COINTEGRATION TESTING USING PSEUDOLIKELIHOOD RATIO TESTS

ANDRÉ LUCAS
Free University Amsterdam

This paper considers pseudomaximum likelihood estimators for vector autoregressive models. These estimators are used to determine the cointegration rank of a multivariate time series process using pseudolikelihood ratio tests. The asymptotic distributions of these tests depend on nuisance parameters if the pseudolikelihood is non-Gaussian. This even holds if the likelihood is correctly specified. The nuisance parameters have a natural interpretation and can be consistently estimated. Some simulation results illustrate the usefulness of the tests: non-Gaussian pseudolikelihood ratio tests generally have a higher power than the Gaussian test of Johansen if the innovations demonstrate leptokurtic behavior.

1. INTRODUCTION

The cointegration literature has been rapidly developing over the last decade. Much effort has been put into the construction of statistics for determining the number of cointegrating relationships and the exact form of these relationships (see, e.g., Engle and Granger, 1987; Phillips and Durlauf, 1986; Park and Phillips, 1988; Phillips, 1988, 1991; Johansen, 1988, 1989, 1991; Park, 1992; Boswijk, 1992; Bierens, 1994; Kleibergen and van Dijk, 1994; Stock, 1994). In all of these procedures the ordinary least-squares (OLS) estimator plays an important role. Skimming the empirical literature for applications of cointegration testing procedures, one finds that the likelihood-based testing procedure of Johansen (1988, 1991) is mostly used. This procedure of Johansen also heavily relies on the OLS estimator through the (OLS) residuals of certain preliminary regressions.

The OLS estimator has several advantages. First, the estimator has a closed form and is, therefore, easy to compute. It is also available in most statistical computer packages. Second, because of its simple form, an asymptotic analysis of the properties of the OLS estimator is analytically tractable.

149

Third, the estimator has some optimality properties (e.g., minimum variance) if the disturbances in the model are normally distributed.

The OLS estimator, however, also has certain disadvantages, especially from a sensitivity point of view. In the standard linear regression model with stationary variables, the variance of the OLS estimator quickly increases if the disturbances become more fat-tailed. Moreover, outliers and influential observations have a large impact on the estimates. For both issues, see Huber (1981) and Hampel, Ronchetti, Rousseeuw, and Stahel (1986).

This sensitivity of OLS can be avoided by considering alternative classes of estimators, such as maximum likelihood–type (M) estimators (Huber, 1981) and pseudomaximum likelihood (PML) estimators (White, 1982; Gouriéroux, Monfort, and Trognon, 1984). These alternative classes contain estimators that are less sensitive than OLS and, at the same time, have a reasonable efficiency if the errors are normally distributed. Some of these estimators outperform the OLS estimator in terms of efficiency if the errors in the model are non-Gaussian. The possible efficiency gain of using non-OLS estimators when the error process is non-Gaussian also holds in a context with nonstationary variables. Moreover, non-OLS estimators can provide protection against outliers and influential observations in the nonstationary setting as well (see, e.g., Lucas, 1995a; Hoek, Lucas, and van Dijk, 1995).

The main objective of this paper is to develop a cointegration testing procedure based on PML estimators (see Gouriéroux et al., 1984) and to study the properties of this procedure by means of an asymptotic analysis and simulations. The considered test is a generalization of the Gaussian likelihood ratio test of Johansen (1988, 1991) and uses the ratio of two, possibly non-Gaussian, pseudolikelihoods. The motivation for this approach is twofold. First, in many economic applications, for example, in finance (de Vries, 1994), the normality assumption for the error term is untenable. This leaves some room for improving the power properties of the cointegration test of Johansen. Second, dealing with outliers and influential observations in the data is a common feature of empirical econometric model building. A test that automatically corrects for some of these atypical observations seems a useful tool for the applied researcher.

This paper only considers vector autoregressive (VAR) time series models. This implies that the parametric approach of Johansen (1988, 1991) is used to test the cointegration hypothesis, in contrast to the semiparametric approach of, for example, Phillips (1987, 1988, 1991) and Phillips and Durlauf (1986). In the parametric framework, one can easily construct test statistics that, like those of Johansen, are based on the (pseudo)likelihood ratio principle.

The results in this paper include the following. First, new cointegration tests are developed based on the (pseudo)likelihood ratio principle. The relation of these tests to the likelihood ratio test of Johansen is established. Moreover, it is shown that the asymptotic distributions of the new cointegra-

tion tests depend on nuisance parameters if the pseudolikelihood is non-Gaussian. This holds even if the pseudolikelihood happens to coincide with the true likelihood. Some results on the optimal choice of the pseudolikelihood are also briefly mentioned. Second, a small simulation experiment is provided, illustrating that the new cointegration tests outperform the likelihood ratio test of Johansen in terms of power if the innovations are fat-tailed.

The paper is set up as follows. Section 2 discusses the model, the hypotheses of interest, and the class of PML estimators. Section 3 develops an asymptotic theory for the test statistic that is formulated in Section 2. It also establishes the relation between the asymptotic distribution of the new test and that of Johansen's test statistic. Section 4 contains a small simulation experiment, illustrating the performance of the different tests. Section 5 briefly discusses the problems of introducing deterministic regressors or additional nuisance parameters in the regression model. Concluding remarks are found in Section 6. The Appendix contains the proofs of all theorems in the main text.

The following notational conventions are adopted. Let $r < k$. Given a $(k \times r)$ matrix A of full column rank, $A_\perp$ denotes a $(k \times (k - r))$ matrix of full column rank, such that $A_\perp^T A = 0$. Moreover, A^T denotes the transpose of A. The limiting distributions of the test statistics in this paper are functionals of stochastic processes. If $W(s)$, $s \in [0,1]$, is a stochastic process (e.g., a Brownian motion), then stochastic integrals of the form $\int_0^1 W(s)\, dW(s)$ are denoted by $\int W\, dW$. Integrals with respect to Lebesgue measure are denoted analogously; for example, $\int WW^T$ denotes $\int_0^1 W(s)W(s)^T\, ds$. Finally, the symbol $\Rightarrow$ denotes weak convergence of probability measures (see Billingsley, 1968).

2. THE MODEL AND TEST STATISTICS

Consider the VAR model of order $p + 1$,

$$\Delta y_t = \Pi y_{t-1} + \Phi_1 \Delta y_{t-1} + \cdots + \Phi_p \Delta y_{t-p} + \varepsilon_t, \tag{1}$$

where y_t and ε_t are column vectors of dimension k, Δ is the first difference operator, $\Delta y_t = y_t - y_{t-1}$, and Π, $\Phi_1, \ldots, \Phi_p$, are parameter matrices. The random vectors ε_t are assumed to satisfy the following assumption.

Assumption 1.

(i) $\{\varepsilon_t\}_{t=0}^{\infty}$ is an independent and identically distributed (i.i.d.) process with density function $f(\varepsilon_t)$.

(ii) $E(\varepsilon_0) = 0$.

(iii) $\Omega_{11} = E(\varepsilon_0 \varepsilon_0^T)$ is positive definite.

In empirical applications, deterministic functions of time are usually added to (1). The complexities induced by these additional regressors are not of prime interest here. Therefore, their introduction is delayed until Section 5.

This paper concentrates on the determination of the rank of Π. The following assumption is maintained throughout the paper.

Assumption 2.

   (i) $|I - (I + \Pi)z - \sum_{i=1}^{p} \Phi_i(1 - z)z^i| = 0$ implies either $|z| > 1$ or $z = 1$.
   (ii) The elements of y_t are integrated to at most order one.

Assumption 2 guarantees that the rank of Π coincides with the number of stationary relations among the elements of y_t (see Johansen, 1988, 1991). Part (i) of Assumption 2 implies that the nonstationary behavior of y_t can be removed by differencing. Part (ii) states that first-order differencing suffices for obtaining stationarity. An important consequence of Assumption 2 is that Π can be decomposed as $\Pi = AB^{\mathsf{T}}$, with A and B two matrices of full column rank r, with r denoting the number of cointegrating relations. The matrices A and B and their orthogonal complements frequently show up in the asymptotic distribution theory in Section 3.

