Faculteit der Economische Wetenschappen en Econometrie

# Serie Research Memoranda

## Semi-nonparametric cointegration testing

H. Peter Boswijk
André Lucas

vrije Universiteit *amsterdam*

# SEMI-NONPARAMETRIC COINTEGRATION TESTING

By H. Peter Boswijk and André Lucas

*of* **Amsterdam and** *Vrije Universiteit* **Amsterdam**

This paper considers a semi-nonparametric cointegration test. The test uses the LM-testing principle. The score function needed for the LM-test is estimated from the data using an expansion of the density around a Student $t$ distribution. In this way, we capture both the possible fat-tailedness and the skewness of the innovation process. Model selection criteria are employed to select the appropriate order of the expansion in finite samples. Using a Monte-Carlo experiment, we show that the semi-nonparametric cointegration test has good size and power properties. The test outperforms previous testing procedures in terms of power over a broad class of distributions for the innovation process.

1. **Introduction.** The low power of **univariate unit** root tests and multivariate unit root (or cointegration) tests has been an ongoing puzzle since the introduction of formal unit root testing procedures. The seminal literature on unit root testing is generally based on the Gaussian distribution, which is evident from the use of estimation techniques like ordinary least-squares and Gaussian maximum likelihood, see, e.g., Fuller (1976), Engle and Granger (1987), and Johansen (1988). In order to improve the power of these tests, two complementary approaches have been put forward in the literature. The first approach concentrates on the fact that the Gaussian based procedures have difficulty in approaching the power envelope dictated by the set of point optimal unit root tests. Power can be improved substantially by considering (exponentially) weighted averages of point optimal tests, see, e.g., Elliot *et* al. (1996). The second approach concentrates on the distributional assumptions underlying conventional unit root tests. Although the assumption of Gaussian disturbances may be appropriate in some circumstances, it is highly questionable in other settings. If the stochastics in the model are inherently non-Gaussian, it is intuitively clear that efficiency and power can be gained by exploiting this non-normality during the estimation and testing stage of the modeling process. Examples of this type of literature in the unit root context are, e.g., Lucas (1996a,b) and Hodgson (1997).

The present article fits in the second strand of literature mentioned above. We study cointegration tests that exploit the possible non-normality of a set of time-series in performing a cointegration test. The non-normality is captured by using

a likelihood based testing procedure where the likelihood function is estimated from the data using semi-nonparametric density expansions as in Gallant and Nychka (1987).

Several non-Gaussian likelihood based cointegration testing procedures like the Likelihood Ratio (LR) test, the Wald test, and the Lagrange Multiplier (LM) test, have been studied in Lucas (1996a,b, 1997). In all these papers, the form of the likelihood is taken as given and specified by the researcher, such that the likelihood based tests are, in fact, pseudo-likelihood or quasi-likelihood based tests, see White (1982) and Gouriéroux et al. (1984). The main conclusion emerging from the papers on non-Gaussian cointegration tests is that power can increase dramatically if one uses a non-Gaussian pseudo-likelihood for estimation in situations with fat-tailed errors. This holds even if the non-Gaussian pseudo-likelihood does not coincide with the true likelihood. The power gain has to be paid for in terms of a power loss if innovations turn out to be Gaussian after all. This power loss can, however, be kept between reasonable bounds by an appropriate choice of the pseudo-likelihood.

Given the above general conclusions, it seems straightforward to look for adaptive procedures for cointegration testing. Such procedures have the advantage that power is gained if innovations are fat-tailed, while at the same time no power is lost (asymptotically) if innovations turn out to be Gaussian. This contrasts both with the results for Gaussian pseudo-likelihood based cointegration tests as Johansen (1988,1991) and cointegration tests based on a        non-Gaussian pseudo-likelihood.

There are several ways to implement adaptive procedures into cointegration tests, varying from fully non-parametric kernel based methods to fully parametric modeling techniques, e.g., using the Student $t$ distribution with estimated degrees of freedom instead of the normal distribution. In this paper we make use of semi-nonparametric techniques. We use polynomial densities like the ones proposed by Gallant, and Nychka (1987) as a pseudo-likelihood function. By letting the degree of the polynomial diverge with the sample size, we are able to let the pseudo-likelihood come arbitrarily close to the true likelihood function. The use of semi-nonparametric techniques has advantages over the use of both the fully non-parametric and the fully parametric approach. The advantage over the parametric approach is evident: less restrictive assumptions have to be made about the distribution of the time-series process. There are also some advantages over the fully non-parametric approach, however. As the semi-nonparametric approach uses parametric elements, this can be more efficient in finite samples than a fully non-parametric procedure. Moreover, the parametric elements in the model might, enhance the interpretability of the final results. Finally, the semi-nonparametric approach more easily allows for the use of standard model selection procedures for model reduction in order to improve efficiency in finite samples.

The original polynomial densities studied by Gallant and Nychka (1987) essentially use the product of the normal density and the square of a polynomial. The normal density can in this setup be considered as the central density or the

leading term in the expansion. In the present article we use a slightly different class of polynomial densities based on a fat-tailed central density. We argue that this is more appropriate in finite samples. Deviations from normality essentially have two important aspects: deviations in the middle of the distribution and deviations in the tails. While the polynomial terms are suited to take care of the former type of deviations, they are less adequate for dealing with the latter, unless high order polynomials are used. Therefore, we use the Student $t$ distribution as the central density. In this way we can parsimoniously capture both departures from normality mentioned above, including phenomena such as skewness and multimodality as well as fat-tailedness.

Given our choice for the semi-nonparametric approach based on polynomial densities with a fat-tailed central density, we are left with the choice of the testing principle. In our case. we choose between the LR, the LM, and the Wald testing principles. The Wald cointegration test is least attractive in the present context, as this test can be very sensitive to trivial transformations of the data, like reordering the variables, see Kleibergen and Van Dijk (1994) and Lucas (1996a, Chapter S). By contrast, both the LR and the LM test have the attractive property that, they are invariant to non-singular reparameterizations of the model. In the present article, we focus on the LM testing principle for three reasons. First. unlike the LR test, the LM test can be corrected for the possible misspecification of the pseudo-likelihood using a White-type information matrix estimator. The failure of the LR. test to make such a correction leads to additional nuisance parameters in the limiting distribution of the LR statistic. In the present context of scm-nonparametric cointegration testing this may seem less important, as the additional nuisance parameters should vanish asymptotically because the SNP estimator envisages to provide a consistent estimate of the true likelihood function. The second reason for preferring the LM test over the LR test is computation time. The LM test requires estimation under the null only, whereas the LR test requires estimation under both the null and the alternative. Although one might, argue that with the present state of computer technology the additional effort, of computing the LR test is only marginal, it is the experience of the authors that the computational advantage of the LM test in the present context is substantial. This is intuitively clear, because the estimation problem for cointegrated time-series models based on polynomial Student $t$ densities is highly non-linear in the parameters. Moreover, as testing for cointegration involves a mult ivariate estimation problem, the computational burden increases rapidly with t he order of the semi-nonparametric density expansion. Therefore, the merits of the LM test over the LR test from a computational point of view remain considerable in the present, context. The third reason for preferring the LM test over the LR test emerges from Lucas (199613) and concerns the stability of the test, if the data exhibit volatility clustering. Lucas shows that the size distortions of' the LM test are generally lower than those of the LR test if the data display autoregressive conditional heteroskedasticity (ARCH), especially if the ARCH effects are highly persistent. Because one possible area of application of our semi-nonparametric cointegration testing procedures is in financial eco-

nomics and as many financial economics time-series display volatility clustering, the reduced size tlistortions of the LM test make the LM test preferable over the LR test.

The contributions of the present article to the contemporary literature are the following. First, we contribute to the literature on non-Gaussian cointegration testing by using a pseudo-likelihood that is dictated by the data. In this way we complement the results of Shin and So (1997), who study adaptive univariate unit root tests, and of Hodgson (1997), who studies adaptive inference procedures for the stationary relations in multivariate cointegrated time-series models. We also generalize the results of Lucas (1996a,b, 1997) for fixed pseudo-likelihoods. Second, we use semi-nonparametric techniques instead of fully non-parametric kernel-based methods to estimate the true likelihood function in the present multivariate context. Third, we derive the limiting distribution of the semi-nonparametric cointegration test and propose an inference procedure that can easily be implemented in practice. Fourth, we present a simulation experiment comparing the proposed method with several likelihood based cointegration tests available in the literature. It turns out that the adaptive cointegration test that uses a traditional model selection criterion to determine the degree of the polynomial in the tlensity expansion has the best overall performance.

The article is set up as follows. Section 2 discusses the model that is used and introduces the pseudo-maximum likelihood estimator and the non-Gaussian LM cointegration test. Section 3 presents the implementation of the semi-nonparametric cointegration LM test. Section 4 gives the asymptotic distribution theory for the proposed testing procedure. Section 5 presents the results of a Monte-Carlo simulation experiment in which the performances of several (old and new) testing procedures are evaluated for a variety of distributional assumptions. Finally, Section 6 concludes the paper with some suggestions for future research.

