Testing for stationarity in a cointegrated system

Robert M. Kunst*

University of Vienna

and

Institute for Advanced Studies, Vienna

Abstract

In systems of variables with a specified or already identified cointegrating rank, stationarity of component variates can be tested by a simple restriction test. The implied decision is often in conflict with the outcome of unit root tests on the same variables. Using Bayes testing and decision contours, this paper searches for a solution to such conflict situations in sample sizes of empirical relevance. The decision contour evaluations suggest that some priority should be given to the decision by the restriction test on self-cointegration. This decision should be re-considered if the univariate Dickey-Fuller statistic or its covariate-augmented version indicate explosive behavior in the observed time series.

Key words: Bayes test, unit roots, cointegration, decision contours.

JEL Code: C11, C12, C15, C32.

*Address: Department of Economics, BWZ, University of Vienna, Bruenner Strasse 72, A-1210 Wien (Vienna), Austria; telephone: +43-1-4277-37479; fax: +43-1-4277-37498; e-mail: robert.kunst@univie.ac.at

1 Introduction

The procedure designed by JOHANSEN (1988, 1995) for the estimation of the cointegrating rank and of the cointegrating vectors in vector autoregressive systems has enjoyed tremendous popularity in economic applications. Usually, the procedure is conducted in sequential steps. First, univariate unit-root tests classify the variables of interest according to their degree of integration. Variables integrated of order zero or one are kept while higher-order integrated variables are eliminated or differenced. Then, the cointegrating rank is determined. Last, linear restriction tests are applied that check whether pre-specified vectors of interest are contained in the cointegrating space.

The cointegrating space may contain unit vectors. Testing for a unit vector to be contained in the cointegrating space is just a particular case of the general restriction test suggested by JOHANSEN (1995). A related procedure in a conditional rather than multivariate framework has been analyzed by HORVATH AND WATSON (1995). Whenever a unit vector cointegrates, the corresponding component variable is stationary. Practitioners often report that the decision resulting from the restriction testing step of JOHANSEN's procedure disagrees with the decision of the preliminary unit-root tests. A specific variable may be classified as stationary according to the preliminary univariate unit-root test and as non-cointegrating according to the restriction test, and *vice versa*. The natural question is then how to combine these contradictory pieces of evidence to reach a statistically well-based classification of these problematic variables. For example, NEUSSER (1991) tests for unit roots in real interest rates from several economies. While univariate tests accept the null of a unit root, the multivariate unit-vector test cannot reject the null that real interest rates cointegrate. NEUSSER (1991) gives more weight to the latter evidence, as "a multivariate approach might turn out to be more informative ... It is then possible to condition on a broader set of variables". However, the advantage of the multivariate approach is not so obvious, as the loss of inference efficiency due to increasing the number of variables might give some advantage to the univariate test.

We note that many researchers tend to avoid including stationary variables in the JOHANSEN framework, although the procedure has been designed to incorporate cases with integration order zero as well. The statement that all individual components must be integrated of order one is erroneous, although it can be found in some sources, including software descriptions. Stationary components imply cointegrating unit vectors, in other words the stationary variable is cointegrating with itself. The hypothesis that a specific unit vector is contained in the cointegration space can be subjected to a restriction test *after* determination of the rank and estimation of the full system ('post-testing'). In contrast with most preliminary unit-root tests, stationarity of the component is the null hypothesis of these restriction tests and its first-order integration is the alternative. The distribution of the corresponding test statistic is *chi-square* and not any of the non-standard mixture distributions that are known from the unit-root testing literature (see also TANAKA, 1996).

This paper attempts to give guidelines on how to act in cases of conflict. Conflict arises from changes in the identified integration order between the pre-testing and the post-testing phase. A variable may be classified as stationary in the pre-testing stage but its unit vector is rejected as a cointegrating vector in post-testing. Conversely, a variable may be classified as first-order integrated in pre-testing but its unit vector is accepted as cointegrating in post-testing. Ideally, any combination of the two test statistics should yield a decision in favor of either stationarity or first-order integration.

The technique that we use for the joint evaluation of pairs of test statistics is relatively new. Drawing on HATANAKA (1996), it was used for the joint evaluation of univariate seasonal unit root tests by KUNST AND REUTTER (2002). The basic idea is to evaluate the probability of a hypothesis of interest, conditional on observed pairs of statistics. A higher probability for a certain hypothesis indicates a preference for this hypothesis. The technique is Bayesian in spirit, as the classical distinction between null and alternative hypothesis as well as the terminology of 'rejection' and 'failing to reject' is abandoned in favor of the concept of 'preference for hypotheses'. It also relies on the specification of prior

distributions. Details will be provided below.

In order to avoid distracting attention from the main focus, important side issues will be ignored in this paper, such as the possible appearance of seasonal unit roots, the complex restriction test for second-order integration within the framework of the multivariate JOHANSEN procedure, or the correct specification of the deterministic features of the system.

In related work, RAHBEK AND MOSCONI (1999) use a slightly different model frame and view cointegration as *conditional* on stationary variables that may develop cointegration among their cumulated sums. For the full system of conditioned integrated and conditioning stationary variables, a vector autoregressive representation does not exist. By contrast, we assume the existence of a multivariate autoregressive representation for the whole vector of variables, in line with the model that was analyzed by JOHANSEN (1995).

