficiency, indeed having a uniformly smaller standard deviation than that of $\hat{\beta}_{OLS}$. The bias of $\hat{\beta}_c$ is also always substantially less than that of $\hat{\beta}_{OLS,c}$, at times by as much as 80%, and its standard deviation is uniformly lower too. To get an idea of the magnitude of the bias and the impact of the bias correction,

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

TABLE 3. Mean and standard deviation of $\hat{\beta}$: $\beta = 0$, $\alpha = 1 - c/n$, $\phi = -0.98$, $\sigma_v^2 = \sigma_u^2 = 1$

Innovations			Gaussian errors				t_5 errors			
n	α		$\hat{\beta}_{OLS}$	$\hat{\beta}_{REML}$	$\hat{\beta}_{OLS,c}$	$\hat{\beta}_{REML,c}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{REML}$	$\hat{\beta}_{OLS,c}$	$\hat{\beta}_{REML,c}$
A. $c = 1$										
50	0.98	bias	0.0945	0.0484	0.0164	0.0097	0.0953	0.0491	0.0173	0.0105
		s.d.	0.0841	0.0789	0.0884	0.0821	0.0839	0.0785	0.0883	0.0817
100	0.99	bias	0.0486	0.0247	0.0094	0.0052	0.0491	0.0254	0.0010	0.0059
		s.d.	0.0436	0.0401	0.0447	0.0409	0.0442	0.0408	0.0453	0.0416
200	0.995	bias	0.0245	0.0124	0.0049	0.0027	0.0247	0.0125	0.0051	0.0028
		s.d.	0.0226	0.0204	0.0228	0.0207	0.0225	0.0206	0.0227	0.0208
400	0.9975	bias	0.0123	0.0061	0.0025	0.0012	0.0126	0.0065	0.0028	0.0016
		s.d.	0.0114	0.0104	0.0114	0.0104	0.0114	0.0104	0.0114	0.0105
B. $c = 5$										
50	0.90	bias	0.0796	0.0370	0.0058	0.0011	0.0811	0.0389	0.0074	0.0031
		s.d.	0.0927	0.0927	0.0981	0.0965	0.0926	0.0933	0.0980	0.0971
100	0.95	bias	0.0424	0.0199	0.0043	0.0011	0.0427	0.0206	0.0047	0.0018
		s.d.	0.0493	0.0482	0.0506	0.0492	0.0492	0.0489	0.0506	0.0499
200	0.975	bias	0.0219	0.0103	0.0026	0.0007	0.0222	0.0109	0.0029	0.0014
		s.d.	0.0255	0.0247	0.0258	0.0249	0.0256	0.0252	0.0260	0.0255
400	0.9875	bias	0.0110	0.0051	0.0013	0.0002	0.0113	0.0056	0.0015	0.0007
		s.d.	0.0132	0.0128	0.0133	0.0128	0.0129	0.0127	0.0129	0.0128

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

we note that Campbell and Yogo (2006, Table 5) report empirical estimates of β (after normalization to ensure unit innovation variances, as before) that range between 0.01 and 0.02 for monthly data and between 0.03 and 0.3 for annual data. Similarly, from page 826 of Amihud and Hurvich (2004), we get an estimated (normalized) β of approximately 0.17 for annual data and 0.02 for monthly data. The simulation results clearly demonstrate the advantage enjoyed by the REML estimate and also its bias-corrected version, in terms of both bias and standard deviation over the corresponding OLS counterparts. To test the robustness of the procedure to nonnormal thick-tailed errors, we also generated data from the same model and parameter configurations but letting (e_t, v_t) be independent t_5 errors. The results are presented in the second set of columns in Table 3 and it is seen that once again, $\hat{\beta}$ and $\hat{\beta}_c$ are consistently better than $\hat{\beta}_{OLS}$, and $\hat{\beta}_{OLS,c}$, respectively, in terms of both bias and standard deviation.

We next turn to studying the quality of the χ_1^2 approximation to the distribution of the RLRT for the preceding model and parameter configurations with both normal and t_5 innovations. The χ_1^2 approximation was assessed by two measures: (i) quantile-quantile (QQ) plots of the RLRT against the theoretical quantiles of a

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

χ^2_1 distribution and (ii) simulation sizes at the 5% and 1% levels. The QQ plots are shown in Figure 3 and the simulation sizes are reported in Table 4. (We present the QQ plots only for the normal innovations because the plots for t_5 errors are qualitatively similar in nature.) The RLRT is seen to be very well approximated by the χ^2_1 distribution according to each of the two measures. Comparing the results of Table 4 with the performance of the usual LRT shown in Table 2 demonstrates the significant advantage that the RLRT provides over the usual LRT. It is also worth stressing again that the performance of the RLRT is not affected by the bias of $\hat{\beta}$ at all.

We also carried out a simulation study to compare the size and power of our procedure with those of Jansson and Moreira (2006) and Campbell and Yogo (2006). We note that the RLRT procedure is very straighforward to use, whereas the Jansson–Moreira procedure is much more complex. The sample size was set to $n = 100, 200$, and 400, and $\alpha = 1 - cn^{-1}$ with $c = 0, 1, 5, 10$, and 20. The innovation variances were set as $\sigma_u^2 = \sigma_v^2 = 1$ and $\phi = \mathrm{Corr}(u_t, v_t) = -0.98$. The initial value X_0 was drawn from $N(0, \sigma_v^2)$, because the Jansson and Moreira (2006) and Campbell and Yogo (2006) procedures require the initial value to be $o_p(n^{1/2})$. As a result, the restricted likelihood used was of the form in (14). Following Jansson and Moreira (2006), the parameter β was defined as $\beta = n^{-1}b\sqrt{1-\phi^2}$, where $b = 0, 25$, and 50. Thus, $b = 0$ corresponds to the size simulations, whereas $b = 25, 50$ corresponds to the power of the procedure against local alternatives.

