Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions using the Restricted Likelihood

Willa W. Chen∗ Rohit S. Deo †

April 21, 2007

Abstract: The restricted likelihood, which has small curvature, is derived for the bivariate predictive regression problem. The bias of the Restricted Maximum Likelihood (REML) estimates is shown to be approximately 50% less than that of the 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 $(3/4 - \rho^2) n^{-1} (G_3 (\cdot) - G_1 (\cdot)) + O (n^{-2})$, where $|\rho| < 1$ is the correlation of the innovation series and $G_s (\cdot)$ is the c.d.f. of a $\chi^2_s$ random variable. This very small error, free of the AR parameter, implies that the RLRT for predictability has very good size properties even when the regressor is nearly integrated, a fact borne out by our simulation study. The Bartlett corrected RLRT achieves an $O (n^{-2})$ error. The power of the RLRT under a sequence of local Pitman alternatives is obtained and shown to be always identical to that of the Wald test and higher than that of the Rao score test for empirically relevant regions. Some extensions to the case of vector AR(1) regressors and more general univariate regressors are provided. The RLRT is found to work very well in simulations and to be robust to non-normal errors.

Keywords: Bartlett correction, likelihood ratio test, curvature

∗ Department of Statistics, Texas A&M University, College Station, Texas 77843, USA
† New York University, 44 W. 4’th Street, New York NY 10012, USA; E-mail: rdeo@stern.nyu.edu
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:

\[
Y_t = c + \beta X_{t-1} + u_t, \quad (1)
\]

\[
X_t = \mu + \alpha X_{t-1} + v_t, \quad (2)
\]

where \( |\alpha| < 1, u_t = \phi v_t + e_t, (e_t, v_t) \sim N(0, \text{diag}(\sigma_e^2, \sigma_v^2)) \) are an 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 parameter \( \alpha \) lying close to the unit root. It is well known (Stambaugh, 1999) that the standard ordinary least squares (OLS) estimate of \( \beta \) is biased when the errors \( \{u_t, v_t\} \) are contemporaneously correlated, with the amount of bias increasing as \( \alpha \) 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 \( \beta \) that allows the researcher to compute a bias-corrected OLS estimate as well as a \( t \)-statistic based on the 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, even providing an example to make this point. Instead, following Fisher (1973), Sprott argued in a series of 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. In brief, one of Sprott’s (1975, 1980) key ideas can be summarised roughly as follows:

(i) There may exist some \( 1 - 1 \) transformation of the parameters such that the likelihood ratio
for the transformed parameters is approximately quadratic in the pivotal statistic based on that transformed parameter.

(ii) The conditions that allow for this quadratic approximation to be adequate also improve the normal approximation to the distribution of the pivotal statistic.

It follows from these two observations that since the likelihood is invariant under $1 - 1$ re-parametrisation, the likelihood ratio will be well approximated by a chi-square variable in finite samples as long as some parametrisation satisfying (i) exists, even if one does not know what that parametrisation is. Sprott (1973, 1975, 1980) showed that such a parametrisation 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 likelihood ratio test (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.

A curvature related approach to tackling the predictive regression problem for the model in (1) and (2) was taken by Jansson and Moreira (2006), assuming 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 $0$. Under these assumptions, Jansson and Moreira (2006) noted that though the joint likelihood of $(y_t - y_{t-1}, x_t)$ was curved exponential, they could remove the curvature and obtain a linear exponential family by conditioning on a statistic that is specific ancillary for $\beta$. The linear exponential family obtained by conditioning enabled them to obtain tests with certain finite sample optimality properties. All the finite sample properties that Jansson and Moreira (2006) obtain, stated in their Theorems 2 and 5 and Lemmas 1, 3 and 4, are under the assumptions stated above, viz. that $(\phi, \sigma_e, \sigma_v)$ 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 $0$. Such assumptions would normally not be satisfied in practice. Also, we will see below in Section 2 that the assumption $\mu = 0$ removes a major source of the problem since the likelihood already has smaller curvature when $\mu = 0$. In their Theorem 6, Jansson and Moreira (2006) allow for an unknown $\mu$ taking

---

\(^1\)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 since that is not the focus of our work here.
potentially non-zero values. However, in this Theorem 6 they do not provide any finite sample results on the size or power properties of their test statistic but pointwise asymptotic results along a sequence of local-to-unity parameter values. This pointwise asymptotic result in their Theorem 6, which allows for arbitrary unknown $\mu$, is obtained under the assumption that the error covariance parameters $(\phi, \sigma_e, \sigma_v)$ are known.

The approach we take in this paper is to work with an unconditional likelihood, called the restricted likelihood, that already has very small curvature without making any assumptions about knowledge of any nuisance parameters. As a result, we are able to obtain theoretical results that demonstrate that the LRT based on the restricted likelihood (RLRT) has good finite sample performance in the predictive regression problem. Interestingly, we will see that the restricted likelihood also has an interpretation as a conditional likelihood, though the conditioning in this case is done with respect to a statistic that is complete sufficient for $(c, \mu)$. In Section 2, we provide motivation for considering the restricted likelihood. We then obtain the restricted likelihood for the bivariate predictive regression model and provide results on the bias of the REML estimates. In Section 3, we state our result on the finite sample behaviour of the RLRT for $\beta$ and compare its power under a sequence of Pitman alternatives to that of the restricted likelihood based Wald and Rao score test. Section 4 provides results on extensions of the REML method to higher order AR processes for the regressor series, as well as multivariate AR(1) regressors. The finite sample performance of the REML estimates and the RLRT is studied through simulations in Section 5. 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 $\alpha$ in the univariate AR(1) model given in (2). If $LRT_{\alpha, \mu}$ denotes the LRT for testing $H_0 : \alpha = \alpha_0$ versus $H_0 : \alpha \neq \alpha_0$ in that model,

\[
P(LRT_{\alpha,\mu} \leq x) - G_1(x) = 0.25 (1 + 7\alpha_0)(1 - \alpha_0)^{-1} n^{-1} (G_3(x) - G_1(x)) + O(n^{-2}) , \tag{3}
\]

where \(G_s(x)\) is the c.d.f. of a \(\chi^2_s\) random variable. On the other hand, if \(LRT_\alpha\) denotes the LRT for testing \(H_0 : \alpha = \alpha_0\) versus \(H_0 : \alpha \neq \alpha_0\) in the univariate AR(1) model given in (2) with \(\mu\) known to be 0, then it follows from Theorem 1 of van Giersbergen (2006) that

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

It is obvious from (3) that \(LRT_{\alpha,\mu}\) is very unstable when the autoregressive parameter \(\alpha_0\) is close to the unit root with the leading error term in the Edgeworth expansion going to infinity. In stark contrast, we see from (4) that \(LRT_\alpha\) is very well behaved with the leading term in its Edgeworth expansion being both very small and free of \(\alpha\). Simulation results in Figure 1 of van Giersbergen (2006) confirm the accuracy of the standard chi-square approximation for \(LRT_\alpha\), even when \(\alpha_0\) is close to unity. Not surprisingly, van Garderen (1999) has found that for the univariate AR(1) model in (2) with \(\mu\) known to be 0, the Efron (1975) curvature (one of the standard measures of curvature of the likelihood) is very small\(^3\), 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 \(\mu\). Since the bivariate prediction model in (1) and (2) is a vector AR(1) (with the first column of the coefficient matrix restricted to zero), one is led to suspect that the LRT for \(\beta\) may perhaps be well behaved if the intercept vector \((c, \mu)\) were

\(^2\)van Giersbergen (2006) provides the expected value of the LRT, while the combined results from the remaining references show that the leading term in the Edgeworth expansion (3) is half that expected value.