To determine the rank of Π, estimates of the parameters in (1) are needed. Johansen (1988, 1991) assumes that the errors ε_t are i.i.d. and follow a Gaussian distribution. One can then use the conditional Gaussian maximum likelihood estimation to obtain the required estimates. In fact, this amounts to running several OLS regressions and solving a generalized eigenvalue problem. Due to both its computational ease and its theoretical foundation in the likelihood principle, the procedure of Johansen has become very popular.

As mentioned in Section 1, the normality assumption for the error term is questionable in many empirical applications. If the errors are non-Gaussian, one expects to gain efficiency by exploiting the distributional properties of the innovations. Therefore, we consider the class of PML estimators as an alternative to the Gaussian maximum likelihood estimator of Johansen. The pseudolikelihood is assumed to be of the form

$$\mathcal{L}_T(\theta) \propto \prod_{t=1}^{T} |\Omega_{11}|^{-1/2} \cdot \exp(-\rho(\Omega_{11}^{-1/2}\varepsilon_t)), \tag{2}$$

where $\rho(\cdot)$ is a function that satisfies Assumption 3 in Section 3, later, Ω_{11} is defined in Assumption 1, $\varepsilon_t = \Delta y_t - \Pi y_{t-1} - \Phi_1 \Delta y_{t-1} - \cdots - \Phi_p \Delta y_{t-p}$, and θ is the vector of unknown parameters. The matrix Ω_{11} is estimated along with the parameters in (1). The pseudolikelihood is allowed to be improper in the sense that $\int \exp(-\rho(\Omega_{11}^{-1/2}\varepsilon_t))d\varepsilon_t$ need not exist. In this way, pseudolikelihoods with a redescending[1] score function are also covered by the results in this paper. The PML estimator is given by the vector $\hat{\theta}_T$ that maximizes $\ell_T(\theta) = \ln(\mathcal{L}_T(\theta))$. Note that (2) comprises most likelihood functions that are used in the literature. The Gaussian maximum likelihood estimator of Johansen (1988, 1991), for example, is obtained by setting $\rho(e) = e^{\mathsf{T}}e/2$. Also, the Student t maximum likelihood estimator, as discussed by Prucha and Kelejian (1984), and the class of maximum likeli-

hood–type, or M, estimators (see Hampel et al., 1986; Huber, 1981) are contained as special cases of (2).

The hypotheses of interest concern the number of cointegrating relationships. As this number can range from 0 to k, there are k hypotheses of interest. The rth hypothesis postulates that there are at most r cointegrating relationships, $H_r: \text{rank}(\Pi) \leq r$, with $r = 0, \ldots, k - 1$. The alternative hypothesis in each case is $H_k: \text{rank}(\Pi) = k$. In the framework of Johansen (1991), testing H_r versus H_k leads to his trace test statistic. Alternatively, the hypothesis H_r could be tested against the alternative H_{r+1}. This would produce the test given in formula (2.14) of Johansen (1991). In the present paper, we only discuss the first set of null hypotheses and alternatives. The results can, however, easily be extended to tests of H_r versus H_{r+1}.

Using the pseudo-log likelihood, $\ell_T(\theta)$, we can use the likelihood ratio principle[2] to test H_r versus H_k. Let $\bar{\theta}_{T,r}$ and $\hat{\theta}_T$ denote the PML estimates under the null hypothesis, H_r, and under the alternative hypothesis, H_k, respectively. The pseudolikelihood ratio (PLR) test is given by

$$PLR_r = 2(\ell_T(\hat{\theta}_T) - \ell_T(\bar{\theta}_{T,r})) \tag{3}$$

(see White, 1982). A subscript r is added to the test statistic in order to indicate the null hypothesis that is tested. If no confusion is caused, this subscript is omitted. The limiting distribution of PLR_r is derived in the next section.

3. ASYMPTOTIC DISTRIBUTION THEORY

In this section, the asymptotic distribution of the PLR test statistic is discussed under the null hypothesis. Apart from the conditions on the behavior of ε_t stated in Assumption 1, some regularity conditions are needed for the function $\rho(\cdot)$ used in the definition of the pseudolikelihood in (2). Assume that the following conditions are satisfied.

Assumption 3.

(i) $\rho(\cdot)$ is twice continuously differentiable; the first- and second-order derivatives with respect to ε_t are denoted by $\psi(\Omega_{11}^{-1/2}\varepsilon_t) = \partial\rho(\Omega_{11}^{-1/2}\varepsilon_t)/\partial\varepsilon_t$ and $\psi'(\Omega_{11}^{-1/2}\varepsilon_t) = \partial\psi(\Omega_{11}^{-1/2}\varepsilon_t)/\partial\varepsilon_t^\top$, respectively.

(ii) $\psi'(\Omega_{11}^{-1/2}\varepsilon_t)$ is first-order Lipschitz.

(iii) $E(\psi(\Omega_{11}^{-1/2}\varepsilon_0)) = 0$.

(iv) $E(\psi'(\Omega_{11}^{-1/2}\varepsilon_0)) = C_1$, with $|C_1| \neq 0$.

(v) The random vector $\psi(\Omega_{11}^{-1/2}\varepsilon_0) \otimes \varepsilon_0$ has finite second moments.

(vi) $E(\psi'(\Omega_{11}^{-1/2}\varepsilon_0) \otimes \varepsilon_0) = 0$.

Parts (i) and (ii) of Assumption 3 impose some smoothness conditions on the pseudolikelihood. These conditions can be weakened. Discontinuities in the function ψ, for example, can be coped with if the density of ε_t is sufficiently smooth. This is seen by comparing the results of Herce (1994) for the least absolute deviations estimator with those of Lucas (1995b) for smooth

M estimators. If allowance is made for these types of discontinuities, the methods of proof have to be changed considerably. Therefore, attention in this paper is restricted to smooth versions of ρ. Part (iii) of Assumption 3 is another centering condition in order to guarantee the consistency of the PML estimator. Part (iv) implies that the PML estimator can be approximated using a first-order Taylor series expansion of the first-order condition that defines the estimator. Part (v) is a moment condition. For the Gaussian PML estimator, it states that fourth-order moments of the errors exist. The condition is somewhat too strict and is mainly used to facilitate the proofs of the theorems. It can, for example, be replaced by the conditions that $\psi(\Omega_{11}^{-1/2}\varepsilon_0)$ has finite second-order moments and that $\partial \ell_T(\theta)/\partial \Omega_{11}$ has finite + first-order moments. Note that part (v) implies that the second-order moment of $\psi(\Omega_{11}^{-1/2}\varepsilon_0)$ exists and is finite. Finally, part (vi) implies that we can abstract from the fact that Ω_{11} is estimated rather than known. If this part of Assumption 3 is not met, the limiting distribution of the *PLR* test changes. Note that part (vi) is satisfied if both $\rho(\cdot)$ and $f(\cdot)$ are even, that is, $f(\varepsilon_t) = f(-\varepsilon_t)$, and if the appropriate moments exist.

The next lemma follows directly from Johansen (1988) and Phillips and Durlauf (1986). Therefore, its proof is omitted.