Table 5 shows the size of the Campbell and Yogo (2006) Bonferroni Q procedure for a variety of parameter values across 10,000 replications. It is seen that their procedure can produce significant size distortions even when local to unity (which is the framework for which their test is designed), with the tests being significantly oversized under the null hypothesis. Because these tests are based on a Bonferroni-style inequality, it is not clear as to how one should compute size-adjusted power for them, and as a result we do not provide a power comparison with the Campbell and Yogo (2006) procedure, simply noting that it can be signifi-

TABLE 4. Rejection rates, simulation mean and variance of RLRT (univariate regressor): $\beta = 0$, $\alpha = 1 - c/n$, $\phi = -0.98$, $\sigma_v^2 = \sigma_u^2 = 1$

	$c = 1$					$c = 5$				
		Gaussian		t_5			Gaussian		t_5	
n	α	5%	1%	5%	1%	α	5%	1%	5%	1%
50	0.98	0.0445	0.0084	0.0449	0.0092	0.90	0.0506	0.0100	0.0519	0.0131
100	0.99	0.0444	0.0095	0.0504	0.0101	0.95	0.0472	0.0093	0.0548	0.0123
200	0.995	0.0429	0.0093	0.0456	0.0104	0.975	0.0460	0.0103	0.0533	0.0134
400	0.9975	0.0443	0.0108	0.0473	0.0097	0.9875	0.0497	0.0103	0.0550	0.0136

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

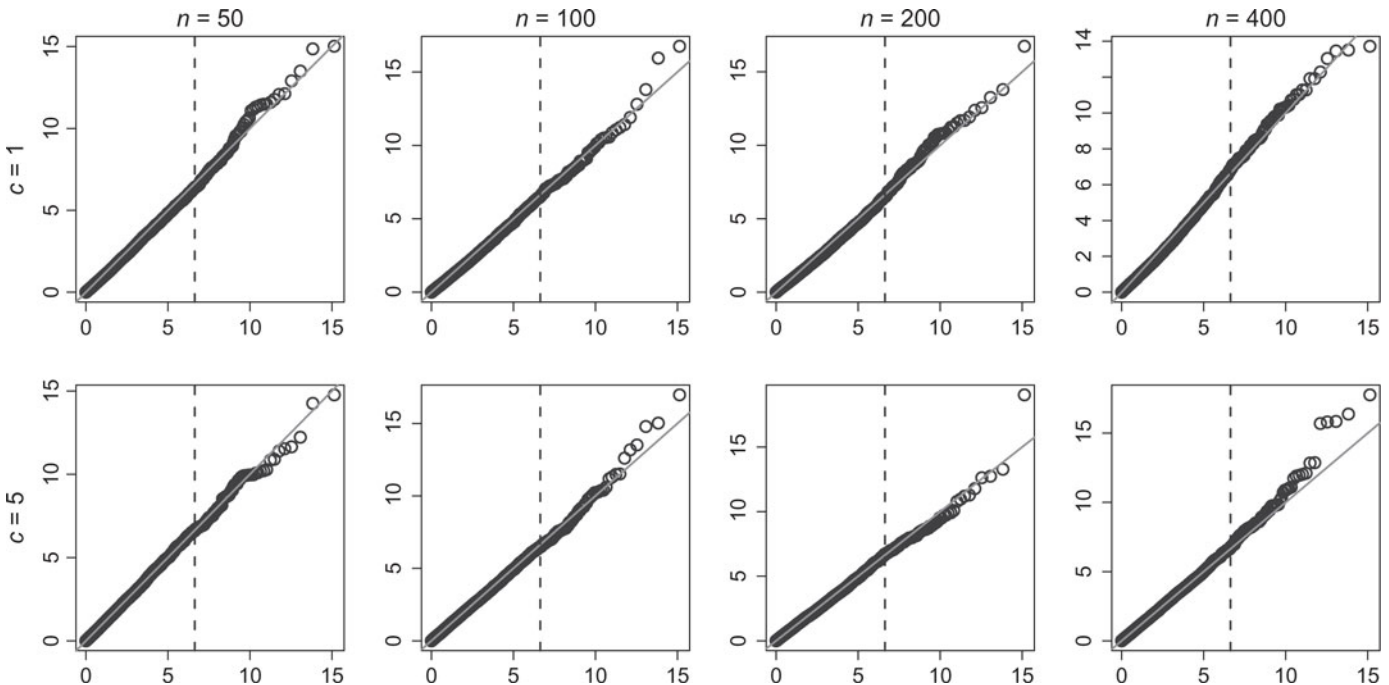


FIGURE 3. QQ plots of RLRT from simulations of 10,000 repetitions with $n = 50, 100, 200,$ and 400. The vertical dashed lines are the 99th percentile of χ^2_1. Data are generated from $Y_t = \beta X_{t-1} + u_t$, where $\beta = 0$, $X_t = \alpha X_{t-1} + v_t$ with $\alpha = 1 - c/n$ with $c = 1,\ 5$, and $\text{corr}(v_t, u_t) = -0.98$.

TABLE 5. Sizes of Campbell and Yogo procedure: $\beta = 0$, $\alpha = 1 - c/n$, $\phi = -0.98$, $\sigma_v^2 = \sigma_u^2 = 1$

	$c = 1$		$c = 5$		$c = 20$	
n	5% 2-side	2.5% 1-side	5% 2-side	2.5% 1-side	5% 2-side	2.5% 1-side
100	0.0647	0.0604	0.0724	0.0590	0.0999	0.0541
200	0.0572	0.0519	0.0664	0.0496	0.0758	0.0325
400	0.0490	0.0433	0.0586	0.0406	0.0670	0.0302

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

cantly oversized. It is somewhat puzzling that a test that is supposed to be based on a Bonferroni-style inequality yields sizes that are significantly larger than the declared nominal size. However, we point out that in the simulation results that Campbell and Yogo (2006) themselves present in their Table 3, the nominal 5% Bonferroni Q-test has rejection percentages as large as 0.117 and 0.09 in 10,000 replications, which is well above what is to be expected even after accounting for simulation error. One potential cause for this poor performance may be the fact that the Campbell–Yogo procedure mean corrects the predictor series by subtracting a generalized least squares (GLS) estimate of the mean that is computed under the assumption that $\alpha = 1 - 7n^{-1}$ (see Campbell and Yogo, 2005). If the true value of α deviates substantially from this assumption, the mean adjustment will be poor and may potentially degrade the size performance of the test.

The results of the comparison between the RLRT and the Jansson and Moreira (2006) procedures are reported in Table 6 for 5,000 replications. As is to be expected from the simulation reported in Table 4, the RLRT maintains its size very well at all the stationary parameter configurations, whereas the Jansson and Moreira (2006) procedure shows a little size distortion when $c = 20$. The RLRT is oversized when the predictor series is nonstationary and $\text{Corr}(u_t, v_t) \simeq -1$, which is not surprising because the chi-square limit distribution will no longer hold at the boundary value of $\alpha = 1$. As we see in Table 7 however, the RLRT is not oversized at the unit root when $\text{Corr}(u_t, v_t) = -0.5$, suggesting that the chi-square approximation is degraded at the boundary only when the innovation correlation is extremely high. Furthermore, as argued in the discussion after Theorem 3, we note again that the case of a nonstationary predictor series is not relevant in financial applications where the predictive regression model is most commonly used.