The outline of this paper is as follows. Section 2 describes three hypothesis tests that are more or less commonly used in discriminating stationary and integrated variables. Section 3 introduces the Bayes-testing method that is suggested for evaluating combinations of any two of these tests. The results of an application of this suggested method are presented and commented in Section 4. Section 5 concludes.

2 Testing for unit roots in cointegrated systems

2.1 A standard test for a unit root

Suppose the vector variable X_t has n components that are individually either I(0) or I(1), and obeys a vector autoregression (VAR) of order p, $\Phi(B)X_t = \mu + \varepsilon_t$, with ε_t an ideally Gaussian white noise and B denoting the lag operator. The system can be transformed into its error-correction representation

$$\Pi(B)\Delta X_t = \mu + \alpha\beta' X_{t-1} + \varepsilon_t \tag{1}$$

with $n \times r$–matrices α, β of full rank that are uniquely determined up to a non-singular matrix factor of dimension $r \times r$. r denotes the cointegrating rank, β

is a matrix with cointegrating column vectors, and α is the so-called loading matrix. $\Pi(z)$ is a polynomial of order $p-1$. In the following, it will be assumed that r has already been determined. H_r denotes the maintained hypothesis of model (1) with $\operatorname{rk}\alpha = \operatorname{rk}\beta = r$. Under H_r, the model is estimated efficiently using reduced-rank regression.

Following most applications of the JOHANSEN procedure, only a constant μ is allowed as the deterministic part of (1). While this specification is at odds with the common usage of trend regressors in univariate unit-root tests, which are included for test similarity at the expense of test power, it is in line with the usual interpretation of error correction. A linear combination of non-stationary variables that is trend-stationary does not correspond to this concept.

If X contains an I(0) component $X^{(j)}$, say, the unit n–vector e_j with 1 at its jth entry and 0 otherwise is contained in the column space of β. This implies that $r \geq n_0$ if n_0 denotes the number of stationary components. If $r = n_0$, there is no non-trivial cointegration in the system, as all cointegrating vectors are unit vectors or linear combinations thereof. If $r = n = n_0$, the whole system is stationary. Without restricting generality, assume that the variable in question is the first one $X^{(1)}$ such that the critical cointegrating vector is e_1. The hypothesis of interest is H_u: $\beta = (e_1, \varphi)$, where φ is an $n \times (r-1)$–matrix. For this problem, JOHANSEN (1995, p. 108) shows that, under H_u, β is estimated by a sequence of conditioning operations. The likelihood-ratio statistic $\xi_2 = T\{\ln(1-\lambda_{\max}) - \ln(1-\lambda_u)\}$ for testing H_u in H_r requires estimating both models. Here, $\lambda_{\max}$ denotes the largest eigenvalue of the unrestricted canonical correlation problem for the n–variables X_{t-1} and ΔX_t, conditional on an intercept and $p-1$ lags of ΔX_t (for all details, see JOHANSEN, 1995). λ_u denotes the largest eigenvalue for the model restricted by H_u. T is sample size. Under H_u, LR_{ur} is (asymptotically) distributed as $\chi^2(n-r)$.

This paper focuses on the case $r = 1$, where inference becomes particularly simple. Because e_1 is the only cointegrating vector and φ disappears, the system restricted by H_u reduces to $\Pi(B)\Delta X_t = \mu + \alpha X_{t-1}^{(1)} + \varepsilon_t$, a multivariate regression problem. The eigenvalue λ_u is the conditional multiple correlation of $X_{t-1}^{(1)}$ and

ΔX_t and can be obtained from first regressing both sides on a constant and on lagged differences. Then, the possibly non-stationary scalar residual from $X_{t-1}^{(1)}$ is regressed on the n–variate residual from ΔX_t. The R^2 of this regression is λ_u. The statistic LR_{ur} is distributed as $\chi^2(n-1)$ under H_u.

Under H_u, the component variable is stationary, as e_1 cointegrates. However, the alternative is not the usual general hypothesis of first-order integration. Rather, the presence of r cointegrating relationships or of $n-r$ unit roots *in the system* is maintained. Therefore, the test is not a valid unit-root test for general purposes, although it is a valid check on univariate unit roots conditional on an already specified cointegrating rank. Note that, *in the system*, the same number of unit roots is present under the null and under the alternative, which explains the validity of the standard distribution for ξ_2.

Some of these issues have been considered by Horvath and Watson (1995) who analyze the general case of testing for given cointegrating vectors, which includes unit vectors as a special case. In their test alternative, the cointegrating rank differs from the null, therefore they obtain non-standard distributions. The use of multivariate VAR analysis for assessing the stationarity of individual components is mentioned by Johansen and Juselius (1992) who assume that the cointegrating rank has been pre-tested and is therefore fixed. Most applications proceed by first identifying the rank and then testing for special vectors, hence the latter approach is in focus here.