LEMMA 1. *Given Assumptions 1–3*,

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} (\varepsilon_t^{\mathsf{T}}, \psi(\Omega_{11}^{-1/2}\varepsilon_t)^{\mathsf{T}}) \Rightarrow (W_1(s)^{\mathsf{T}}, W_2(s)^{\mathsf{T}}),$$

with $(W_1(s)^{\mathsf{T}}, W_2(s)^{\mathsf{T}})^{\mathsf{T}}$ *a multivariate Brownian motion with covariance matrix*

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

with $s \in [0,1]$, $\Omega_{22} = E(\psi(\Omega_{11}^{-1/2}\varepsilon_0)\psi(\Omega_{11}^{-1/2}\varepsilon_0)^{\mathsf{T}})$, *and* $\Omega_{12} = \Omega_{21}^{\mathsf{T}} = E(\varepsilon_0 \psi(\Omega_{11}^{-1/2}\varepsilon_0)^{\mathsf{T}})$. *Moreover, if we define* $\Psi = \sum_{i=1}^{p+1} i \Pi_i, \Pi_1 = I_k + \Pi + \Phi_1$, $\Pi_{p+1} = -\Phi_p$, *and* $\Pi_i = \Phi_i - \Phi_{i-1}$ *for* $i = 2, \ldots, p$, *then*

$$T^{-1/2}(B_\perp^{\mathsf{T}} B_\perp)^{-1} B_\perp^{\mathsf{T}} y_{\lfloor sT \rfloor} \Rightarrow U(s) \equiv (A_\perp^{\mathsf{T}} \Psi B_\perp)^{-1} A_\perp^{\mathsf{T}} W_1(s).$$

To establish the limiting behavior of the *PLR* statistic, it is useful to define the matrix

$$S_0 = (A_\perp^{\mathsf{T}} \Omega_{11} A_\perp)^{-1/2} A_\perp^{\mathsf{T}} \Omega_{12} C_1^{-1} A_\perp (A_\perp^{\mathsf{T}} C_1^{-1} \Omega_{22} C_1^{-1} A_\perp)^{-1/2}, \tag{4}$$

which is the correlation matrix between $A_\perp^{\mathsf{T}} \varepsilon_t$ and $A_\perp^{\mathsf{T}} C_1^{-1} \psi(\Omega_{11}^{-1/2}\varepsilon_t)$. Let $S_1 R S_2^{\mathsf{T}}$ denote the singular value decomposition of S_0, with S_1 and S_2 two orthogonal matrices and R a diagonal matrix containing the absolute values of the canonical correlations between $A_\perp^{\mathsf{T}} \varepsilon_t$ and $A_\perp^{\mathsf{T}} C_1^{-1} \psi(\Omega_{11}^{-1/2}\varepsilon_t)$. We can now obtain the limiting behavior of the *PLR* statistic.

THEOREM 1. *Let $\hat{\varepsilon}_t$ and $\tilde{\varepsilon}_t$ denote the residuals calculated at $\hat{\theta}_T$ and $\bar{\theta}_{r,T}$, respectively. Given Assumptions 1–3, and $\hat{\varepsilon}_t - \varepsilon_t = o_p(1)$ and $\tilde{\varepsilon}_t - \varepsilon_t = o_p(1)$ uniformly in t, then $PLR \Rightarrow \overline{PLR}$, with*

$$\overline{PLR} = \mathrm{tr}\left(\bar{K}_0\left(\int V_1 \, dV_2^\top\right)^\top \left(\int V_1 V_1^\top\right)^{-1} \left(\int V_1 \, dV_2^\top\right)\right),$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $V_1(s)$ and $V_2(s)$ are two standard Brownian motions with diagonal correlation matrix R (defined immediately after (4)), and

$$K_0 = (A_\perp^\top C_1^{-1} A_\perp)^{-1},$$

$$\bar{K}_0 = (A_\perp^\top C_1^{-1} \Omega_{22} C_1^{-1} A_\perp),$$

$$\bar{K}_0 = S_2^\top \bar{K}_0^{1/2} K_0 \bar{K}_0^{1/2} S_2.$$

Remark 1. The additional two conditions in Theorem 1 ensure that the correct optimum of (2) is chosen from the (possibly large) set of local optima. In fact, the conditions imply that A, $\Phi_1, \ldots, \Phi_p$, and Ω_{11} be consistently estimated, while the actual unit root parameters and the cointegrating vectors (α_{22} and β in the Appendix) are consistently estimated at a rate higher than $T^{1/2}$. Low-level conditions for consistency in a (possibly) nonlinear context can be found in, for example, Gallant (1987). Note that it is nontrivial to check the additional conditions in practice. One way to circumvent the whole consistency problem associated with local optima is to use a simple consistent estimator as a starting value for a one-step Newton–Raphson improvement of the objective function in (2). Such a one-step estimator would have the same asymptotic properties as its fully iterated counterpart. As a starting value, one could, for example, use the Gaussian-based estimator of Johansen (1988).

Remark 2. Theorem 1 gives the asymptotic distribution under the null hypothesis H_r. It is also possible to derive the asymptotic distribution of the *PLR* test under local alternatives as in Johansen (1989) and Rahbek (1994). This is done in Lucas (1996).

Remark 3. A question that is not addressed explicitly in the present paper concerns the optimal choice of the pseudolikelihood. In the present context of cointegration testing, it is nontrivial to find a satisfactory definition of optimality that allows us to solve for the optimal pseudolikelihood. Following Cox and Llatas (1991), Lucas (1996) uses the criterion of asymptotic mean squared error (AMSE) of the estimators for the unit root parameters (α_{22} in the Appendix) in order to determine the optimal pseudolikelihood. His results show that the optimal pseudoscore from an AMSE perspective is a linear combination of the Gaussian score and the true likelihood score.

The weight of the Gaussian pseudoscore in this linear combination decreases if either the innovations ε_t become more fat-tailed or one gets farther away from the null hypothesis of no cointegration. This clearly suggests that the use of non-Gaussian pseudolikelihoods can improve the power properties of the *PLR* cointegration test in situations with fat-tailed innovations.

It is illustrative to consider the two main differences between Theorem 1, earlier, and Theorem 3 of Johansen (1988). Johansen assumes Gaussian error terms, thus imposing $\psi(\Omega_{11}^{-1/2}\varepsilon_t) = \Omega_{11}^{-1}\varepsilon_t$. It then follows that the Brownian motion $W_2(s)$ is a linear transformation of the Brownian motion $W_1(s)$; in particular, $W_2(s) = \Omega_{11}^{-1}W_1(s)$. Moreover, $S_0 = R = I_{k-r}$, where I_{k-r} denotes the unit matrix of order $(k-r)$. The two Brownian motions V_1 and V_2 of Theorem 1 are then perfectly correlated. If a different specification is chosen for ψ, V_1 and V_2 are imperfectly correlated, which leads to a more complicated expression for the asymptotic distribution of *PLR*. Note that for $\psi(\Omega_{11}^{-1/2}\varepsilon_t) = \Omega_{11}^{-1}\varepsilon_t$ it also follows that $C_1 = \Omega_{11}^{-1}$, and thus $\bar{K}_0 = I_{k-r}$.

Another way to explicate the difference between the general *PLR* test and the Gaussian one follows by defining the Brownian motion $V_3(s) = V_2(s) - RV_1(s)$. Note that $V_3(s)$ and $V_1(s)$ are independent. The stochastic integral $\int V_1 \, dV_2^\top$ in Theorem 1 can now be split into two parts. The first part is $\int V_1 \, dV_1^\top R$, which is a Gaussian functional (see Phillips, 1991). The second part is $\int V_1 \, dV_3^\top$ and is mixed normally distributed with zero mean. Given $V_1(s)$, the variance of $\int V_1 \, dV_3^\top$ is $(I - R^2) \otimes (\int V_1 V_1^\top)$. From this decomposition of $\int V_1 \, dV_2^\top$, it can be seen that for smaller values of R the Gaussian functional becomes less important in the limiting distribution of the *PLR* test, while the mixed normal random variate becomes more important. As mentioned in the previous paragraph, for the Gaussian *PLR* test R is at its maximum: $R = I$. By definition, it then follows that $\int V_1 \, dV_3(s)^\top$ is identically equal to zero. In contrast, if R is at its minimum $R = 0$, then the Gaussian functional vanishes and the limiting distribution of the *PLR* test becomes χ^2 with $(k-r)^2$ degrees of freedom. This happens, for example, if $\psi(\cdot)$ is bounded and if ε_t has infinite variance (for the univariate case, cf. Knight, 1989).

