Faculteit der Economische Wetenschappen en Econometrie

SERIE RESEARCH MEMORANDA

A Comparison of Parametric, Semi-nonparametric, Adaptive, and Nonparametric Cointegration Tests

H. Peter Boswijk
André Lucas
Nick **Taylor**

Research Memorandum 1998-62

December 1998

vrije **Universiteit** *amsterdam*

A Comparison of Parametric, Semi-nonparametric, Adaptive, and Nonparametric Cointegration Tests*

H. Peter Boswijk[†] **André Lucas**[‡] **Nick Taylor**[§]

Abstract

This paper provides an extensive Monte-Carlo comparison of several contemporary cointegration tests. Apart from the familiar Gaussian based tests of Johansen, we also consider tests based on non-Gaussian quasi-likelihoods. Moreover, we compare the performance of these parametric tests with tests that estimate the score function from the data using either kernel estimation or semi-nonparametric density approximations. The comparison is completed with a fully nonparametric cointegration test. In small samples, the overall performance of the semi-nonparametric approach appears best in terms of size and power. The main cost of the semi-nonparametric approach is the increased computation time. In large samples and for heavily skewed or multimodal distributions, the kernel based adaptive method dominates. For near-Gaussian distributions, however, the semi-nonparametric approach is preferable again.

Key words: cointegration testing, adaptive estimation, nonparametrics, semi-nonparametrics, Monte-Carlo simulation.
JEL **Codes: C14, C32.**

1 Introduction

The last decade has witnessed an explosively growing interest in the long-run properties of economic time series. Key words in this area of research are non-stationarity, unit roots, and cointegration. The concept of cointegration has witnessed a particularly widespread popularity in the applied literature. See,

*André Lucas thanks the Dutch Funding Organization for Scientific Research (N.W.O.) for financial support.

[†]Dept. Quantitative Economics and Tinbergen Institute, University of Amsterdam, Roetersstraat 11, NL-1018WB Amsterdam, the Netherlands, peterb@fee.uva.nl

[‡]Dept. Finance and Tinbergen Institute, ECO/BFS, Vrije Universiteit, De Boelelaan 1105, NL-1081HV Amsterdam, the Netherlands, alucas@econ.vu.nl

[§]School of Accounting and Finance, University of Manchester, Oxford Road, Manchester, Ml3 9PL, United Kingdom, msrysnt@fs1.acc.man.ac.uk

e.g., Franses et al. (1998) for marketing applications, and Clarida and Taylor (1997) for an application to spot and forward exchange rates. Other areas of application include stock markets, the term structure of interest rates, international trade and purchasing power parity, consumption and income, and the demand for money.

Cointegrating relations are often given an economic interpretation relating to market equilibrium and/or market efficiency. As such, it is useful to test whether different economic time series are cointegrated. Statistical tests for cointegration have gone through several stages of development, see, e.g., Hamilton (1994). Here we will focus on cointegration tests based on the systems approach and the likelihood principle. The seminal reference is Johansen (1988). Johansen tests for the presence of cointegrating relations in the framework of a vector autoregressive (VAR) time series model. By assuming a normal distribution for the innovations to the VAR, Johansen is able to derive a closed-form expression for the likelihood ratio test statistics for the null of no cointegration against the alternative of cointegration, as well as its limiting distribution under the null. Johansen (1991a) extends the test procedure to allow for deterministic trends in the time series.

The assumption of normally distributed innovations is crucial in deriving the form of the Johansen test statistic. It is much less important for the applicability of the limiting distribution of the, test. If, however, innovations are non-normal, e.g., fat-tailed or skewed, then the power of the test can be increased by exploiting the non-normality in the estimation and testing stage. This can be done by extending the methodology of Johansen to non-Gaussian likelihoods or quasi-likelihoods as in Lucas (1997a,1998). This is a parametric approach which gives satisfactory results if the salient features of the true likelihood are adequately captured by the postulated quasi-likelihood. If this is not the case, then the parametric approach adopted in Lucas may result in poor power performance, compare the simulations in Shin and So (1998) for the univariate case.

To avoid the loss in power due to an inappropriate choice of the quasi-likelihood, it might seem a good idea to estimate the likelihood function from the data using (semi-)nonparametric techniques. Hodgson (1998) discusses adaptive estimation of long-run parameters in an error-correction framework. His ideas can easily be extended to the estimation of the unit root parameters in such models. For the univariate case, this has been done by Shin and So (1998). We extend their test procedure in the present paper to the multivariate setting, thus constructing an adaptive version of Johansen's cointegration test. Boswijk and Lucas (1997) follow a similar route and use semi-nonparametric (SNP) density expansions following the ideas of Gallant and Nychka (1987) instead of kernel estimation as in Hodgson (1998) and

Shin and So (1998). Both the adaptive and the SNP approach claim a power gain over the Johansen method. Moreover, Boswijk and Lucas also claim a power gain with respect to the parametric tests proposed in Lucas (1998). No formal comparison between the adaptive and SNP methods in the framework of cointegration testing has yet been performed, such that it is difficult to say which of the two methods performs best.

The aim of the present paper is to provide a Monte-Carlo comparison of several cointegration tests available in the contemporary literature. We investigate the size and power properties of the tests under various distributional assumptions, sample sizes, and data generating processes. We are mainly interested in three comparisons: (i) the Gaussian versus the non-Gaussian tests, (ii) the parametric versus the (semi)-nonparametric tests, and (iii) the adaptive versus the SNP tests. To complete the comparisons, we also include a fully nonparametric cointegration test, in particular the test proposed by Bierens (1997). This tests builds on a similar generalized eigenvalue problem as Johansen (1988,1991a).

The simulations reveal that the SNP approach is a clear winner in small samples. The power gain under non-normal innovations does not come at the cost of a power loss under Gaussian innovations. This stands in sharp contrast to the adaptive approach. For large sample sizes, however, the SNP tests have much more difficulty in picking up skewness and/or multimodality compared to the adaptive approach. Again, however, the adaptive approach suffers from a substantial power loss if the innovations are (near) normal. The main cost of using the SNP approach over the adaptive approach is the required computation time.

The paper is set up as follows. Section 2 gives details on the model and cointegration test procedures. Section 3 presents the estimation principles used. This section also provides some computational details pertaining to the available Gauss code to compute the tests. The simulation set-up and results are presented in Section 4. Section 5 concludes.

2 Model and test statistics

Consider the VAR of order p,

$$\Delta y_t = \Pi y_{t-1} + \Phi_1 \Delta y_{t-1} + \ldots + \Phi_{p-1} y_{t-p+1} + \mu + \varepsilon_t, \qquad (1)$$

where $y_t, \mu, \varepsilon_t \in \mathbb{R}^k$, $\Pi, \Phi_i \in \mathbb{R}^{k \times k}$ for $i = 1, \ldots, p-1$, y_t is an observed time series, μ is a vector of constants, II and $\Phi_1, \ldots, \Phi_{p-1}$ are parameter matrices, and ε_t is an unobserved innovation process. Model (1) can be augmented with additional deterministic components such as seasonal dummies and linear

trends. These additional complexities are not the prime focus of our paper, however, and are therefore omitted from the analysis.

We make the following standard assumption for the process in (1), see also Johansen (1988,1991a).

Assumption 1 *(i) The roots of* $|(1-z)I-\Pi z-\Phi_1(1-z)-\ldots-\Phi_{p-1}z^{p-1}(1-z)|$ *lie outside the unit circle or at* $+1$*; (ii) the series* Δy_t *is stationary; (iii) the innovations* ε_t *are independent and identically distributed (i.i.d.) with finite variance-covariance matrix* $\Omega_{\varepsilon\varepsilon}$ *.*

Parts (i) and (ii) of Assumption 1 ensure that the time series y_t is integrated of at most order one, i.e., the first differences of the time series are stationary, while the levels are possibly non-stationary. Part (iii) of Assumption 1 allows us to invoke a multivariate invariance principle to establish the limiting distribution of our test statistics. This assumption can be relaxed to the assumption of a finite variance martingale difference sequence at the cost of additional complexities in the proofs. As the focus of the present paper is not on the limiting distribution theory, but more on the finite sample performance of the cointegration tests, we stick to the requirement in part (iii).