\(^3\)Interestingly and somewhat surprisingly, van Garderen (1999) found that if the innovation variance \(\sigma_v^2\) is known the Efron curvature of the model is \(2n^{-1} + O(n^{-2})\), which, though still small, is larger than when \(\sigma_v^2\) 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 (4) increases from \(-0.25n^{-1}\) to \(-1n^{-1}\) when \(\sigma_v^2\) 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.
known. Since the assumption that \((c, \mu)\) 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 general, the restricted likelihood is defined as the likelihood of linearly transformed data where the linear transformation is chosen to be orthogonal to the regression design matrix (and orthogonal to the vector of ones in the case of a location parameter). Harville (1974) provided a simple formula for computing the restricted likelihood and pointed out that the restricted likelihood would change only by a multiplicative constant for different choices of the linear transformation. 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, while 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, while Chen and Deo (2006 a, b) showed that confidence intervals for the sum of the autoregressive 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 restricted maximum likelihood (REML) estimates also are less biased than regular ML estimates in nearly integrated univariate \(AR\) models with intercept (Cheang and Reinsel, 2000) and with trend (Kang, Shin and Lee, 2003).

When \(X = (X_0, ..., X_n)'\) follows the univariate AR(1) model in (2), it follows from Harville...
(1974), that the restricted log-likelihood is given by

\[
L_R (\sigma^2, \alpha, \mathbf{X}) = -\left(\frac{n}{2}\right) \log \sigma^2 + \frac{1}{2} \log \left(\frac{1 + \alpha}{(n-1)(1-\alpha) + 2}\right) - \frac{1}{2\sigma^2} Q(\alpha),
\]

where

\[
Q(\alpha) = \mathbf{X}'\Sigma_X^{-1}\mathbf{X} - \left(\frac{(\mathbf{X}'\Sigma_X^{-1}\mathbf{X})}{\mathbf{1}'\Sigma_X^{-1}\mathbf{X}}\right)^2
\]

and \( \Sigma_X \equiv \text{Var}(\mathbf{X}) \), whereas the regular likelihood of \( \mathbf{X} \) for model (2) with known \( \mu \) (set, w.l.o.g., to zero) is

\[
L (\sigma^2, \alpha, \mathbf{X}) = -\left(\frac{n+1}{2}\right) \log \sigma^2 + \frac{1}{2} \log (1-\alpha^2) - \frac{1}{2\sigma^2} \mathbf{X}'\Sigma_X^{-1}\mathbf{X}.
\]

On comparing (5) and (7) and noting that the second term in \( Q(\alpha) \) is \( O(1) \) whereas \( \mathbf{X}'\Sigma_X^{-1}\mathbf{X} \) 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 \( \mu \). Due to this, it can be shown that the curvature properties of the restricted likelihood for the AR(1) as well as the Edgeworth expansion of the LRT are the same as those of the AR(1) model with known intercept, up to order \( O(n^{-1}) \). As a matter of fact, the restricted likelihood for the AR(1) also provides REML estimates of \( \alpha \) whose bias is \( -2\alpha n^{-1} + O(n^{-2}) \) and thus is identical, up to order \( O(n^{-1}) \), to the bias of the maximum likelihood estimate when the intercept is known. See Marriott and Pope (1954) and Cheang and Reinsel (2000).

Since the bias of the maximum likelihood estimate of \( \alpha \) 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 autoregressive parameter is close to the unit root. All of these results show 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. We now show that the restricted likelihood continues to do so in the case of the bivariate predictive model.

We first define some quantities that will be useful for 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

\[
S(\phi, \beta, \alpha) = (\mathbf{Y}_c - \phi\mathbf{X}_{c,1} - (\beta - \phi\alpha)\mathbf{X}_{c,2})'(\mathbf{Y}_c - \phi\mathbf{X}_{c,1} - (\beta - \phi\alpha)\mathbf{X}_{c,2})
\]
and note that for computational purposes $Q(\alpha)$ given in (6) can be written as

$$Q(\alpha) = \sum_{t=0}^{n} (X_t - \bar{X})^2 + \alpha^2 \sum_{t=1}^{n-1} (X_t - \bar{X})^2 - 2\alpha \sum_{t=0}^{n-1} (X_t - \bar{X})(X_{t+1} - \bar{X})$$

$$- \frac{(1-\alpha)^2 (n-1)}{(n-1)(1-\alpha) + 2} [X_0 + X_n - 2\bar{X}]^2.$$

We now can state the following theorem.

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

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

The REML estimates $\hat{\psi} = (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\sigma_v^2}, \hat{\sigma_e^2})$ are given by

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

$$(\hat{\phi}, \hat{\beta}) = \begin{bmatrix} 1 & 0 \\ \hat{\alpha} & 1 \end{bmatrix} (X_c'X_c)^{-1} X_c'Y_c,$$

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

and

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

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

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

and

$$E\left(\hat{\beta} - \beta\right) = \phi E(\hat{\alpha} - \alpha)$$

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

and

$$E\left(\hat{\phi} - \phi\right) = 0.$$
Remark 1 It is obvious that obtaining the REML estimates is computationally easy since all the estimates are obtained in succession after the optimisation of a one-dimensional function which is almost quadratic.

Remark 2 Note that the restricted likelihood in (9) 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)

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

where $\hat{\alpha}_{OLS}$ is the OLS estimate of $\alpha$. 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 1 also suggests a bias corrected version of the REML estimate of $\beta$ that may be computed as

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

where

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

is the bias corrected REML estimate of $\alpha$. Since 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,

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

particularly since the parameter $\phi$ can take any value in $(-\infty, \infty)$. (Amihud and Hurvich (2004) find empirical estimates of $\phi$ that are as large as $-94$). This is indeed the case, as we see in the simulations reported in Section 5 below.

The restricted likelihood given in (9) is derived for the situation where we assume that the initial value $X_0$ comes from the stationary distribution $N \left( \mu (1 - \alpha)^{-1}, \sigma^2_v (1 - \alpha^2)^{-1} \right)$. 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 \( X_t = \mu + X_t \), where \( X_t = \alpha X_{t-1} + v_t \) for \( t \geq 1 \) and \( X_0 = v_1 \). Under this assumption, the restricted likelihood is given by

\[
\tilde{L} (\beta, \alpha, \phi, \sigma_v^2, \sigma_e^2) = - \left( \frac{n-1}{2} \right) \log \sigma_v^2 - \frac{1}{2\sigma_e^2} S (\phi, \beta, \alpha) - \frac{n}{2} \log \sigma_e^2 \left( 1 + \frac{1}{n(1-\alpha)^2 + 1} \right) - \frac{1}{2\sigma_e^2} \tilde{Q} (\alpha),
\]

where

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

and

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

The bias results of Theorem 1 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 test.

### 3 REML Likelihood Ratio Test

One standard method for testing the composite hypothesis \( H_0 : \beta = 0 \) vs. \( H_a : \beta \neq 0 \) is the likelihood ratio test (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 (8) and just above it, it can be easily verified that under \( H_0 : \beta = 0 \) the restricted estimates \( \hat{\psi}_0 = \left( 0, \hat{\alpha}_0, \hat{\phi}_0, \hat{\sigma}_e^2, \hat{\sigma}_v^2 \right) \) are obtained as

\[
\hat{\alpha}_0 = \arg \min_{\alpha} \ n \log Q (\alpha) - \log \left( \frac{1 + \alpha}{(n-1)(1-\alpha) + 2} \right) + (n-1) \log R (\alpha),
\]
\[ R(\alpha) = Y_c' \left( I - Z_c(\alpha) (Z'_c(\alpha) Z_c(\alpha))^{-1} Z'_c(\alpha) \right) Y_c, \]
\[ Z_c(\alpha) = X_1 - 1\bar{X}_1 - \alpha (X_0 - 1\bar{X}_0), \]
\[ \hat{\phi}_0 = (Z'_c(\hat{\alpha}_0) Z_c(\hat{\alpha}_0))^{-1} Z'_c(\hat{\alpha}_0) Y_c, \]
\[ \hat{\sigma}^2_{v,0} = \frac{Q(\hat{\alpha}_0)}{n}, \quad \hat{\sigma}^2_{e,0} = \frac{1}{n - 1} R(\hat{\alpha}_0). \]

Just as in the unrestricted case, it is obvious that obtaining the restricted estimates requires only the optimisation of a nearly quadratic one dimensional function. The REML based likelihood ratio test (RLRT) for testing \( H_0 : \beta = 0 \) vs. \( H_a : \beta \neq 0 \) is now given by
\[ R_T = -2 L(\hat{\psi}_0) + 2 L(\hat{\psi}), \quad (11) \]
where \( L(\cdot) \) is the REML log-likelihood presented in (9). Under \( H_0 : \beta = 0 \), the asymptotic distribution of \( R_T \) is \( \chi^2_1 \), the chi-square distribution with one degree of freedom. In finite samples, however, the distribution of a \( LRT \) may deviate from the limiting \( \chi^2 \) distribution, particularly in the presence of nuisance parameters. The first step towards attempting to improve this approximation of the finite sample distribution by the limiting one was taken by Bartlett (1937).