An important result that follows from Theorem 1 is that the limiting distribution of the *PLR* test depends on nuisance parameters, namely, $\bar{K}_0$ and R. The nuisance parameters in $\bar{K}_0$ reflect the discrepancy between the pseudolikelihood and the true likelihood. As was already noted in White (1982), misspecification of the likelihood causes a breakdown of the information matrix equality. In the present setting, this implies $C_1 \neq \Omega_{22}$ if the pseudolikelihood does not coincide with the true likelihood. Therefore, the matrix $\bar{K}_0$ reduces to the identity matrix if the likelihood is correctly specified. The second set of nuisance parameters, present in the matrix R, reflects the effect of using a non-Gaussian PML estimator. It is interesting to note

that both sets of nuisance parameters disappear if one uses the Gaussian PML estimator of Johansen (1988). One of the peculiar findings in this paper is that nuisance parameters remain present in the limiting distribution of *PLR* even if the pseudolikelihood coincides with the true likelihood. This result is presented in the following corollary.

COROLLARY 1. *If* ε_t *has a density* $f(\varepsilon_t) = c|\Omega_{11}|^{-1/2} \exp(-\rho(\Omega_{11}^{-1/2}\varepsilon_t))$, *where c is such that* $\int f(\varepsilon_t)\, d\varepsilon_t = 1$, *then*

$$PLR \Rightarrow \mathrm{tr}\left(\left(\int V_1\, dV_2^\top\right)^\top \left(\int V_1 V_1^\top\right)^{-1} \left(\int V_1\, dV_2^\top\right)\right),$$

with V_1 *and* V_2 *two standard Brownian motions such that* $E(V_1(s)V_2(s)^\top) = sR$, *where R is defined immediately after* (4).

Corollary 1 states that if the pseudolikelihood is correctly specified, then the asymptotic distribution of the *PLR* statistic depends on nuisance parameters only through the canonical correlations between $A_\perp^\top \varepsilon_t$ and $A_\perp^\top C_1^{-1} \psi(\Omega_{11}^{-1/2}\varepsilon_t)$. For the Gaussian pseudolikelihood, these correlations are equal to unity. In most other circumstances, however, the correlations are less than unity, which results in a more complicated asymptotic distribution of the test statistic.[3]

Corollary 1 can be used to simulate critical values of *LR* cointegration tests for (correctly specified) non-Gaussian likelihoods. A procedure for obtaining consistent estimates of these critical values is fairly straightforward. For given parameter estimates, the matrix S_0 in (4) can be consistently estimated: replace Ω_{11} by $T^{-1}\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t^\top$, C_1 by $T^{-1}\sum_{t=1}^T \psi'(\hat{\Omega}_{11}^{-1/2}\hat{\varepsilon}_t)$, and so on, with $\hat{\varepsilon}_t$ denoting the *t*th regression residual. The estimate of S_0 can then be used to estimate the canonical correlations R by means of a singular value decomposition. Let $\hat{R}$ denote a diagonal matrix containing the estimated singular values of S_0. Then the critical values of $\overline{PLR}$ can be simulated in the usual way by generating random walks $\hat{V}_1$ and $\hat{V}_2$ with correlation matrix $\hat{R}$ and replacing the integrals in Corollary 1 by sums. Note that this methodology can be extended to simulate the critical values of the *PLR* test for misspecified pseudolikelihoods. In that case, a consistent estimate of $\tilde{K}_0$ is also needed. Such an estimate can be constructed in the same way as above using the residuals $\hat{\varepsilon}_t$.

The above procedure for computing critical values has two major drawbacks. First, critical values have to be simulated for every estimate of R. This might prove too time consuming for useful practical purposes. Second and more important, the critical values of $\overline{PLR}$ provide poor approximations to the critical values of the *PLR* test in finite samples (see Lucas, 1996). Therefore, Section 4 proceeds by directly simulating the *PLR* test in order to obtain the critical values.

4. SIMULATION RESULTS

In this section, we present the results of a small simulation experiment. The experiment only serves as an illustration of the properties of some simple *PLR* tests relative to the Johansen test.

We start by simulating the distribution of the *PLR* statistic under the null hypothesis in order to obtain the critical values of the test. As was mentioned in Section 3, this can be done by approximating the distribution of $\overline{PLR}$ in Theorem 1. Lucas (1996), however, shows that the critical values of $\overline{PLR}$ provide poor approximations to the critical values of the *PLR* test in finite samples. Therefore, we compute the *PLR* test directly for many simulated time series and use the computed values to approximate the distribution of the test. Note that this involves solving two (possibly) nonlinear maximization problems for each simulation, which makes the whole experiment very time consuming. The setup is as follows. For several values of k, we generate a k-variate random walk y_t, $t = 0, \ldots, T$, with standard Gaussian innovations. Using the generated time series y_t, we compute PLR_0, which tests the hypothesis of 0 cointegrating relations versus k cointegrating relations. This is done over N Monte Carlo simulations. We set $T = 100$ and $N = 1,000$.

To illustrate the properties of the *PLR* test, we only consider a very simple pseudolikelihood, namely, the multivariate Student t:

$$\rho(\Omega_{11}^{-1/2}\varepsilon_t) = \tfrac{1}{2}(\nu + k)\ln(1 + \varepsilon_t^{\mathsf{T}}\Omega_{11}^{-1}\varepsilon_t/(\nu - 2)).$$

We restrict our attention to $\nu = 1, 3, 5, 10, \infty$. Note that setting $\nu = \infty$ yields the Gaussian PML estimator of Johansen. Also note that setting $\nu = 1$ and $\nu = 3$ only determines the form of the pseudolikelihood that is used and not the distribution of the innovations. The results of the level simulations are summarized in Table 1.

One feature that appears from Table 1 is that the distribution of the *PLR* test shifts to the right if either the degrees of freedom parameter, ν, decreases or the dimension of the time series, $k - r$, increases. The effect of a decrease in ν is larger in higher dimensions.

Note that the quantiles of the non-Gaussian *PLR* tests depend on the nuisance parameters in $\bar{K}_0$, because the true (Gaussian) likelihood does not coincide with the postulated (Student t) pseudolikelihood. This problem can be solved by generating the random walks for the Monte Carlo simulations using Student t–distributed innovations. This would generally shift the quantiles of the non-Gaussian *PLR* tests to the left. As is argued next, however, the critical values in Table 1 are safer to use in empirical analyses than critical values based on Student t random walks. When using the critical values from Table 1, the tests have approximately the correct size for thin-tailed observations. For fat-tailed observations, the tests are generally conservative, although their power behavior is still superior to that of the Gaussian *PLR* test for a large range of (local) alternatives.

TABLE 1. Critical values of the *PLR* test for
the Student t pseudolikelihood[a]

ν	$k - r$	Quantile							
		0.500	0.600	0.700	0.800	0.900	0.950	0.975	0.990
∞	1	0.59	0.91	1.38	1.98	3.05	4.10	5.61	7.82
	2	5.81	6.66	7.60	8.71	10.76	12.68	14.55	17.09
	3	14.89	16.23	17.89	20.02	22.63	25.03	27.56	29.26
	4	28.16	29.99	32.06	34.38	38.65	41.57	43.96	47.45
10	1	0.59	0.90	1.33	2.02	3.28	4.20	5.52	8.21
	2	5.85	6.89	7.91	9.10	11.10	13.37	15.17	17.65
	3	15.46	16.73	18.42	20.44	23.27	25.88	28.56	32.03
	4	28.72	30.60	32.95	35.58	39.70	43.00	46.22	49.21
5	1	0.59	0.94	1.38	2.15	3.38	4.65	5.99	8.62
	2	6.05	7.11	8.25	9.54	11.81	13.88	16.07	18.22
	3	16.09	17.48	19.17	21.35	24.42	26.97	29.79	33.79
	4	29.80	31.68	33.82	36.79	41.36	45.06	48.00	51.09
3	1	0.62	0.99	1.47	2.34	3.59	4.85	6.87	9.17
	2	6.39	7.36	8.66	10.36	12.71	14.81	17.50	19.22
	3	16.81	18.32	20.12	22.33	25.42	28.34	31.65	35.03
	4	30.96	33.00	35.20	38.25	43.53	46.83	50.56	54.03
1	1	0.88	1.36	2.02	2.96	5.14	6.88	8.47	12.06
	2	7.50	9.03	10.73	12.88	15.63	18.84	21.28	24.65
	3	19.28	21.22	23.75	26.42	30.11	33.26	37.43	42.10
	4	34.33	36.85	39.85	43.58	48.98	54.27	57.86	61.61

[a]The degrees of freedom parameter of the Student t pseudolikelihood is denoted by ν. k is the dimension of the time series and r is the cointegrating rank. The critical values were obtained using 1,000 Monte Carlo simulations with multivariate Gaussian random walks of length 100.