The power comparison between the two procedures in Table 6 shows that the RLRT provides uniformly higher power than the Jansson and Moreira (2006) procedure at all the alternatives considered, with significant power gains ranging from twice as much to almost seven times as much. These simulations provide strong evidence that the restricted likelihood procedure performs significantly better than

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

TABLE 6. Rejection rates for testing $H_0 : \beta = 0$: True $\beta = b/n\sqrt{1-\phi^2}$, $\phi = -0.98$, $\alpha = 1 - c/n$ (5,000 replications)

		$b = 0$		$b = 25$		$b = 50$	
c	n	RLRT	J&M	RLRT	J&M	RLRT	J&M
0	100	0.0852	0.0462	1.000	0.7718	1.0000	0.9852
	200	0.0880	0.0544	1.000	0.7846	1.0000	0.9940
	400	0.0860	0.0448	1.000	0.7906	1.0000	0.9992
1	100	0.0556	0.0506	0.9982	0.4796	1.0000	0.9990
	200	0.0530	0.0522	0.9984	0.4688	1.0000	0.9990
	400	0.0496	0.0500	0.9980	0.4680	1.0000	0.9996
5	100	0.0552	0.0492	0.3300	0.0782	1.0000	0.5454
	200	0.0500	0.0500	0.3386	0.0828	1.0000	0.5352
	400	0.0486	0.0568	0.3328	0.0886	0.9988	0.5124
10	100	0.0566	0.0560	0.1612	0.0724	0.6508	0.1110
	200	0.0488	0.0500	0.1598	0.0650	0.6932	0.1016
	400	0.0456	0.0570	0.1690	0.0728	0.7082	0.1044
20	100	0.0530	0.0610	0.1162	0.0646	0.3180	0.0784
	200	0.0496	0.0570	0.1072	0.0634	0.3066	0.0770
	400	0.0474	0.0622	0.1088	0.0684	0.3196	0.0804

TABLE 7. Rejection rates for testing $H_0 : \beta = 0$: True $\beta = b/n\sqrt{1-\phi^2}$, $\alpha = 1 - c/n$, $n = 1{,}000$, (500 replications)

	c	$b = 0$		$b = 5$		$b = 10$		$b = 15$	
ϕ	b	RLRT	J&M	RLRT	J&M	RLRT	J&M	RLRT	J&M
−0.5	0	0.058	0.054	0.046	0.058	0.042	0.042	0.048	0.066
	5	0.516	0.424	0.240	0.102	0.170	0.076	0.130	0.080
	10	0.898	0.826	0.644	0.338	0.450	0.180	0.340	0.154
	15	0.984	0.946	0.914	0.702	0.782	0.392	0.650	0.230
0.5	0	0.052	0.046	0.050	0.036	0.044	0.018	0.048	0.022
	5	0.368	0.412	0.195	0.132	0.146	0.054	0.110	0.040
	10	0.754	0.602	0.526	0.220	0.402	0.122	0.306	0.088
	15	0.922	0.720	0.780	0.274	0.634	0.196	0.546	0.082

the Jansson and Moreira procedure in terms of both size and power even when the predictor series is nearly integrated, while also yielding estimates with low (and estimable) bias. The RLRT also does not impose any restrictions on the specification of the initial value and extends easily to the multivariate regressor case.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

We also include a comparison between the RLRT and the Jansson and Moreira (2006) procedure using the parameter configuration in Table II on page 702 of Jansson and Moreira (2006). This configuration sets the innovation correlation at -0.5 and 0.5 and allows us to compare the RLRT against the various procedures reported in Table II in Jansson and Moreira (2006). The results of this comparison, based on $n = 1{,}000$ and 500 replications to be consistent with the design in Jansson and Moreira (2006), are reported in Table 7. We first note that the RLRT is now no longer oversized at the unit root, suggesting that the chi-square approximation works even at the boundary value $\alpha = 1$ if the innovation correlation is not very high. Furthermore, the RLRT maintains nominal size at all the configurations while once again almost uniformly dominating the Jansson and Moreira procedure in terms of power.

We finally generate data from a model in which the regressors are a bivariate AR(1) model. More specifically, we use the model $Y_t = \boldsymbol{\beta}'\mathbf{X}_{t-1} + u_t$ and $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{v}_t$, where $u_t = \boldsymbol{\phi}'\mathbf{v}_t + e_t$. The sample size was set at $n = 200$, the slope vector $\boldsymbol{\beta}$ was set to zero, and $\boldsymbol{\phi} = (-80, -80)$. Two configurations of the AR coefficient matrix $\mathbf{A}$ were considered, $\text{diag}\,(0.95, 0.8)$ and $\text{diag}\,(0.95, 0.95)$. The innovation matrix $\boldsymbol{\Sigma}_v$ was set to one of the following three configurations: (a) the identity matrix, (b) variances equal to 2 and correlation $\rho_v = 0.5$, (c) variances equal to 10 and correlation $\rho_v = 0.9$. In all cases, the innovation variance σ_e^2 was set to unity. This design (except for $\boldsymbol{\Sigma}_v = \mathbf{I}$) matches that used in Amihud and Hurvich (2004). The number of replications for each parameter configuration was 5,000 in the vector case. As before, the quality of the χ_2^2 approximation to the distribution of the RLRT was assessed via two measures (i) QQ plots of the RLRT against the theoretical quantiles of a χ_2^2 distribution (Figure 4) and (ii) simulation sizes at the 5% and 1% levels. The results are provided in Tables 8 and 9. As noted in Section 4.1, OLS is no longer asymptotically efficient if $\mathbf{A}$ is a diagonal matrix because the system is no longer a SUR. Hence, not only does REML afford

TABLE 8. Mean and variance of the $\hat{\beta}$ (bivariate regressor): $\beta = 0$, $\phi = c(-80, -80)$, $\sigma_e^2 = 1$

diag $\mathbf{A}$		(0.95, 0.80)				(0.95, 0.95)			
ρ_v		$\hat{\beta}_{1,OLS}$	$\hat{\beta}_{2,OLS}$	$\hat{\beta}_{1,REML}$	$\hat{\beta}_{2,REML}$	$\hat{\beta}_{1,OLS}$	$\hat{\beta}_{2,OLS}$	$\hat{\beta}_{1,REML}$	$\hat{\beta}_{2,REML}$
0	bias	1.8375	2.0377	0.7480	0.6828	2.1131	2.1399	0.7579	0.7653
	s.d.	3.3518	5.4121	2.3268	3.6798	3.5761	3.6521	2.3144	2.4042
0.5	bias	2.4394	1.2751	0.6125	0.7532	2.1007	2.1479	0.7012	0.7411
	s.d.	4.3870	7.1407	2.0053	3.2747	5.0234	5.1596	2.0516	2.1255
0.9	bias	3.9979	−1.2468	0.3587	0.5306	2.0745	2.1766	0.6291	0.6384
	s.d.	6.5038	10.1373	1.2865	2.1064	11.4054	11.5401	1.6685	1.6961

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

TABLE 9. Rejection rates, simulation mean and variance of the RLRT (bivariate regressor): $\beta = 0$, $\phi = c(-80, -80)$, $\sigma_e^2 = 1$

diag **A**	(0.95, 0.80)		(0.95, 0.95)	
ρ_v	5%	1%	5%	1%
0	0.0494	0.0106	0.0554	0.0124
0.5	0.0536	0.0112	0.0558	0.0132
0.9	0.0572	0.0122	0.0548	0.0104

Note: Tables 1–5, 8, and 9 are based on 10,000 replications.