2.2 Traditional non-standard tests for unit roots

For a scalar variable x_t, the popular DF test for unit roots (Dickey and Fuller, 1979) is based on the t–statistic on β in the regression

$$\Delta x_t = a + bt + \beta x_{t-1} + \sum_{j=1}^{p-1} \pi_j \Delta x_{t-j} + \varepsilon_t \quad . \tag{2}$$

The null hypothesis is one unit root in the autoregressive operator for x_t, i.e., $\varphi(1) = 0$ for $\varphi(z) = (1 - \sum_{j=1}^{p} \pi_j z^j)(1-z) - \beta z$ or, equivalently, $\beta = 0$. The alternative is that $\varphi(z)$ does not have a unit root. While the DF test has

been criticized in the literature (for a critical review, see MADDALA AND KIM, 1998), its apparent simplicity is attractive. Also note that it exactly corresponds to a univariate version of JOHANSEN's cointegration model (1). Hence, one of the key arguments against the DF test, i.e., doubts on the autoregressive nature of the generating mechanism, is misplaced in the setting of the JOHANSEN procedure, which assumes an autoregression for the system variable X_t. If (1) is the true model and a component is stationary, the implied autoregressive model for that component will usually not have finite order. Therefore, the equation (2) approximates the JOHANSEN model in two aspects: firstly, the lags of other components are omitted from the right-hand side; secondly, the lag order of the univariate autoregressive representation is truncated at p. However, in short data sets the Dickey-Fuller test may be competitive due to its less profligate parameterization.[1]

Like the multivariate JOHANSEN procedure, the univariate DF test can be used with several combinations of deterministic terms. Because the test is often used to discriminate drifting integrated from trend-stationary variables, it makes sense to use the test as in (2), though the set-up of hypotheses is then non-standard, as b is implicitly restricted under the null. In the model (1) that is investigated here, trend-stationary variables cannot appear. Hence, the specification without the trend regressor deserves consideration.

For the pre-test stage in the Johansen procedure, unit-root tests are commonly applied with the aim of classifying the variables into one out of three classes: I(0), I(1), and I(2) variables. According to the so-called Pantula principle (after PANTULA, 1989), a two-stage sequence of DF tests starts with testing the I(2) null hypothesis against an I(0)∪I(1) alternative 'I(0/1)', then in case of first-stage rejection an I(1) null is tested against an I(0) alternative. As outlined above, the application of the Johansen procedure does not require the second test, as I(0) variables are treated correctly in the multivariate model. The DF test statistic ξ_1 of this paper corresponds to this second test. The first test serves to eliminate objects outside the focus of the analysis. Unless the re-

[1] I am grateful to a referee for pointing out this argument.

searcher decides to proceed with the first difference of the original variables, it is really a *specification* test and is comparable to tests for breaks, non-normality etc. Specification tests are not in focus here (see Section 3.1).

2.3 Multivariate augmentation of the Dickey-Fuller test

Like NEUSSER (1990), some researchers argue that the post-test of JOHANSEN has an inherent advantage in discriminatory power over the DICKEY-FULLER test, as it processes multivariate information. However, multivariate information can be incorporated into the DF test directly. For example, assuming a second variable y_t to be I(0/1), the t–statistic on β, $t(\beta)$, in the regression

$$\Delta x_t = a + bt + \beta x_{t-1} + \sum_{j=1}^{p-1} \pi_j \Delta x_{t-j} + \sum_{j=1}^{p-1} \tilde{\pi}_j \Delta y_{t-j} + \varepsilon_t \tag{3}$$

will have similar asymptotic properties to the original DF test. Using certain assumptions, HANSEN (1995) shows that the null distribution of $t(\beta)$ is a mixture of Dickey-Fuller and standard distributions. Although HANSEN develops his results in a univariate regression framework conditional on Δy_t, (3) can also be viewed as a component in a VAR. For demonstration, assume a first-order VAR for the variables $(\Delta x_t, \Delta y_t)$ augmented by a lag of x_t, i.e.,

$$\begin{aligned} \Delta x_t &= \alpha_1 x_{t-1} + \pi_{11} \Delta x_{t-1} + \pi_{12} \Delta y_{t-1} + \varepsilon_t^{(1)} \\ \Delta y_t &= \alpha_2 x_{t-1} + \pi_{21} \Delta x_{t-1} + \pi_{22} \Delta y_{t-1} + \varepsilon_t^{(2)} \quad . \end{aligned} \tag{4}$$

This is the general form for a VAR on $(x_t, \Delta y_t)$ with mixed lag orders of two and one for the variables. This is also an error-correction representation for a second-order VAR on (x_t, y_t) with the potential cointegrating vector $(1, 0)'$ assumed as known. The basic assumption that the system contains no variables with unit roots different from one, no variables with explosive behavior, and no variables that are integrated of order two or higher imposes some conditions on the admissible parameter space for the model (4). These conditions are complex constraints in the coefficient space, though they are relatively simple conditions for the eigenvalues of the matrix when the system is re-written as a four-dimensional first-order VAR. This idea will be taken up in Section 3.

The cases $\alpha_1 = 0$ or $\alpha_2 = 0$ are of special interest. If $(\alpha_1, \alpha_2) = (0, 0)$, both variables are I(1) and there is no cointegration. For $\alpha_1 \neq 0$, x_t is self-cointegrating and stationary while y_t is I(1), whether $\alpha_2 = 0$ or $\alpha_2 \neq 0$. The case $\alpha_1 = 0$ and $\alpha_2 \neq 0$ is not admissible, as it violates the assumption that both variables are I(0) or I(1). Honoring HANSEN, the t–test for $\alpha_1 = 0$ will be called the CADF (covariate-augmented Dickey-Fuller) test in the following. The symbol ξ_1 will denote the value of a univariate unit-root test statistic, i.e., either the DF or the CADF version.