2. **The testing procedure.** We consider the vector autoregressive (VAR) model of order $p$.

$$(1) \qquad \Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \mu + \varepsilon_t,$$

where $\Pi$ and $\Phi_i$, $i = 1, \ldots, p$ are $k \times k$ parameter matrices, $\mu$ is a constant term, and $\varepsilon_t$ is an innovation. We assume that $\{\varepsilon_t\}$ is an independently and identically distributed (i.i.d.) process and that $y_t$ is observed for $t = -p+1, \ldots, T$. The key parameter of interest in the present article is $r = \mathrm{rank}(\Pi)$. If $r < k$, the matrix $\Pi$ can be decomposed into $\bar{\alpha}\bar{\beta}^\top$, with $\bar{\alpha}$ and $\beta$ two $k \times r$ matrices of full column rank $r$, see, e.g., Johansen (1991).

Following the decomposition of $\Pi$ of Kleibergen and Van Dijk (1994), we

obtain that (1) can be rewritten as[1]

$$(2) \qquad \Delta y_t = \alpha(I_r, \beta^\top) y_{t-1} + E_2 \alpha_{22} y_{2,t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \mu + \varepsilon_t,$$

where $\alpha$ is again a $k \times r$ matrix of full column rank, $I_r$ denotes the identity matrix of order r, $\beta$ is a $(k-r) \times r$ matrix, $\alpha_{22}$ is a $(k-r) \times (k-r)$ matrix, $E_2 = (0, I_{k-r})^\top$ is a $k \times (k-r)$ matrix, and $y_{2,t-1}$ denotes a vector containing the last $(k-r)$ elements of $y_{t-1}$. Under the null hypothesis $H_0 : \text{rank}(\Pi) = r$, we have $\alpha_{22} = 0$. So the null of no cointegration can be tested by testing whether $\alpha_{22} = 0$ or not. In fact, the trace cointegration test of Johansen (1991) amounts to performing a likelihood ratio (LR) test for $H_0 : \alpha_{22} = 0$ versus the alternative $\alpha_{22} \neq 0$, see also Lucas (1997). A Wald and a Lagrange Multiplier (LM) test for $H_0 : \alpha_{22} = 0$ can be found in Kleibergen and Van Dijk (1994) and Lucas (1996a,b).

In the present paper we confine ourselves to the LM cointegration test for the reasons outlined in the introduction. The LM test for cointegration using non-Gaussian (quasi- or pseudo-)likelihoods was introduced by Lucas (199613). Our test extends the analysis of Lucas in that we allow the (quasi-)score to be estimated from the data. The precise procedure used to achieve this objective is outlined in Section 3. In the remainder of the present section, we provide some more details on the LM cointegration test.

Let

$$(3) \qquad \mathcal{L}(\theta) \propto \prod_{t=1}^{T} |R| . \exp(-\rho(R^\top \varepsilon_t(\theta)))$$

denote a quasi- or pseudo-likelihood function (cf. White, 1982, and Gouriéroux *et al.,* 1984), with $\theta$ the vector of unknown parameters, with $R$ a $k \times k$ matrix such that $(R^\top R)^{-1}$ is a positive definite scaling matrix (the covariance matrix of $\varepsilon_t$ in the Gaussian case), and $\varepsilon_t(\theta) = \Delta y_t - \alpha(I_r, \beta^\top) y_{t-1} - E_2 \alpha_{22} y_{2,t-1} - \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} - \mu$. Let $\theta^\top = (\theta_1^\top, \theta_2^\top, \theta_3^\top)$, where $\theta_1 = \text{vec}(\alpha_{22})$ (with $\text{vec}(\cdot)$ the column-stacking operator), $\theta_2$ contains the parameters $\Phi_i, \mu, \alpha,$ and $\beta$, and $\theta_3$ is a vector of unknown parameters characterizing $R$ and the function $\rho(\cdot)$. Define the log-likelihood function $\ell(\theta) = \log \mathcal{L}(\theta) = \sum_{t=1}^{T} \ell_t(\theta)$, the gradient $G(\theta) = \partial \ell(\theta)/\partial \theta = \sum_{t=1}^{T} G_t(\theta)$, minus the Hessian matrix $H_C(\theta) = -\partial^2 \ell(\theta)/\partial \theta \partial \theta^\top$, and the outer-product-of-gradient matrix $H_\Omega(\theta) = \sum_{t=1}^{T} G_t(\theta) G_t(\theta)^\top$. Then the

---

[1] A decomposition of (1) as in (2) is not always possible. The decomposition is invalid if the leading $r \times r$ submatrix of the cointegrating vectors $\bar{\beta}$ is singular, in which case it cannot be normalized to the identity matrix. This complication is irrelevant for most situations of practical interest, because for empirical data singularity of the leading submatrix of $\bar{\beta}$ occurs with probability zero under suitable assumptions for the innovation process $\{\varepsilon_t\}$. Even if the singularity occurs, a somewhat more complicated decomposition applies for which all the results in this paper hold true, see Lucas (1996b).

LM statistic for $H_0 : \alpha_{22} = 0$ with a possibly misspecified likelihood is given by

(4)  $$\boldsymbol{LM} = (S_1 H_C^{-1} G)^\top (S_1 H_C^{-1} H_\Omega (H_C^{-1})^\top S_1^\top)^{-1} S_1 H_C^{-1} G,$$

with $S_1 = (I_{(k-r)^2}, 0)$ a selection matrix such that $S_1 \theta = \theta_1 = \mathrm{vec}(\alpha_{22})$, and with G, $H_C$, and $H_\Omega$ evaluated at the pseudo-maximum likelihood estimator $\tilde{\theta}$ under the restriction $\alpha_{22} = 0$. The form of the LM test statistic in (4) is standard, see, e.g., Gallant (1987, p. 230). We consider that part of the gradient (normalized by the information matrix) that corresponds to $\alpha_{22}$. Next, we test whether this part of the gradient is sufficiently close to zero at the estimates under the null. To cope with the possible misspecification of the likelihood, the matrix in the middle of (4) is needed, see also White (1982). If the likelihood were correctly specified, the matrices $H_C$ and $H_\Omega$ would have the same expected value, and after suitable normalization, the same limit (in distribution), such that the LM statistic could be simplified somewhat further.

Suppose that $\theta_3$ is void, so that both the scaling matrix and the form of the pseudo-likelihood are known. Defining $\psi_t(\theta) = \psi(\varepsilon_t(\theta)) = \partial\rho(R^\top \varepsilon_t(\theta))/\partial\varepsilon_t$ and $Z_t(\theta) = -\partial\varepsilon_t^\top(\theta)/\partial\theta$, the gradient becomes G(B) $= \sum_{t=1}^T Z_t(\theta)\psi_t(\theta)$, whereas asymptotically valid choices for $H_C$ and $H_\Omega$ **are** given by

$$H_C(\theta) = \sum_{t=1}^T Z_t(\theta) C_{1T}(\theta) Z_t(\theta)^\top, \qquad H_\Omega(\theta) = \sum_{t=1}^T Z_t(\theta) \Omega_{22T}(\theta) Z_t(\theta)^\top,$$

where $C_{1T}(\theta) = T^{-1} \sum_{t=1}^T \partial\psi(\varepsilon_t(\theta))/\partial\varepsilon_t^\top$, and $\Omega_{22T}(\theta) = \sum_{t=1}^T \psi_t(\theta)\psi_t(\theta)^\top$. In this form, the LM statistic was first proposed and analyzed by Lucas (1996b), albeit with the scaling matrix $(R^\top R)^{-1}$ estimated jointly with the other parameters. In the next section, we extend his analysis to the case where the form of the pseudo-likelihood (i.e., the function $\rho(\cdot)$) is estimated as well.

**3. Semi-nonparametric estimation of the score function.** The LM cointegration test statistic based on a non-Gaussian likelihood was already introduced in Lucas (1996b). This test has superior power compared to the Johansen (1991) trace test if the innovations are fat-tailed. If the innovations are Gaussian, however, the LM cointegration test based on a fixed non-Gaussian pseudo-likelihood has less power than Johansen's test. A natural extension of previous work in this area, therefore, is to let the data decide on which pseudo-likelihood has to be used. If the data are nearly normally distributed, Johansen's test can be used best. By contrast, if the data are fat-tailed or heavily skewed, the use of a non-Gaussian pseudo-likelihood is called for. The form of the pseudo-likelihood is determined by the choice of $\rho(\cdot)$ in (3). So by varying $\rho(\cdot)$, we can tune the performance of the LM cointegration test under different distributional assumptions for the innovation process.

Related work in this area consists of papers by, e.g., Hodgson (1997) on the adaptive estimation of the long-run parameters in vector error-correction models, and Shin and So (1997) on adaptive univariate unit root testing. In both papers, kernel-based methods are used to estimate the pseudo-log-likelihood $\rho(\cdot)$ in (3).

In the present paper we use the semi-nonparametric (SNP) techniques as intro-
duced in Gallant and Nychka (1987). The advantage of the SNP approach over
the kernel-approach lies in the fact that the semi-nonparametric approach allows
one to employ certain parametric elements. In particular, one can choose a para-
metric family of densities as the leading term in a Hermite expansion of the true
likelihood, see the comments in Gallant and Nychka (1987). This is especially
useful if one has some information about the shape of the distribution under
study. By a clever choice of the leading density, one can improve the efficiency of
the estimation procedure in finite samples compared to the fully non-parametric
kernel-based approach.