Our main interest is in the rank of the matrix II. If the rank of Π equals r, we say there are r cointegrating relations. In that case the matrix II can be decomposed as $\Pi = \alpha\beta^\top$, with α and β two k x r matrices of full column rank. The columns of β are called the cointegrating vectors, while r is called the cointegration rank of the system. In order to test whether the rank of Π is equal to r, we introduce the LU-type decomposition of Π as proposed by Kleibergen and van Dijk (1994). Let

$$\Pi = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} I_r & \beta_{21}^\top \\ 0 & I_{k-r} \end{pmatrix}, \tag{2}$$

with α_{11}, α_{21}, α_{22}, and β_{21} r x r, $(k-r)$ x r, $(k-r)$ x $(k-r)$, and $(k-r)$ x r matrices respectively. Moreover, we define $\alpha^\top = (\alpha_{11}^\top, \alpha_{21}^\top)$ and $\beta^\top = (I_r, \beta_{21}^\top)$, and assume that α and β have full column rank. The hypothesis H_0 : $\mathrm{rank}(\Pi) = r$ now boils down to the hypothesis H_0' : $\alpha_{22} = 0$, *see* Kleibergen and van Dijk (1994) and Lucas (1997a,1998) for more details.

Let $\mathrm{vec}(\cdot)$ denote the operator that stacks the columns of a matrix into a vector. Define $\theta^\top = (\mathrm{vec}(\alpha)^\top, \mathrm{vec}(\Phi_1)^\top, \ldots, \mathrm{vec}(\Phi_{p-1})^\top, \mu^\top)$. Most of the tests in this paper build on the parametric model (1), combined with a (possibly misspecified) family of densities $f(\varepsilon_t; \nu)$, where ν is an additional parameter vector determining the shape of the density. Defining $\varepsilon_t(\theta, \alpha_{22})$ as

the residual from (1) under (2) for a particular value of the parameters, the (possibly nonparametrically estimated) quasi-likelihood $\mathcal{L}(\cdot)$ becomes

$$\mathcal{L}(Y_T; \theta, \alpha_{22}, \nu) = \prod_{t=1}^{T} f(\varepsilon_t(\theta, \alpha_{22}); \nu), \tag{3}$$

with $Y_T^\top = (y_1^\top, \ldots, y_T^\top)$, and T the sample size. The dimension of ν may range from zero to infinity. For example, if a Gaussian quasi-likelihood with known covariance matrix is used, ν is empty. By contrast, if the density of ε_t in (1) is estimated nonparametrically, ν is infinite dimensional. Normally, the parameter vector ν at least contains the nonredundant elements of the covariance matrix of the errors, $\Omega_{\varepsilon\varepsilon}$, see Assumption 1. Our main interest in the present paper is to compare contemporary cointegration tests that use different specifications of ν in (3) and/or different methods to estimate ν.

To test the null hypothesis H_0 : $\alpha_{22} = 0$ against the alternative H_0 : $\alpha_{22} \neq 0$, we consider two types of tests, namely the (quasi)-likelihood ratio (QLR) test and the (quasi)-Lagrange multiplier (QLM) test. A Wald variant of the test is not considered, as the Wald test for $\alpha_{22} = 0$ turns out to depend on the ordering of the variables in y_t, see Kleibergen and van Dijk (1994) and Lucas (1996). Let θ and $\tilde{\theta}$ denote the estimates of θ under the alternative and under the null, respectively. A similar definition holds for $\hat{\nu}$ and $\tilde{\nu}$. We also need the additional notation $\bar{\theta}$ and $\bar{\nu}$ to denote estimates of θ and ν respectively, based on a preliminary estimation procedure. We now obtain

$$\text{QLR} = -2 \ln[\mathcal{L}(Y_T; \tilde{\theta}, 0, \tilde{\nu}) / \mathcal{L}(Y_T; \hat{\theta}, \hat{\alpha}_{22}, \hat{\nu})], \tag{4}$$

as one of the tests considered. The QLM tests are somewhat more difficult to present. Let

$$G_T = G(Y_T; \theta, \alpha_{22}, \nu), \tag{5}$$

$$J_T = J(Y_T; \theta, \alpha_{22}, \nu), \tag{6}$$

and

$$H_T = H(Y_T; \theta, \alpha_{22}, \nu), \tag{7}$$

denote the gradient, the outer product of gradients, and the Hessian of the quasi-likelihood with respect to $(\theta^\top, \text{vec}(\alpha_{22})^\top, \text{vec}(\nu)^\top)$, respectively. Furthermore, let S denote a selection matrix such that

$$S(\theta^\top, \text{vec}(\alpha_{22})^\top, \text{vec}(\nu)^\top)^\top = \text{vec}(\alpha_{22}). \tag{8}$$

Then the QLM tests we consider are either of the form

$$\left(S\tilde{H}_T^{-1}\tilde{G}_T\right)^\top \left(S\tilde{H}_T^{-1}\tilde{J}_T\tilde{H}_T^{-1}S^\top\right)^{-1} \left(S\tilde{H}_T^{-1}\tilde{G}_T\right), \tag{9}$$

or

$$\left(S\bar{H}_T^{-1}\bar{G}_T\right)^{\top}\left(S\bar{H}_T^{-1}\bar{J}_T\bar{H}_T^{-1}S^{\top}\right)^{-1}\left(S\bar{H}_T^{-1}\bar{G}_T\right), \tag{10}$$

where the notation with a tilde, e.g., $\tilde{G}_T$, denotes evaluation in $(\tilde{\theta}, 0, \tilde{\nu})$, while the notation with the upper bar, e.g., $\bar{G}_T$, denotes evaluation in $(\bar{\theta}, 0, \bar{\nu})$. The formulation in (9) will be used for the parametric and semi-nonparametric cointegration tests, while (10) is useful for the adaptive cointegration test.

For inference, the computed cointegration tests have to be confronted with a critical value. It is common practice to use the critical values from the limiting distribution of QLR and QLM. In the present paper, we do not want to focus on the different regularity conditions needed to establish the limiting distribution for the parametric, semi-nonparametric, and the adaptive tests. For the technical details, the reader is referred to Johansen (1988,1991a), Lucas (1997a,1998), Boswijk and Lucas (1997), Hodgson (1998), and Shin and So (1998). The limiting distribution, however, is the same for all testing procedures considered. This is summarized 'informally' in Theorem 1.

Define $\alpha_{\perp}$ as any k x (k-r) matrix of full column rank such that $\alpha_{\perp}^{\top}\alpha = 0$. Furthermore, define the quasi-score ψ_t and the quasi-information $\mathcal{I}$ as

$$\psi_t = -\frac{\partial \ln f(\varepsilon_t;\nu)}{\partial \varepsilon_t}, \qquad \mathcal{I} = -E\left(\frac{\partial^2 \ln f(\varepsilon_t;\nu)}{\partial \varepsilon_t \partial \varepsilon_t^{\top}}\right). \tag{11}$$

Theorem 1 *Under 'suitable'* **regularity conditions (see** *the references mentioned above), the QLM tests for model (1) under the additional assumption* $\mu = \alpha\mu_0$ *with* $\mu_0 \in \mathbb{R}^r$, *converge weakly to the functional*

$$\text{trace}\left\{\left(\int \bar{B}_1 dB_2^{\top}\right)^{\top}\left(\int \bar{B}_1\bar{B}_1^{\top}\right)^{-1}\left(\int \bar{B}_1 dB_2^{\top}\right)\right\}, \tag{12}$$

with $\bar{B}_1 = B_1 - \int B_1$, B_1 *and* B_2 *two standard* $(k-r)$*-vector Brownian motions with diagonal correlation matrix* R, and R *containing the canonical correlations* between $\alpha_{\perp}^{\top}\varepsilon_t$ and $\alpha_{\perp}^{\top}\mathcal{I}^{-1}\psi_t$.