Any further stationary augmentation is possible. Augmentation by lagged 'level' I(0/1) variables is not possible, as these may be cointegrated with the x_{t-1} regressor and may therefore impair the evidence on stationarity in x_t. Therefore, even the CADF test uses less conditioning information than the JOHANSEN test in the general model (1) for $r > 1$ and $n > 2$.

3 Joint evidence from two tests

The debate on the correct way of assessing the merits of a joint application of two hypothesis tests with exchanged null and alternative hypotheses remains unresolved in statistics. So-called confirmatory analysis (see CHAREMZA AND SYCZEWSKA, 1998, or a similar concept used in DHRYMES, 1998) is shunned in the literature (see MADDALA AND KIM, 1998). This technique, although of interest in its own right, is still 'local' in the sense that it evaluates test power and size at specified points of the parameter space. Thus, the prescription of the joint picture may be of little help to the practitioner, as this point of the parameter space is unknown or testing would otherwise not be necessary. By contrast, Bayes testing is 'global' in the sense that the points of the parameter space are weighted by a prior distribution. Critics of Bayes testing point out the sensitivity of the global decision to the choice of prior distributions, while practitioners are often reluctant to conduct the lengthy computation that is involved in the calculation of posterior odds by way of numerical integration.

Inspired by the work of HATANAKA (1996), KUNST AND REUTTER (2002)

constructed a Bayes test to evaluate the performance of a comparable joint test globally. This Bayes test directly answers the question whether, for a given value of the two test statistics (ξ_1, ξ_2), it is more probable that this value was generated from Θ_A or from Θ_B. If, for example, Θ_A has a higher probability, the researcher who observes (ξ_1, ξ_2) will decide for Θ_A. This decision rule minimizes the probability of an incorrect decision, assuming that (ξ_1, ξ_2) is the only available information. The need to determine probabilities of hypotheses requires a Bayesian approach. It is plausible to assign a prior weight of 0.5 to each of the hypotheses, unless there is a reason for preferring one of the hypotheses, for example if $I(1)$ processes are more frequently observed than $I(0)$ processes. In our application, we assume no such prior information, thus the principle of insufficient reason applies and prior probabilities are equal. If the hypotheses are not simple, prior distributions for θ are also required on the parameter subspaces Θ_A and Θ_B. To avoid biasing the test outcome by arbitrary choice of priors, the prior distributions are standardized across hypotheses, such that the weighting of comparable regions of the parameter space is identical under both hypotheses. By comparing the posterior probabilities of $\Theta_A | (\xi_1, \xi_2)$ and $\Theta_B | (\xi_1, \xi_2)$, an optimal decision is assigned to any observable point (ξ_1, ξ_2). The image space of (ξ_1, ξ_2) is decomposed into two regions, the preference area for Θ_A and the preference area for Θ_B.

In the current problem, we define Θ_A by the hypothesis that the first unit vector e_1 cointegrates and that the first component is stationary. We define Θ_B by the hypothesis that e_1 does not cointegrate. The parameter subspaces Θ_A and Θ_B vaguely correspond to the hypotheses H_u and H_r introduced in Section 2.1. Precise definitions are given in Section 3. The distribution of ξ_2 is asymptotically constant over Θ_A with *c.d.f.* $G_A(.)$, while the distribution of ξ_1 is asymptotically constant over Θ_B with *c.d.f.* $G_B(.)$. It is convenient to report the preference areas in the fractile space as $\left(G_B^{-1}(\xi_1), G_A^{-1}(\xi_2)\right)$, instead of the original image space of the test statistics (ξ_1, ξ_2). This enables depicting the preference areas in a unit square. Exact validity of the asymptotic distributions G_A and G_B is not necessary for this step. Even distributions that are entirely

different from G_A and G_B imply valid decompositions of the two preference areas, though a direct comparison with classical testing is impaired. In finite samples, the distributions of $G_B^{-1}(\xi_1)\,|\theta \in \Theta_B$ and $G_A^{-1}(\xi_2)\,|\theta \in \Theta_A$ will not be uniform and will depend on θ within the hypotheses.

In detail, a parameter θ is drawn either from the prior on Θ_A or from the prior on Θ_B, with identical probability. Corresponding to θ, a finite trajectory of the time-series process is generated using a T–vector of random numbers drawn from an n–variate normal distribution. From each trajectory, statistics ξ_1, ξ_2 are calculated and are transformed into (approximate) fractiles $G_B^{-1}(\xi_1)$ and $G_A^{-1}(\xi_2)$. Even after generating a high number of replications, a comparison of probabilities for a given point in $[0,1] \times [0,1]$ will not be possible. This is, however, possible for a small area in $[0,1] \times [0,1]$, a 'bin' defined as $((j-1)\Delta, j\Delta) \times ((k-1)\Delta, k\Delta)$, where $\Delta = n_g^{-1}$ denotes the resolution of the grid with n_g grid values and the pair of integers (j,k) indexes the bin for $j,k = 1,\ldots,n_g$. Each bin contains a large quantity of similar values of fractiles that correspond to statistics that, in turn, stem from either a Θ_A trajectory or a Θ_B trajectory. The relative frequency of observations from Θ_A is an estimate of $P(\Theta_A|G_B^{-1}(\xi_1) \in (j\Delta,(j+1)\Delta),\, G_A^{-1}(\xi_2) \in (k\Delta,(k+1)\Delta)$. If most trajectories stem from Θ_A, the researcher, who observes the statistics but does not know θ, will then decide for Θ_A. Otherwise she will decide for Θ_B.