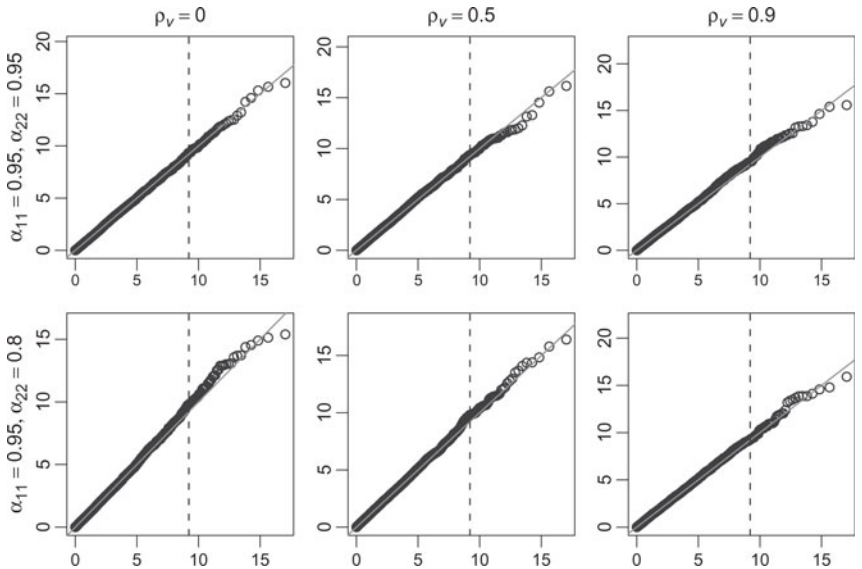


FIGURE 4. QQ plots of LRT from simulations of 5,000 repetitions with $n = 200$. The vertical dashed lines are the 99th percentile of χ_2^2. Data are generated from $Y_t = \boldsymbol{\beta}'\mathbf{X}_{t-1} + u_t$, where $\boldsymbol{\beta} = 0$, $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{v}_t$ with $\mathbf{A} = \text{diag}(\alpha_{11}, \alpha_{22})'$ and $\rho_v = 0, 0.5$, and 0.9.

a dramatic reduction in bias over OLS, it also provides a great reduction in the standard deviation, as can be seen in Table 8 with the reduction being as much as 80%. Furthermore, for the inference problem it is once again seen that the RLRT is very well approximated by the χ^2 distribution in all the cases we consider.

The overall conclusion to be had from the simulations is that the REML procedure yields point estimates that are much less biased than their OLS counterparts and also an RLRT that is very well behaved, even when the regressors are close to being integrated.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

NOTES

1. There are several formal measures of curvature based on higher order derivatives of the likelihood in the literature. See, for example, Kass and Slate (1994). We will not define any such measure explicitly because that is not the focus of our work here.

2. van Giersbergen (2006) provides the expected value of the LRT, whereas the combined results from the remaining references show that the leading term in the expansion (4) is half that expected value.

3. All the expansions that we provide in this paper are formal. Though we do not attempt to do so here, it may be possible to show that these expansions are valid by using the work of Chandra and Ghosh (1979).

4. Interestingly and somewhat surprisingly, van Garderen (1999) found that if the innovation variance σ_v^2 is known the Efron curvature of the model is $2n^{-1} + O(n^{-2})$, which, though still small, is larger than when σ_v^2 is unknown, and he provided a geometrical explanation for this phenomenon. Correspondingly, results in van Giersbergen (2006) imply that the coefficient of the leading term in (5) increases from $-0.25n^{-1}$ to $-1n^{-1}$ when σ_v^2 is known, indicating that the LRT is better approximated by the limiting chi-square distribution when the innovation variance is unknown than when it is known, though, of course, both approximations are very good by themselves.

5. Based on monthly data with a sample size of $n = 379$, Amihud and Hurvich (2004, Table 3) find the empirical estimate of ϕ to be approximately -92. The corresponding estimate of β is approximately 2 and that of α is 0.99. In their other empirical study based on annual data with $n = 45$, they find that $\phi \simeq -95$, $\beta \simeq 20$, and $\alpha \simeq 0.9$.

6. This finding is very much in keeping with the results of van Giersbergen (2006) for the LRT in the univariate AR(1) model that we described in note 2.

REFERENCES

Amihud, Y. & C. Hurvich (2004) Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* 39, 813–841.

Baker, M., R. Taliaferro, & J. Wurgler (2006) Predicting returns with managerial decision variables: Is there a small sample bias? *Journal of Finance* 61, 1711–1730.

Barndorff-Nielsen, O.E. & P. Hall (1988) On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* 75, 374–378.

Bartlett, M. (1953) Approximate confidence intervals III. More than one unknown parameter. *Biometrika* 40, 306–317.

Campbell, J. & M. Yogo (2005) Implementing the Econometric Methods in "Efficient Tests of Stock Return Predictability." Available at http://finance.wharton.upenn.edu/~yogo/.

Campbell, J. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.

Chandra, T.K. & J.K. Ghosh (1979) Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhya, Series A*. 41, 22–47.

Cheang, W.K. & G. Reinsel (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. & R. Deo (2006) A Smooth Transition to the Unit Root Distribution via the Chi-Square Distribution with Interval Estimation for Nearly Integrated Autoregressive Processes. Working paper, Texas A & M and New York University.

Chen, W. & R. Deo (2007) The Chi-Square Approximation of the Restricted Likelihood Ratio Test for the Sum of Autoregressive Coefficients with Interval Estimation. Working paper, Texas A & M and New York University.