Analogously to Gallant and Nychka (1987), we define the Kth order approx-
imation to the true density of the innovations as

$$(5) \qquad f_K(e) = c_K^{-1} \left( \sum_{i_1,\dots,i_k=0}^{K} \gamma_{i_1 \dots i_k} \prod_{h=1}^{k} e_h^{i_h} \right)^2 t_\nu(e),$$

where the &dimensional vector e denotes $R_K^\top(\varepsilon_t - m_K)$, with $R_K$ a Choleski
decomposition of the inverse of the scaling matrix. Furthermore, $t_\nu(\cdot)$ is the
Student $t$ distribution with zero mean, unit scaling[2] matrix, and $\nu$ degrees of
freedom. The constant $c_K$ and the vector $m_K$ are such that $f_K(R_K^\top(\varepsilon_t - m_K))$
integrates to one and $\varepsilon_t f_K(R_K^\top(\varepsilon_t - m_K))$ integrates to zero such that $\varepsilon_t$ has
zero mean.

To fix ideas, consider the following univariate ($k = 1$) example with a third
order ($K = 3$) SNP expansion:

$$(6) \qquad f_3(e) = c_3^{-1}\left(\gamma_0 + \gamma_1 e + \gamma_2 e^2 + \gamma_3 e^3\right)^2 t_\nu(e),$$

with e = $R_3(\varepsilon_t - m_3)$. (6) clearly shows that the density expansion is always
positive. In fact, the expansion amounts to multiplying the central density, which
in our case is $t_\nu(\cdot)$, by the square of a polynomial. These polynomial terms "mop
up" the differences between the true likelihood and the postulated central density.
By an appropriate choice of the central density, the number of additional terms
in the SNP expansion can be kept to a minimum.

Before proceeding, a few remarks are in order. First, the expansion in (5) is
different from that in Gallant and Nychka (1987) in that the Student $t$ distri-
bution is used as the leading term in the expansion, as opposed to the normal
distribution. Although this difference is irrelevant if we consider the asymptotics
of the procedure (T, $K \to$ co), in finite samples a low order Hermite-type ex-
pansion around the normal might prove insufficient to describe the degree of
leptokurtosis usually encountered in, e.g., financial data. By modeling the lep-
tokurtosis directly through the choice of the central density, we can substantially
limit the number of terms needed in the SNP expansion when the procedure is

---

[2]By this we mean that the covariance matrix corresponding to $t_\nu(\cdot)$ for $\nu > 2$ is equal to
$\nu I_k/(\nu \quad 2)$.

applied in finite samples. Note, however, that the use of the Student $t$ distribution also gives rise to certain additional complications. For example, the degrees of freedom parameter $\nu$ must now be linked to the number of terms in the density expansion in order for the SNP expansion in (5) to be integrable. These technical issues are dealt with below.

A second remark concerns the degree $K$ of the approximation. The elegance of semi-nonparametric maximum likelihood estimation as proposed by Gallant and Nychka (1987) consists of the fact that the procedure can be as efficient as maximum likelihood if we let the order $K$ of the SNP expansion rise with the number of observations $T$. In that case the approximation becomes more accurate if more data are available. Given the comments in the previous paragraph, the lower bound on the value of $\nu$ must then also rise with the number of observations.

The third remark concerns the need for the vector $m_K$ in (5). As already noted by Johansen (1994), the asymptotic distribution of cointegration tests is extremely sensitive to the deterministic components included in the regression model and the drift terms present in the data generating process. In order to guarantee a similar interpretation of these components over different orders of the SNP expansion, we impose the restriction that the mean of the innovations $\varepsilon_t$ is equal to zero. Note that this complicates the estimation, as the mean generally is a function of the remaining parameters used in the SNP expansion. More on this topic can be found in the remainder of this section.

The final remark before proceeding concerns the identifiability of (5). In its present form, the parameters in the approximation are not identified. This is easily seen by multiplying all y-parameters by a constant factor. As a result, the integrating constant $c_K$ has to be multiplied by the square of this constant factor, resulting in an identical density approximation. To overcome the identifiability problem, we impose the restriction $\gamma_{0\ldots0} = 1$, such that the leading term in the SNP expansion becomes the Student $t$ density with $\nu$ degrees of freedom, zero mean, and scaling matrix $(R_K^\top R_K)^{-1}$.

The parameters that have to be estimated for the Kth order SNP expansion consist of $\nu, \alpha, \beta, \Phi_1, \ldots, \Phi_{p-1}, \mu, R_K$, and all $\gamma$ parameters. The parameters $c_K$ and $m_K$ depend on the remaining parameters in the following way.

LEMMA 1.    Given *the density* approximation *in (5), and* given *that* $\nu >$ $2 \cdot k \cdot K,$ *then*

$$(7) \qquad \int_{-\infty}^{\infty} c_K \cdot f_K(e)de =$$

$$\sum_{i_1,\ldots,i_k,j_1,\ldots,j_k=0}^{K} \gamma_{i_1\ldots i_k} \cdot \gamma_{j_1\ldots j_k} \cdot \left( \prod_{h=1}^{k} C_0(i_h + j_h) \right) \cdot C_1 \left( \nu, \sum_{h=1}^{k} (i_h + j_h) \right),$$

*with*

$$C_0(i) \;=\; \frac{2^{i/2}}{\sqrt{\pi}} \Gamma\!\left( \frac{i+1}{2} \right) \frac{(1 + (-1)^i)}{2}$$

*and*

$$C_1(\nu, i) = \left(\frac{\nu}{2}\right)^{i/2} \frac{\Gamma((\nu - i)/2)}{\Gamma(\nu/2)}.$$

*Furthermore, for* $\nu > 2 \cdot k \cdot K + 1$

$$(8) \qquad \int_{-\infty}^{\infty} c_K \cdot e_n \cdot f_K(e) de = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_k=0}^{K} \gamma_{i_1\ldots i_k} \cdot \gamma_{j_1\ldots j_k} \cdot$$

$$\left(\prod_{h=1}^{k} C_0(i_h + j_h + 1_{\{h=n\}})\right) \cdot C_1\left(\nu, \sum_{h=1}^{k}(i_h + j_h) + 1\right),$$

*with* $e_n$ *the nth element of the k-dimensional vector* e, *and* $1_A$ *the* indicator function of *the set A.*

COROLLARY 1. *Let* $\xi_1$ *denote the right-hand side of (7), and let* $\xi_2$ *denote* a k-dimensional *vector with the nth element being equal to the right-hand side of* (8). *Then given the* restrictions

$$\int_{-\infty}^{\infty} f_K(R_K^{\top}(\varepsilon_t - m_K)) d\varepsilon_t = 1$$

and

$$\int_{-\infty}^{\infty} \varepsilon_t f_K(R_K^{\top}(\varepsilon_t - m_K)) d\varepsilon_t = 0,$$

*it follows that* $c_K = \xi_1/|R_K|$ *and* $m_K = -(R_K^{\top})^{-1}\xi_2/\xi_1$.

So for a fixed value of $K$, we can view the SNP expansion as a simple pseudo-likelihood with additional nuisance parameters to be estimated. To be somewhat more precise, the function $\rho(\cdot)$ in (3) now also depends on the parameters $\nu$ and $\gamma$.... The asymptotic distribution theory of the LM cointegration test for $K$ fixed and $T \to \infty$ is not affected by the presence of additional nuisance parameters in the estimation stage. Evidently, however, the asymptotic distribution does depend on the order $K$ of the SNP expansion. More on this issue can be found in Section 4. For now it suffices to note that the Kth order SNP expansion can be used in a straightforward way to obtain estimates of the parameters of interest and to test the cointegration hypothesis in a pseudo-maximum-likelihood framework.

To conclude this section, we briefly comment on the form of the SNP expansion as provided in (5). In our view, the expansion used in (5) is not the most natural one to use in applications. In fact, the polynomial term in parentheses is a restricted $(k . K)$th order polynomial. This is easily seen by considering the case $k = 2$ and K = 1, in which case the polynomial can be written as

$$(9) \qquad \gamma_{00} + \gamma_{10}e_1 + \gamma_{01}e_2 + \gamma_{11}e_1e_2,$$

with $\gamma_{00} = 1$. (9) is a second order polynomial, with zero coefficients for the terms $e_1^2$ and $e_2^2$. This is not important if we consider the asymptotics $K, T \to co$. **In** applications with finite samples, however, it seems more natural to let the Kth order SNP expansion be based on an unrestricted Kth order polynomial instead of a restricted $(k \cdot K)$th order polynomial. To achieve this, we impose the condition that

$$(10) \qquad\qquad \gamma_{i_1 \ldots i_k} = 0 \quad \forall \sum_{h=1}^{k} i_h > K.$$

In (9), this would amount, to setting $\gamma_{11} = 0$. If $T$, $K \to \infty$, it does not matter asymptotically whether (10) is imposed or not. In finite samples, however, restricting the Kth order SNP expansion to a Kth order polynomial may matter a great deal.

**4. Asymptotic  distribution  theory.**  In this section we discuss the asymptotic properties of the semi-nonparametric cointegration test as discussed in the previous section. The most appropriate limit theory would be one in which the order of the expansion K and the sample size $T$ diverge to infinity simultaneously. This, however, proved outside the scope of the present paper. Therefore, we provide a different limit theory. First, we discuss the limiting result if $K$ is kept fixed, while the sample size $T$ tends to infinity. These results are heavily based on the work of Lucas (1996a,b, 1997) for fixed pseudo-likelihoods. Next, we discuss the effect on the limiting distribution if we let $K$ tend to infinity.