Remark 1 The notation in Theorem 1 is standard, $\int B_1 dB_2^{\top}$ denoting the stochastic integral $\int_0^1 B_1(s)dB_2(s)^{\top}$, and $\int B_1$ and $\int B_1 B_1^{\top}$ denoting standard Riemann integrals, $\int_0^1 B_1(s)ds$ and $\int_0^1 B_1(s)B_1(s)^{\top}ds$, respectively.

Remark 2 Theorem 1 only gives the results for the QLM test. The only QLR test we consider in this paper is that of Johansen (1991a) based on the Gaussian quasi-likelihood. The limiting distribution of that test statistic is given by (12) with B_2 replaced by B_1, see Johansen (1991a). If a non-Gaussian quasi-likelihood were to be used for the QLR test, additional nuisance parameters would enter the limiting distribution if the quasi-likelihood and the true likelihood did not coincide, see Lucas (1997a). This is why we concentrate on the use of the QLM rather than the QLR type tests.

Remark 3 Theorem 1 states the additional assumption $\mu = \alpha\mu_0$. This implies that there are no linear deterministic trends in the data generating process. If this assumption does not hold, the limiting distribution changes in the familiar way, compare Johansen (1991a). In particular, the first element of the vector $\bar{B}_1(s)$ then has to be replaced by s – 0.5. Similar changes to the limiting distribution have to be carried through in case additional deterministic components are present either in the regression model or in the data generating process.

Before proceeding with the presentation of the different estimators used in this paper, we first pay some more attention to the way to conduct inference in the present framework. As mentioned in Theorem 1, the limiting distribution of the cointegration tests considered depends on the nuisance parameter R, containing the canonical correlations between $\alpha_\perp^\top \varepsilon_t$ and $\alpha_\perp^\top \mathcal{I}^{-1}\psi_t$. It is therefore clear that the critical values of the QLM test depend both on the true likelihood and the quasi-likelihood. This means that new critical values have to be tabulated for each new estimation principle chosen and each new distribution of the innovations. There are at least two ways to deal with this problem. The first approach was suggested and implemented by Lucas (1998). One obtains a consistent estimate of R using the regression residuals $\hat{\varepsilon}_t$, $\tilde{\varepsilon}_t$, or $\bar{\varepsilon}_t$, and the estimated quasi-scores $\hat{\psi}_t$, $\tilde{\psi}_t$, or $\bar{\psi}_t$ (and corresponding estimates of the quasi-information $\mathcal{I}$). Using the estimated value of R, the integrals and Brownian motions in (12) can be approximated by sums and by random walks, respectively. The random walks have standard normal innovations with correlation matrix $\hat{R}$, $\tilde{R}$, or $\bar{R}$. By drawing a large number of (correlated) random walks for a given estimate of R and using the discretized version of (12), one can obtain an estimate of the appropriate asymptotic critical value or p-value of the test. The computation time required for these simulations is feasible for practical purposes. We adopt this simulation based method in the Monte-Carlo comparison in Section 4. As an alternative to the simulation based method, one can use Gamma approximations to the usual limiting distribution of the Johansen cointegration test and mix these with an independent stochastic term depending only on the (estimated) matrix R. For the univariate case, this approach was suggested by Abadir and Lucas (1996), while the multivariate case has very recently been addressed by Boswijk and Doornik (1998) and Doornik (1998).

3 Estimators and implementation of test statistics

In this section we consider the different choices for the quasi-likelihood (3). We also discuss how the parameters of the quasi-likelihood can be chosen, and how these estimates can be used in the construction of the cointegration tests. Each choice of the quasi-likelihood and estimation principle is treated in a separate subsection. The final subsection contains some details on the non-parametric cointegration test procedure of Bierens (1997). This test is included for completeness. It can be used to contrast the results for the parametric model (1) combined with a possibly nonparametrically estimated quasi-likelihood, with the results one obtains by a fully nonparametric approach.

3.1 Gaussian quasi-likelihood

As an obvious benchmark case, we consider the Gaussian QLR and QLM cointegration tests as proposed by Johansen (1991a) and Kleibergen and van Dijk (1994), respectively. In this case the parameter ν only contains the nonredundant elements of the variance-covariance matrix $\Omega_{\varepsilon\varepsilon}$ of the innovations. Maximum likelihood estimates of ν under both the null and the alternative can be obtained explicitly once the parameters θ and α_{22} are known, see Johansen (1991a). For the Gaussian quasi-likelihood, R in Theorem 1 reduces to the identity matrix, such that $B_2 \equiv B_1$, see also Remark 2.

3.2 Student t quasi-likelihood

A first parametric alternative to the Gaussian quasi-likelihood is the Student t quasi-likelihood. We use a Student t with 5 degrees of freedom. Cointegration tests based on this quasi-likelihood were studied in, e.g., Lucas (1997a,1998), and successfully applied in, e.g., Franses and Lucas (1998), Franses et al. (1998). Fixing the degrees of freedom parameter a priori has some advantages from a statistical robustness point of view, see Lucas (1997b). Again, ν only contains the nonredundant parameters of the covariance matrix of the innovations. Though no explicit form is available for the estimates of θ, α_{22}, and ν, they can be obtained straightforwardly by standard maximization techniques. Estimated residuals and quasi-scores can be used to estimate the nuisance parameter R and to conduct inference as described towards the end of Section 2.

3.3 Semi-nonparametric approach

The semi-nonparametric (SNP) approach centers around the following specification of the quasi-likelihood for the t-th observation:

$$p_n(L\varepsilon_t + m; \xi_1)^2 \cdot t(L\varepsilon_t + m; \xi_2), \tag{13}$$

where $p_n(\cdot; \xi_1)$ is a nth order polynomial with coefficients given in ξ_1, $t(\cdot; \xi_2)$ is the standard Student t distribution with zero mean and unit scaling matrix, and degrees of freedom parameter ξ_2. The lower-triangular matrix L captures the covariances between the innovations, while the vector m is added to ensure that the expectation of ε_t following from (13) equals zero. For $\xi_2 \to \infty$, we obtain the SNP approach as introduced by Gallant and Nychka (1987). The extension to the Student t distribution ($\nu < \infty$) was proposed by Boswijk and Lucas (1997) in order to obtain a parsimonious parameterization of skewness and fat-tailedness. The skewness is mainly captured by the premultiplication with the polynomial factor, while the fat-tailedness is captured by both the polynomial factor and the leptokurtosis of the Student t distribution.

For the quasi-likelihood in (13), we have $\nu = (\xi_1, \xi_2, \xi_3)$, where ξ_3 contains the non-redundant elements of L. The vector m does not enter, as it is a known function of ν, see Boswijk and Lucas (1997). Moreover, there is a constraint on the value of ξ_2 with respect to the order n of the polynomial, see also Boswijk and Lucas (1997). This constraint is needed to ensure that (13) is a proper density, i.e., integrates out to 1.