Bartlett argued that if a \( LRT \) is asymptotically distributed as \( \chi^2_0 \) and \( E(LRT) = p(1 + cn^{-1}) + O(n^{-2}) \) for some constant \( c \), then \( (1 + cn^{-1})^{-1} LRT \) would be better approximated by the limiting \( \chi^2_0 \) distribution. The quantity \( 1 + cn^{-1} \), which may depend on the underlying parameter values, has since become known as the Bartlett correction to the likelihood ratio test. There is a large literature that has developed subsequently that deals with such corrections. This literature includes Lawley (1956), who derived a general formula for the Bartlett correction, Hayakawa (1977, 1987), who provided an asymptotic expansion of the distribution of the \( LRT \) and Barndorff-Nielsen and Hall (1988), who showed that for regular models the Bartlett corrected \( LRT \) is approximated by the limiting \( \chi^2 \) distribution with an error of \( O(n^{-2}) \) compared to an error of \( O(n^{-1}) \) for the uncorrected test statistic. For the predictive regression model in this paper, we are able to obtain the following theorem on the finite sample behaviour of the \( RLRT \).

**Theorem 2** Under \( H_0 : \beta = 0 \) in the model given by (1) and (2), we have
\[ P(R_T \leq x) = P(\chi^2_1 \leq x) + (3/4 - \rho^2) n^{-1} \left[ P(\chi^2_3 \leq x) - P(\chi^2_1 \leq x) \right] + O(n^{-2}), \quad (12) \]
where $\rho = \text{Corr} (u_t, v_t)$.

Theorem 2 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

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

where

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

then

$$P (R_{TB} \leq x) = P (\chi^2_1 \leq x) + O (n^{-2}).$$

**Remark 3** The results of Theorem 2 and Corollary 1 continue to hold for the restricted likelihood given in (10) when the initial condition $X_0$ is not from the stationary distribution.

The results of Theorem 2 obviously imply that the $\chi^2_1$ 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 tests particularly when it is close to unity. Theorem 2 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 (12), with $\rho^2$ replaced by $\hat{\rho}^2$. The other is to use the Bartlett corrected statistic $R_{TB}$. However, (12) suggests that the original test $R_T$ used in conjunction with the standard $\chi^2_1$ 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. It is worth noting that though the REML likelihood does not provide an unbiased estimate of $\beta$ (indeed, the bias of $\hat{\beta}$ can be arbitrarily large due to the fact that $\phi$ is unbounded, as noted below Theorem 1), the REML likelihood yields a very well behaved test for $\beta$, irrespective of how large $\phi$ is. This result serves to illustrate the point that it may be more desirable at times
to carry out tests of hypothesis 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 interesting to compare the quality of the chi-square approximation in Theorem 2, where all four parameters \((\alpha, \phi, \sigma_v^2, \sigma_e^2)\) are nuisance parameters, with the approximation in the case where the parameters \((\phi, \sigma_v^2, \sigma_e^2)\) are known and \(\alpha\) is the only nuisance parameter. From equation (31) and the discussion below it in the proof of Theorem 2, 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 satisfy

\[
P(R_{T,1} \leq x) = P(\chi_1^2 \leq x) - \rho^2 n^{-1} \left[ P(\chi_3^2 \leq x) - P(\chi_1^2 \leq x) \right] + O(n^{-2}).
\]

By comparing this result with that in Theorem 2, one sees that the quantity \(3/4\) is a measure of the extent to which lack of knowledge of the innovation parameters \((\phi, \tau_v, \tau_e)\) affects the finite sample distribution of \(R_T\). Furthermore, since \(\sup \rho^2 > \sup |\rho^2 - 3/4|\) for \(\rho \in (-1, 1)\), the chi-square approximation to the RLRT is better in a "minimax" sense for the case when the innovation covariance parameters are unknown than when they are known. 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 the footnote at the beginning of section 2.

The result in Theorem 2 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 test 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, in order 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 test against such local alternatives.

**Theorem 3** 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 (9) of the model in (1) and (2). Assume that
the true value of $\beta$ is given by $\beta = \beta_0 + \xi n^{-1/2}$. Define

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

and

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

Let $\bar{G}_{s,\Delta}(x)$ denote the survival function of a non-central $\chi^2$ random variable with $s$ degrees of freedom and non-centrality parameter $\Delta$. Then,

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

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

and

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

From Theorem 3 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 $O\left(n^{-1}\right)$) against local Pitman alternatives. Since $\bar{G}_{s,\Delta}(x) - \bar{G}_{l,\Delta}(x) < 0$ for all $x > 0$ when $l > s$, it follows from (13) that RLRT will be guaranteed to be more powerful 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 which is of relevance in empirical applications in finance and economics. It is also interesting to note that the non-centrality parameter $\Delta$, which will be the main source of power, increases as $\alpha$ gets closer to the unit root. The non-centrality parameter also depends inversely on $\phi$, but does not depend on the correlation between the two innovation series $(u_t, v_t)$.

In the next Section we derive the REML likelihood under more general models for the regressor series as well as for multiple regressors and discuss some efficiency and computational issues.
4 REML for more general regressor models

It is easy to generalise the REML likelihood in two directions that are both of practical interest. One generalisation is to the case where the predictor series is a multivariate AR(1) process. Applications of such models can be found, for example, in Amihud and Hurvich (2004), who considered dividend yield and earnings to price ratio as bivariate predictors of market returns. The other generalisation 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 1.

4.1 Multivariate regressors

Assume that the data \((Y_1, ..., Y_n, X'_0, ..., X'_n)\) follows

\[
Y_t = c + \beta'X_{t-1} + u_t,
\]

\[
X_t = \mu + \lambda X_{t-1} + v_t,
\]

where \(u_t = \phi'v_t + e_t, (e_t, v_t)' \sim N(0, \text{diag}(\sigma^2_e, \Sigma_v))\) is an i.i.d. series and \(A\) is a \(k \times k\) matrix with all eigenvalues less than unity in absolute value. Let \(\Sigma_v \equiv Var(v_t)\) and \(\Sigma_X \equiv Var(X_t)\), given by

\[
vec(\Sigma_X) = (I_{k^2} - A \otimes A)^{-1} vec(\Sigma_v)
\]

and define

\[
\hat{\tau} = \left[ \Sigma_X^{-1} + n (I - A)' \Sigma_v^{-1} (I - A) \right]^{-1} \left[ \Sigma_X^{-1} X_0 + (I - A)' \Sigma_v^{-1} (I - A) \sum_{t=1}^n X_t \right].
\]