For perfect resolution $\Delta \downarrow 0$, a decision for hypothesis Θ_A (Θ_B) is suggested at point (ξ_1,ξ_2) or $\left(G_B^{-1}(\xi_1), G_A^{-1}(\xi_2)\right)$ whenever $P(\Theta_A|(\xi_1,\xi_2)) > (<)$ $P(\Theta_B|(\xi_1,\xi_2))$. The curve

$$C = \left\{\left(G_B^{-1}(\xi_1), G_A^{-1}(\xi_2)\right) : \mathrm{P}(\Theta_A|(\xi_1,\xi_2)) = \mathrm{P}(\Theta_B|(\xi_1,\xi_2))\right\}$$

was named the *decision contour* for the testing problem in Kunst and Reutter (2002). A referee has suggested the alternative name *iso-odds curve*, as it contains those points where the decision maker is undecided between hypotheses and where posterior odds for both hypotheses coincide. Apart from the main contour, other iso-odds curves may be of interest, as these may represent the decision in the presence of prior information in favor of one of the two hypotheses.

The maps will show such secondary curves for odds of 0.25 and of 0.75.

The advantage of evaluating iso-odds curves over a fully classical approach is that it allows the assessment of a decision conditional on an observed pair of test statistics. The advantage relative to a traditional Bayesian approach—with the decision taken conditional on the complete sample information and not just two statistics—is that it does not require complex evaluations within an empirical application. The decision prescribed by the iso-odds curves for all similar problems can be used again for the data at hand. It is optimal in the sense that, under the given prior design and restricting available information to (ξ_1, ξ_2), any deviation from it incurs a larger probability of classifying the observed series incorrectly.

The difference between the 'local' and 'global' approach can also be seen as follows. The local (or classical) approach conditions the analysis and all simulations on fixed parameter values. This is helpful for studying theoretical properties but does not directly help the practitioner who does not know the true parameter in an application. The global approach conditions all analysis and simulations on the *observed* statistics. The simulation design is varied over virtually 'all' possible data-generating processes. A given value of the observed statistic may have been produced by any value of the parameter space but it may be connected more frequently to one of the two subspaces (hypotheses). This then helps the practitioner who also observes a pair of test statistics and knows that the more probable hypothesis is the preferred decision.

3.1 Prior distributions within a frame

While there is some diversity regarding the use of the word 'model selection' in econometrics, its traditional meaning in statistics describes a situation where one out of two or more specified model classes is selected on the basis of sample information. In parametric modeling, a decision is sought whether a parameter θ belongs to Θ_j, where Θ is partitioned as $\bigcup_{j=1}^{J} \Theta_j$. The parameter may be interpreted as indicating the data-generating or true model or as its best ap-

proximation within Θ or quasi-true model. Here, $\Theta = \Theta_A \cup \Theta_B$ and there are only two model classes.

In model selection, the general or maintained model is put beyond doubt. This may be justified as a working assumption, as a result of previous specification searches, or by economic theory. It follows that, for the purpose of model selection, other model classes are not considered and pure specification tests are inappropriate. For example, in our selection problem it does not make sense to test for I(2) models, as these models are not within Θ. Kunst and Reutter (2002) refer to the maintained statistical model as a *frame* of the decision problem. The choice of the frame is crucial for the selection decision, as different frames may lead to different selections. In the current problem, the model frame consists of vector autoregressions (1) with given cointegrating dimension, given stability restrictions, and given lag order.

In line with classical statistics, we focus on estimating the parameter θ from the data via maximum likelihood as a point estimate $\hat{\theta}$. In Bayesian statistics, usage of $\hat{\theta}$ can be motivated as an approximate distributional center of the posterior distribution from a non-informative prior distribution. Focusing on $\hat{\theta}$ is natural, as the statistics $\xi_1$and ξ_2 are derived using the maximum-likelihood method. In order to determine whether the observed data is more likely to have been generated by Θ_A or by Θ_B, it does not suffice to look whether $\hat{\theta} \in \Theta_A$ or $\hat{\theta} \in \Theta_B$, particularly if one of the two hypothesis sets is lower-dimensional. In many cases, such a decision is inspired by a high *a priori* probability of θ being a member of the lower-dimensional part. We express this *a priori* probability by assigning a weight of 0.5 to either hypothesis. Assigning identical prior weights to all hypotheses is justified by the principle of insufficient reason, unless prior knowledge favors a hypothesis. For simple applications with a univariate test statistic, this approach yields decisions that are comparable to a classical test, except that the significance level is implied by the prior distribution and the sample size, instead of being fixed exogenously. In large samples, the Bayesian setup usually attains a fully consistent test in the sense that decisions are correct with probability one.

The elicitation will be highlighted on the basis of an assumed cointegrating rank of one. The cases of first- and second-order autoregression are considered.