Chesher, A. & R. Smith (1995) Bartlett corrections to likelihood ratio tests. *Biometrika* 82, 433–436.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Cordeiro, G. (1987) On the corrections to the likelihood ratio statistics. *Biometrika* 74, 265–274.
Cordeiro, G., D. Botter, & S. L. Ferrari (1994) Nonnull asymptotic distributions of three classic criteria in generalised linear models. *Biometrika* 81, 709–720.
Cribari-Neto, F. & G. Cordeiro (1996) On Bartlett and Bartlett-type corrections. *Econometric Reviews* 15, 339–367.
Dufour, J.-M. & M.L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or non-stationary AR(1) errors. *Journal of Econometrics* 47, 115–143.
Efron, B. (1975) Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics* 3, 1189–1242.
Fisher, R. (1973) *Statistical Methods and Scientific Inference.* Haffner.
Francke, M. & A. de Vos (2006) Marginal likelihood and unit roots. *Journal of Econometrics* 137, 708–728.
Giraitis, L. & P.C.B. Phillips (2006) Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis* 27, 51–60.
Harris, P. & H.W. Peers (1980) The local power of the efficient scores test statistic. *Biometrika* 67, 525–529.
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.
Hayakawa, T. (1975) The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62, 451–460.
Hayakawa, T. (1977) The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* 29, 359–378.
Hayakawa, T. (1987) Correction. *Annals of the Institute of Statistical Mathematics* 39, 681.
Jansson, M. & M. Moreira (2006) Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.
Kalbfleisch, J. & D. Sprott (1970) Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society, Series B* 32, 175–194.
Kang, W., D. Shin, & Y. Lee (2003) Biases of the restricted maximum likelihood estimators for ARMA processes with polynomial time trend. *Journal of Statistical Planning and Inference* 116, 163–176.
Kass, R.E. & E.H. Slate (1994) Some diagnostics of maximum likelihood and posterior nonnormality. *Annals of Statistics* 22, 668–695.
Lewellen, J. (2004) Predicting returns with financial ratios. *Journal of Financial Economics* 74, 209–235.
Lehmann, E. & J. Romano (2005) *Testing Statistical Hypotheses, 3rd ed.* Springer-Verlag.
Marriott, F. & J. Pope (1954) Bias in the estimation of autocorrelations. *Biometrika* 41, 390–402.
McCullagh, P. & D.R. Cox (1986) Invariants and likelihood ratio statistics. *Annals of Statistics* 14, 1419–1430.
Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
Rahman, S. & M. King (1997) Marginal-likelihood score-based tests of regression disturbances in the presence of nuisance parameters. *Journal of Econometrics* 82, 81–106.
Smyth, G. & Verbyla, A. (1996) A conditional likelihood approach to residual maximum likelihood estimation in generalized linear models. *Journal of the Royal Statistical Society, Series B* 58, 565–572.
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*, *Series B* 37, 259-270.
Sprott, D. (1980) Maximum likelihood in small samples: Estimation in the presence of nuisance parameters. *Biometrika* 67, 515–523.

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Sprott, D. (1990) Inferential estimation, likelihood, and linear pivotals. *Canadian Journal of Statistics* 18, 1–10.

Sprott, D. & R. Viveros-Aguilera (1984) The interpretation of maximum-likelihood estimation. *Canadian Journal of Statistics* 12, 27–38.

Stambaugh, R. (1999) Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Tunnicliffe Wilson, G. (1989) On the use of marginal likelihood in time series model estimation. *Journal of the Royal Statistical Society, Series B* 51, 15–27.

van Garderen, K. (1999) Exact geometry of autoregressive models. *Journal of Time Series Analysis* 20, 1–21.

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

APPENDIX

Proof of Theorem 2. As noted at the start of Section 2, the REML likelihood corresponds to the likelihood of $\mathbf{T}(\mathbf{Y}', \mathbf{X}')'$, where $\mathbf{T}$ is any full row rank matrix such that $\mathbf{T1} = \mathbf{0}$. We will obtain this likelihood by choosing $\mathbf{T}$ to have the form

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{bmatrix},$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ are full row rank matrices of dimension $(n-1) \times n$ and $n \times (n+1)$, respectively, satisfying $\mathbf{T}_1\mathbf{1} = \mathbf{0}$, $\mathbf{T}_2\mathbf{1} = \mathbf{0}$, $\mathbf{T}_1\mathbf{T}_1' = \mathbf{I}$, and $\mathbf{T}_2\mathbf{T}_2' = \mathbf{I}$, and using the fact that

$$L\left(\mathbf{T}(\mathbf{Y}', \mathbf{X}')'\right) = L\left(\mathbf{T}_1\mathbf{Y} \mid \mathbf{T}_2\mathbf{X}\right) L\left(\mathbf{T}_2\mathbf{X}\right).$$

We first obtain $L\left(\mathbf{T}_1\mathbf{Y} \mid \mathbf{T}_2\mathbf{X}\right)$. Because $u_t = \phi v_t + e_t$, where $\phi = \sigma_{uv}/\sigma_v^2$ and $e_t \sim N\left(0, \sigma_e^2\right)$ is a series independent of $\{v_t\}$, we get

$$\begin{aligned} Y_t &= \eta + \beta X_{t-1} + \phi\left(X_t - \mu - \alpha X_{t-1}\right) + e_t \qquad \textbf{(A.1)} \\ &= \eta + \phi X_t + (\beta - \phi\alpha) X_{t-1} + e_t. \end{aligned}$$

Let $\tilde{\mathbf{Y}} = \mathbf{T}_1\mathbf{Y}$, $\tilde{\mathbf{X}}_1 = \mathbf{T}_1\mathbf{X}_1$, and $\tilde{\mathbf{X}}_0 = \mathbf{T}_1\mathbf{X}_0$. From (A.1) it then follows that

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\theta} + \tilde{\mathbf{e}}, \qquad \textbf{(A.2)}$$

where $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1,\ \tilde{\mathbf{X}}_0]$, $\tilde{\mathbf{e}} = \mathbf{T}_1\mathbf{e}$, $\mathbf{e} = (e_1, \ldots, e_n)'$, and $\boldsymbol{\theta} = (\phi,\ \beta - \phi\alpha)'$. Because X_t is a function only of $\{v_t, v_{t-1}, \ldots\}$, the series $\{X_t\}$ is independent of $\{e_t\}$. Furthermore, knowledge of $\mathbf{T}_2\mathbf{X}$, where $\mathbf{T}_2$ is any full row rank matrix such that $\mathbf{T}_2\mathbf{1} = \mathbf{0}$, implies knowledge of $\tilde{\mathbf{X}}$. Hence, from (A.2) the conditional distribution of $\tilde{\mathbf{Y}}$ given $\mathbf{T}_2\mathbf{X}$ is $N(\tilde{\mathbf{X}}\boldsymbol{\theta}, \sigma_e^2\mathbf{I})$, because $\mathbf{T}_1\mathbf{T}_1' = \mathbf{I}$. It follows that the conditional log-likelihood of $(\tilde{\mathbf{Y}} \mid \mathbf{T}_2\mathbf{X})$ up to an additive constant is given by

$$l_1\left(\tilde{\mathbf{Y}} \mid \mathbf{T}_2\mathbf{X}, \boldsymbol{\theta}, \sigma_e^2\right) = -\left(\frac{n-1}{2}\right)\log\sigma_e^2 - \frac{1}{2\sigma_e^2}\left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right)^*\left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right).$$

Note, however, that $\tilde{\mathbf{X}}\boldsymbol{\theta} = \mathbf{T}_1\left[\mathbf{X}_1,\ \mathbf{X}_0\right]\boldsymbol{\theta}$. Thus,

$$\left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right)^*\left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right) = \left(\mathbf{Y} - \left[\mathbf{X}_1,\ \mathbf{X}_0\right]\boldsymbol{\theta}\right)'\mathbf{T}_1^*\mathbf{T}_1\left(\mathbf{Y} - \left[\mathbf{X}_1,\ \mathbf{X}_0\right]\boldsymbol{\theta}\right).$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Because the matrix $\mathbf{T}_1$ when augmented by the row $n^{-1/2}\mathbf{1}'$ is an orthogonal matrix, it follows that $\mathbf{T}_1^*\mathbf{T}_1 = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$ and hence

$$\begin{aligned}\left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right)^* \left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\right) &= \left(\mathbf{Y} - [\mathbf{X}_1,\ \mathbf{X}_0]\,\boldsymbol{\theta}\right)' \left[\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'\right] \left(\mathbf{Y} - [\mathbf{X}_1,\ \mathbf{X}_0]\,\boldsymbol{\theta}\right) \\ &= S(\phi, \beta, \alpha).\end{aligned}$$