To obtain somewhat more insight into the form of the polynomial, consider the bivariate case, $k = 2$. We then have

$$\begin{aligned} \mathbf{P}\quad \mathbf{n}\quad ((x_1, x_2); \xi_1) = 1 + \\ \xi_{1,1}x_1 + \xi_{1,2}x_2 + \\ \xi_{1,3}x_1^2 + \xi_{1,5}x_1x_2 + \xi_{1,6}x_2^2 + \\ \\ \xi_{1,N-n}x_1^n + \xi_{1,N-n+1}x_1^{n-1}x_2 + \ldots + \xi_{1,N-1}x_1x_2^{n-1} + \xi_{1,N}x_2^n, \end{aligned} \tag{14}$$

where N denotes the number of elements in ξ_1, i.e., $N = n(n+1)/2 - 1$, and $\xi_1 = (\xi_{1,1}, \ldots, \xi_{1,N})$. BY an appropriate choice of the elements of ξ_1, one can model several types of skewness and leptokurtosis. Note that for n = 0, (13) reduces to the Student t quasi-likelihood with *estimated* degrees of freedom parameter.

Gallant and Nychka (1987) formally prove for $\xi_2 = \infty$ that the set of densities characterized by (13) forms a dense set in a larger class of densities

that comprises most familiar densities used in econometric applications. This holds a *fortiori* if $\xi_2 \in \mathbb{R}_+$. It is therefore possible to approximate most familiar densities arbitrarily closely by a quasi-likelihood (13) that has a sufficiently large value of n. By letting the degree n of the polynomial diverge to infinity with the sample size at the appropriate rate, one can under suitable regularity conditions consistently estimate the true likelihood from the data. Little is known, however, on the precise rate required. We therefore adopt a different approach in the Monte-Carlo simulations in Section 4. First, we estimate the model parameters for several choices of n = 0, 1, . . . , ii. For each of these choices, we compute the cointegration test. Next we use the Akaike Information Criterion (AIC) to select the most suitable value of n and the corresponding cointegration test statistic. This approach can be compared with kernel estimation of the innovations' density with data dependent bandwidth selection, see also Subsection 3.4. Similar approaches are adopted in the literature, see, e.g., Gallant and Tauchen (1997). We have no formal proof that the AIC results in an admissible rate of divergence of the polynomial degree n with the sample size T. Given our simulation results in Section 4, however, we conjecture that the use of the AIC for the SNP approach does not invalidate the inference procedure suggested in Section 2.

The quasi-likelihood maximization problem based on (13) is highly non-linear. Apart from this, however, there are no conceptual difficulties with obtaining the parameter estimates of θ and ν under the null for given n or for the AIC selected value of n. Again, these estimates are used to construct an estimate of R that can be used for inference purposes.

3.4 Adaptive approach

Whereas in the previous subsection we used semi-nonparametric density expansions in order to estimate the complete true likelihood from the data, here we use kernel estimation. We label this approach the adaptive one. Adaptive estimation has a long history, see, e.g., Manski (1984). It has been applied in the non-stationary time series context by Hodgson (1998) for the estimation of the parameters in (1) for known value of r, and by Shin and So (1998) for testing the rank of II in (1) in the univariate (k = 1) case. In the present paper we extend the results of the previous papers by constructing an adaptive cointegration test for the multivariate case. Though a formal proof of the validity of Theorem 1 for this adaptive test is beyond the scope of the present paper, such a proof can quite straightforwardly be constructed using the results of Hodgson (1998) and Shin and So (1998). This claim is strongly supported by the simulation results in Section 4.

The adaptive cointegration testing procedure is effectively a two-step

procedure. First, consistent estimates of the model parameters are constructed. For this we use the standard estimates based on the Gaussian quasi-likelihood. Let ε_t^* denote the t-th regression residual implied by these preliminary parameter estimates. Then the density of the innovations is estimated as

$$f^*(\varepsilon_t) = \frac{1}{T-1}\sum_{i \neq t} K_h(\varepsilon_i^* - \varepsilon_t), \tag{15}$$

where $K_h(\cdot)$ is a kernel, see further below for more details. The delete-one $(i \neq t)$ kernel estimate in (15) is needed to establish the limiting distribution of the test, see Hodgson (1998) and Shin and So (1998). The parameter ν is now equal to the true density of the innovations ε_t, and it is estimated using (15). Denote this estimate by ν^*. Given ν^*, one can update the preliminary estimate of 19 by doing a one-step Newton-Raphson improvement. The gradient needed for this one-step improvement can be obtained directly from (15) by straightforward differentiation. The Hessian, however, is replaced by the outer-product-of-gradient matrix. In this way, we avoid the explicit computation of second order derivatives. The replacement of the Hessian by the outer-product-of-gradient is valid asymptotically due to the information matrix equality and the consistency of the kernel estimator. Denote the up dated estimate of θ by $\bar{\theta}$. This $\bar{\theta}$ can used to construct residuals under the null $\alpha_{22} = 0$. Call these residuals $\bar{\varepsilon}_t$. The estimate of ν used to construct the cointegration test is now given by

$$\bar{\nu} = \bar{f}(\varepsilon_t) = \frac{1}{T-1}\sum_{i \neq t} \bar{K}_h(\bar{\varepsilon}_i - \varepsilon_t). \tag{16}$$

Again, (16) is used to construct the gradient of the likelihood and the outer-product-of-gradients. The test statistic is then given by (10), with $\bar{H}_T$ set equal to $\bar{J}_T$. Moreover, the canonical correlations between the gradients based on (15) and the residuals $\bar{\varepsilon}_t$ are used to estimate the nuisance parameter R needed for inference.

To complete the description of the adaptive approach, we have to give some more details on the choice of the kernel functions $K_h(\cdot)$ and $\bar{K}_h(\cdot)$. We only discuss $K_h(\cdot)$, as the definition of $\bar{K}_h(\cdot)$ is completely analogous. Let $\phi(\cdot)$ be the multivariate standard normal density function. Then

$$K_h(x) = h^{-1} \cdot |V|^{-1/2} \cdot \phi((V^*)^{-1/2}x/h), \tag{17}$$

where

$$V^* = \frac{1}{T-1}\sum_{t=1}^{T} \varepsilon_t^*(\varepsilon_t^*)^{\top}. \tag{18}$$

The scalar h denotes the bandwidth parameter. We choose Silverman's (1986) rule of thumb to select the bandwidth,

$$h = 0.96/T^{1/(4+k)}. \tag{19}$$

Note that no scale parameter is needed in the expression for h, as the scale parameter is already present through the matrix V in (18).

Following Hodgson (1998) and Shin and So (1998), we also consider a symmetrized version of the density estimates (15) and (16), e.g.,

$$f^{*,s}(\varepsilon_t) = \frac{1}{2(T-1)} \sum_{i \neq t} [K_h(\varepsilon_i^* - \varepsilon_t) + K_h(\varepsilon_i^* + \varepsilon_t)], \tag{20}$$

such that $f^{*,s}(\varepsilon_t) = f^{*,s}(-\varepsilon_t)$. A claim made by Hodgson is that the symmetrized version of the kernel estimate also has satisfactory properties for non-symmetric distributions. We investigate this claim in Section 4 by focusing on the properties of the adaptive cointegration test.

3.5 Fully nonparametric approach

All cointegration tests so far are centered around the parametric model (1). In this subsection, we briefly discuss the fully 'nonparametric cointegration testing procedure of Bierens (1997). Though the motivation underlying the test of Bierens differs from that underlying the adaptive approach, it is useful to confront the cointegration tests based on (1) with a test that does not use any parametric model at all.