3.2 Imposing stationarity as a cointegration restriction in an AR(1) model

The first-order autoregression with a cointegrating rank of one can be written as

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \varepsilon_t \tag{5}$$

with the n–vectors μ, α, β and the covariance matrix $\mathrm{E}\,(\varepsilon_t \varepsilon_t') = \mathbf{\Sigma}_\varepsilon$. If the system is to be stable apart from the integrating directions, the eigenvalues of $\mathbf{\Pi} = \mathbf{I} + \alpha\beta'$ must be in the range $[-1, 1]$. More particularly, $n-1$ eigenvalues are 1 and one eigenvalue is in the open interval $(-1, 1)$.

A simple and good *a priori* distribution over $\Theta = \{(\alpha',\, \beta',\, \mu',\, \mathrm{vech}\,(\mathbf{\Sigma}_\varepsilon)')' :$ $n-1$ roots are 1, 1 root is stable$\}$ assumes the following specifications:

$$\mathbf{\Pi} = \mathbf{Z} \begin{bmatrix} \lambda & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & \mathbf{I}_{(n-1)\times(n-1)} \end{bmatrix} \mathbf{Z}^{-1} \tag{6}$$

with

$$\begin{aligned} Z_{jj} &= 1, j = 1, \ldots, n, \\ Z_{jk} &\sim N(0, \sigma_z^2), j \neq k, \\ \mu &\sim N(0, \sigma_\mu^2 \mathbf{I}_n) \\ \lambda &\sim U(-1, 1) \end{aligned} \tag{7}$$

The variance parameters $\mathbf{\Sigma}_\varepsilon, \sigma_z^2, \sigma_\mu^2$ cannot be given improper prior distributions as it is common in Bayesian estimation theory, as the model will be simulated and one cannot draw from an improper distribution. Because the test decision is invariant in $\mathbf{\Sigma}_\varepsilon$, $\mathbf{\Sigma}_\varepsilon \equiv \mathbf{I}_n$ is set and the possible reaction to modifications of the hyperparameters σ_μ^2 and σ_z^2 is then studied. In the basic experiments, we set $\sigma_\mu^2 = 0$ and $\sigma_z^2 = 1$. The implied distribution for the matrix $\mathbf{\Pi}$ is a member of the Jordan distribution family that was also used in Kunst and

REUTTER (2002). Assuming $\mathbf{\Sigma}_\varepsilon$ as fixed, the definition of Θ excludes some lower-dimensional manifolds from $\mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^n$, such as the case of $n = 2$, $\alpha = (1, 0)'$, $\beta = (0, 1)'$, which would give rise to second-order integration. $\mathbb{S}^n$ is the surface of the n–dimensional unit sphere and has dimension $n - 1$. The parameter space Θ has dimension $3n - 1$. The drawn random parameter has dimension $n^2 + 1$ and this may imply some inefficiency for $n > 2$, as not all matrix elements of $\mathbf{Z}$ are needed to determine $\mathbf{\Pi}$. In practice, this feature is not costly unless n is very large.

The hypotheses of interest Θ_A and Θ_B correspond to $\beta = \zeta e_1$ and $\beta \neq \zeta e_1$ for $\zeta \neq 0$. Analogously, one may consider $\beta = \zeta e_j$ for any specific $j \in \{1, ..., n\}$, though we focus on the first variable without restricting generality. Θ_A has lower dimension $2n$, whereas $\Theta_B = \Theta - \Theta_A$ has full dimension $3n - 1$. Thus, stationarity of $X^{(1)}$ forms a typical null hypothesis of classical statistics. Within Θ_B, $X^{(1)}$ is first-order integrated and there is (non-trivial) cointegration in the system. Cases of stationary $X^{(j)}$ for $j \neq 1$ also belong to Θ_B and are not treated separately.

Assuming $n = 2$ for simplicity of exposition, the hypothesis Θ_A implies that

$$\begin{aligned}
\mathbf{\Pi} &= \mathbf{I} + \alpha\beta' = \begin{bmatrix} 1 + \alpha_1\zeta & 0 \\ \alpha_2\zeta & 1 \end{bmatrix} \\
&= (1 - z_{12}z_{21})^{-1} \begin{bmatrix} 1 & z_{12} \\ z_{21} & 1 \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -z_{12} \\ -z_{21} & 1 \end{bmatrix} \\
&= (1 - z_{12}z_{21})^{-1} \begin{bmatrix} \lambda - z_{12}z_{21} & (1 - \lambda) z_{12} \\ (1 - \lambda) z_{21} & 1 - \lambda z_{12}z_{21} \end{bmatrix} . \qquad (8)
\end{aligned}$$

Therefore, $z_{12}(1 - \lambda) = 0$ and $\lambda = 1$ or $z_{12} = 0$. $\lambda = 1$ is excluded by assumption, such that $z_{12} = 0$. The priors for the hypotheses Θ_A and Θ_B differ by restricting the element $z_{12} = 0$ for Θ_A, while $z_{12} \sim N(0, 1)$ under Θ_B.

3.3 Imposing stationarity as a cointegration restriction in an AR(2) model

In higher-order systems, imposing stationarity conditions on the coefficient matrices becomes a more complex issue. Bivariate second-order autoregressions will be treated in detail. Extensions to $n > 2$ and $p > 2$ are straightforward.