**Lemma 1** Then the REML log-likelihood up to an additive constant for the model in (14) and
(15) is given by

\[ L_M = -\left(\frac{n-1}{2}\right) \log \sigma_e^2 - \frac{1}{2\sigma_e^2} S(\phi, \beta, A) - \frac{1}{2} \log |\Sigma_X| - \frac{n}{2} \log |\Sigma_v| \]

\[ -\frac{1}{2} \log |\Sigma_X^{-1} + n (I - A)' \Sigma_v^{-1} (I - A)| \]

\[ -\frac{1}{2} \left\{ (X_0 - \hat{\tau})' \Sigma_X^{-1} (X_0 - \hat{\tau}) + \sum_{t=1}^{n} (X_t - \hat{\tau} - A (X_{t-1} - \hat{\tau}))' \Sigma_v^{-1} (X_t - \hat{\tau} - A (X_{t-1} - \hat{\tau})) \right\}, \]

where

\[ S(\phi, \beta, A) = \sum_{t=1}^{n} (Y_{t,c} - \phi' X_{t,c} - (\beta' - \phi' A) X_{t-1,c})^2, \]

\[ X_{t,c} = X_t - n^{-1} \sum_{t=1}^{n} X_t \text{ and } X_{t-1,c} = X_{t-1} - n^{-1} \sum_{t=1}^{n} X_{t-1}. \]

To ease the computational burden during optimisation, the likelihood can be defined in terms of the re-parametrised set \((\Sigma_v, \sigma_e^2, A, \phi, \gamma)\), where \(\gamma = \beta - A \phi\). This re-parametrisation allows us to concentrate \((\sigma_e^2, \phi, \gamma)\) out of the likelihood, thus reducing the dimensionality of the optimisation problem. The likelihood can then be sequentially optimised, first over \((\Sigma_v, A)\), with the REML estimates of \((\sigma_e^2, \phi, \gamma)\) being then obtained by OLS through the minimisation of \(S(\phi, \gamma)\). A further simplification of the above likelihood occurs if the coefficient matrix \(A\) is diagonal, given by \(A = \text{diag}(\alpha_1, ..., \alpha_k)\) with \(\max_i |\alpha_i| < 1\), in which case one gets

\[ \text{Var} (X_t) \equiv \Sigma_X = \begin{pmatrix} \frac{\sigma_{v,ij}}{1 - \alpha_i \alpha_j} \end{pmatrix}, \]

where \(\Sigma_v = ((\sigma_{v,ij}))\). 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 \(A\) can be assumed to be a diagonal matrix, the predictive regression model is no longer a SUR system and hence OLS will no longer be efficient. However, REML will clearly retain efficiency, no matter what the form of \(A\) is, thus giving it an advantage both in terms of asymptotic efficiency and power over any OLS based procedure.

Since the dimension of the parameter space is very large in the vector case, it is not feasible to obtain a result such as Theorem 2 in the most general case. However, in the case where \(A\) is a diagonal matrix and where \((\sigma_e^2, \phi, \Sigma_v)\) are assumed known with \(\Sigma_v\) diagonal, we are able to obtain the following result on the finite sample behaviour of the RLRT for testing \(H_0 : \beta = 0\). The proof follows along lines similar to those for Theorem 2 and is omitted.
Theorem 4 In the model given by (14) and (15), assume that \( A = \text{diag}(\alpha_1, \ldots, \alpha_k) \), with \( \max_i |\alpha_i| < 1 \), and that \((\sigma^2_e, \phi, \Sigma_v)\) are known with \( \Sigma_v = \text{diag}(\sigma^2_{v,11}, \ldots, \sigma^2_{v,kk}) \). Let \( R_M \) denote the RLRT based on the restricted likelihood in (16) for testing \( H_0 : \beta = 0 \). Then,

\[
P(R_M \leq x) = P\left(\chi^2_k \leq x\right) - \frac{n^{-1}}{\sum_{i=1}^{k} \frac{\phi^2_i \sigma^2_{v,ii} \sigma^2_e}{1 + \phi^2_i \sigma^2_{v,ii} \sigma^2_e}} \left[ P\left(\chi^2_{k+2} \leq x\right) - P\left(\chi^2_k \leq x\right)\right] + O\left(n^{-2}\right).
\]

Since \(0 < n^{-1} \sum_{i=1}^{k} \frac{\phi^2_i \sigma^2_{v,ii} \sigma^2_e}{1 + \phi^2_i \sigma^2_{v,ii} \sigma^2_e} < n^{-1}k\) trivially, the result shows that the \( \chi^2 \) distribution once again provides a very good approximation to the RLRT in this situation. It is useful to note that Theorem 4 shows that the quality of the \( \chi^2 \) approximation to the RLRT is affected only minimally by the dependence between \( u_t \) and \( v_t \), 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 \( \Sigma_v \) is diagonal is not very unreasonable. Finally, from the discussion below equation (4) and below Theorem 2, one would expect the \( \chi^2 \) approximation to continue to work well when \((\sigma^2_e, \phi, \Sigma_v)\) are unknown (with \( \Sigma_v \) diagonal) and, indeed, we find this to be the case in our simulations in Section 5 below. As a matter of fact, the simulations show that the RLRT behaves very well even when the cross-correlation in the \( X \) 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

\[
Y_t = c + \beta X_{t-1} + u_t
\]

and

\[
X_t = \mu + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + v_t,
\]

17
where \( u_t = \phi v_t + e_t \) and \((e_t, v_t) \sim N(0, \text{diag}(\sigma_e^2, \sigma_v^2))\) 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

\[
Y_c = Y - 1\bar{Y}, \quad X_c = [X_1 - 1\bar{X}_1, ..., X_{p+1} - 1\bar{X}_{p+1}],
\]

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

**Lemma 2** The REML log-likelihood up to an additive constant for the model in (17) and (18) is given by

\[
L(\alpha, \beta, \phi, \sigma_e^2, \sigma_v^2) = - \left( \frac{n-1}{2} \right) \log \sigma_e^2 - \frac{1}{2\sigma_e^2} S(\phi, \beta, \alpha_1, ..., \alpha_p)
- \left( \frac{n+p-1}{2} \right) \log \sigma_v^2 - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |1'\Sigma^{-1}1| - \frac{1}{2\sigma_v^2} (X - 1\hat{\tau})' \Sigma^{-1} (X - 1\hat{\tau})
\]

where \( \text{Var}(X) = \sigma_v^{-2}\Sigma, \hat{\tau} = (1'\Sigma^{-1}1)^{-1} 1'\Sigma^{-1}X \) and \( S(\phi, \beta, \alpha_1, ..., \alpha_p) = Z_c'Z_c \) where

\[
Z_c = \left( Y_c - \phi X_{c,1} - (\beta - \phi \alpha_1)X_{c,2} + \sum_{i=2}^{p} \alpha_{c,i} X_{c,i+1} \right).
\]

Though at first glance the expression in (19) looks formidable, it is actually very easy to compute. The quantity \( S(\phi, \beta, \alpha_1, ..., \alpha_p) \) is, of course, just a quadratic form. It is also well known that both the determinant \( |\Sigma| \) and bilinear/quadratic forms of the type \( y'\Sigma^{-1}x \) are very easy to compute for AR\((p)\) models, since for such models \( \Sigma^{-1} \) can be easily expressed as \( B'B \) for some lower triangular matrix \( B \). (See equation 25 below). After a re-parametrisation \( \gamma = \beta - \phi \alpha_1 \), the parameters \((\sigma_e^2, \sigma_v^2, \phi, \gamma)\) can be concentrated out of (19) and the concentrated log-likelihood optimised over the remaining parameters \((\alpha_2, ..., \alpha_p)\).