Thus, we get

$$l_1\left(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\theta}, \sigma_e^2\right) = -\left(\frac{n-1}{2}\right)\log\sigma_e^2 - \frac{1}{2\sigma_e^2}S(\phi, \beta, \alpha). \tag{A.3}$$

The log-likelihood of $\mathbf{T}_2\mathbf{X}$ is obtained from Harville (1974) and up to an additive constant is given by

$$\begin{aligned}l_2\left(\mathbf{T}_2\mathbf{X},\ \alpha, \sigma_v^2\right) = &-\left(\frac{n}{2}\right)\log\sigma_v^2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log\left|\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\right| \\ &-\frac{1}{2\sigma_v^2}\left(\mathbf{X} - \mathbf{1}\hat{\tau}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{X} - \mathbf{1}\hat{\tau}\right),\end{aligned} \tag{A.4}$$

where $\hat{\tau} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ and $\boldsymbol{\Sigma} = \sigma_v^{-2}\mathrm{Var}(\mathbf{X})$. From (A.3) and (A.4), we obtain the log-likelihood of $(\mathbf{T}_1\mathbf{Y},\ \mathbf{T}_2\mathbf{X})$ up to an additive constant to be

$$\begin{aligned}L\left(\mathbf{T}_1\mathbf{Y},\ \mathbf{T}_2\mathbf{X},\ \alpha, \beta, \phi, \sigma_v^2, \sigma_e^2\right) = &-\left(\frac{n-1}{2}\right)\log\sigma_e^2 - \frac{1}{2\sigma_e^2}S(\phi, \beta, \alpha) \\ &-\left(\frac{n}{2}\right)\log\sigma_v^2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log\left|\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\right| \\ &-\frac{1}{2\sigma_v^2}Q(\alpha).\end{aligned} \tag{A.5}$$

The final form, as stated in (13), is obtained by algebraic simplification, using the fact that $\boldsymbol{\Sigma}^{-1} = \mathbf{B}'\mathbf{B}$, where $\mathbf{B}$ is the $(n+1)\times(n+1)$ matrix given by

$$\mathbf{B} = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & -\alpha & 1 \end{bmatrix} \tag{A.6}$$

and noting that

$$\hat{\tau} = \frac{(1-\alpha)\sum_{i=0}^{n-2} X_i + X_{-1} + X_{n-1}}{(n-1)(1-\alpha)+2}.$$

To obtain the REML estimates of $(\alpha, \beta, \phi, \sigma_e^2, \sigma_v^2)$, it helps to consider the reparametrized set of parameters $(\alpha, \gamma, \phi, \sigma_e^2, \sigma_v^2)$, where $\gamma = \beta - \phi\alpha$. It is then immediately obvious from

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

inspecting (A.5) that the REML estimates of (α, σ_v^2) are obtained by simply maximizing just

$$l_2\left(\mathbf{T}_2\mathbf{X},\ \alpha, \sigma_v^2\right) = -\left(\frac{n}{2}\right)\log\sigma_v^2 - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\log\left|\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}\right| - \frac{1}{2\sigma_v^2}\left(\mathbf{X}-\mathbf{1}\hat{\tau}\right)'\mathbf{\Sigma}^{-1}\left(\mathbf{X}-\mathbf{1}\hat{\tau}\right).$$

In other words, the REML estimates of (α, σ_v^2) are just those estimates that would have been obtained by maximizing the REML likelihood function of $(X_1, \ldots, X_n)$. It thus follows immediately from Cheang and Reinsel (2000) that the bias of $\hat{\alpha}$ is

$$\mathrm{E}\left(\hat{\alpha}-\alpha\right) = \frac{2\alpha}{n-1} + o\left(n^{-1}\right).$$

The REML estimates of (ϕ, γ) are obtained as the least squares estimates

$$\left(\hat{\phi}, \hat{\gamma}\right)' = \left(\mathbf{X}_c'\mathbf{X}_c\right)^{-1}\mathbf{X}_c'\mathbf{Y}_c,$$

which then yields the REML estimate $\hat{\beta} = \hat{\gamma} + \hat{\phi}\hat{\alpha}$. Because

$$\begin{pmatrix}\hat{\phi}\\ \hat{\beta}\end{pmatrix} = \begin{bmatrix}1 & 0\\ \hat{\alpha} & 1\end{bmatrix}\begin{pmatrix}\hat{\phi}\\ \hat{\gamma}\end{pmatrix},$$

it follows that the REML estimates of the original parameters (ϕ, β) can also be obtained in a direct regression of $\mathbf{Y}_c$ on $(\mathbf{X}_1 - \mathbf{1}\bar{X}_1 - \hat{\alpha}(\mathbf{X}_0 - \mathbf{1}\bar{X}_0),\ \mathbf{X}_0 - \mathbf{1}\bar{X}_0)$. Thus, $\hat{\beta}$ is identical to the augmented regression method (ARM) estimate considered by Amihud and Hurvich (2004) using the REML estimate $\hat{\alpha}$ as a proxy for α. Hence, the bias of $\hat{\beta}$ can be obtained from Theorem 2 of Amihud and Hurvich (2004) and is

$$\mathrm{E}(\hat{\beta}-\beta) = \phi\mathrm{E}(\hat{\alpha}-\alpha) = -\frac{2\alpha\phi}{n-1} + o(n^{-1}).$$

Finally, Lemma 1 of Amihud and Hurvich (2004) implies that $\mathrm{E}(\hat{\phi}) = \phi$. ■

Proof of Theorem 3. Because the LRT is invariant to reparametrization we choose to work with the reparametrization $\boldsymbol{\lambda} = (\beta, \alpha, \phi, \tau_v, \tau_e) \equiv (\beta, \boldsymbol{\lambda}_2)$, where $\tau_v = \sigma_v^{-2}$ and $\tau_e = \sigma_e^{-2}$ because this greatly reduces the burden of our computations. For the REML log-likelihood given in (13), we will denote expectations of the log-likelihood derivatives as

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

Letting $\delta = \tau_e/\tau_v$, it is easily seen that the information matrix $\mathbf{K} = ((-\kappa_{rs}))$ is given by

$$\mathbf{K} = \begin{bmatrix}\mathbf{K}_{\beta,\alpha} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_{\phi,\tau_v,\tau_e}\end{bmatrix} + O\left(n^{-1}\right), \tag{A.7}$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

where

$$\mathbf{K}_{\beta,\alpha} = \frac{\delta}{1-\alpha^2}\begin{bmatrix} 1 & -\phi \\ -\phi & \left(\phi^2+\delta^{-1}\right) \end{bmatrix}$$

and

$$\mathbf{K}_{\phi,\tau_v,\tau_e} = \text{diag}\left\{\delta, \frac{1}{2\tau_v^2}, \frac{1}{2\tau_e^2}\right\}.$$

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 4×4 submatrix of $\mathbf{K}$.