We first introduce some assumptions for the error process in (1).

ASSUMPTION  1.  The innovations $\varepsilon_t$ are independently and identically distributed (i.i.d.) with zero mean and finite variance-covariance matrix $\Omega_{11}$.

The results in this section can probably be generalized to sequences of martingale differences, but this is not pursued in the present paper. Note that Assumption 1 formally excludes phenomena like volatility clustering, which are dominantly present in, e.g., financial data. It appears from Lucas (1996b), however, that techniques like the ones studies in the present paper are still useful for analyzing the cointegrating properties of multivariate systems exhibiting ARCH-type behavior.

We now treat the order K of the SNP expansion as fixed and consider the limiting distribution as the sample size $T$ tends to infinity. Recall that $\theta^\top = (\theta_1^\top, \theta_2^\top, \theta_3^\top)$, where $\theta_1 = \text{vec}(\alpha_{22})^\top$, $\theta_2$ contains the remaining regression parameters, and $\theta_3$ contains $\nu$, the $\gamma$ parameters, and the non-redundant parameters in $R_K$. Let $\theta_a^\top = \theta_1^\top, \theta_2^\top)$, the full vector of regression parameters. Define

$$\rho_K(\varepsilon_t(\theta_a), \theta_3) = -\log f_K(R_K^\top(\varepsilon_t(\theta_a) - m_K)),$$

with $f_K(\cdot)$ as in (5) with constants $c_K$ and $m_K$ as in Corollary 1, so that the log-likelihood is given by $\ell_K(\theta) = - \sum_{t=1}^{T} \rho_K(\varepsilon_t(\theta_a), \theta_3)$. **Define** $\psi_{K,t}(\theta) = \psi_K(\varepsilon_t(\theta_a), \theta_3) = \partial \rho_K(\varepsilon_t(\theta_a), \theta_3)/\partial \varepsilon_t$ and $\chi_{K,t}(\theta) = \qquad \chi_K(\varepsilon_t(\theta_a), \theta_3) =$

$-\partial \rho_K(\varepsilon_t(\theta_a), \theta_3)/\partial \theta_3$ . We need the following conditions on $\rho_K$ and the true distribution of $\varepsilon_t$, compare Lucas (1996b, 1997).

ASSUMPTION  2.   There is a pseudo-true value $\theta_K$ satisfying $S_1 \theta_K = 0$, such that

1. the vector $(\varepsilon_t(\theta_K)^\top, \psi_{K,t}(\theta_K)^\top, \chi_{K,t}(\theta_K)^\top)^\top$ has mean zero and finite positive semi-definite covariance matrix $\Omega_K = (\Omega_{ij,K}), i, j, = 1, 2, 3$, and $\Omega_{ii,K}$ positive definite for $i = 1, 2$;

2. the following expectations exist:

$$C_{1,K} = E\left(\frac{\partial \psi_{K,t}(\theta_K)}{\partial \varepsilon_t^\top}\right), \qquad C_{2,K} = E\left(\frac{\partial \chi_{K,t}(\theta_K)}{\partial \varepsilon_t^\top}\right),$$

$$C_{3,K} = E\left(\frac{\partial \chi_{K,t}(\theta_K)}{\partial \theta_3^\top}\right),$$

with $|C_{1,K}| \neq 0$ and $|C_{3,K}| \neq 0$.

We introduce the following notation. If $B(s)$ denotes a Brownian motion, then $\int B$ is a short-hand notation for $\int_0^1 B(s)ds$. Similarly, $\int_0^1 B(s)dB(s)$ is denoted by $\int BdB$. We can now prove the following theorem.

THEOREM  1.    *If the* data is generated *by (1) with $\varepsilon_t$ satisfying Assumption 1, and the pseudo-likelihood $\rho_K(\cdot)$ is used for estimation, with $\rho_K(\cdot)$ satisfying Assumption 2, then the LM cointegration test given in* **(4)**  *weakly converges to*

$$(11) \qquad \text{trace}\left(\left[\int \bar{F}dB_2^\top\right]^\top \left[\int \bar{F}\bar{F}^\top\right]^{-1} \left[\int \bar{F}dB_2^\top\right]\right),$$

*where $\bar{F} = F \quad \int F$, $F = B_1$ **i f** $\alpha_\perp^\top \mu = 0$, $B_1$ and $B_2$ **are**  two correlated standard Brownian motions with $E(F(s)B_2(s)^\top) = P_K$ with $P_K$ a diagonal* matrix *containing the canonical correlations between $\alpha_\perp^\top \varepsilon_t$ and $\alpha_\perp^\top C_{1,K}^{-1}\psi_K(\varepsilon_t)$, with $\alpha_\perp$ denoting the orthogonal complement **of** $\alpha$. If $\alpha_\perp^\top \mu \neq 0$, the first element of F(s) is replaced by $s$. Moreover, in that case the* matrix $P_K$ *contains a* **zero** *in the* upper-left *element, while the* remaining *diagonal elements contain the canonical correlations between $(\alpha_\perp^\top \mu)_\perp \alpha_\perp^\top \varepsilon_t$ and $\alpha_\perp^\top C_{1,K}^{-1}\psi_K$ (Et).*

The result in Theorem 1 is identical to the result in Lucas (1996b). Limiting results for more general deterministic functions of time like the ones in Johansen (1994) can be found in Lucas (1996a).

The limiting distribution of the non-Gaussian LM tests depends both on the postulated pseudo-likelihood and on the true likelihood through the correlations in the matrix $P_K$. This makes it difficult to perform inference in practical settings, as new critical values have to be computed for each new choice of the pseudo-likelihood and for each underlying true likelihood. Lucas (199613) proposes an

inference procedure that can be implemented in practice. Using the parameter estimates under the null, one can obtain a consistent estimate of the matrix $P_K$. This estimate of $P_K$ can be used to simulate from the asymptotic distribution in (11), where the (correlated) Brownian motions are replaced by correlated random walks of length $T$. The simulations can be used to compute an asymptotic $p$-value of the LM test. As simulating from the (approximation to the) asymptotic distribution does not take much time, this way of performing inference seems well-suited for practical applications of non-Gaussian cointegration tests. The results in Lucas (1996b) and in Section 5 illustrate that the procedure produces good results in samples as small as $T = 100$ with only 500 drawings from the asymptotic distribution used for estimating the pvalue of the LM test.

The result in Theorem 1 is not entirely new, albeit that in the proof of the theorem we have to deal explicitly with the additional nuisance parameters due to the SNP expansion used as a pseudo-likelihood. It is more interesting, however, to discuss the effect on the limiting distribution if $K \to \infty$. If $K$ tends to infinity, the SNP expansion approaches (in Sobolev norm, see Gallant and Nychka, 1987) the true likelihood, such that the function $\psi_K(\cdot)$ approaches the true score function. Let $\psi_\infty(\cdot)$ denote the true score function, then it holds automatically that $E(\psi_\infty(\varepsilon_t)) = 0$. Moreover, the covariance between $\varepsilon_t$ and $\psi_K(\varepsilon_t)$ tends to the covariance between $\varepsilon_t$ and $\psi_\infty(\varepsilon_t)$, i.e., $-I_k$. By a similar argument, the information matrix equality holds for $K \to \infty$. In particular, let $\Omega_{22,K} = E(\psi_K(\varepsilon_t)\psi_K(\varepsilon_t)^\top)$, then $\Omega_{22,\infty} = C_{1,\infty}$. As a result, the correlations for the case $\alpha_\perp^\top \mu = 0$ mentioned in Theorem 1 reduce to the singular values of the matrix $(\alpha_\perp^\top \Omega_{11} \alpha_\perp)^{-1/2}(\alpha_\perp^\top \Omega_{22,\infty}^{-1} \alpha_\perp)^{1/2}$. All these results follow from the fact that we estimate the true density of the innovations $\varepsilon_t$ adaptively. Another advantage of adapting for the true innovation density is that the pseudo LM test automatically approaches a true LM test. The true LM test is approximately optimal from an (asymptotic) expected mean-squared error perspective if the innovations are sufficiently fat-tailed or if sufficiently distant (local) alternatives to the null hypothesis are considered, see Lucas (1996a, Section 7.4). Finally, note that a simplification of the LM test statistic can be carried through as for $K \to \infty$ it holds that $H_C = H_\Omega$ in (4).