It is beyond the scope of the present paper to give a detailed exposition of the implementation of Bierens' test. The essential idea is based on the following properties. Let $\{F_k(x), k = 1, 2, \ldots\}$ denote a set of functions on $[0, 1]$ satisfying $\int_0^1 F_k(x)dx = 0$ and $\int_0^1 F_j(x)F_k(x)dx = 0, j \neq k$; these conditions are satisfied, e.g., by $F_k(x) = \cos(2k\pi x)$. Next, define for any time series $\{z_t, t = 1, \ldots, T\}$, the weighted average

$$M_k(z) = \frac{1}{T} \sum_{t=1}^{T} F_k\left(\frac{t}{T}\right) z_t. \tag{21}$$

If $y_t \sim I(1)$, then it follows that $M_k(y) = O_p(T^{1/2})$. However, for a stationary linear combination $\beta' y_t$, we find $M_k(\beta' y) = \beta' M_k(y) = O_p(T^{-1/2})$. This different rate of convergence in the stationary and non-stationary directions is used to obtain a consistent test for the null hypothesis of no cointegration against the alternative of cointegration. The reader is referred

to the original paper, Bierens (1997), for more details on the test statistic and its null distribution. Here we only mention that we have implemented two versions of Bierens test. The first variant of the test (using $F_k(x) = \cos(2k\pi x)$) only provides the correct limiting distribution in case the assumption $\mu = \alpha\mu_0$ in Theorem 1 is satisfied. The second variant of the test uses $F_k(x) = \cos(2k\pi[Tx - \frac{1}{2}]/T)$, and is correct even if this assumption fails to hold, e.g., in case there are deterministic trends in the data.

4 Monte-Carlo set-up and results

Following the simulation experiment in Boswijk and Lucas (1997), we consider the following bivariate ($k = 2$) data generating process (DGP):

$$\Delta \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -c/T \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}. \tag{22}$$

More complicated DGP's involving genuine cointegrating relations and/or endogeneity of regressors can also be used without altering the results of the present paper. Of course, such alterations will affect the absolute rejection frequencies of our test procedures, but they will not affect the *ordering* of the different testing principles in terms of power performance, see also Lucas (1998). Dividing c in (22) by the sample size T follows the local alternatives formulation of Phillips (1988), see also Johansen (1991b). It allows us to investigate the effect of the sample size on the performance of the tests in an elegant way.

We consider 4 values for c, namely c = $0, 5, 10, 20$, and two sample sizes, namely $T = 100, 1000$. To limit the (heavy) computational burden of the simulations, we conduct 2000 and 1200 Monte-Carlo simulations for sample sizes $T = 100$ and $T = 1000$, respectively. Note that within each Monte-Carlo simulation, we need an additional simulation in order to determine the critical values or pvalues of the test, see the end of Section 2. We use 500 simulations to estimate the asymptotic pvalues. In our experience, this is sufficient from a practical point of view. We test the null of no cointegration against the alternative of at least one cointegrating relation. We use a 5% significance level. We set $\bar{n} = 3$ as the upper bound on the order of the SNP polynomial used in the simulations. This is done in order to limit the computational burden.

In order to investigate the performance of the tests under alternative conditions, we use several distributions for the innovations ε_t. Following Boswijk and Lucas (1997), we use:

- the standard bivariate normal distribution;

- the standard bivariate Student t distribution with 3 degrees of freedom;
- the standard bivariate Student t distribution with 1 degree of freedom, truncated such that 95% of the original probability mass is preserved;
- a χ^2 distribution with 3 degrees of freedom for each of the (independent) components of ε_t;
- an F distribution with degrees of freedom parameters equal to 3 and 3, respectively, for each of the (independent) components of ε_t;
- a mixture of three normals, each with unit covariance matrix, and with means $(0, -3/2)$ with probability 0.5, $(3/2, 7/6)$ with probability 0.3, and $(-9/4, 2)$ with probability 0.2;
- a mixture of four normals, each with unit covariance matrix, and with means $(\pm 3, \pm 3)$, each mixture component receiving probability 0.25.

This comprises a variety of different distributions, displaying skewness, fat-tailedness, and multimodality. We extend the above set of distributions for i.i.d. ε_t by considering bivariate ARCH and GARCH processes. ARCH and GARCH processes exhibit volatility clustering, a phenomenon that is important for financial time series. We concentrate on two processes:

- the components of ε_t are individually ARCH(l) with parameter 0.95, i.e., $\varepsilon_{it} = \sqrt{h_{it}}\eta_{it}$ for $i = 1, 2$ with $h_{it} = 1 + \alpha\varepsilon^2_{i,t-1}$, where $\alpha = 0.95$ and η_{it} is i.i.d. standard normal (over i as well as t);
- the components of ε_t are individually GARCH(1,1) with parameters 0.15 and 0.8, i.e., $\varepsilon_{it} = \sqrt{h_{it}}\eta_{it}$ for i = 1, 2 with $h_{it} = 1+\alpha\varepsilon^2_{i,t-1}+\beta h_{i,t-1}$, where $\alpha = 0.15$, $\beta = 0.8$, and η_{it} is i.i.d. standard normal (over i as well as t);

The parameters of the GARCH process are typically found in empirical applications for financial time series using daily data. The persistence of the volatility process for our GARCH(1,1) specification is $\alpha + \beta = 0.95$, which is also the value taken for the ARCH(l) parameter. By considering the volatility clustering processes, we can investigate the robustness of our results with respect to realistic deviations from Assumption 1.

The results of the simulations are presented in Tables 1 and 2.

First note that the sizes of the tests seem acceptable in all cases considered, excepting the ARCH(l) process. For the ARCH(l), all tests appear oversized, but the size distortion for the Gaussian tests is much greater than for the non-Gaussian tests. This is in accordance with results of Caner (1998).

Table 1:
Cointegration Test Performance Comparison, Sample Size $T = 100$

c	Joh	G	t(5)	SNP				A-S	A-NS	NP1	NP2
				0	12	3	AIC				
	Normal										
0	0.06	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.04
5	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.05	0.04
10	0.15	0.13	0.12	0.13	0.13	0.13	0.11	0.13	0.11	0.09	0.06	0.05
20	0.48	0.43	0.35	0.42	0.43	0.38	0.35	0.42	0.33	0.27	0.05	0.04
	t(3)										
0	0.06	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.04	0.03
5	0.08	0.07	0.16	0.16	0.16	0.15	0.14	0.16	0.11	0.12	0.06	0.05
10	0.16	0.14	0.39	0.39	0.38	0.36	0.33	0.38	0.28	0.27	0.06	0.05
20	0.48	0.42	0.78	0.78	0.79	0.77	0.70	0.77	0.63	0.60	0.05	0.04
	Truncated (95%) Cauchy										
0	0.07	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.03	0.04	0.05	0.04
5	0.08	0.07	0.45	0.48	0.38	0.36	0.30	0.48	0.22	0.28	0.05	0.05
10	0.17	0.15	0.84	0.84	0.80	0.75	0.67	0.84	0.57	0.60	0.05	0.04
20	0.48	0.42	0.99	0.99	0.99	0.98	0.94	0.99	0.91	0.90	0.06	0.05
	$\chi^2(3)$										
0	0.05	0.04	0.05	0.05	0.06	0.05	0.05	0.05	0.04	0.03	0.05	0.04
5	0.07	0.07	0.11	0.10	0.12	0.13	0.13	0.13	0.12	0.20	0.05	0.05
10	0.15	0.13	0.22	0.22	0.30	0.30	0.30	0.32	0.26	0.46	0.05	0.04
20	0.46	0.40	0.56	0.59	0.71	0.72	0.74	0.74	0.61	0.85	0.06	0.05
	$F(3,3)$										
0	0.05	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.06	0.04
5	0.06	0.05	0.85	0.87	0.93	0.92	0.86	0.93	0.84	0.89	0.06	0.04
10	0.10	0.09	0.95	0.96	0.98	0.98	0.92	0.98	0.97	0.99	0.04	0.04
20	0.37	0.30	0.98	0.97	0.97	0.96	0.91	0.97	1.00	1.00	0.07	0.05
	Mixture of normals, I										
0	0.06	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.04	0.04	0.05	0.04
5	0.08	0.07	0.07	0.06	0.06	0.12	0.16	0.16	0.06	0.14	0.04	0.04
10	0.15	0.13	0.13	0.12	0.14	0.27	0.40	0.40	0.12	0.34	0.05	0.04
20	0.49	0.43	0.38	0.43	0.47	0.68	0.82	0.81	0.38	0.74	0.06	0.06