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

### 5 Simulations

We first study the performance of four estimators of \( \beta \): (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 below Theorem 1. The data was simulated from the model given by equations (1) and (2) with three sample sizes, $n = 50, 100$ and 150. For each sample size, the predictive regression slope coefficient $\beta$ was set to 0 while the autoregressive coefficient $\alpha$ was varied over three values, 0.95, 0.975 and 0.99. Since the bias in the estimate of $\beta$ is proportional to $\phi$, we chose $\phi = -80$ (as noted earlier, Amihud and Hurvich, 2004, find empirical estimates of $\phi$ as high as -90 ) to highlight the bias gains to be had by using the bias corrected REML estimate. The values of $\sigma^2_v$ and $\sigma^2_e$ were set to be 0.00040681 and 1 respectively, implying that $\text{Corr}(u_t, v_t) = -0.85$, which is representative of the innovation correlation found in empirical examples. Tables I-III report the simulation means and standard deviations of these four estimates based on 10,000 replications for each of the three values of $\alpha$.

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 $\alpha = 0.99$. 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 70% and its standard deviation is uniformly lower too. The simulation results clearly demonstrate the advantage enjoyed by the REML estimate in terms of both bias and standard deviation over the OLS estimate. To test the robustness of the procedure to non-normal 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 Tables I-III 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 $\chi^2_1$ approximation to the distribution of the $RLRT$ for the above model and parameter configurations with both normal and $t_5$ innovations. The $\chi^2_1$ approximation was assessed by three measures (i) QQ plots of the $RLRT$ against the theoretical quantiles of a $\chi^2_1$ distribution (ii) simulation means and standard deviations of the $RLRT$ (which are theoretically 1 and 2 respectively for a $\chi^2_1$ distribution) and (iii) simulation sizes at the 5% and 1% level. The simulation mean, standard deviation and sizes are reported in Table IV while the QQ plots are shown in Figure 1. (We present the QQ plots only for the normal innovations since the plots for $t_5$ errors are qualitatively similar in nature). Apart from
a small distortion at \( n = 50 \) and \( \alpha = 0.99 \), the RLRT is seen to be very well approximated by the \( \chi^2 \) distribution according to each of the three measures. It is also worth stressing again that the performance of the RLRT is not affected by the bias of \( \hat{\beta} \) at all.

We finally generate data from a model in which the regressors are a bivariate AR(1) model. More specifically, we use the model \( Y_t = \beta'X_{t-1} + u_t \) and \( X_t = AX_{t-1} + v_t \), where \( u_t = \phi'v_t + e_t \). The sample size was set at \( n = 200 \), the slope vector \( \beta \) was set to zero, while \( \phi = (-80, -80) \). Two configurations of the autoregressive coefficient matrix \( A \) were considered, \( \text{diag} \ (0.95, 0.8) \) and \( \text{diag} \ (0.95, 0.95) \). The innovation matrix \( \Sigma_v \) 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 \( \sigma^2_e \) was set to unity. This design (except for \( \Sigma_v = I \)) 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 \( \chi^2 \) approximation to the distribution of the RLRT was assessed via three measures (i) QQ plots of the RLRT against the theoretical quantiles of a \( \chi^2 \) distribution (ii) simulation means and standard deviations of the RLRT (which are theoretically 2 and 4 respectively for a \( \chi^2 \) distribution) and (iii) simulation sizes at the 5% and 1% level. The results are provided in Tables V and VI and Figure 2. As noted in sub-section 4.1 above, OLS is no longer asymptotically efficient if \( A \) is a diagonal matrix since the system is no longer a SUR. Hence, not only does REML afford a dramatic reduction in bias over OLS, it also provides a great reduction in variance, as can be seen in Table V. Furthermore, for the inference problem it is once again seen that the RLRT is very well approximated by the \( \chi^2 \) 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.

6 Appendix

Proof of Theorem 1:
As noted at the start of Section 2) above, the REML likelihood corresponds to the likelihood of $TZ$, where $T$ is any full row rank matrix such that $T1 = 0$. We will obtain this likelihood by choosing $T$ to have the form

$$T = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix},$$

where $T_1$ and $T_2$ are full row rank matrices of dimension $(n - 1) \times n$ and $n \times (n + 1)$ respectively, satisfying $T_11 = 0$, $T_21 = 0$, $T_11' = I$ and $T_2T_2' = I$ and using the fact that

$$L(TZ) = L(T_1Y / T_2X) L(T_2X).$$

We first obtain $L(T_1Y / T_2X)$. Since $u_t = \phi v_t + \epsilon_t$, where $\phi = \sigma_{uv}/\sigma_v^2$ and $\epsilon_t \sim N(0, \sigma^2)$ is a series independent of $\{v_t\}$, we get

$$Y_t = c + \beta X_{t-1} + \phi (X_t - \mu - \alpha X_{t-1}) + \epsilon_t$$

$$= c + \phi X_t + (\beta - \phi \alpha) X_{t-1} + \epsilon_t. \tag{20}$$

Let $\tilde{Y} = T_1Y$, $\tilde{X}_1 = T_1X_1$ and $\tilde{X}_0 = T_1X_0$. From (20) it then follows that

$$\tilde{Y} = \tilde{X}\theta + \tilde{\epsilon}, \tag{21}$$

where $\tilde{X} = [\tilde{X}_1, \tilde{X}_0]$, $\tilde{\epsilon} = T_1\epsilon$, $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ and $\theta = (\phi, \beta, -\phi \alpha)'$. Since $X_t$ is a function only of $\{v_t, v_{t-1}, ..., \}$, the series $\{X_t\}$ is independent of $\{\epsilon_t\}$. Furthermore, knowledge of $T_2X$, where $T_2$ is any full row rank matrix such that $T_21 = 0$, implies knowledge of $\tilde{X}$. Hence, from (21) the conditional distribution of $\tilde{Y}$ given $T_2X$ is $N(\tilde{X}\theta, \sigma^2 I_2)$, since $T_1T_1' = I$. It follows that the conditional log-likelihood of $\tilde{Y} / T_2X$ up to an additive constant is given by

$$l_1 (\tilde{Y} / T_2X, \theta, \sigma^2) = -\left(\frac{n-1}{2}\right) \log \sigma^2 - \frac{1}{2\sigma^2} \left(\tilde{Y} - \tilde{X}\theta\right)^* \left(\tilde{Y} - \tilde{X}\theta\right).$$

Note, however, that $\tilde{X}\theta = T_1[X_1, X_0] \theta$. Thus,

$$\left(\tilde{Y} - \tilde{X}\theta\right)^* \left(\tilde{Y} - \tilde{X}\theta\right) = (Y - [X_1, X_0] \theta)' T_1' T_1 (Y - [X_1, X_0] \theta).$$

Since the matrix $T_1$ when augmented by the row $n^{-1/2}1'$ is an orthogonal matrix, it follows that $T_1' T_1 = I - n^{-1}11'$ and hence

$$\left(\tilde{Y} - \tilde{X}\theta\right)^* \left(\tilde{Y} - \tilde{X}\theta\right) = \left(\tilde{Y} - [X_1, X_0] \theta\right)' \left(I - n^{-1}11'\right) \left(Y - [X_1, X_0] \theta\right)$$

$$= S(\phi, \beta, \alpha).$$

21
Thus, we get
\[ l_1 \left( \bar{Y} / \bar{X}, \theta, \sigma_e^2 \right) = -\left( \frac{n-1}{2} \right) \log \sigma_e^2 - \frac{1}{2\sigma_e^2} S(\phi, \beta, \alpha). \] (22)

The log-likelihood of \( T_2X \) is obtained from Harville (1974) and up to an additive constant is given by
\[ l_2 \left( T_2X, \alpha, \sigma_e^2 \right) = -\left( \frac{n}{2} \right) \log \sigma_e^2 - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |1'\Sigma^{-1}1| - \frac{1}{2\sigma_e^2} (X - 1\hat{\tau})'\Sigma^{-1}(X - 1\hat{\tau}), \] (23)

where \( \hat{\tau} = (1'\Sigma^{-1}1)^{-1}1'\Sigma^{-1}X \) and \( \Sigma = \sigma_v^2 Var(X) \). From (22) and (23), we obtain the log-likelihood of \( (T_1Y, T_2X) \) up to an additive constant to be
\[ L \left( T_1Y, T_2X, \alpha, \beta, \phi, \sigma_v^2, \sigma_e^2 \right) = -\left( \frac{n-1}{2} \right) \log \sigma_e^2 - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |1'\Sigma^{-1}1| - \frac{1}{2\sigma_v^2} Q(\alpha). \] (24)