To illustrate the power of the Student t-based *PLR* test relative to the Gaussian one, consider the following simple data-generating process:

$$\Delta \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} -\tilde{c}_2/T \\ 0 \end{pmatrix} (y_{1,t-1} - y_{2,t-1}) + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{5}$$

where $\tilde{c}_2$ is a positive constant. The two roots of the VAR polynomial in (5) are 1 and $(1 - \tilde{c}_2/T)^{-1}$, respectively. For $0 \le \tilde{c}_2 \le 2T$, both roots lie on or outside the unit circle. If $\tilde{c}_2 = 0$, then the system has two unit roots and there is no cointegration. If $0 < \tilde{c}_2 < 2T$, there is one cointegrating relationship. Models such as (5) are often used in an analysis of cointegration tests under local alternatives (see Phillips, 1988; Johansen, 1989; Rahbek, 1994; Lucas, 1996).

We consider three different test statistics, namely, the Gaussian *PLR* test and the Student *t*–based *PLR* test with 5 degrees and 1 degree of freedom, respectively. We simulate the rejection frequencies of these tests in the usual way. After generating a time series using (5), we compute each of the preceding test statistics and compare them with their 5% and 10% critical values from Table 1. The simulations use time series of length $T = 100$, while $N = 1,000$ Monte Carlo replications are performed. Therefore, the standard errors of the rejection frequencies are smaller than or equal to $0.5N^{-1/2} \approx 0.016$.

Using the data-generating process in (5), two experiments are performed. In the first experiment, the ε_t are drawn from a bivariate normal distribution with mean zero and covariance matrix I_2. The restriction of the covariance matrix to be the unit matrix is unimportant in the present setup because of the presence of the scaling matrix Ω_{11} in the pseudolikelihood. For Gaussian ε_t, it follows from Lucas (1996) that the optimal pseudoscore function from a minimum mean squared error perspective is the Gaussian score function, $\psi(\Omega_{11}^{-1/2}\varepsilon_t) = \Omega_{11}^{-1}\varepsilon_t$. Therefore, we expect the Johansen or Gaussian *PLR* test to have the highest power in this case.

In the second experiment, the Gaussian distribution for ε_t is replaced by the truncated bivariate Cauchy distribution. The Cauchy was truncated to the set $\{\varepsilon_{1t}^2 + \varepsilon_{2t}^2 \leq F_{0.95}(2,1)\}$, where $F_{0.95}(2,1)$ is the 95th percentile of the F distribution with 2 degrees and 1 degree of freedom, respectively. The truncation is introduced in order to guarantee the existence of a sufficient number of moments (cf. Assumption 1). As mentioned in Remark 3, one can expect that a power gain for the *PLR* test can be realized by exploiting the non-Gaussian form of the truncated Cauchy distribution. Note, however, that for the truncated Cauchy distribution the critical values in Table 1 are in fact inappropriate, as they were generated using the Gaussian distribution. This is discussed in more detail later.

The results of these two experiments are given in Table 2. We first discuss the experiment with the Gaussian innovations. At $\tilde{c}_2 = 0$, the rejection frequency should be equal to the size of the test. It appears that all tests are a bit undersized, but this is probably due to the limited number of Monte Carlo replications used for both Tables 1 and 2. For departures from the null hypothesis, the number of rejections increases. As expected, the rejection frequencies for the Gaussian *PLR* test are higher than those of the other tests. Furthermore, for $\tilde{c}_2 \geq 5$, the rejection frequencies are increasing in the degrees of freedom parameter ν. This follows from the fact that the efficiency of the Student t PML estimators relative to the optimal (Gaussian) estimator is increasing in the degrees of freedom parameter ν.

We now turn to the results of the second experiment, the one with the truncated Cauchy innovations. The first thing to notice is that the non-Gaussian *PLR* tests have an actual size below the nominal size. This is due to the fact that the critical values in Table 1 are based on Gaussian innovations. As

TABLE 2. Rejection frequencies of the *PLR* tests[a]

$\bar{c}_2$	Gaussian			Truncated Cauchy		
	PLR_G	PLR_5	PLR_1	PLR_G	PLR_5	PLR_1
			10% Level			
0	0.077	0.065	0.069	0.103	0.026	0.021
1	0.103	0.110	0.123	0.103	0.031	0.041
5	0.295	0.280	0.243	0.287	0.458	0.594
10	0.734	0.662	0.502	0.689	0.905	0.921
20	0.994	0.982	0.865	0.984	0.999	0.999
			5% Level			
0	0.031	0.034	0.024	0.044	0.010	0.003
1	0.048	0.053	0.060	0.052	0.009	0.018
5	0.185	0.177	0.133	0.170	0.302	0.441
10	0.571	0.502	0.337	0.549	0.795	0.860
20	0.973	0.937	0.763	0.957	0.998	0.998

[a]The table contains the rejection frequencies of the Gaussian-based *PLR* test (PLR_G) and the Student t-based *PLR* tests with 5 degrees (PLR_5) and 1 degree (PLR_1) of freedom. The hypothesis of no cointegrating relationships (H_0) is tested against the alternative of stationarity (H_2). The data-generating process is (5) with either Gaussian or truncated Cauchy innovations ε_t. The entries are based on 1,000 Monte Carlo simulations with time series of length $T = 100$.

was mentioned earlier for the level simulations, the critical values of non-Gaussian *PLR* tests in situations with fat-tailed innovations are to the left of the ones in Table 1. As a result, the use of the critical values from Table 1 in this case leads to a conservative cointegration testing procedure.

Despite the low rejection frequencies under the null, the power of the non-Gaussian *PLR* tests very rapidly exceeds the power of the Johansen test if we consider local departures from the null hypothesis. For $\bar{c}_2 = 5$, the rejection frequencies of PLR_5 and PLR_1 are already, respectively, 1.5 and 2.5 times as high as that of PLR_G. This demonstrates that it is worthwhile to exploit the nonnormality of the innovations in order to increase the power of cointegration tests. The power can be increased further if the discrepancy between the actual and nominal sizes of the tests is eliminated, by, for example, generating critical values using Student t innovations. The use of such alternative critical values, however, would also severely distort the size of the non-Gaussian tests in situations with thin-tailed observations. Therefore, we choose to use the values in Table 1. This gives us testing procedures with approximately the correct size for thin-tailed innovations. For fat-tailed innovations, the tests are conservative, but the power behavior of the tests is still better than that of the Gaussian test for a large range of local alternatives.

5. MODEL EXTENSIONS

This section briefly discusses two possible model extensions and their effect on the asymptotic distribution of the *PLR* statistic. First, we treat the consequences of including deterministic functions of time as additional regressors in (1). Second, we discuss the effect of incorporating additional unknown nuisance parameters in the pseudolikelihood (2).

It is well-known that the incorporation of deterministic time trends in (1) complicates the asymptotic analysis. For example, if the data-generating process is (1) and if we use a regression model that contains a constant term in addition to the regressors in (1), then the Brownian motion $V_i(s)$ in Theorem 1 has to be replaced by the demeaned stochastic process $V_1(s) - \int_0^1 V_1(s)\,ds$. Similarly, the presence of a linear time trend as an additional regressor causes the entrance of detrended stochastic processes in the limiting distribution of *PLR*. The results get even more complicated if we allow for deterministic components to be present in the data-generating process in (1) instead of in the fitted regression model only. A well-known example of such a process is the random walk with nonzero drift. For such processes, the interpretation of the deterministic components and their effect on the asymptotic distributions is a delicate matter (see Johansen, 1994).