Table 1: (continued)

c	Joh	G	$t(5)$	SNP					A-S	A-NS	NP1	NP2
				0	1	2	3	AIC				
	Mixture of normals, II											
0	0.06	0.04	0.05	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.05	0.04
5	0.09	0.08	0.05	0.07	0.08	0.11	0.53	0.52	0.39	0.66	0.04	0.04
10	0.16	0.13	0.05	0.13	0.16	0.21	0.84	0.84	0.79	0.96	0.06	0.05
20	0.46	0.41	0.16	0.41	0.49	0.49	0.94	0.94	0.98	1.00	0.05	0.05
	ARCH(l), α = 0.95											
0	0.24	0.21	0.06	0.06	0.07	0.08	0.09	0.07	0.08	0.08	0.07	0.06
5	0.24	0.23	0.14	0.15	0.15	0.16	0.15	0.15	0.13	0.15	0.07	0.06
10	0.33	0.30	0.24	0.24	0.25	0.26	0.25	0.23	0.21	0.22	0.09	0.06
20	0.61	0.56	0.35	0.36	0.41	0.44	0.40	0.35	0.31	0.30	0.13	0.08
	GARCH(1,1), α = 0.15, β = 0.8											
0	0.08	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04
5	0.11	0.10	0.07	0.09	0.08	0.09	0.09	0.09	0.09	0.08	0.06	0.05
10	0.18	0.16	0.11	0.15	0.15	0.14	0.14	0.14	0.14	0.13	0.06	0.04
20	0.52	0.46	0.29	0.41	0.40	0.40	0.38	0.40	0.34	0.29	0.06	0.05

Note: The table contains empirical rejection frequencies over 2000 Monte Carlo replications of several cointegration testing procedures. The data generating process is $\Delta y_{1t} = \varepsilon_{1t}$, $\Delta y_{2t} = c \cdot y_{2,t-1}/T + \varepsilon_{2t}$. c is the non-centrality parameter in the data generating process, such that c = 0 gives an indication of the size of the test. Joh is the Johansen (1991) test statistic. G and *t(5)* are the cointegration LM test statistics based on a Gaussian and a Student *t(5)* quasi-likelihood, respectively. SNP is the semi-nonparametric approach, with fixed order of the polynomial equal to 0, 1, 2, and 3. The column AIC indicates that the order of the SNP expansion is determined through the Akaike information criterion. A-S and A-NS give the results for the Hodgson (1997) type adaptive cointegration test with symmetrized and non-symmetrized kernel density estimator, respectively. NP1 and NP2 give the results for Bierens (1997) nonparametric test, without and with taking care of deterministic trends in the data, respectively. The table has nine panels, corresponding to different distributions for the innovations $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})^\top$. Normal is the standard normal distribution. *t(3)* is a Student t distribution with 3 degrees of freedom. Truncated Cauchy gives drawings from a standard Cauchy distribution. The drawings are discarded if $\varepsilon_t^\top \varepsilon_t$ exceeds the 95th percentile of the $F(1,1)$ distribution. For the $\chi^2(3)$ and the $F(3,3)$ distribution, ε_{1t} and ε_{2t} are drawn independently from the mentioned distributions. The first mixture of normals consists of 3 normals with unit covariance matrix. The means are $(0, -3/2)$ with probability 0.5, $(3/2, 7/6)$ with probability 0.3, and $(-9/4, 2)$ with probability 0.2. The second mixture of normals has four normals with unit variance matrix and means $(\pm 3, \pm 3)$, all selected with equal probability 0.25. For ARCH(l), $\varepsilon_{it} = \sqrt{h_{it}}\eta_{it}$, $h_{it} = 1 + \alpha\varepsilon_{i,t-1}^2$, with α = 0.95 and η_{it} i.i.d. standard normal (over i as well as t). For GARCH, similarly $h_{it} = 1 + \alpha\varepsilon_{i,t-1}^2 + \beta h_{i,t-1}$, with α = 0.15 and 4 = 0.8.

Table 2:
Cointegration Test Performance Comparison, Sample Size $T = 1000$

c	Joh	G	t(5)	SNP				A-S	A-NS	NP1	NP2
				0	12	3	AIC				
	Normal										
0	0.06	0.06	0.06	0.05	0.05	0.06	0.06	0.05	0.06	0.06	0.07	0.07
5	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.07	0.06	0.06
10	0.14	0.14	0.13	0.13	0.15	0.14	0.14	0.13	0.13	0.12	0.05	0.05
20	0.44	0.43	0.35	0.43	0.43	0.43	0.41	0.43	0.34	0.33	0.05	0.05
	$t(3)$										
0	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.05	0.06	0.07	0.07
5	0.08	0.08	0.21	0.21	0.20	0.18	0.16	0.21	0.20	0.17	0.06	0.06
10	0.17	0.16	0.52	0.53	0.50	0.46	0.36	0.52	0.47	0.43	0.05	0.05
20	0.42	0.42	0.92	0.93	0.92	0.91	0.79	0.92	0.89	0.88	0.06	0.06
	Truncated (95%) Cauchy										
0	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.04	0.06	0.05	0.05
5	0.07	0.07	0.53	0.57	0.45	0.43	0.25	0.57	0.50	0.51	0.06	0.05
10	0.17	0.17	0.94	0.95	0.91	0.87	0.60	0.95	0.90	0.90	0.05	0.05
20	0.43	0.42	1.00	1.00	1.00	0.98	0.85	1.00	1.00	1.00	0.06	0.06
	$\chi^2(3)$										
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.08	0.08	0.11	0.12	0.14	0.13	0.13	0.14	0.22	0.36	0.07	0.07
10	0.16	0.16	0.26	0.25	0.35	0.32	0.30	0.34	0.56	0.81	0.06	0.06
20	0.43	0.43	0.65	0.68	0.83	0.81	0.75	0.80	0.95	0.99	0.08	0.08
	$F(3,3)$										
0	0.05	0.05	0.04	0.05	0.04	0.05	0.04	0.04	0.04	0.03	0.06	0.06
5	0.06	0.06	0.99	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.06	0.06
10	0.10	0.10	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.06	0.06
20	0.37	0.36	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.05	0.05
	Mixture of normals, I										
0	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.06	0.05	0.05
5	0.08	0.08	0.07	0.08	0.08	0.15	0.16	0.16	0.07	0.19	0.06	0.06
10	0.15	0.15	0.15	0.15	0.17	0.38	0.46	0.46	0.14	0.47	0.06	0.06
20	0.46	0.45	0.41	0.46	0.50	0.82	0.86	0.86	0.43	0.92	0.05	0.05

Table 2: (continued)

c	Joh	G	$t(5)$	SNP					A-S	A-NS	NP1	NP2
				0	1	2	3	AIC				
	Mixture of normals, II											
0	0.06	0.06	0.06	0.06	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.06
5	0.09	0.09	0.06	0.09	0.09	0.10	0.43	0.43	0.47	0.71	0.05	0.05
10	0.15	0.14	0.07	0.15	0.17	0.17	0.71	0.71	0.91	0.97	0.05	0.05
20	0.44	0.43	0.14	0.43	0.55	0.52	0.89	0.89	1.00	1.00	0.06	0.06
	ARCH(l), $\alpha = 0.95$											
0	0.14	0.13	0.07	0.06	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06
5	0.18	0.18	0.26	0.26	0.26	0.25	0.19	0.26	0.25	0.28	0.05	0.05
10	0.28	0.27	0.54	0.53	0.53	0.52	0.39	0.53	0.52	0.53	0.05	0.05
20	0.54	0.54	0.88	0.88	0.89	0.87	0.67	0.87	0.85	0.83	0.06	0.06
	GARCH(1,1), $\alpha = 0.15, \beta = 0.8$											
0	0.08	0.08	0.06	0.07	0.07	0.08	0.08	0.07	0.07	0.07	0.05	0.05
5	0.10	0.10	0.09	0.10	0.10	0.09	0.10	0.10	0.09	0.10	0.05	0.05
10	0.18	0.18	0.15	0.16	0.16	0.16	0.15	0.16	0.14	0.17	0.05	0.05
20	0.46	0.45	0.37	0.43	0.43	0.44	0.40	0.42	0.36	0.34	0.06	0.06