The final form, as stated in (9) is obtained by algebraic simplification, using the fact that \( \Sigma^{-1} = B'B \), where \( B \) is the \((n+1) \times (n+1)\) matrix given by
\[
\begin{bmatrix}
\sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 & 0 \\
-\alpha & 1 & 0 & \cdots & 0 & 0 \\
0 & -\alpha & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -\alpha & 1
\end{bmatrix}
\] (25)

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 \( (\alpha, \beta, \phi, \sigma_v^2, \sigma_e^2) \), it helps to consider the re-parametrised set of parameters \( (\alpha, \gamma, \phi, \sigma_e^2, \sigma_v^2) \), where \( \gamma = \beta - \phi \alpha \). It is then immediately obvious from inspecting (24) that the REML estimates of \( (\alpha, \sigma_v^2) \) are obtained by simply maximising just
\[ l_2 \left( T_2X, \alpha, \sigma_v^2 \right) = -\left( \frac{n}{2} \right) \log \sigma_v^2 - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |1'\Sigma^{-1}1| - \frac{1}{2\sigma_v^2} (X - 1\hat{\tau})'\Sigma^{-1}(X - 1\hat{\tau}). \]

In other words, the REML estimates of \( (\alpha, \sigma_v^2) \) are just those estimates that would have been obtained by maximising the REML likelihood function of \((X_1, ..., X_n)\). It thus follows immediately from Cheang and Reinsel (2000) that the bias of \( \hat{\alpha} \) is
\[ E(\hat{\alpha} - \alpha) - \frac{2\alpha}{n-1} + o(n-1). \]
The REML estimates of \((\phi, \gamma)\) are obtained as the least squares estimates
\[
\begin{pmatrix}
\hat{\phi} \\
\hat{\beta}
\end{pmatrix} = \left( X_c'X_c \right)^{-1} X_c'Y_c,
\]
which then yields the REML estimate \(\hat{\beta} = \hat{\gamma} + \hat{\phi} \hat{\alpha}\). Since
\[
\begin{pmatrix}
\hat{\phi} \\
\hat{\beta}
\end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \hat{\alpha} & 1 \end{pmatrix} \begin{pmatrix} \hat{\phi} \\ \hat{\gamma} \end{pmatrix},
\]
it follows that the REML estimates of the original parameters \((\phi, \beta)\) can also be obtained in a direct regression of \(Y_c\) on \(\left( X_1 - 1X_1 - \hat{\alpha}(X_0 - 1X_0), X_0 - 1X_0)\). Thus, \(\hat{\beta}\) is identical to the ARM estimate considered by Amihud and Hurvich (2004) using the REML estimate \(\hat{\alpha}\) as a proxy for \(\alpha\). Hence, the bias of \(\hat{\beta}\) can be obtained from Theorem 2 of Amihud and Hurvich (2004) and is
\[
E \left( \hat{\beta} - \beta \right) = \phi E (\hat{\alpha} - \alpha)
\]
\[
= -\frac{2\alpha\phi}{n-1} + o(n^{-1}).
\]
Finally, Lemma 1 of Amihud and Hurvich (2004) implies that \(E \left( \hat{\phi} \right) = \phi\).

Proof of Theorem 2:

Since the LRT is invariant to re-parametrisation we choose to work with the re-parametrisation \(\lambda = (\beta, \alpha, \phi, \tau_v, \tau_e) \equiv (\beta, \lambda_2)\), where \(\tau_v = \sigma_v^{-2}\) and \(\tau_e = \sigma_e^{-2}\) since this greatly reduces the burden of our computations. For the REML log-likelihood given in (9), we will denote expectations of the log-likelihood derivatives as
\[
\kappa_{rs} = n^{-1} E \left( \partial^2 L / \partial \lambda_r \lambda_s \right), \quad \kappa_{rst} = n^{-1} E \left( \partial^3 L / \partial \lambda_r \lambda_s \lambda_t \right), \quad \kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \lambda_t.
\]
Letting \(\delta = \tau_e / \tau_v\), it is easily seen that the information matrix \(K = (-\kappa_{rs})\) is given by
\[
K = \begin{bmatrix}
K_{\beta,\alpha} & 0 \\
0 & K_{\phi,\tau_v,\tau_e}
\end{bmatrix} + O(n^{-1}), \quad (26)
\]
where
\[
K_{\beta,\alpha} = \frac{\delta}{1 - \alpha^2} \begin{bmatrix} 1 & -\phi \\
-\phi & (\phi^2 + \delta^{-1}) \end{bmatrix},
\]
and
\[
K_{\phi,\tau_v,\tau_e} = \begin{bmatrix}
\kappa_{\phi,\tau_v,\tau_e} \\
0
\end{bmatrix}.
\]
and

\[ K_{\phi,\tau_v,\tau_e} = \text{diag} \left\{ \delta, \frac{1}{2\tau_v^2}, \frac{1}{2\tau_e^2} \right\}. \]

Let \( \kappa^{rs} \) denote the entries of \(-K^{-1}\) and \( \tilde{\kappa}^{rs} \) denote the entries of \(-K^{-1}_{22}\), where \( K_{22} \) is the lower right \( 4 \times 4 \) sub-matrix of \( K \). From the results of Hayakawa (1977, 1987), Cordeiro (1987), Barndorff-Nielsen and Hall (1988) and Chesher and Smith (1995), we have

\[
P(R_T \leq x) = P(\chi_1^2 \leq x) + A n^{-1} [ P(\chi_3^2 \leq x) - P(\chi_1^2 \leq x) ] + O(n^{-2}),
\]

where

\[
A = \frac{1}{2} \left\{ \sum_{\lambda} (l_{rstu} - l_{rstuvw}) - \sum_{\lambda_2} (\tilde{l}_{rstu} - \tilde{l}_{rstuvw}) \right\},
\]

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

\[
l_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left( \frac{1}{6} \kappa_{rstuv} - \frac{1}{4} \kappa_{rstw} - \kappa_{rstv}^{(u)} - \kappa_{rstw}^{(v)} + \kappa_{rst}^{(wu)} + \kappa_{rt}^{(uw)} + \kappa_{rt}^{(sw)} + \kappa_{rt}^{(vw)} \right)
\]

and

\[
\tilde{l}_{rstuvw} = \tilde{\kappa}^{rs} \tilde{\kappa}^{tu} \tilde{\kappa}^{vw} \left( \frac{1}{6} \kappa_{rstuv} - \frac{1}{4} \kappa_{rstw} - \kappa_{rstv}^{(u)} - \kappa_{rstw}^{(v)} + \kappa_{rst}^{(wu)} + \kappa_{rt}^{(uw)} + \kappa_{rt}^{(sw)} + \kappa_{rt}^{(vw)} \right).
\]

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

\[
\sum_{\lambda} (l_{rstu} - l_{rstuvw}) = \sum_{(\beta,\alpha)} (l_{rstu} - l_{rstuvw}) + \sum_{(\phi,\tau_v,\tau_e)} (l_{rstu} - l_{rstuvw})
+ \sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))} (l_{rstu} - l_{rstuvw}) + O(n^{-1}),
\]

where \( \sum_{((\beta,\alpha),(\phi,\tau_v,\tau_e))} \) denotes that at least one index in the summand must come from \( (\beta,\alpha) \) and at least one index from \( (\phi,\tau_v,\tau_e) \). By the same logic, we can simplify the term \( \sum_{\lambda_2} (\tilde{l}_{rstu} - \tilde{l}_{rstuvw}) \)
in A as

\[
\sum_{\lambda_2} (\bar{l}_{rstu} - \bar{l}_{rstuvw}) = \sum_{\alpha} (\bar{l}_{rstu} - \bar{l}_{rstuvw}) + \sum_{(\phi, \tau_v, \tau_e)} \left( \bar{l}_{rstu} - \bar{l}_{rstuvw} \right) + \sum_{(\alpha, (\phi, \tau_v, \tau_e))} \left( \bar{l}_{rstu} - \bar{l}_{rstuvw} \right) + O(n^{-1})
\]

\[
= \sum_{\alpha} (\bar{l}_{rstu} - \bar{l}_{rstuvw}) + \sum_{(\phi, \tau_v, \tau_e)} (l_{rstu} - l_{rstuvw})
\]

\[
+ \sum_{(\alpha, (\phi, \tau_v, \tau_e))} \left( \bar{l}_{rstu} - \bar{l}_{rstuvw} \right) + O(n^{-1}),
\]