**5. Simulation experiment.** In this section we consider a simple simulation experiment to uncover the main characteristics of the SNP cointegration LM test in finite samples. We compare the new test with the performance of Johansen's (1991) trace test and Lucas' (1996b) LM test based on a Student $t$ pseudo-likelihood with five degrees of freedom. We consider the following data generating process (DGP),

$$(12) \qquad \begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -c/T \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \varepsilon_t,$$

with $T$ denoting the sample size and c denoting a constant. A DGP as in (12) was also used in Lucas (1996b, 1997). Note that (12) does not contain a drift term, such that the limit distribution with $\alpha_\perp^\top \mu = 0$ of Theorem 1 applies. Simulations

with a drift term were also performed, but resulted in identical conclusions. Despite the model's simplicity, it suits the present purposes of demonstrating the main differences between the cointegration testing procedures. Using more complicated DGP's involving, e.g., contemporaneously correlated errors or non-trivial cointegrating vectors, does not alter the general conclusions of this paper, see also the simulations in Lucas (1996b). For c $\neq$ 0, the DGP in (12) contains one cointegrating relation, namely $(0,1)^\top$. We test the null hypothesis of no cointegration, $H_0 : r = 0$, versus the unrestricted alternative, $H_1 : r = 2.$

The regression model used to estimate the parameters is given in (1). Note that the data generating process satisfies $\alpha_\perp^\top \mu = 0$, such that there are two nuisance parameters entering the limiting distribution of the cointegration test. These parameters are estimated consistently using the methods described in Section 4. Next, p-values are simulated using 500 simulations. If the simulated pvalue is below 0.05, we conclude that the null hypothesis of no cointegration is rejected.

The sample size used in the simulations is $T = 100$. Especially if for example financial data are used, one expects to have a much larger number of observations. By using $T = 100$, we can check whether the SNP cointegration test also works well in small samples. In particular, we are interested in whether the SNP approach can already outperform Johansen's (1991) LR test in samples of this size, or whether we have to pay for the increased flexibility of the SNP approach by a substantial power loss.

We consider two values of c in (12), namely c = 0 and c = 20. If c = 0, we should obtain the null distribution. The rejection frequency in this case should, therefore, be approximately equal to the nominal size of the test: 5%. The power of the test is investigated by looking at the rejection frequencies for c = 20.

We consider eight different test statistics. Test one is Johansen's (1991) LR trace test for cointegration. Test two is an LM version of Johansen's trace test. It uses the LM test discussed in Section 2 based on a 0th order SNP expansion with $\nu$ fixed at infinity. Test three is the LM test of Lucas (1996b) based on the fixed Student $t$ pseudo-likelihood with five degrees of freedom. This corresponds to a 0th order SNP expansion with $\nu$ fixed at five. Tests four through seven use the LM test based on SNP expansions of orders $K = 0$ through $K = 3$, respectively, as given in (5). Finally, test eight uses the Akaike Information Criterion (AIC) to select the order of the SNP expansion. For $K = 0$ through $K = 3,$ the AIC is computed. The value of $K$ with the highest AIC is used in the computation of the LM cointegration test. Note that the highest order of the SNP expansion used in the present setup is three. Given the comments at the end of Section 3, this expansion results in five $\gamma$ parameters to be estimated. Given the limited number of observations, it seems difficult to go beyond this order of the SNP expansion without running into severe convergence problems during the estimation stage. As explained in Section 3, there is a link between the order of the SNP expansion and the value of $\nu$. If $\nu \leq 2K$, then (5) no longer represents an integrable density function. In the simulation experiments, we impose $\nu \geq 3 + 2K$, such that the first two moments of the density given in (5) always exist.

We consider 9 different Monte-Carlo experiments. For each experiment, we

use 100 Monte Carlo replications. The experiments use different distributions for the i.i.d. sequence $\{\varepsilon_t\}$, namely

1. $\varepsilon_t$ is drawn from the standard normal

2. $\varepsilon_t$ is drawn from the Student $t$ distribution with 3 degrees of freedom;

3. $\varepsilon_t$ is drawn from the Student $t$ distribution with 2 degrees of freedom;

4. $\varepsilon_t$ is drawn from the Cauchy distribution, which is truncated such that 95% of the original probability mass is retained;

5. $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are independent drawings from a (mirrored) $\chi^2$ distribution with 3 degrees of freedom, recentered to have zero mean;

6. $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are independent drawings from a (mirrored) $F$ distribution with (6,6) degrees of freedom, again recentered to have zero mean;

7. $\varepsilon_t$ is drawn from a mixture of three normals; only the means of the mixture components differ; the first component has mean $(0, -1\frac{1}{2})^\top$ and is drawn with probability 0.5; the second component has mean $(1\frac{1}{2}, 1\frac{1}{6})^\top$ and is drawn with probability 0.3; the third component has mean $(-2\frac{1}{4}, 2)^\top$ and is drawn with probability 0.2;

8. $\varepsilon_t$ is drawn from a mixture of three normals, each drawn with equal probability and each having a unit covariance matrix; the means of the normals are $(3, \sqrt{3}), (-3, \sqrt{3})$, and $(0, -2\sqrt{3})$;

9. $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are independent; $\varepsilon_{2t}$ is standard normally distributed, while the density of $\varepsilon_{1t}$ is given by $(450\pi)^{-1/2}x^6 \exp(-x^2/2)$.

The first experiment provides a benchmark. One of the aims of the SNP cointegration test is have a good level and power performance for both thin-tailed and fat-tailed innovations. For Gaussian $\varepsilon_t$, therefore, we hope that the SNP test does not much worse than the test based on the Gaussian pseudo-likelihood. The next three experiments are used to investigate the performance of the tests under increasing degrees of leptokurtosis. Note that the $t(2)$ distribution does not satisfy Assumption 1. It is included to study the effect on the tests of extreme forms of leptokurtosis. The $\chi^2$ and the $F$ distribution are used to investigate the effect of skewness and the combination of skewness and leptokurtosis. The experiment with the mixtures of normals shows the performance of the tests if the underlying distribution is heavily skewed and multi-modal. The main difference between the two mixtures is the distance between the means of the mixture components. Especially if the distance between the mixture components becomes large, the performance of cointegration tests based on a fixed non-Gaussian pseudo likelihood can deteriorate substantially, see the univariate simulations in Shin and So (1997). The final experiment shows the performance of the tests if the true distribution is an SNP expansion of order $K = 3$. Some plots of the relevant distributions are given in Figure 1.
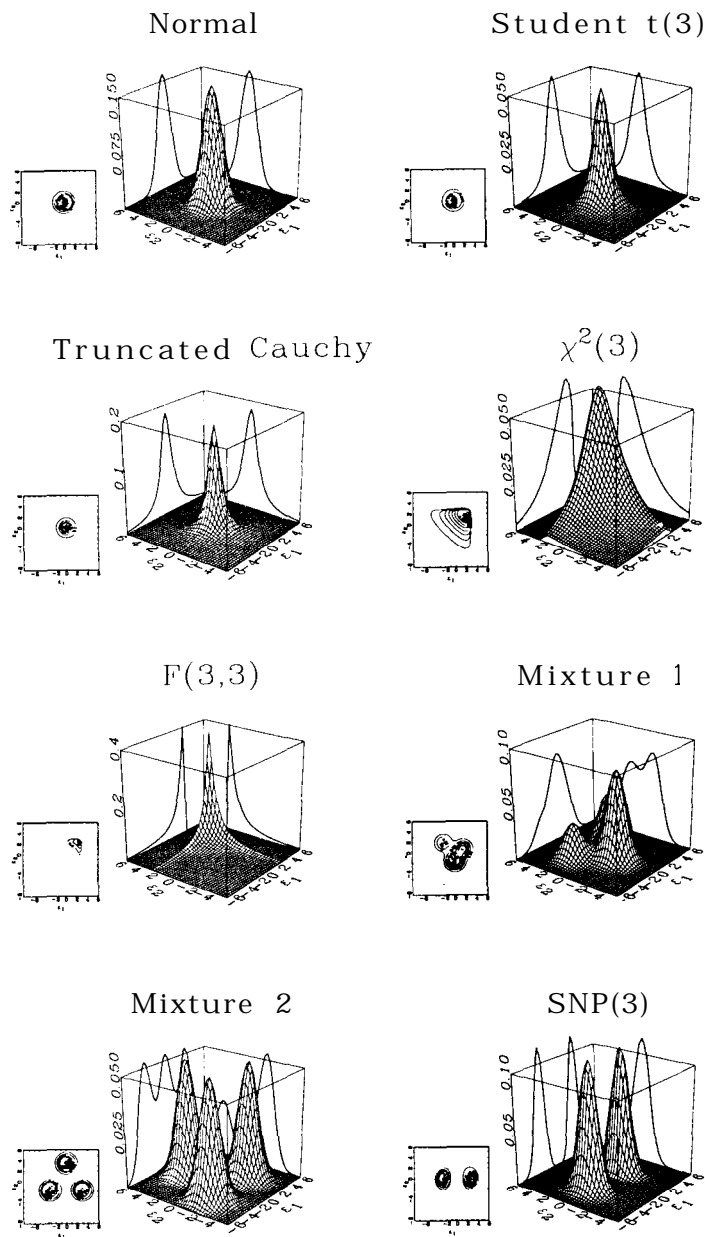
Normal

Student t(3)

Truncated Cauchy

$\chi^2(3)$

F(3,3)

Mixture 1

Mixture 2

SNP(3)

FIG. *1.   Some of the densities used in the Monte-Carlo experiments.*

The rejection frequencies of the different tests in the Monte-Carlo experiments can be found in Table 1.

The first, thing to note in Table 1 is that the rejection frequencies under the null (c = 0) reflect the nominal level of 5% quite accurately for all test statistics and distributions given the limited number of Monte-Carlo replications. Further note that the rejection frequencies of the Gaussian-based tests (Joh and G) under the alternative (c = 20) are fairly stable, with only the $F$ distribution giving rise to a lower power given the average performance. When we compare the rejection frequencies of the LR-type test (Joh) and the LM-type test (G) under the alternative (c = 20), we find that in terms of finite sample power the LR-test outperforms the LM test in terms of power.