Theorem 1 of Hayakawa (1977) provides an expression for the formal expansion of the distribution of R_T. This expression is in notation that is not easy to use and also contains a term, A_2, that was subsequently shown (Chesher and Smith, 1995) to be identically zero in regular models. As a result, we use the notation of Cordeiro (1987) and from the discussion on page 342 of Cribari-Neto and Cordeiro (1996) obtain the expansion

$$P\left(R_T \le x\right) = P\left(\chi_1^2 \le x\right) + An^{-1}\left[P\left(\chi_3^2 \le x\right) - P\left(\chi_1^2 \le x\right)\right] + O\left(n^{-2}\right), \qquad \textbf{(A.8)}$$

where

$$A = \frac{1}{2}\left\{\sum_{\boldsymbol{\lambda}}\left(l_{rstu} - l_{rstuvw}\right) - \sum_{\boldsymbol{\lambda}_2}\left(\tilde{l}_{rstu} - \tilde{l}_{rstuvw}\right)\right\}, \qquad \textbf{(A.9)}$$

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

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

and

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

Exploiting the near diagonal $\left(\text{up to } O\left(n^{-1}\right)\right)$ structure of $\mathbf{K}$, we can simplify the term $\sum_{\boldsymbol{\lambda}}\left(l_{rstu} - l_{rstuvw}\right)$ in A as

$$\sum_{\boldsymbol{\lambda}}\left(l_{rstu} - l_{rstuvw}\right) = \sum_{(\beta,\alpha)}\left(l_{rstu} - l_{rstuvw}\right) + \sum_{(\boldsymbol{\phi},\tau_v,\tau_e)}\left(l_{rstu} - l_{rstuvw}\right) + \sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))}\left(l_{rstu} - l_{rstuvw}\right) + O\left(n^{-1}\right), \qquad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

where $\sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))}$ denotes that at least one index in the summand must come from (β,α) and at least one index from (ϕ,τ_v,τ_e). By the same logic, we can simplify the term $\sum_{\boldsymbol{\lambda}_2}(\tilde{l}_{rstu}-\tilde{l}_{rstuvw})$ in A as

$$\begin{aligned}\sum_{\boldsymbol{\lambda}_2}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right) &= \sum_{\boldsymbol{\alpha}}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)+\sum_{(\phi,\tau_v,\tau_e)}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)\\ &\quad+\sum_{(\alpha,(\phi,\tau_v,\tau_e))}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)+O\left(n^{-1}\right)\\ &= \sum_{\boldsymbol{\alpha}}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)+\sum_{(\phi,\tau_v,\tau_e)}\left(l_{rstu}-l_{rstuvw}\right)\\ &\quad+\sum_{(\alpha,(\phi,\tau_v,\tau_e))}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)+O\left(n^{-1}\right), \qquad \textbf{(A.11)}\end{aligned}$$

where the last step in (A.11) follows from the fact that the entries of $\mathbf{K}$ for (ϕ,τ_v,τ_e) are diagonal up to $O(n^{-1})$. It thus follows from (A.9)–(A.11) that

$$\begin{aligned}A &= \frac{1}{2}\left\{\sum_{(\beta,\alpha)}\left(l_{rstu}-l_{rstuvw}\right)-\sum_{\boldsymbol{\alpha}}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)\right\}\\ &\quad+\frac{1}{2}\left\{\sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))}\left(l_{rstu}-l_{rstuvw}\right)-\sum_{(\alpha,(\phi,\tau_v,\tau_e))}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)\right\}+O\left(n^{-1}\right)\\ &\equiv A_{(\beta,\alpha)}+C_{\boldsymbol{\lambda}}+O\left(n^{-1}\right). \qquad \textbf{(A.12)}\end{aligned}$$

It is obvious from the structure of $A_{(\beta,\alpha)}$ in (A.12) that $A_{(\beta,\alpha)}$ would be the leading remainder term in the expansion of the distribution of R_T of the form in (A.8) if (ϕ,τ_v,τ_e) were known and α were the only nuisance parameter. The term $C_{\boldsymbol{\lambda}}$, thus, is a measure of the extent to which lack of knowledge of the parameters (ϕ,τ_v,τ_e) affects the finite-sample distribution of R_T. We now compute the two terms $A_{(\beta,\alpha)}$ and $C_{\boldsymbol{\lambda}}$, beginning with $A_{(\beta,\alpha)}$.

Though (β,α) are not orthogonal, the computation of $\sum_{(\beta,\alpha)}(l_{rstu}-l_{rstuvw})$ can be simplified by working with the transformed parameters (γ,α), where $\gamma=\beta-\phi\alpha$, because γ is orthogonal to α. (Note that as stated previously, when computing $A_{(\beta,\alpha)}$ the remaining parameters (ϕ,τ_v,τ_e) are fixed.) Because (γ,α) is an affine transformation of (β,α), we can exploit the fact that the first term of $A_{(\beta,\alpha)}$ is invariant under such transformations (Hayakawa, 1977, p. 371) and thus get

$$\frac{1}{2}\sum_{(\beta,\alpha)}\left(l_{rstu}-l_{rstuvw}\right)=\frac{1}{2}\sum_{(\gamma,\alpha)}\left(l_{rstu,p}-l_{rstuvw,p}\right), \qquad \textbf{(A.13)}$$

where the extra subscript p will mean that the computation is being carried out for the re-parametrized form of (13) with $\gamma=\beta-\phi\alpha$. The right-hand side of (A.13) is much simpler to compute because of the fact that $\kappa_{\gamma\alpha,p}=\kappa^{(\gamma)}_{\gamma\gamma,p}=\kappa^{(\gamma\gamma)}_{\gamma\gamma,p}=0$ and all the terms $\kappa_{rst,p}$ and $\kappa_{rstu,p}$ are at most $O(n^{-1})$, for all permutations of the subscripts. Because $\kappa_{\gamma\gamma,p}=-\delta(1-\alpha^2)^{-1}$ and $\kappa_{\alpha\alpha,p}=-(1-\alpha^2)^{-1}$, it follows that