The bivariate AR(2) system allows the representation

$$\begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \\ \mathbf{I}_2 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{9}$$

which is written in compact notation as

$$z_t = \mathbf{A} z_{t-1} + e_t \quad . \tag{10}$$

The system is stable if all eigenvalues of $\mathbf{A}$ are inside the unit circle. For simplicity, we do not consider complex eigenvalues, hence all eigenvalues of $\mathbf{A}$ are assumed to be real and lying within the interval $(-1, 1)$, excepting one eigenvalue of unity.[2] Then, the system is cointegrated with cointegrating rank one. Considering the Jordan representation of $\mathbf{A}$, again assuming a diagonal non-derogatory form,

$$\mathbf{A} = \mathbf{Z}\mathbf{D}\mathbf{Z}^{-1} \quad , \tag{11}$$

the special block form of $\mathbf{A}$ must be imposed on the generating elements in $\mathbf{Z}$. Suppose the 4×4–matrices $\mathbf{Z}$ and $\mathbf{D}$ are split into 2×2–submatrices $\mathbf{Z}_{ij}$ for $i, j = 1, 2$ and $\mathbf{D}_i$ for $i = 1, 2$. The conditions (9) and (11) yield

$$\begin{aligned} \mathbf{Z}_{11}\mathbf{D}_1 &= \mathbf{\Phi}_1\mathbf{Z}_{11} + \mathbf{\Phi}_2\mathbf{Z}_{21} \quad , \\ \mathbf{Z}_{12}\mathbf{D}_2 &= \mathbf{\Phi}_1\mathbf{Z}_{12} + \mathbf{\Phi}_2\mathbf{Z}_{22} \quad , \end{aligned} \tag{12}$$

[2] It is straightforward to generalize the prior distribution such that it may contain complex roots. A good argument for assigning discrete priors to complex and all-real cases is more difficult. Some of the experiments of this paper were also conducted allowing for complex roots. Effects on the main decision contours were relatively small.

and

$$\begin{aligned} \mathbf{Z}_{21}\mathbf{D}_1 &= \mathbf{Z}_{11} \quad , \\ \mathbf{Z}_{22}\mathbf{D}_2 &= \mathbf{Z}_{12} \quad . \end{aligned} \tag{13}$$

Therefore, while the three elements in $\mathbf{D}$ that are not 1 can be drawn from a uniform distribution over $(-1, 1)$, only one out of two submatrices is filled with normal elements, while the other one follows from (13). To retain the standardization from the AR(1) model, first the diagonal elements of $\mathbf{Z}_{11}$ and $\mathbf{Z}_{22}$ are set at 1 and the off-diagonal entries are drawn from $N(0, 1)$. Then, $\mathbf{Z}_{12}$ and $\mathbf{Z}_{21}$ are obtained from (13). This procedure requires only 4 draws from a Gaussian distribution and serves as the reference prior for the unrestricted model Θ_B.

For Θ_A, the form of $\mathbf{\Phi}_1 + \mathbf{\Phi}_2$ must also be restricted, as the second column of the impact matrix $\mathbf{I}_2 - \mathbf{\Phi}_1 - \mathbf{\Phi}_2$ must be $(0, 0)'$. This restriction can also be written as

$$\begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \\ \mathbf{I}_2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad . \tag{14}$$

The vector $(0, 1, 0, 1)'$ is an eigenvector for the eigenvalue 1 in $\mathbf{A}$. $\mathbf{Z}$ contains eigenvectors for the respective eigenvalues at the positions defined in $\mathbf{D}$. The corresponding column of $\mathbf{Z}$ is replaced by $(0, 1, 0, 1)$, whence (13) is easily seen to be fulfilled. Only three draws from a Gaussian distribution are necessary. From (14), note that the second column of the impact matrix is $(0, 0)'$ and therefore x_t is stationary.

For both Θ_A and Θ_B, the original $\mathbf{D}$ matrix was shuffled in the beginning in order to avoid asymmetries. For Θ_A, the eigenvector $(0, 1, 0, 1)'$ was inserted at the position of the unit element of $\mathbf{D}$. All univariate unit-root tests were conducted for x_t.

3.4 Decision boundaries

The aim of the maps is to establish areas where Θ_A and Θ_B are preferred, given the observed test statistics, for example the DICKEY-FULLER statistic ξ_1 and the JOHANSEN statistic ξ_2. In line with usual Bayes testing, a hypothesis is preferred whenever its probability given the data, or rather the pair (ξ_1, ξ_2), exceeds 1/2. Based on a suggestion by HATANAKA (1996), both statistics are coded by the respective fractiles under their null distributions. For the calculation of these fractiles, two options are available. Firstly, asymptotic distributions can be used, such as χ^2 for the JOHANSEN test, which is particularly attractive if closed forms of the distribution functions exist, or alternatively simulated theoretical distributions for the specified sample size. This is the solution that is more representative of the situation of a fictitious empirical researcher who is using tests to reach a decision. Secondly, simulated finite-sample fractiles can be obtained directly from the part of the simulated 'posteriors' that have been drawn from the respective null model. Because theoretical and asymptotic null distributions may not be valid in finite samples and in the presence of a various nuisance parameters that are randomized for both hypothesis priors, the latter option may be attractive. We tentatively used both specifications and found the deviations between them to be acceptably small. Finally, the former option was adopted, particularly as a map for the sample fractiles would require any potential user of the map to re-run our specific simulation design. On the other hand, fractiles of the χ^2 distribution exist in a closed form and fractiles of the DICKEY-FULLER distribution can be simulated quickly.