All of the preceding points have been addressed in the literature for multivariate time series in the context of Gaussian PML estimators. These results carry over in a straightforward manner to the present context of non-Gaussian PML estimators. This is illustrated by the results in Lucas (1996). Consequently, the results of Rahbek (1994) for the power of the Gaussian *PLR* test in the presence of nonzero drift terms in (1) also go through. This leaves us with the dilemma of choosing the appropriate additional deterministic regressors. If one chooses too few of them, inference is, in general, asymptotically biased. If one chooses the correct regressors, the test statistics are not asymptotically similar (see Johansen, 1991, Theorems 2.1 and 2.2). Finally, if one incorporates too many deterministic functions of time as additional regressors, the power of the *PLR* test diminishes (see Rahbek, 1994).

A second type of model extension concerns the presence of additional nuisance parameters in the pseudolikelihood. So far, we have only dealt with the presence of a scaling matrix Ω_{11}. This matrix could be estimated along with the other parameters under suitable regularity conditions (see Assumption 3 and the Appendix). From the proof of Theorem 1, it is seen that the appropriately normalized Hessian of the pseudolikelihood is asymptotically block diagonal between Ω_{11} and the parameters of interest for constructing the *PLR* test. Consequently, we could also use a consistent preliminary estimate of Ω_{11} in the construction of the *PLR* test without altering the asymptotic distribution of the test. This finding can easily be generalized toward cases where additional nuisance parameters are present in the pseudolikelihood. A simple example is given by the Student *t* pseudolikelihood, where

the degrees of freedom parameter ν is unknown. In the extreme case, one can treat the density of the innovations as an (infinite-dimensional) nuisance parameter. This would result in an adaptive cointegration test (cf. Manski, 1984, and note 3).

One can think of three strategies for dealing with unknown nuisance parameters. First, one can set the nuisance parameters to some user-defined values. This strategy may prove useful if one only uses the *PLR* test for protection against outliers and leptokurtosis or for checking the results of a Gaussian-based modeling exercise (see Franses and Lucas, 1995). The nuisance parameters can then be regarded as a type of tuning constants. This way of tackling the problem is often encountered in robust statistics. Second, one can use preliminary consistent estimates of the parameters in order to eliminate them. Third, one may want to estimate the nuisance parameters along with the other parameters in (1) by formulating the relevant pseudo-score equations. Such estimators are consistent under conventional regularity conditions (cf. Assumption 3 and the Appendix).

If one uses the first of these three strategies, the asymptotic distribution of the *PLR* test remains unaltered. For each of the other two strategies, it is often sufficient that $E(\partial^2 \psi (\Omega_{11}^{-1/2} \varepsilon_t)/\partial \varepsilon_t^T \partial \nu) = 0$, where ν now denotes the *complete* vector of nuisance parameters. This condition is met for a large class of pseudolikelihoods of the form in (2). For example, the Student t pseudolikelihood obviously satisfies this condition if the distribution of ε_t is spherically symmetric.

6. CONCLUSIONS

In this paper, we have studied the properties of likelihood ratio–type tests for testing the cointegration hypothesis. Instead of using the Gaussian likelihood, we have based our inference on a certain class of pseudolikelihoods. This class contained several well-known densities — for example, the Gaussian and the Student t. The asymptotic distribution of the *PLR* test was derived and was shown to depend on two types of nuisance parameters, arising from the possible misspecification of the pseudolikelihood and from the use of a non-Gaussian pseudolikelihood. Even if the likelihood was correctly specified, nuisance parameters remained present if the likelihood was non-Gaussian.

Using a simulation experiment, we investigated the properties of the tests and found that the choice of the pseudolikelihood can have a great influence on both the level and power of the *PLR* test. The power simulations in this paper demonstrated that the Johansen trace test is optimal if the innovations are Gaussian, whereas for innovations that are truncated Cauchy the Student t–based *PLR* tests perform better in terms of power.

Several interesting directions for future research remain. First, one can try to get rid of the nuisance parameters in the asymptotic distribution of the

PLR test that are due to the misspecification of the likelihood. One solution would be to construct adaptive cointegration tests, as mentioned in note 3. Alternatively, following White (1982), one could investigate generalizations of the Wald and Lagrange multiplier–type tests instead of likelihood ratio-type tests. This line is followed in Lucas (1996). Second, one can investigate the possibilities of bootstrap procedures for constructing accurate estimates of *p*-values of the *PLR* test in finite samples. The construction of automatic procedures for obtaining critical values and *p*-values of the *PLR* tests seems a valuable contribution for the applied researcher. In this respect, the Bartlett corrections put forward in Lucas (1996) might also prove useful. Third, more simulation evidence can be gathered in order to demonstrate the advantages and disadvantages of the non-Gaussian *PLR* test over the Gaussian one in situations that are of practical interest. This especially concerns the inclusion of deterministic functions of time in the regression model as well as nonzero drift terms in the data-generating process. Fourth, it is interesting to study the effects of dynamic misspecification of the regression model on the asymptotic distribution of the *PLR* test. One can then proceed by designing methods for correcting for these effects. Some interesting possibilities for this approach can be found in Phillips (1991), who uses the Whittle likelihood for the Gaussian PML estimator, and in Bierens (1994), who constructs a nonparametric cointegration test. As a final point for future research, it remains to be shown how well non-Gaussian *PLR* tests perform on empirical data. One of the chief difficulties is to construct fast iteration schemes in order to maximize the pseudolikelihood. As this likelihood is, in general, highly nonlinear, this might prove a nontrivial task.

NOTES

1. A function $\psi(\cdot)$ is called redescending if $\lim_{x\to\infty} \psi(x) = 0$ and/or $\lim_{x\to-\infty} \psi(x) = 0$. It is called strongly redescending if there exists a finite vector of constants c, such that, for all $x \leq -c$ and/or for all $x \geq c$, $\psi(x) = 0$.

2. There are, of course, two other well-known testing principles, namely, the Wald and the Lagrange multiplier (LM) test. The Wald and the LM test for cointegration based on Gaussian PML estimators are dealt with in, for example, Kleibergen and van Dijk (1994). Lucas (1996) deals with the Wald and LM test for a class of M estimators.

3. As one of the referees pointed out, Corollary 1 shows that one set of nuisance parameters present in the asymptotic distribution of the *PLR* test in Theorem 1, namely, $\bar{K}_0$, can be eliminated using a nonparametric density estimator. Such an estimator can be used to estimate the density of ε_t. Based on this density estimate, one can estimate the model parameters in (1) by means of maximum likelihood and compute a likelihood ratio test. If the density estimate converges sufficiently fast to the true density, Corollary 1 shows that the $k - r$ elements of the matrix R are the only nuisance parameters that enter the limiting distribution of this *adaptive* cointegration testing procedure. More research in this area is needed.

REFERENCES

Bierens, H.J. (1994) Nonparametric Cointegration Analysis. Working paper, Southern Methodist University, Dallas, Texas.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Boswijk, P. (1992) *Cointegration, Identification and Exogeneity. Inference in Structural Error Correction Models.* Ph.D. Thesis, Tinbergen Institute, Amsterdam. Amsterdam: Thesis Publishers.

Cox, D.D. & I. Llatas (1991) Maximum likelihood type estimation for nearly nonstationary auto-regressive time series. *Annals of Statistics* 19, 1109–1128.

de Vries, C.G. (1994) Stylized facts of nominal exchange rate returns. In F. van der Ploeg (ed.), *The Handbook of International Macroeconomics,* pp. 348–389. Oxford: Blackwell.

Engle, R.F. & C.W.J. Granger (1987) Co-integration and error correction: Representation, esti-mation and testing. *Econometrica* 55, 251–276.

Franses, P.H. & A. Lucas (1995) Outlier Robust Cointegration Analysis. Econometric Institute report 9529/A, Erasmus University, Rotterdam.

Gallant, A.R. (1987) *Nonlinear Statistical Models.* New York: Wiley.

Gouriéroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, & W.A. Stahel (1986) *Robust Statistics. The Approach Based on Influence Functions.* New York: Wiley.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Herce, M.A. (1994) Asymptotic Theory of LAD Estimation in a Unit Root Process with Finite Variance Errors. Mimeo, University of North Carolina at Chapel Hill.