(30)

where the last step in (30) follows from the fact that the entries of \(K\) for \((\phi, \tau_v, \tau_e)\) are diagonal up to \(O(n^{-1})\). It thus follows from (28), (29) and (30) that

\[
A = \frac{1}{2} \left\{ \sum_{(\beta, \alpha)} (l_{rstu} - l_{rstuvw}) - \sum_{\alpha} \left( \bar{l}_{rstu} - \bar{l}_{rstuvw} \right) \right\}
\]

\[
+ \frac{1}{2} \left\{ \sum_{((\beta, \alpha), (\phi, \tau_v, \tau_e))} (l_{rstu} - l_{rstuvw}) - \sum_{(\alpha, (\phi, \tau_v, \tau_e))} \left( \bar{l}_{rstu} - \bar{l}_{rstuvw} \right) \right\} + O(n^{-1})
\]

\[
\equiv A_{(\beta, \alpha)} + C_\lambda + O(n^{-1}),
\]

(31)

It is obvious from the structure of \(A_{(\beta, \alpha)}\) in (31) that \(A_{(\beta, \alpha)}\) would be the leading remainder term in the expansion of the distribution of \(R_T\) of the form in (27) if \((\phi, \tau_v, \tau_e)\) were known and \(\alpha\) were the only nuisance parameter. The term \(C_\lambda\), thus, is a measure of the extent to which lack of knowledge of the parameters \((\phi, \tau_v, \tau_e)\) affects the finite sample distribution of \(R_T\). We now compute the two terms \(A_{(\beta, \alpha)}\) and \(C_\lambda\), beginning with \(A_{(\beta, \alpha)}\).

Though \((\beta, \alpha)\) are not orthogonal, the computation of \(\sum_{(\beta, \alpha)} (l_{rstu} - l_{rstuvw})\) can be simplified by working with the transformed parameters \((\gamma, \alpha)\), where \(\gamma = \beta - \phi \alpha\), since \(\gamma\) is orthogonal to \(\alpha\) (Note that as stated above, when computing \(A_{(\beta, \alpha)}\) the remaining parameters \((\phi, \tau_v, \tau_e)\) are fixed). Since \((\gamma, \alpha)\) is an affine transformation of \((\beta, \alpha)\), we can exploit the fact that the first term of \(A_{(\beta, \alpha)}\) is invariant under such transformations (see page 371 of Hayakawa 1977) and thus get

\[
\frac{1}{2} \sum_{(\beta, \alpha)} (l_{rstu} - l_{rstuvw}) = \frac{1}{2} \sum_{(\gamma, \alpha)} (l_{rstu,p} - l_{rstuvw,p}),
\]

(32)
where the extra subscript $p$ will mean that the computation is being carried out for the re-parameterised form of (9) with $\gamma = \beta - \phi \alpha$. The right hand side of (32) is much simpler to compute due to the fact that $\kappa_{\gamma \alpha, p} = \kappa_{\gamma \gamma, p} = \kappa_{\gamma \gamma, p} = 0$ and all the terms $\kappa_{rst, p}$ and $\kappa_{rstu, p}$ are at most $O(n^{-1})$, for all permutations of the subscripts. Since $\kappa_{\gamma \gamma, p} = -\delta (1 - \alpha^2)^{-1}$ and $\kappa_{\alpha \alpha, p} = - (1 - \alpha^2)^{-1}$, it follows that

\[
\frac{1}{2} \sum_{(\gamma, \alpha)} l_{rstu, p} = \frac{1}{2} \kappa_{\gamma \gamma, p} \kappa_{\alpha \alpha, p} (\kappa_{\gamma \alpha, p})^2 + O(n^{-1}) = -(1 + 3\alpha^2) (1 - \alpha^2)^{-1} + O(n^{-1}) \tag{33}
\]

and

\[
\frac{1}{2} \sum_{(\gamma, \alpha)} l_{rstuvw, p} = \frac{1}{2} \kappa_{\gamma \gamma, p} \kappa_{\alpha \alpha, p}^2 \kappa_{\alpha \alpha, p} (\kappa_{\gamma \alpha, p})^2 + O(n^{-1}) = -4\alpha^2 (1 - \alpha^2)^{-1} + O(n^{-1}). \tag{34}
\]

From (32), (33) and (34), we get

\[
\frac{1}{2} \sum_{(\beta, \alpha)} (l_{rstu} - l_{rstuvw}) = -1 + O(n^{-1}). \tag{35}
\]

The second term in $A_{(\beta, \alpha)}$ is not invariant under affine transformations and we revert to the original log-likelihood (9) to compute it. Noting that $\kappa_{\alpha \alpha} = - (\phi^2 \delta + 1) (1 - \alpha^2)^{-1}$ and $\tilde{\kappa}_{\alpha \alpha} = - (\phi^2 \delta + 1)^{-1} (1 - \alpha^2)$, we have

\[
\frac{1}{2} \sum_{\alpha} (\tilde{l}_{rstu} - \tilde{l}_{rstuvw}) = \frac{1}{2} \tilde{\kappa}_{\alpha \alpha} \tilde{\kappa}_{\alpha \alpha} (\kappa_{\gamma \alpha, p})^2 - \frac{1}{2} \tilde{\kappa}_{\alpha \alpha} \tilde{\kappa}_{\alpha \alpha} (\kappa_{\gamma \alpha, p})^2 + O(n^{-1})
\]

\[
= -\frac{1}{\phi^2 \delta + 1} + O(n^{-1}). \tag{36}
\]

From (35) and (36) we conclude that

\[
A_{(\beta, \alpha)} = -\frac{\phi^2 \delta}{\phi^2 \delta + 1} + O(n^{-1}) = -\rho^2 + O(n^{-1}). \tag{37}
\]

We now turn our attention to computing the second term, $C_{\lambda}$, in (31). We first note from the near diagonal structure of $K$ that

\[
\frac{1}{2} \sum_{(\beta, \alpha, (\phi, \tau_\nu, \tau_\kappa))} (l_{rstu} - l_{rstuvw}) = \frac{1}{2} \sum_{(\beta, (\phi, \tau_\nu, \tau_\kappa))} l_{rrtt} + \frac{1}{2} \sum_{(\alpha, (\phi, \tau_\nu, \tau_\kappa))} l_{rrtt}
\]

\[
- \frac{1}{2} \sum_{(\beta, (\phi, \tau_\nu, \tau_\kappa))} l_{rrtvu} - \frac{1}{2} \sum_{(\alpha, (\phi, \tau_\nu, \tau_\kappa))} l_{rrtvu}
\]

\[
- \frac{1}{2} \sum_{(\beta, \alpha, (\phi, \tau_\nu, \tau_\kappa))} l_{rstuw}. \tag{38}
\]

26
Each of the terms in (38) are now computed. The details are not provided here, both to save space and also because the computation does not afford any special insight into the problem. The detailed calculations, however, are available from the authors. The terms in \( \sum_{(\alpha, (\phi, \tau_v, \tau_e))} (l_{rstu} - \hat{l}_{rstuvw}) \) can be decomposed in a manner similar to that in (38), except that in this case there are no terms in \( \beta \). When all these terms are put together, one gets

\[
C_\lambda = \frac{3}{4} + O\left(n^{-1}\right).
\]

(39)

The theorem now follows from (27), (31), (37) and (39).