NOTE: The table contains empirical rejection frequencies over 1200 Monte Carlo replications of several cointegration testing procedures. The data generating process is $\Delta y_{1t} = \varepsilon_{1t}$, $\Delta y_{2t} = c \cdot y_{2,t-1}/T + \varepsilon_{2t}$. c is the non-centrality parameter in the data generating process, such that c = 0 gives an indication of the size of the test. Joh is the Johansen (1991) test statistic. G and t(5) are the cointegration LM test statistics based on a Gaussian and a Student *t(5)* quasi-likelihood, respectively. SNP is the semi-nonparametric approach, with fixed order of the polynomial equal to 0, 1, 2, and 3. The column AIC indicates that the order of the SNP expansion is determined through the Akaike information criterion. A-S and A-NS give the results for the Hodgson (1997) type adaptive cointegration test with symmetrized and non-symmetrized kernel density estimator, respectively. NP1 and NP2 give the results for Bierens (1997) nonparametric test, without and with taking care of deterministic trends in the data, respectively. The table has nine panels, corresponding to different distributions for the innovations $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})^\top$. Normal is the standard normal distribution. *t(3)* is a Student t distribution with 3 degrees of freedom. Truncated Cauchy gives drawings from a standard Cauchy distribution. The drawings are discarded if $\varepsilon_t^\top \varepsilon_t$ exceeds the 95th percentile of the F(l, 1) distribution. For the $\chi^2(3)$ and the $F(3,3)$ distribution, ε_{1t} and ε_{2t} are drawn independently from the mentioned distributions. The first mixture of normals consists of 3 normals with unit covariance matrix. The means are (0, $-3/2$) with probability 0.5, $(3/2, 7/6)$ with probability 0.3, and $(-9/4, 2)$ with probability 0.2. The second mixture of normals has four normals with unit variance matrix and means $(\pm 3, \pm 3)$, all selected with equal probability 0.25. For ARCH(l), $\varepsilon_{it} = \sqrt{h_{it}}\eta_{it}$, $h_{it} = 1 + \alpha\varepsilon_{i,t-1}^2$, with $\alpha = 0.95$ and η_{it} i.i.d. standard normal (over i as well as t). For GARCH, similarly $h_{it} = 1 + \alpha\varepsilon_{i,t-1}^2 + \beta h_{i,t-1}$, with $\alpha = 0.15$ and $\beta = 0.8$.

Caner proves that the Gaussian cointegration test with infinite variance errors has critical values which lie to the left of those of Johansen (1988,1991a). As the ARCH(l) considered is close the the infinite variance region, the high rejection frequencies for c = 0 can be expected. Note, however, that these size distortions do not appear if the ARCH effect is not dominant, though the volatility persistence may be just as high, see the GARCH(1,1) results. Smaller size distortions in case of ARCH(l) for the robust tests and the SNP tests can also be expected, see also Lucas (1998), as the critical values of these tests generally also lie to the left of those of Johansen. Note that the SNP and adaptive density estimates will generally be fat-tailed, because the unconditional distribution of ε_t for GARCH processes is fat-tailed, see Nelson (1990).

Given the satisfactory result for the level of the tests, we concentrate the remaining discussion on the power properties of the tests. We first discuss the case of small sample sizes, T = 100. Next, we deal with larger samples, T = 1000.

For the Gaussian distribution, we see that the Gaussian QLR test is optimal, closely followed by the Gaussian *QLM* test and the SNP(0) based test, SNP(0) denoting the 0th order SNP expansion, i.e., the Student t with estimated degrees of freedom. As expected in this case, the power of the SNP test generally decreases if the degree of the polynomial is (unnecessarily) increased. The AIC based SNP approach, however, succeeds in selecting the appropriate order of the polynomial, such that the power behavior of SNP(AIC) almost coincides with that of the Gaussian *QLR* and *QLM* tests. We also note the familiar power loss of the Student t approach with fixed degrees of freedom parameter *(t(5))*, see, e.g., Lucas (1998). Furthermore, the power behavior of the adaptive approach is much worse than that of the SNP(AIC) approach, especially for the non-symmetrized kernel density estimator. Finally, note that the nonparametric test has virtually no power whatsoever. This holds consistently throughout the simulations. We therefore refrain from further comments on the fully nonparametric approach in the subsequent discussion.

For the Student *t(3)* distribution and the truncated Cauchy, some of the results are changed dramatically. The power of the Gaussian tests is similar to the case of Gaussian ε_t. The power of the SNP, *t(5),* and adaptive approaches, however, are increased substantially. Again we note a decrease in power for the SNP approach if the degree of the polynomial is increased. Also, the performance of the *t(5)* based *QLM* test and of the SNP(AIC) test are indistinguishable. The power of the adaptive tests clearly falls below that of the SNP(AIC) approach. Note that absolute power substantially increases with the degree of leptokurtosis.

We now turn to the skewed distributions. If the distribution is thin-tailed (χ^2 (3)), we note a similar power behavior of the Gaussian tests as for normally distributed innovations. The robust test based on the t(5) quasi-likelihood clearly does better. However, the robust test is outperformed for distant alternatives by the adaptive procedure based on the incorrect symmetrized kernel density estimate. This, in turn, is outperformed by the SNP(AIC) approach, while finally the adaptive procedure based on the non-symmetrized kernel density estimator performs best. Note that the power of the SNP tests is generally increasing now in the degree of the polynomial. This is to be expected, as the quasi-likelihoods with the higher order polynomials are better suited at capturing the skewness. Quite similar results hold if the skewed distribution is also fat-tailed ($F(3,3)$). Note that the non-Gaussian cointegration tests reveal a substantial power increase with respect to the situation with thin-tailed skewed innovations ($\chi^2(3)$). By contrast, the Gaussian based procedures display a power loss.

We now turn to the multimodal distributions, i.e., the mixtures of normals. If multimodality is fairly limited and if there is also skewness (mixture I), we see the familiar behavior for the Gaussian tests. Also the Student t(5) based test and the adaptive (symmetrized) test display about the same behavior as for normally distributed innovations. By contrast, the SNP and the non-symmetrized adaptive approaches are able to detect this form of non-normality. For the SNP approach, however, the order of the polynomial has to be set high enough. It is comforting to note that the AIC generally succeeds in choosing sufficiently high polynomial orders. This results in a doubling of power for this sample size for distant alternatives. It is important to notice the strong effect of the incorrect imposition of a symmetrized kernel density estimate. Clearly, if skewness is expected, the non-symmetrized approach seems preferable to the symmetrized adaptive approach. If the multimodality is much stronger and if skewness is absent (mixture II), the results change. The power of the Gaussian based tests is still constant. By contrast, the power of the Student t(5) quasi-likelihood based test is very poor. This confirms simulation results for the univariate case of Shin and So (1998). The power of the SNP test is again increasing in the order of the polynomial, with the AIC selecting the appropriate order for power maximization. The symmetrized adaptive test also works quite well in this setting, though the power lags somewhat compared to the SNP(AIC) approach for not too distant alternatives. Surprisingly, the non-symmetrized adaptive approach performs even better than the symmetrized one.

The final two processes for the innovations exhibit volatility clustering. First note that there are substantial size distortions for the Gaussian tests, and to a lesser extent for the non-Gaussian based tests. This was explained at

the beginning of the present section. For the remainder, the results for both types of volatility clustering look very similar to the case of i.i.d. Gaussian innovations. The only striking difference is the low power of the Student t(5) based test for the GARCH process. Moreover, the AIC seems less able to pick the appropriate orders of the polynomial from a power maximization perspective.