Hoek, H., A. Lucas, & H.K. van Dijk (1995) Classical and Bayesian aspects of robust unit root inference. *Journal of Econometrics* 69, 27–59.

Huber, P.J. (1981) *Robust Statistics.* New York: Wiley.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynam-ics and Control* 12, 231–254.

Johansen, S. (1989) The power function of the likelihood ratio test for cointegration. *Lecture Notes in Economics and Mathematical Systems, Econometric Decision Models* 366, 323–335.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vec-tor autoregressive models. *Econometrica* 59, 1551–1580.

Johansen, S. (1994) The role of the constant and linear terms in cointegration analysis of non-stationary variables. *Econometric Reviews* 13, 205–229.

Kleibergen, F.R. & H.K. van Dijk (1994) Direct cointegration testing in error correction mod-els. *Journal of Econometrics* 63, 61–103.

Knight, K. (1989) Limit theory for autoregressive parameters in an infinite variance random walk. *Canadian Journal of Statistics* 17, 261–278.

Lucas, A. (1995a) An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics* 66, 153–174.

Lucas, A. (1995b) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.

Lucas, A. (1996) *Outlier Robust Unit Root Analysis.* Ph.D. Thesis, Tinbergen Institute, Amster-dam. Amsterdam: Thesis Publishers.

Manski, C.F. (1984) Adaptive estimation of non-linear regression models. *Econometric Reviews* 3, 145–194.

Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part I. *Econometric Theory* 4, 468–497.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Prucha, I.R. & H.H. Kelejian (1984) The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances. *Econometrica* 52, 721–736.
Rahbek, A.C. (1994) The Power of Some Multivariate Cointegration Tests. Working paper 6, University of Copenhagen.
Stock, J.H. (1994) Unit Roots and Trend Breaks. Mimeo, Harvard University, Cambridge, Massachusetts.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

APPENDIX

This Appendix contains the proofs of Theorem 1 and Corollary 1.

To prove Theorem 1, we introduce some further notation. First, we normalize the matrix B of cointegrating vectors, such that $B^\mathsf{T} = (I_r, \beta^\mathsf{T})$, with β a $((k - r) \times r)$ matrix. Note that under the null hypothesis, H_r, such a normalization is always possible, because rank$(B) = r$. The choice of the leading submatrix in B^T to be the unit matrix, however, may require a reordering of the elements of y_t. As the (pseudo)likelihood ratio test is invariant under such reparameterizations, no generality is lost by imposing this condition. Next, we let $A^\mathsf{T} = (\alpha_{11}^\mathsf{T}, \alpha_{21}^\mathsf{T})$. We also introduce the $(k \times (k - r))$ matrix K_6, which has the property that $\bar{A} = (A, K_6)$ has full rank. Under the hypothesis H_k, the matrix Π can be decomposed as

$$\Pi = AB^\mathsf{T} + K_6 \alpha_{22}(0, I_{k-r}) = A(I_r, 0) + \bar{A}(\beta, \alpha_{22}^\mathsf{T})^\mathsf{T}(0, I_{k-r}),　\text{(A.1)}$$

with α_{22} a $((k - r) \times (k - r))$ matrix. The number of parameters in A, β, and α_{22} equals the number of elements in Π, namely, k^2. Therefore, the parametric decomposition of Π in (A.1) can be used to estimate model (1) under the hypothesis H_k. Note that (A.1) can also be used to estimate model (1) under the null hypothesis H_r. This is seen by setting $\alpha_{22} = 0$, which results in $\Pi = AB^\mathsf{T}$, with A and B of full column rank. Therefore, (A.1) can be used to reformulate the hypotheses of interest as $H_r': \alpha_{22} = 0$ versus $H_k': \alpha_{22} \neq 0$. Similar decompositions are found in Phillips (1991) and Kleibergen and van Dijk (1994), who both use $K_6^\mathsf{T} = (0, I_{k-r})$.

Define the vector of parameters θ to be vec$((\beta, \alpha_{22}^\mathsf{T})^\mathsf{T}, A, \Gamma, \Omega_{11})$, where $\Gamma = (\Phi_1, \ldots, \Phi_p)$. Now the hypothesis $\alpha_{22} = 0$ can be formulated as $H\theta = 0$, with $H = (I_{k-r}, 0) \otimes (0, I_{k-r})$. Let $\bar{\theta}_{r,T}$ denote the estimator of θ under the hypothesis $\alpha_{22} = 0$, and let $\hat{\theta}_T$ denote the estimator under the alternative. Furthermore, let θ_0 denote the true parameter vector. The key convergence results are given in the following lemma.

LEMMA A.1. *Given the conditions of Theorem 1,*

$$D\frac{\partial \ell_T(\theta_0)}{\partial \theta} \Rightarrow -\begin{pmatrix} \int U \otimes d\bar{A}^\mathsf{T} W_2 \\ \\ \xi_1 \end{pmatrix},$$

$$D\frac{\partial^2 \ell_T(\theta_0)}{\partial\theta\,\partial\theta^\mathsf{T}}D \Rightarrow \begin{pmatrix} \int UU^\mathsf{T} \otimes \bar{A}^\mathsf{T} C_1 \bar{A} & 0 \\ \\ 0 & \Xi_1 \end{pmatrix},$$

where $\xi_1 = O_p(1)$, $\Xi_1 = O_p(1)$, $U(s)$ is the Brownian motion defined in Lemma 1, and

$$D = \begin{pmatrix} I_{k(k-r)}/T & 0 \\ 0 & I/T^{1/2} \end{pmatrix}.$$

Proof. It is straightforward to verify that

$$\frac{\partial \ell_T(\theta_0)}{\partial \theta^\mathsf{T}} = \sum_{t=1}^{T} (y_{2,t-1}^\mathsf{T} \otimes \psi_t^\mathsf{T} \bar{A}, Z_{1t}^\mathsf{T} \otimes \psi_t^\mathsf{T}, Z_{2t}^\mathsf{T}), \tag{A.2}$$

where $\psi_t = \psi(\Omega_{11}^{-1/2}\varepsilon_t)$, $y_{2,t-1}$ contains the last $k - r$ rows of y_{t-1}, $Z_{1t}^\mathsf{T} = (y_{t-1}^\mathsf{T} B, \Delta y_{t-1}^\mathsf{T}, \ldots, \Delta y_{t-p}^\mathsf{T})$, and

$$Z_{2t}^\mathsf{T} = -\frac{1}{2} (\mathrm{vec}(\Omega_{11}^{-1}))^\mathsf{T} - \psi_t^\mathsf{T}\Omega_{11}^{1/2}(\varepsilon_t^\mathsf{T} \otimes I_k)\frac{\partial \, \mathrm{vec}(\Omega_{11}^{-1/2})}{\partial (\mathrm{vec}(\Omega_{11}))^\mathsf{T}}.$$

Note that Z_{1t} and Z_{2t} only contain stationary elements, which, together with the i.i.d. assumption for ε_t and the existence of the appropriate moments, implies that $T^{-1/2}\sum_{t=1}^{T}(Z_{1t}^\mathsf{T} \otimes \psi_t^\mathsf{T}, Z_{2t}^\mathsf{T}) = O_p(1)$. Furthermore, we have that

$$y_{2,t} = (0, I_{k-r})y_t = (0, I_{k-r})(B(B^\mathsf{T}B)^{-1}B^\mathsf{T} + B_\perp(B_\perp^\mathsf{T}B_\perp)^{-1}B_\perp^\mathsf{T})y_t,$$

where $B_\perp^\mathsf{T} = (-\beta, I_{k-r})$. From Lemma 1 and the stationarity of $B^\mathsf{T}y_t$, it follows that $y_{2,\lfloor sT \rfloor}/T^{1/2} \Rightarrow U(s)$. The first part of the lemma now follows directly from Phillips and Durlauf (1986) and Hansen (1992).