As noted in the earlier literature on non-Gaussian cointegration tests, the LM test based on the Student t(5) pseudo-likelihood outperforms the Gaussian-based test if innovations are fat-tailed. For thin-tailed innovations, however, we note that the Student t(5) based test does worse than the Gaussian test if innovations are heavily skewed and/or multi-modal (see Mix2 and SNP(3)). The notable exception to this statement is given by the $\chi^2(3)$ distribution, which is heavily skewed, but still results in an outperformance of the Student t(5) based test compared to the Gaussian tests. It is also worthwhile to note the relatively good performance of the Student t(5) test for the $F(3,3)$ distribution. The skewness of this distribution seems to have no major adverse effects on the power of the test.

Given the performance of the tests based on a fixed pseudo-likelihood, we now turn to the performance of the semi-nonparametric cointegration tests. Ideally, we would like the SNP-based tests to be at least as powerful as both the Gaussian-based tests and the Student t(5) test. We concentrate our discussion on the SNP test where the order of the SNP expansion is determined by the Akaike Incormation Criterion (S(A)). For normal innovations, S(A) has similar power behavior as the Gaussian LM test (G). The optimal order of the SNP expansion (K) is low on average. Note that for Gaussian innovations, the LR test works better than the LM test. Therefore, although S(A) performs about as well as the Gaussian LM test, it is still preferable in this context to use the LR test of Johansen (1991). If we consider the fat-tailed distributions t(3) through t(l), we see that S(A) has about the same power as the Student t(5) based LM test. The average order $K$ of the SNP expansion is generally low. Note that for t(2) and t(l), one always chooses $K = 0$ in the simulations, such that the S(A) test is in these cases based on a Student $t$ pseudo-likelihood with estimated degrees of freedom parameter. Both S(A) and the t(5)-based test outperform the Gaussian based tests (Joh and G) for the t(3) through t(1) distributions.

If we now turn to the skewed distributions, we note the improvement obtained for the $\chi^2(3)$ innovations when using S(A) instead of the Gaussian-based or t(5)-based LM tests. Although the Student *t(5)* pseudo-likelihood already outperforms the Gaussian based tests, the *t(5)* pseudo-likelihood is not flexible enough to capture the skewness. The pseudo-likelihood based on the SNP expansion, however, is able to capture some of the skewness, which results in a higher

TABLE 1
*Rejection  Frequencies  of  Cointegration  Tests*

| Distr. | $c$ | Joh | G | t(5) | S(0) | S(1) | S(2) | S(3) | S(A) | $K$ | $\sigma_K$ |
|--------|-----|-----|---|------|------|------|------|------|------|-----|------------|
| Normal | 0 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.03 | 0.04 | 0.03 | 0.56 | 0.94 |
|        | 20 | 0.55 | 0.46 | 0.31 | 0.45 | 0.47 | 0.40 | 0.35 | 0.45 | 0.42 | 0.77 |
| $t(3)$ | 0 | 0.08 | 0.07 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.04 | 0.16 | 0.66 |
|        | 20 | 0.42 | 0.35 | 0.75 | 0.72 | 0.76 | 0.76 | 0.67 | 0.69 | 0.24 | 0.82 |
| $t(2)$ | 0 | 0.05 | 0.03 | 0.06 | 0.06 | 0.06 | 0.05 | 0.04 | 0.06 | 0.00 | 0.00 |
|        | 20 | 0.51 | 0.47 | 0.94 | 0.93 | 0.93 | 0.92 | 0.88 | 0.93 | 0.00 | 0.00 |
| $t(1)$ | 0 | 0.06 | 0.06 | 0.03 | 0.03 | 0.01 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 |
|        | 20 | 0.46 | 0.40 | 0.96 | 0.98 | 0.97 | 0.94 | 0.93 | 0.98 | 0.00 | 0.00 |
| $\chi^2(3)$ | 0 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 1.54 | 0.89 |
|        | 20 | 0.42 | 0.38 | 0.56 | 0.55 | 0.71 | 0.73 | 0.64 | 0.70 | 1.46 | 0.88 |
| $F(3,3)$ | 0 | 0.05 | 0.05 | 0.04 | 0.01 | 0.01 | 0.02 | 0.04 | 0.01 | 0.99 | 0.10 |
|        | 20 | 0.26 | 0.21 | 0.97 | 0.96 | 0.96 | 0.95 | 0.93 | 0.96 | 0.24 | 0.55 |
| Mix1 | 0 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.06 | 0.06 | 0.06 | 2.93 | 0.29 |
|      | 20 | 0.44 | 0.39 | 0.41 | 0.39 | 0.42 | 0.62 | 0.82 | 0.81 | 2.87 | 0.46 |
| Mix2 | 0 | 0.04 | 0.03 | 0.03 | 0.02 | 0.04 | 0.06 | 0.07 | 0.07 | 2.81 | 0.46 |
|      | 20 | 0.52 | 0.43 | 0.17 | 0.45 | 0.51 | 0.41 | 0.94 | 0.94 | 2.85 | 0.39 |
| SNP(3) | 0 | 0.05 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.04 | 0.03 | 1.12 | 1.30 |
|        | 20 | 0.44 | 0.39 | 0.32 | 0.35 | 0.38 | 0.47 | 0.47 | 0.46 | 1.57 | 1.22 |

Note: the table contains the rejection frequencies over 100 Monte-Carlo replications of 8 cointegration tests. These tests use different pseudo-likelihoods to estimate the parameters in model (1) with p = 1. Joh is the Johansen (1991) trace statistic (critical taken from Johansen and Juselius (1990)). S(0) through S(3) use the LM test based on SNP expansions of orders 0 through 3, respectively, as a pseudo-likelihood. G and t(5) use the Gaussian and the Student t(5) pseudo-likelihood, respectively. S(A) uses the AIC to select the optimal order of the SNP expansion. The mean value of $K$ selected with the AIC and its standard deviation $\sigma_K$ are also provided. Inference is conducted using simulated asymptotic p-values based on 500 simulations as described in Section 4. Eight different distributions are considered for the innovations: the standard normal, Student *t(3),* Student t(2), truncated Cauchy (or Student t(l)) retaining 95% of the original probability mass, recentered $\chi^2(3)$ (zero mean), recentered $F(3,3)$ (zero mean), two mixture of 3 normals (Mix1 and Mix2, both described in the main text), and a simple density falling in the SNP(3) class. The nominal level of the tests is 5% in all cases. c = 0 gives the rejection frequency under the null, while $c = 20$ gives the rejection frequency under the alternative. The data generating process in all cases is (12). The number of observations is $T$ = 100. The table took about 72 hours of computation time on a Pentium PC using the Gauss programming language.

rejection frequency of the test under the alternative (c = 20). If the innovations are both fat-tailed and skewed as is the case for the $F(3,3)$ distribution, S(A) and the Student t(5) based test have again the same performance. It seems that fat-tails have a more prominent effect on the performance of the cointegration tests than skewness. This also appears from the average order of the SNP expansion, which is lower for the $F(3,3)$ distribution than for the $\chi^2(3)$ distribution.

The mixtures of normals (Mix1 and Mix2) deviate significantly from the familiar bell-shaped density curves. The power of the Gaussian based tests for these distributions is similar to the power for the Gaussian distribution. This is in accordance with asymptotic theory, which states that the distribution of the Gaussian-based tests does not depend on the form of the distribution, as long as second moments exist. The power behavior of the test based on the Student t(5) pseudo-likelihood for the first mixture is about the same as for the Gaussian distribution. For the second mixture, however, the Student t(5) test suffers from a dramatic power loss. This is in accordance with univariate results obtained by Shin and So (1997). The SNP expansion is able to capture most of the skewness of the first mixture distribution, and some of the peculiar shape of the second mixture distribution. Note the relatively high average orders of the SNP expansion $(K)$. The power increase of S(A) for Mix1 and Mix2 is quite dramatic compared to the power of Johansen test, the Gaussian LM test, or the Student t(5) test. Note that the power of S(A) for Mix2 is mainly due to the consideration of the $K = 3$ expansion. Lower order expansions do not result in substantial power gains. This indicates that one should consider high enough orders for the SNP expansion if severe deviations from normality are suspected. This corroborates the unreported results for a mixture of four normals located at $(\pm 3, \pm 3)$, for which no significant power gains were found if only SNP orders below 4 were considered.

The results for the SNP(3) distribution are as expected. The S(A) test and the S(2) and S(3) tests, i.e., the tests with high-order SNP expansions, have the highest power of the LM-type tests. which is what we would expect. The LR test of Johansen, however, has about the same performance as the LM tests in this context. Note that although the true distribution is a third order SNP expansion, lower order expansions $(K)$ are often chosen by the AIC. This does not result in a substantial power loss.

To summarize, we obtain the following conclusions from the simulation experiment .

1. If innovations are fat-tailed, the Student t(5) pseudo-likelihood often results in a good power performance for the cointegration tests.

2. If innovations are thin-tailed and heavily skewed, the cointegration tests based on SNP expansions outperform the tests based on a Gaussian or Student t(5) pseudo-likelihood.