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

$$\frac{1}{2}\sum_{(\gamma,\alpha)} l_{rstu,p} = \frac{1}{2}\kappa_p^{\alpha\alpha}\kappa_p^{\alpha\alpha}\left(\kappa_{\alpha\alpha,p}^{(\alpha\alpha)}\right) + O\left(n^{-1}\right) = -\left(1+3\alpha^2\right)\left(1-\alpha^2\right)^{-1} + O\left(n^{-1}\right) \tag{A.14}$$

and

$$\frac{1}{2}\sum_{(\gamma,\alpha)} l_{rstuvw,p} = \frac{1}{2}\kappa_p^{\alpha\alpha}\kappa_p^{\alpha\alpha}\kappa_p^{\alpha\alpha}2\left(\kappa_{\alpha\alpha}^{(\alpha)}\right)^2 + O\left(n^{-1}\right) = -4\alpha^2\left(1-\alpha^2\right)^{-1} + O\left(n^{-1}\right). \tag{A.15}$$

From (A.13)–(A.15), we get

$$\frac{1}{2}\sum_{(\beta,\alpha)} \left(l_{rstu} - l_{rstuvw}\right) = -1 + O\left(n^{-1}\right). \tag{A.16}$$

The second term in $A_{(\beta,\alpha)}$ is not invariant under affine transformations, and we revert to the original log-likelihood (13) to compute it. Noting that $\kappa_{\alpha\alpha} = -\left(\phi^2\delta+1\right)\left(1-\alpha^2\right)^{-1}$ and $\tilde{\kappa}^{\alpha\alpha} = -\left(\phi^2\delta+1\right)^{-1}\left(1-\alpha^2\right)$, we have

$$\begin{aligned}\frac{1}{2}\sum_{\boldsymbol{\alpha}}\left(\tilde{l}_{rstu} - \tilde{l}_{rstuvw}\right) &= \frac{1}{2}\tilde{\kappa}^{\alpha\alpha}\tilde{\kappa}^{\alpha\alpha}\left(\kappa_{\alpha\alpha}^{(\alpha\alpha)}\right) - \frac{1}{2}\tilde{\kappa}^{\alpha\alpha}\tilde{\kappa}^{\alpha\alpha}\tilde{\kappa}^{\alpha\alpha}2\left(\kappa_{\alpha\alpha}^{(\alpha)}\right)^2 + O\left(n^{-1}\right) \\ &= -\frac{1}{\phi^2\delta+1} + O\left(n^{-1}\right).\end{aligned} \tag{A.17}$$

From (A.16) and (A.17) we conclude that

$$A_{(\beta,\alpha)} = -\frac{\phi^2\delta}{\phi^2\delta+1} + O\left(n^{-1}\right) = -\rho^2 + O\left(n^{-1}\right). \tag{A.18}$$

We now turn our attention to computing the second term, $C_{\boldsymbol{\lambda}}$, in (A.12). We first note from the near diagonal structure of $\mathbf{K}$ that

$$\begin{aligned}\frac{1}{2}\sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))}\left(l_{rstu} - l_{rstuvw}\right) &= \frac{1}{2}\sum_{(\beta,(\phi,\tau_v,\tau_e))} l_{rrtt} + \frac{1}{2}\sum_{(\alpha,(\phi,\tau_v,\tau_e))} l_{rrtt} \\ &\quad -\frac{1}{2}\sum_{(\beta,(\phi,\tau_v,\tau_e))} l_{rrttvv} - \frac{1}{2}\sum_{(a,(\phi,\tau_v,\tau_e))} l_{rrttvv} \\ &\quad -\frac{1}{2}\sum_{(\beta,\alpha,(\phi,\tau_v,\tau_e))} l_{rstuvw}.\end{aligned} \tag{A.19}$$

All of the terms in (A.19) are now computed. The details are not provided here, both to save space and because the computation does not afford any special insight into the problem. The detailed calculations, however, are available from the authors. The terms in $\sum_{(\alpha,(\phi,\tau_v,\tau_e))}\left(\tilde{l}_{rstu} - \tilde{l}_{rstuvw}\right)$ can be decomposed in a manner similar to that in (A.19), except that in this case there are no terms in β. When all these terms are put together, one gets

$$C_{\boldsymbol{\lambda}} = \frac{3}{4} + O\left(n^{-1}\right). \tag{A.20}$$

The theorem now follows from (A.8), (A.12), (A.18), and (A.20). ■

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press

Proof of Theorem 4. The expansion of the distribution of the LRT and the Wald test under local Pitman alternatives is given in Hayakawa (1975), and that of the distribution of the Rao score test is given by Harris and Peers (1980). These results are consolidated using simpler notation in Cordeiro, Botter, and Ferrari (1994), and we follow the notation used in their work. To obtain the results of Theorem 4, we calculate the quantities in equations (1)–(5b) on page 711 of Cordeiro et al. (1994). Letting $\lambda = (\beta, \alpha, \phi, \tau_v, \tau_e)$, where $\tau_v = \sigma_v^{-2}$ and $\tau_e = \sigma_e^{-2}$, we note that the quantities $\kappa_{ij} = n^{-1}\mathrm{E}(\partial^2 L/\partial\lambda_i\partial\lambda_j)$ have been already obtained in (A.7). To obtain the quantities $\kappa_{i,jk} = n^{-1}\mathrm{E}[(\partial L/\partial\lambda_i)(\partial^2 L/\partial\lambda_j\partial\lambda_k)]$ and $\kappa_{i,j,k} = n^{-1}\mathrm{E}[(\partial L/\partial\lambda_i)(\partial L/\partial\lambda_j)(\partial L/\partial\lambda_k)]$, we exploit the Bartlett identities (Bartlett, 1953) to obtain

$$\kappa_{i,j,k} = 2\kappa_{ijk} - \frac{\partial\kappa_{jk}}{\partial\lambda_i} - \frac{\partial\kappa_{ik}}{\partial\lambda_j} - \frac{\partial\kappa_{ij}}{\partial\lambda_k}$$

and

$$\kappa_{i,jk} = -\kappa_{ijk} + \frac{\partial\kappa_{jk}}{\partial\lambda_i}.$$

The quantities $\partial\kappa_{jk}/\partial\lambda_i$ are then calculated using (A.7). Using the same notation (note, however, that we define all our cumulants κ to be $O(1)$ whereas Cordeiro et al. (1994) define them to be $O(n)$) as in equations (3a)–(5b) of Cordeiro et al. (1994), we get

$$b_{10} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{11} = n^{-1/2}C_1 + O\left(n^{-1}\right),$$

$$b_{12} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{13} = 0,$$

$$b_{20} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{21} = n^{-1/2}C_1 + O\left(n^{-1}\right),$$

$$b_{22} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{23} = O\left(n^{-1}\right),$$

and

$$b_{30} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{31} = n^{-1/2}(C_1 + C_2) + O\left(n^{-1}\right),$$

$$b_{32} = -n^{-1/2}C_1 + O\left(n^{-1}\right), \qquad b_{33} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right).$$

The result now follows from equation (2) of Cordeiro et al. (1994). ■

https://doi.org/10.1017/S0266466608090464 Published online by Cambridge University Press