Let $\psi_t' = \psi'(\Omega_{11}^{-1/2}\varepsilon_t)$; then,

$$D\frac{\partial^2 \ell_T(\theta_0)}{\partial\theta\,\partial\theta^\mathsf{T}}D = \sum_{t=1}^{T} \begin{bmatrix} Q_{11,t} & Q_{12,t} & Q_{13,t} \\ Q_{12,t}^\mathsf{T} & Q_{22,t} & Q_{23,t} \\ Q_{13,t}^\mathsf{T} & Q_{23,t}^\mathsf{T} & Q_{33,t} \end{bmatrix}, \tag{A.3}$$

with

$$Q_{11,t} = -\frac{y_{2,t-1}y_{2,t-1}^\mathsf{T}}{T^2} \otimes \bar{A}^\mathsf{T}\psi_t'\bar{A}$$

$$Q_{12,t} = -\frac{y_{2,t-1}Z_{1t}^\mathsf{T}}{T^{3/2}} \otimes \bar{A}^\mathsf{T}\psi_t' + \left(\frac{y_{2,t-1}\psi_t^\mathsf{T}}{T^{3/2}} \otimes I\right)\frac{\mathrm{vec}\bar{A}^\mathsf{T})}{\mathrm{vec}(A,\Gamma)^\mathsf{T}}$$

$$Q_{13,t} = \left(\frac{y_{2,t-1}\varepsilon_t^\mathsf{T}}{T^{3/2}} \otimes \bar{A}^\mathsf{T}\psi_t'\Omega_{11}^{1/2}\right)\frac{\mathrm{vec}(\Omega_{11}^{-1/2})}{\mathrm{vec}(\Omega_{11})^\mathsf{T}}$$

$$Q_{22,t} = -\frac{Z_{1t}Z_{1t}^\mathsf{T}}{T} \otimes \psi_t'$$

$$Q_{23,t} = \left(\frac{Z_{1t}\varepsilon_t^\mathsf{T}}{T} \otimes \psi_t'\Omega_{11}^{1/2}\right)\frac{\mathrm{vec}(\Omega_{11}^{-1/2})}{\mathrm{vec}(\Omega_{11})^\mathsf{T}}$$

$$Q_{33,t} = -(\partial Z_{2t}/\partial \, \mathrm{vec}(\Omega_{11})^\mathsf{T})/T.$$

It is easily checked that under the present conditions $\sum_{t=1}^{T}(Q_{12,t}, Q_{13,t})$ converges to zero for $T \to \infty$. Furthermore, the weak convergence of $\sum_{t=1}^{T}Q_{11,t}$ follows from Phillips and Durlauf (1986) and Hansen (1992). The convergence of the remaining blocks in (A.3) follows directly by applying the law of large numbers. Joint convergence obviously also holds. ∎

The next lemma gives the appropriate convergence for the Hessian of the pseudolikelihood.

LEMMA A.2. *Let $\hat{\theta}_T$ with corresponding residuals $\hat{\varepsilon}_t$ be such that $\hat{\Omega}_{11}^{-1/2}\hat{\varepsilon}_t - \Omega_{11}^{-1/2}\varepsilon_t = o_p(1)$ uniformly in t; then,*

$$D\left\{ \frac{\partial^2 \ell_T(\hat{\theta}_T)}{\partial\theta\,\partial\theta^\top} - \frac{\partial^2 \ell_T(\theta_0)}{\partial\theta\,\partial\theta^\top} \right\} D \xrightarrow{p} 0,$$

with D as defined in Lemma A.1.

Proof. This follows straightforwardly from the Lipschitz continuity of ψ' and the convergence of (A.3) (cf. Lucas, 1995b). ∎

One of the major results of Lemmas A.1 and A.2 is that the appropriately normalized Hessian of the pseudo-log likelihood is asymptotically block diagonal between the parameters β and α_{22}, on the one hand, and the parameters A, Γ, and Ω_{11}, on the other hand. This simplifies the proof of Theorem 1, because we can now disregard the effect of estimating Ω_{11}. Moreover, without loss of generality, we can restrict our attention to the case $p = 0$, that is, to the VAR(1) model. With a slight abuse of notation, we therefore consider the case $p = 0$ with fixed and known scaling matrix Ω_{11}.

Proof of Theorem 1. We have $\text{vec}(\alpha_{22}) = H\theta$, with θ as defined above Lemma A.1. Following Gallant (1987, Ch. 3, Theorems 13 and 15), we obtain

$$2(\ell_T(\tilde{\theta}) - \ell_T(\hat{\theta})) = \frac{\partial\ell_T(\theta_0)}{\partial\theta^\top} J^{-1}H^\top(HJ^{-1}H^\top)^{-1}HJ^{-1} \frac{\partial\ell_T(\theta_0)}{\partial\theta} + o_p(1), \quad \text{(A.4)}$$

where $J = \partial^2\ell_T(\theta_0)/(\partial\theta\,\partial\theta^\top)$. Note that $H\cdot D = H/T$. Using this fact and Lemmas A.1 and A.2, we establish that $-H(DJD)^{-1}D\partial\ell_T(\theta_0)/\partial\theta$ converges weakly to

$$(I_{k-r} \otimes (0, I_{k-r}))\left(\left(\int UU^\top\right) \otimes \bar{A}^\top C_1 \bar{A}\right)^{-1}\left(\int U \otimes d\bar{A}^\top W_2\right)$$

$$= \left(\left(\int UU^\top\right)^{-1} \otimes (0, I_{k-r})\bar{A}^{-1}C_1^{-1}\right)\left(\int U \otimes dW_2\right).$$

As a result, the right-hand side in (A.4) weakly converges to

$$\left(\int U \otimes dW_2\right)^\top \left(\left(\int UU^\top\right)^{-1} \otimes (0, I_{k-r})\bar{A}^{-1}C_1^{-1}\right)^\top$$

$$\cdot \left(\left(\int UU^\top\right) \otimes ((0, I_{k-r})(\bar{A}^\top C_1 \bar{A})^{-1}(0, I_{k-r})^\top)^{-1}\right)$$

$$\cdot \left(\left(\int UU^\top\right)^{-1} \otimes (0, I_{k-r})\bar{A}^{-1}C_1^{-1}\right)\left(\int U \otimes dW_2\right)$$

$$= \left(\int U \otimes dA_\perp^\top C_1^{-1} W_2\right)^\top \left(\left(\int UU^\top\right)^{-1} \otimes K_0\right)\left(\int U \otimes dA_\perp^\top C_1^{-1} W_2\right)$$

$$= \text{tr}\left(K_0\left(\int Ud(A_\perp^\top C_1^{-1} W_2)^\top\right)^\top \left(\int UU^\top\right)^{-1}\left(\int Ud(A_\perp^\top C_1^{-1} W_2)^\top\right)\right). \quad \text{(A.5)}$$

Now replacing U and $A_\perp^T C_1^{-1} W_2$ in (A.5) by

$$V_1(s) = S_1^T (A_\perp^T \Omega_{11} A_\perp)^{-1/2} (A_\perp^T \Psi B_\perp) U(s),$$

and

$$V_2(s) = S_2^T (A_\perp^T C_1^{-1} \Omega_{22} C_1^{-1} A_\perp)^{-1/2} A_\perp^T C_1^{-1} W_2(s),$$

respectively, we obtain

$$\operatorname{tr}\left(\tilde{K}_0 \left(\int V_1 \, dV_2^T \right)^T \left(\int V_1 V_1^T \right)^{-1} \left(\int V_1 \, dV_2^T \right) \right). \tag{A.6}$$

It is easily checked that $E(V_1(s) V_2(s)^T) = S_1^T S_0 S_2 = R$. ∎

   Proof of Corollary 1. Under the conditions stated in Corollary 1, the information matrix equality holds, meaning that

$$E(d^2 \ln(f(\varepsilon_t))/(d\varepsilon_t \, d\varepsilon_t^T)) = -E((d \ln(f(\varepsilon_t))/d\varepsilon_t)(d \ln(f(\varepsilon_t))/d\varepsilon_t)^T).$$

For the specific form of $f(\cdot)$ given in the corollary, this implies that $C_1 = \Omega_{22}$. The result then follows easily from Theorem 1. ∎