3. For thin-tailed, multi-modal distributions, the Student t(5) based test can

perform very poorly. The SNP LM test clearly outperforms the other testing principles considered in this context.

**6. Concluding remarks.** In this paper we have developed a semi-nonparametric LM-type test for cointegration. The asymptotic distribution of this test depends on nuisance parameters, which can be estimated consistenly. Using the consistent estimates, one can easily simulate asymptotic pvalues of the cointegration test. Using a set of Monte-Carlo experiments, we demonstrated that the new test has good size and power properties. If innovations are fat-tailed, the test based on the SNP approach leads to considerable power gains with respect to tests based on the Gaussian distribution. Also if there is apparent skewness in the distribution, the SNP approach leads to power gains with respect to tests based on a Gaussian pseudo-likelihood. In such cases, the SNP approach also improves upon the power of non-Gaussian LM cointegration tests like the Student t(5) based test of Lucas (1996b, 1997).

Although the model used in the present paper only allowed for a constant in the regression model and (at most) linear trending in the data generating process, we conjecture that our findings also hold for more general deterministing trend functions in either the regression model and/or the data generating process. The limit theory of the SNP cointegration test for these cases can for familiar cases as the ones in Johansen (1994) be deduced from Lucas (1996a). Some further simulation experiments could be used to corroborate the above conjecture.

Another line of future research concerns the relaxation of the i.i.d. assumption used in the present paper. Most financial time series are characterized not only by fat-tailedness, but also by volatility clustering. The SNP approach can easily be extended to handle general forms of volatility clustering in a semi-nonparametric fashion. In a subsequent paper, we plan to extend the present cointegration testing procedures to situations with conditional heteroskedasticity.

**Appendix**

PROOF OF LEMMA 1. By staightforwardly applying the substitution $z = y^2/2$ and the definition of the I-function, we obtain

**(A1)** $$C_0(i) = \int_{-\infty}^{\infty} \frac{y^i}{\sqrt{2\pi}} \exp(-y^2/2)dy = \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) \frac{(1+(-1)^i)}{2}.$$

Moreover, let, y denote a univariate Student $t$ variate with location zero, unit scale, and $\nu$ degrees of freedom. Then from Abramowitz and Stegun (1970, Section 26.7.2) we obtain $E(y^{2i+1}) = 0$ for $\nu > 2i + 1$ and

**(A2)** $$E(y^{2i}) = \frac{\nu^i}{\sqrt{\pi}} \cdot \frac{\Gamma((2i+1)/2)\Gamma((\nu-2i)/2)}{\Gamma(\nu/2)},$$

for $\nu > 2i$. Define

**(A3)** $$C_1(\nu, i) = \left(\frac{\nu}{2}\right)^{i/2} \frac{\Gamma((\nu-i)/2)}{\Gamma(\nu/2)},$$

then $E(y^t) = C_0(i)\ C_1\ (\nu,\ i)$.

Let  $i_1, \ldots, i_k$ be a set of $k$ non-negative integers, and let the nth element of e be denoted by $e_n$. Moreover, let $t_\nu(e)$ denote the multivariate standard Student $t$ distribution. Then using the above results, we have

$$
(A4) \qquad \int_{-\infty}^{\infty} \left( \prod_{h=1}^{k} t_\infty(e) \right) de = \prod_{h=1}^{k} C_0(i_h)
$$

and

$$
(A5) \qquad \int_{-\infty}^{\infty} \left( \prod_{h=1}^{k} e_h^{i_h} \right) t_\nu(e) de = \left( \prod_{h=1}^{k} C_0(i_h) \right) \cdot C_1 \left( \nu, \sum_{h=1}^{k} i_h \right).
$$

Now consider the $K$th order SNP expansion given in (5). Set $c_K = 1$. Then

$$
\int_{-\infty}^{\infty} f_K(e) de = \int_{-\infty}^{\infty} \left( \sum_{i_1,\ldots,i_k=0}^{K} \gamma_{i_1\cdots i_k} \prod_{h=1}^{k} e_h^{i_h} \right)^2 t_\nu(e) de =
$$

$$
(A6) \qquad \int_{-\infty}^{\infty} \left( \sum_{i_1,\ldots,i_k,j_1,\ldots,j_k=0}^{K} \gamma_{i_1\cdots i_k} \gamma_{j_1\cdots j_k} \prod_{h=1}^{k} e_h^{i_h+j_h} \right) t_\nu(e) de.
$$

with $j_1, \ldots, j_k$ another set of non-negative integers. Using (A5) we can simplify (A6) to the following expression:

$$
(A7) \qquad \sum_{i_1,\ldots,i_k,j_1,\ldots,j_k=0}^{K} \gamma_{i_1\cdots i_k} \gamma_{j_1\cdots j_k} \left( \prod_{h=1}^{k} C_0(i_h + j_h) \right) \cdot C_1 \left( \nu, \sum_{h=1}^{k} (i_h + j_h) \right).
$$

The proof of $\int_{-\infty}^{\infty} e f_K(e) de$ runs completely analogously.

PROOF OF COROLLARY 1.   Define $f_{K,1}(\cdot)$ as in (5), only with $c_K = 1$. Using Lemma 1, we obtain

$$
1 = \int_{-\infty}^{\infty} f_K(R_K^{\top}(\varepsilon_t - m_K)) d\varepsilon_t = \int_{-\infty}^{\infty} (c_K |R_K|)^{-1} f_{K,1}(e) de = \xi_1/(c_K |R_K|),
$$

such that $c_K = \xi_1/|R_K|$. Similarly,

$$
0 = \int_{-\infty}^{\infty} \varepsilon_t f_K(R_K^{\top}(\varepsilon_t - m_K)) d\varepsilon_t = \int_{-\infty}^{\infty} (c_K |R_K|)^{-1}((R_K^{\top})^{-1}e + m_K) f_{K,1}(e) de \Leftrightarrow
$$

$$
0 = (c_K |R_K|)^{-1}((R_K^{\top})^{-1}\xi_2 + \xi_1 m_K) \Leftrightarrow m_K = -(R_K^{\top})^{-1}\xi_2/\xi_1.
$$

PROOF  OF  THEOREM  1. The  proof  mainly  follows  the  lines  of  Lucas (1996a,1997). As usual in this type of asymptotic analysis, we restrict attention to the VAR model of order 1, $p = 1$. Higher order VAR models do not affect the limiting result for the unit root parameters. Moreover, we only consider the

case $\alpha_\perp^\top \mu = 0$, or for simplicity $\mu = 0$. The case $\alpha_\perp^\top \mu \neq 0$ is proved similarly. Finally, we only discuss the case $H_0 : r = 0$ versus the alternative $r = k$. Nonzero cointegrating ranks can be dealt with by considering the asymptotic behavior of $y_t$ in the directions dictated by $\bar{\beta}$ and $\bar{\beta}_\perp$, respectively, and using the block-diagonality of the information matrix between the stationary and non-stationary regressors in the model.

Because in the simplified model, $Z_t = -\partial \varepsilon_t^\top(\theta_a)/\partial\theta_a = (y_{t-1}, 1)^\top \otimes I_k$, the score vector is given by

$$G_K(\theta) = \sum_{t=1}^T \begin{pmatrix} y_{t-1} \otimes \psi_{K,t}(\theta) \\ \psi_{K,t}(\theta) \\ \chi_{K,t}(\theta) \end{pmatrix}.$$

Furthermore, an (asymptotically valid) expression for minus the Hessian matrix is

$$H_{C,K}(\theta) = \sum_{t=1}^T \begin{pmatrix} y_{t-1}y_{t-1}^\top \otimes C_{1T}(\theta) & y_{t-1} \otimes C_{1T}(\theta) & y_{t-1} \otimes C_{2T}(\theta)^\top \\ y_{t-1}^\top \otimes C_{1T}(\theta) & C_{1T}(\theta) & C_{2T}(\theta)^\top \\ y_{t-1}^\top \otimes C_{2T}(\theta) & C_{2T}(\theta) & C_{3T}(\theta) \end{pmatrix},$$

with $C_{1T}(\theta) = T^{-1}\sum_{t=1}^T \partial\psi_{K,t}(\theta)/\partial\varepsilon_t^\top$, $C_{2T}(\theta) = T^{-1}\sum_{t=1}^T \partial\chi_{K,t}(\theta)/\partial\varepsilon_t^\top$, and $C_{3T}(\theta) = T^{-1}\sum_{t=1}^T \partial\chi_{K,t}(\theta)/\partial\theta_3^\top$. Finally, the outer-product-of-gradient matrix $H_{\Omega,K}$ is the same as $H_{C,K}$, but with $C_{1T}$, $C_{2T}$ and $C_{3T}$ replaced by $\Omega_{22T}(\theta) = T^{-1}\sum_{t=1}^T \psi_{K,t}(\theta)\psi_{K,t}(\theta)^\top$, $\Omega_{23T}(\theta) = T^{-1}\sum_{t=1}^T \psi_{K,t}(\theta)\chi_{K,t}(\theta)^\top$ and $\Omega_{33T}(\theta) = T^{-1}\sum_{t=1}^T \chi_{K,t}(\theta)\chi_{K,t}(\theta)^\top$, respectively.