We now investigate the effect of a larger sample size by discussing Table 2. For the Gaussian situation, increasing the sample size does not have a substantial impact on the results. Some patterns are emerging for the other distributions, however. Generally, power seems to increase if the unconditional distribution of the innovations is fat-tailed and heavily multimodal, while little is changed otherwise. Another interesting effect of an increased sample size is the relative ordering of SNP(AIC) and the adaptive approach in terms of power. Whereas the SNP(AIC) approach dominates for small sample sizes, for larger sample sizes the adaptive approach seems preferable if the distribution is thin-tailed and skewed or heavily multimodal. This holds at least if we only consider SNP expansions up to order 3. The results might change if higher orders were incorporated in the analysis. This is not unreasonable, as we can link the maximum order of the SNP polynomial (inversely) to the bandwidth parameter of the kernel estimator. Whereas the latter automatically decreases with the sample size, see (19), the former does not automatically increase in our present set-up. More parameters or higher order polynomials are needed to adequately capture skewness and multimodality if more observations are available. We leave this for further research.

We also note that increasing the sample size has no effect on the inability of the symmetrized adaptive approach to exploit moderate departures from normality in the form of moderate skewness and multimodality, see mixture I. By contrast, the SNP(AIC) and non-symmetrized adaptive approaches succeed in gaining power with respect to the Gaussian based test for the mixture I distribution, both for samples of size $T = 100$ and $T = 1000$.

Some final remarks are due concerning the computation time. For $T = 100$, the computation time of the SNP approach of order 3 is dramatically much higher than that of the adaptive approach. It appears, however, that the required computation time for the SNP approach increases approximately linearly, while that of the adaptive approach increases (at least) quadratically. For $T = 1000$, the computation time for SNP(3) is still higher than for the adaptive approach, but the percentage difference has decreased by a factor between 5 and 30, depending on the distribution considered. Note that the difference in computational burden would be reduced considerably if the adaptive approach is augmented with a cross-validation procedure for

bandwidth selection. In that case, the SNP approach might well become less computationally intensive than the adaptive approach.

We summarize the main findings of the simulations in the next section.

5 Conclusions

In this paper we have contributed to the literature on cointegration testing and the application of (semi)-nonparametric techniques to non-stationary data. We have constructed an adaptive multivariate cointegration test and confronted its performance to old and new alternative cointegration test procedures under a wide variety of different conditions. Several conclusions emerge.

First, it turns out to be possible to use semi-nonparametric (SNP) and nonparametric techniques efficiently in the construction of cointegration tests. By using these techniques, we can avoid the arbitrary specification of a quasi-likelihood. In finite samples, this can be done at either a substantial power loss with respect to statistically optimal procedures, or (almost) no loss at all, depending on the method used. The advantages already take effect for samples as small as 100, at least for bivariate processes. We have also shown that the power losses of cointegration tests based on arbitrarily chosen quasi-likelihoods, e.g., a Student t(5), can perform quite poorly if they fail to capture salient characteristics of the true likelihood.

As a second conclusion, for small sample sizes (T = 100) estimating the likelihood through SNP density expansions with the order of the expansion determined by the AIC is clearly preferable to kernel estimation in terms of overall performance of the corresponding cointegration test statistic. If the kernel estimate is symmetrized, it is not suited for picking up mild forms of skewness in contrast to the SNP approach. If it is not symmetrized, there is a substantial power loss with respect to the SNP approach for near-Gaussian innovations. For large sample sizes (T = 1000), the reverse result emerges from our simulation experiments. This is due to the fact that for both sample sizes we use the same upper bound on the order of the SNP expansion that is considered. The general pattern that emerges, therefore, is that the order of the SNP expansion must not be set to low. Otherwise, the potential power advantages of using SNP techniques will not materialize. It is then preferable to use kernel estimators for the density of the innovations.

A third clear conclusion from our simulation is that the fully nonparametric cointegration testing approach of Bierens has almost no power against simple reasonable alternatives. This implies there is further room for the development of fully nonparametric cointegration tests.

As always, several interesting directions for further research remain. First, the tests can be applied to real data. For example, when conducting a standard modeling exercise using the Johansen (1988,1991a) cointegration test, one can check the robustness of the results to the choice of the Gaussian quasi-likelihood by running one of the tests discussed in the present paper. If the results differ, closer inspection of the data and/or the model is warranted. As a second extension, one can further refine the adaptive cointegration test with cross-validation for the bandwidth selection. Some preliminary simulation results revealed, however, that little is gained. Also a formal proof of the validity of the AIC (or some other criterion) for the SNP approach would be very welcome. Finally, it is interesting to see how the tests perform if the data generating process is more complicated, especially if the dynamics in the system are more complicated than that of the simple VAR(l) used in the present paper.

References

[1] Abadir, K.M., and A. Lucas (1996). Approximate normality of t ratios based on M-estimators for the unit root. York University Research Memorandum 96/32.

[2] Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics 77, 379-404.*

[3] Boswijk, H.P., and J.A. Doornik (1998). Distribution approximations for cointegration tests with stationary exogenous regressors. Manuscript, University of Amsterdam.

[4] Boswijk, H.P., and A. Lucas (1997). Semi-nonparametric cointegration testing. VU Research Memorandum 1997-41, Vrije Universiteit, Amsterdam.

[5] Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics 86,* 155-175.

[6] Clarida, R.H. and M.P. Taylor (1997). The term structure of forward exchange premiums and the forecastability of spot exchange markets: correcting the errors. *Review of Economics and Statistics 74,* 353-361.

[7] Doornik, J.A. (1998). Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* **12, 573-594.**

[8] Franses, P.H., T. Kloek, and A. Lucas (1998). Outlier robust analysis of long-run marketing effects for weekly scanning data. *Journal of Econometrics 89,* 293-315.

[9] Franses, P.H., and A. Lucas (1998). Outlier detection in cointegration analysis. *Journal of Business and* Economic *Statistics* 16, 459-468.

[10] Gallant, A.R., and D.W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55, 363-390.*

[11] Gallant, A.R., and G. Tauchen (1997). Reprojecting partially observed systems with application to interest rate diffusions. Working paper, 1997.

[12] Hamilton, J.D. (1994). *Time* Series Analysis. Princeton: Princeton University Press.

[13] Hodgson, D.J. (1998). Adaptive estimation of error correction models. *Econometric Theory* 14, 44-69.

[14] Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

[15] Johansen, S. (1991a). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59,* 1551-1580.

[16] Johansen, S. (1991b). The power function of the likelihood ratio test for cointegration. J. Gruwer (ed.), *Econometric Decision Models: New Methods* and *Applications,* Berlin: Springer Verlag, pp. 323-335.

[17] Kleibergen, F.R. , and H.K. van Dijk (1994). Direct cointegration testing in error correction models. *Journal of Econometrics 63, 61-103.*

[18] Lucas, A. *(1996). Outlier Robust Unit Root* Analysis. Ph.D. Thesis, Erasmus University Rotterdam. Amsterdam: Thesis Publishers. http://www.econ.vu.nl/medewerkers/alucas/default.htm.

[19] Lucas, A. (1997a). Cointegration testing using pseudo likelihood ratio tests. *Econometric* Theory 13, 149-169.

[20] Lucas, A. (199713). Robustness of the Student t based M-estimator. Com*munications in Statistics; Theory and Methods 26,* 1165-1182.

[21] Lucas, A. (1998). Inference on cointegrating ranks using LR and LM tests based on pseudo-likelihoods. *Econometric Reviews* 17, 185-214.

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

[23] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory 6,* 318-334.

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

[25] Shin, D.W., and B.S. So (1998). Unit root tests based on adaptive maximum likelihood estimation. *Econometric Theory*, forthcoming.

[26] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.