**Proof of Theorem 3:**

The expansion of the distribution of the LRT and the Wald test under local Pitman alternatives is given in Hayakawa (1975), while 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 3, 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 \( \kappa_{ij} = n^{-1}E(\partial^2L/\partial\lambda_i\lambda_j) \) have been already obtained in (26). To obtain the quantities \( \kappa_{i,j,k} = n^{-1}E[(\partial L/\partial\lambda_i) (\partial^2 L/\partial\lambda_j\lambda_k)] \) and \( \kappa_{i,j,k} = n^{-1}E[(\partial L/\partial\lambda_i) (\partial L/\partial\lambda_j) (\partial L/\partial\lambda_k)] \), we exploit the Bartlett identities (Bartlett, 1953) to obtain

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

and

\[
\kappa_{i,j,k} = -\kappa_{ijk} + \frac{\partial \kappa_{jk}}{\partial \lambda_i}.
\]

The quantities \( \partial \kappa_{jk}/\partial \lambda_i \) are then calculated using (26). Using the same notation (note, however, that we define all our cumulants \( \kappa \) 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

\[
b_{10} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{11} = n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{12} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{13} = 0,
\]

\[
b_{20} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{21} = n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{22} = -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \ b_{23} = O\left(n^{-1}\right)
\]
and

\begin{align*}
    b_{30} &= -0.5n^{-1/2}C_1 + O\left(n^{-1}\right), \\
    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), \\
    b_{33} &= -0.5n^{-1/2}C_1 + O\left(n^{-1}\right).
\end{align*}

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

References


### Table I. Simulations Results $^3$: $\alpha = .95$

<table>
<thead>
<tr>
<th>innovations Gaussian errors</th>
<th>$t_5$ errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\hat{\beta}_{OLS}$</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
</tr>
<tr>
<td>50 bias</td>
<td>7.0933</td>
</tr>
<tr>
<td>s.d.</td>
<td>7.9388</td>
</tr>
<tr>
<td>100 bias</td>
<td>3.4734</td>
</tr>
<tr>
<td>s.d.</td>
<td>4.5113</td>
</tr>
<tr>
<td>150 bias</td>
<td>2.2471</td>
</tr>
<tr>
<td>s.d.</td>
<td>3.3079</td>
</tr>
</tbody>
</table>

### Table II. Simulations Results: $\alpha = .975$

<table>
<thead>
<tr>
<th>innovations Gaussian errors</th>
<th>$t_5$ errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\hat{\beta}_{OLS}$</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
</tr>
<tr>
<td>50 bias</td>
<td>7.5665</td>
</tr>
<tr>
<td>s.d.</td>
<td>7.6500</td>
</tr>
<tr>
<td>100 bias</td>
<td>3.7411</td>
</tr>
<tr>
<td>s.d.</td>
<td>4.1445</td>
</tr>
<tr>
<td>150 bias</td>
<td>2.4245</td>
</tr>
<tr>
<td>s.d.</td>
<td>2.9387</td>
</tr>
</tbody>
</table>

### Table III. Simulations Results: $\alpha = .99$

<table>
<thead>
<tr>
<th>innovations Gaussian errors</th>
<th>$t_5$ errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\hat{\beta}_{OLS}$</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
</tr>
<tr>
<td>50 bias</td>
<td>7.9383</td>
</tr>
<tr>
<td>s.d.</td>
<td>7.4703</td>
</tr>
<tr>
<td>100 bias</td>
<td>3.9753</td>
</tr>
<tr>
<td>s.d.</td>
<td>3.9270</td>
</tr>
<tr>
<td>150 bias</td>
<td>2.6113</td>
</tr>
<tr>
<td>s.d.</td>
<td>2.7212</td>
</tr>
</tbody>
</table>

$^3$In Tables I, II,III and IV, the true values of parameters are: $\beta = 0$, $\phi = -80$, $\sigma^2_v = 4.0681 \times 10^{-4}$, and $\sigma^2_e = 1$. 

32
### Table IV. Rejection Rates, Simulation Mean and Variance of RLRT – univariate regressor

<table>
<thead>
<tr>
<th>innovations</th>
<th>Gaussian errors</th>
<th>t&lt;sub&gt;5&lt;/sub&gt; errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>n</td>
<td>5%</td>
</tr>
<tr>
<td>.95</td>
<td>50</td>
<td>.0481</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.0477</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>.0496</td>
</tr>
<tr>
<td>.975</td>
<td>50</td>
<td>.0507</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.0494</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>.0489</td>
</tr>
<tr>
<td>.99</td>
<td>50</td>
<td>.0591</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.0507</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>.0504</td>
</tr>
</tbody>
</table>

### Table V. Simulation Results<sup>4</sup> – bivariate regressor

<table>
<thead>
<tr>
<th>diag A</th>
<th>(.95, .80)</th>
<th>(.95, .95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ&lt;sub&gt;v&lt;/sub&gt;</td>
<td>ß&lt;sub&gt;1, OLS&lt;/sub&gt;</td>
<td>ß&lt;sub&gt;2, OLS&lt;/sub&gt;</td>
</tr>
<tr>
<td>0 bias</td>
<td>1.8375</td>
<td>2.0377</td>
</tr>
<tr>
<td>s.d.</td>
<td>3.3518</td>
<td>5.4121</td>
</tr>
<tr>
<td>.5 bias</td>
<td>2.4394</td>
<td>1.2751</td>
</tr>
<tr>
<td>s.d.</td>
<td>4.3870</td>
<td>7.1407</td>
</tr>
<tr>
<td>.9 bias</td>
<td>3.9979</td>
<td>-1.2468</td>
</tr>
<tr>
<td>s.d.</td>
<td>6.5038</td>
<td>10.1373</td>
</tr>
</tbody>
</table>

### Table VI. Rejection Rates, Simulation Mean and Variance of RLRT– bivariate regressor

<table>
<thead>
<tr>
<th>diag A</th>
<th>(.95, .80)</th>
<th>(.95, .95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ&lt;sub&gt;v&lt;/sub&gt;</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>0</td>
<td>.0494</td>
<td>.0106</td>
</tr>
<tr>
<td>.5</td>
<td>.0536</td>
<td>.0112</td>
</tr>
<tr>
<td>.9</td>
<td>.0572</td>
<td>.0122</td>
</tr>
</tbody>
</table>

<sup>4</sup>In Tables V and VI, the true values of parameters are: β = 0, φ = c(−80, −80) and σ<sup>2</sup> = 1.
Figure 1: QQ plots of LRT from simulations of 10,000 repetitions with \( n = 50, 100 \) and 150. The vertical dashed lines are the 99th percentile of \( \chi_1^2 \). Data are generated from \( Y_t = \beta X_{t-1} + u_t \), where \( \beta = 0 \), \( X_t = \alpha X_{t-1} + v_t \) with \( \alpha = .95, .97 \) and .99, and \( \text{corr}(v_t, u_t) = -.85 \).
Q–Q Plots: simulated RLRT vs. theoretical $\chi^2$
5,000 repetitions, $n = 200$

$\alpha_{11} = 0.95, \alpha_{22} = 0.8$

$\rho_v = 0$

$\rho_v = 0.5$

$\rho_v = 0.9$

Figure 2: QQ plots of LRT from simulations of 5,000 repetitions with $n = 200$. The vertical dashed lines are the 99th percentile of $\chi^2$. Data are generated from $Y_t = \beta'X_{t-1} + u_t$, where $\beta = 0$, $X_t = A X_{t-1} + v_t$ with $A = \text{diag}(\alpha_{11}, \alpha_{22})'$ and $\rho_v = 0, .5 \text{ and } .9$.  

35