Let $D_T = \text{diag}(T^{-1}I_{k^2}, T^{-1/2}I_k, T^{-1/2}I_m)$, with m the dimension of $\theta_3$. From the functional central limit theorem for i.i.d. sequences, together with the continuous mapping theorem, it follows that

$$D_T G_K(\theta_K) \Rightarrow \begin{pmatrix} \int W_1 \otimes dW_2 \\ \int dW_2 \\ \int dW_3 \end{pmatrix} = \bar{G}_K,$$

where $W = (W_1^\top, W_2^\top, W_3^\top)^\top$ is a vector Brownian motion process with covariance matrix $\Omega_K$. Furthermore, we have

$$D_T H_{C,K}(\theta_K)D_T \Rightarrow \bar{H}_{C,K}, \quad D_T H_{\Omega,K}(\theta_K)D_T \Rightarrow \bar{H}_{\Omega,K},$$

with

$$\bar{H}_{C,K} = \begin{pmatrix} \int W_1 W_1^\top \otimes C_{1,K} & \int W_1 \otimes C_{1,K} & \int W_1 \otimes C_{2,K}^\top \\ \int W_1^\top \otimes C_{1,K} & C_{1,K} & C_{2,K}^\top \\ \int W_1^\top \otimes C_{2,K} & C_{2,K} & C_{3,K} \end{pmatrix},$$

and

$$\bar{H}_{\Omega,K} = \begin{pmatrix} \int W_1 W_1^\top \otimes \Omega_{22,K} & \int W_1 \otimes \Omega_{22,K} & \int W_1 \otimes \Omega_{23,K}^\top \\ \int W_1^\top \otimes \Omega_{22,K} & \Omega_{22,K} & \Omega_{23,K}^\top \\ \int W_1^\top \otimes \Omega_{23,K} & \Omega_{23,K} & \Omega_{33,K} \end{pmatrix}.$$

Let $\tilde{\theta}$ denote the pseudo-maximum likelihood estimator under the restriction $\alpha_{22} = 0$. Note that the restricted model is simply the i.i.d. model $\Delta y_t = \mu + \varepsilon_t$, so that $\tilde{\theta}$ is e-consistent, which implies $D_T(H_{C,K}(\tilde{\theta}) - H_{C,K}(\theta_K))D_T = o_p(1)$ and $D_T(H_{\Omega,K}(\tilde{\theta}) - H_{\Omega,K}(\theta_K))D_T = o_p(1)$. Using $S_1 D_T = T^{-1} S_1$ and a Taylor series approximation $D_T G_K(\tilde{\theta}) = D_T G_K(\theta_K) + D_T H_{C,K}(\theta_K)(\tilde{\theta} - \theta_K) + o_p(1)$, we obtain

$$
\begin{aligned}
T^{-1} S_1 H_{C,K}^{-1}(\tilde{\theta}) G_K(\tilde{\theta}) &= S_1 (D_T H_{C,K}(\tilde{\theta}) D_T)^{-1} D_T G_K(\tilde{\theta}) \\
&= S_1 (D_T H_{C,K}(\theta_K) D_T)^{-1} D_T G_K(\theta_K) + \\
&\quad S_1 D_T^{-1}(\tilde{\theta} - \theta_K) + o_p(1) \\
&\Rightarrow S_1 \bar{H}_{C,K}^{-1} \bar{G}_K .
\end{aligned}
$$

Similarly, we have

$$
T^{-2} S_1 H_{C,K}^{-1}(\tilde{\theta}) H_{\Omega,K}(\tilde{\theta})(H_{C,K}^{-1}(\tilde{\theta}))^{\top} S_1^{\top} \Rightarrow S_1 \bar{H}_{C,K}^{-1} \bar{H}_{\Omega,K}(\bar{H}_{C,K}^{-1})^{\top} S_1^{\top},
$$

so that

$$
LM \Rightarrow (S_1 \bar{H}_{C,K}^{-1} \bar{G}_K)^{\top} (S_1 \bar{H}_{C,K}^{-1} \bar{H}_{\Omega,K}(\bar{H}_{C,K}^{-1})^{\top} S_1^{\top})^{-1} S_1 \bar{H}_{C,K}^{-1} \bar{G}_K .
$$

From the partitioned inverse of $\bar{H}_{C,K}$ we obtain, using its specific structure,

$$
S_1 \bar{H}_{C,K}^{-1} = \left[ \left( \int W_1 W_1^{\top} - \int W_1 \int W_1^{\top} \right) \otimes C_{1,K} \right]^{-1} \left( I_{k^2} : - \int W_1 \otimes I_k : 0 \right).
$$

Defining $\bar{W}_1 = W_1 - \int W_1$, this leads to

$$
\begin{aligned}
LM &\Rightarrow \left( \int \bar{W}_1 \otimes C_{1,K}^{-1} dW_2 \right)^{\top} \left[ \int \bar{W}_1 \bar{W}_1^{\top} \otimes C_{1,K}^{-1} \Omega_{22,K} C_{1,K}^{-1} \right]^{-1} \cdot \\
&\qquad\qquad\qquad\qquad \left( \int \bar{W}_1 \otimes C_{1,K}^{-1} dW_2 \right) \\
&= \left( \int \bar{B}_1 \otimes dB_2 \right)^{\top} \left[ \int \bar{B}_1 \bar{B}_1^{\top} \otimes I_k \right]^{-1} \left( \int \bar{B}_1 \otimes dB_2 \right) \\
&= \text{trace} \left( \left[ \int \bar{B}_1 dB_2^{\top} \right]^{\top} \left[ \int \bar{B}_1 \bar{B}_1^{\top} \right]^{-1} \left[ \int \bar{B}_1 dB_2^{\top} \right] \right),
\end{aligned}
$$

where $B_1 = \Omega_{11}^{-1/2} W_1$ and $B_2 = (C_{1,K}^{-1} \Omega_{22,K} C_{1,K}^{-1})^{-1/2} C_{1,K}^{-1} W_2$ are standard Brownian motion processes, with correlation matrix $P_K = E(B_1 B_2^{\top}) = \Omega_{11}^{-1/2} \Omega_{12,K} C_{1,K}^{-1} (C_{1,K}^{-1} \Omega_{22,K} C_{1,K}^{-1})^{-1/2}$. By an appropriate choice of the inverse square root matrices, the matrix $P_K$ is diagonalized, with the canonical correlations between $\varepsilon_t$ and $C_{1,K}^{-1} \psi_K(\varepsilon_t)$ on the diagonal.

Observe that, although the limiting Hessian and outer-product-of-gradient matrices are not block-diagonal, the nuisance parameters $\theta_3$ have no effect on the limiting distribution of the test statistic: the same result would be obtained

if $\theta_3$ were known, which is the case analyzed in Lucas (1996b). His proof for more general models (with lagged differences, a possible drift, and with cointegrating ranks greater than zero under the null hypothesis) can be readily extended to the current situation of unknown $\theta_3$

## REFERENCES

M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1970.

G. Elliot, T.J. Rothenberg, and J.H. Stock. Efficient tests for an autoregressive unit root. *Econometrica,* 64:813–836, 1996.

R.F. Engle and C.W.J. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica,* 55:251–276, 1987.

W.A. Fuller. *Introduction to Statistical Time Series.* Wiley, New York, 1976.

A.R. Gallant and D.W. Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica.* 55:363–390, 1987.

C. Gouriéroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: theory. *Econometrica,* 52:681–700, 1984.

D.J. Hodgson. Adaptive estimation of error correction models. Econometric *Theory,* 1997 (forthcoming).

S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control,* 12:231–254, 1988.

S. Johansen. Estimation and hypothesis testing of cointegration vectors in **gaussian** vector autoregressive models. *Econometrica,* 59:1551–1580, 1991.

S. Johansen. The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews,* 13:205–229, 1994.

S. Johansen and K. Juselius. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics,* 52:169–210, 1990.

F. Kleibergen and H.K. Van Dijk. Direct cointegration testing in error correction models. *Journal of Econometrics,* 63:61–103, 1994.

A. Lucas. *Outlier Robust Unit Root Analysis.* Thesis Publishers, Amsterdam, 1996a. http://www.econ.vu.nl/vakgroep/bfs/alm/lucas/thesis/Welcome.html.

A. Lucas. Inference on cointegrating ranks using pseudo LM tests. Technical Report 42, Vrije Universiteit, Fac. Economics, Amsterdam, 1996b. Revision re-submitted.

A. Lucas. Cointegration testing using pseudo likelihood ratio tests. *Econometric Theory,* 13:149–169, 1997.

D.W. Shin and B.S. So. Unit root tests based on adaptive maximum likelihood estimation. Working Paper in Statistics, Ewha Womans University, Seoul, 1997.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica,* 50:1–25, 1982.

DEPT. OF ECONOMETRICS
UNIVERSITY OF AMSTERDAM
ROETERSSTRAAT 11
1018WB AMSTERDAM
THE NETHERLANDS
EMAIL: PETERB@FEE.UVA.NL

FINANCIAL SECTOR MANAGEMENT
ECO/BFS, VRIJE UNIVERSITEIT
DE BOELELAAN 1105
1081HV AMSTERDAM
THE NETHERLANDS
EMAIL: ALUCAS@ECON.VU.NL