As a first step, fractile points for ξ_2 are determined analytically for χ^2 in order to approximate the empirical distribution of $\xi_2|\Theta_A$, while Monte Carlo draws of random walks serve to determine fractile points for ξ_1 that approximate the empirical distribution of $\xi_1|\Theta_B$. Fractiles are stored at a grid of $\Delta = 0.01$, which gives $100^2 = 10,000$ discretized bins. As a second step, a large number of trajectories are randomly drawn, that is, with randomized nuisance, from Θ_A as well as from Θ_B, and are allocated to the bins by comparison to

the previously determined fractiles. If more of these pairs within a 'bin' stem from a certain hypothesis, the conditional probability of that hypothesis exceeds the conditional probability of the rival hypothesis. The bin is then marked as 'belonging to Θ_j' with $j = A, B$.

In the presented charts, which were generated using the GAUSS software package, the two preference areas for hypothesis Θ_A and Θ_B are separated by the *decision boundary*. The preference areas are indicated by different shading. Furthermore, we chose to divide the decision area for Θ_A by a further boundary at the location where the posterior probability for Θ_A is 0.75. In the region to the southwest of this secondary boundary, support for Θ_A can be regarded as 'strong'. This area is marked in the darkest shade. In some cases, we also show the analogous boundary for the probability for Θ_B of 0.75. Then the area to the northeast with strong preference for Θ_B is marked in pure white. In summary, all maps show three or four areas: the areas in dark grey and medium grey prefer Θ_A and the areas in light grey and (possibly) white support Θ_B.

The visual impression from decision contour curves may permit a tentative evaluation of the relative strength of tests. If in the largest part of the diagram the boundary curve runs parallel to an axis, the statistic displayed on that axis is of little value as compared to the rival statistic. If the boundary curve runs in a different direction, both statistics should be combined in order to improve upon the overall decision. Particularly, if the contour shows a $-45°$ slope, a direct comparison of p–values is recommended. If the 'corners of conflict', in this case the north-west and the south-east corner, are intersected by the boundary curve, usage of a joint test is supported. If these corners are allotted clearly on the basis of one of the two statistics, usage of just one test statistic suffices.

For a large number of replications, the areas are typically connected and are separated by smooth boundary curves. Denoting the (approximate) null fractiles for the statistics ξ_1 and ξ_2 by $p_{1,x}$ and $p_{2,y}$ for $0 \leq x, y \leq 1$, $P\{(\xi_1, \xi_2) \in \left(p_{1,0.01(k-1)}, p_{1,0.01k}\right) \times \left(p_{2,0.01(l-1)}, p_{2,0.01l}\right) | \Theta_j\}$ may be small for both $j = A, B$ for some k, l. In other words, some bins are poorly populated. Then, the simulated posteriors of interest $P\{\Theta_j | \, (\xi_1, \xi_2) \in \left(p_{1,0.01(k-1)}, p_{1,0.01k}\right) \times$

$\left(p_{2,0.01(l-1)}, p_{2,0.01l}\right)\}$ will be unreliable and, therefore, there will be little information on whether the posterior probability of Θ_A or Θ_B is larger. Consequently, the simulated boundaries may look blurred and unreliable in areas where the marginal density of $\left(G_B^{-1}(\xi_1), G_A^{-1}(\xi_2)\right)$ or, equivalently, of (ξ_1, ξ_2) is low. In analogy to usual density estimation, *kernel smoothing* was invoked to mitigate this problem.

With kernel smoothing, the value in the bin (k_0, l_0) is replaced by a weighted average over an area of neighboring bins that are centered at (k_0, l_0). Formally, the function value $f(k_0, l_0)$ for a function defined on $\{1, \ldots, n_g\} \times \{1, \ldots, n_g\}$ is replaced by its smoothed version

$$f_s(k_0, l_0) = \sum_{k=k_0-n_w}^{k_0+n_w} \sum_{l=l_0-n_w}^{l_0+n_w} w(k,l) f(k,l) \quad . \tag{15}$$

Here, n_g denotes the number of grid values, in this case $n_g = 100$. Some modifications have to be adopted for indices outside the range $\{1, \ldots, n_g\}$. After some experimentation with kernel functions and areas, we decided to use the kernel function

$$w(k,l) = \frac{W}{1 + |k - k_0| + |l - l_0|}, \quad k_0 - n_w \le k \le k_0 + n_w, \quad l_0 - n_w \le l \le l_0 + n_w \tag{16}$$

with the area size parameter n_w, which was set at the minimum value that achieved smooth boundaries. The value W is set according to the requirement

$$1 = \sum_{k=k_0-n_w}^{k_0+n_w} \sum_{l=l_0-n_w}^{l_0+n_w} w(k,l). \tag{17}$$

Whenever the indices violate the range, the weights of non-existing observations are set to zero and hence an asymmetric version of the kernel is used, with W properly defined.

Kernel smoothing can be applied to the original counts of entries in the bins or to the empirical posterior probabilities of hypotheses. In benchmark cases with known outcome, it was found that smoothing the probabilities created less distortions than smoothing the entry numbers, hence this option was chosen.

4 The simulated contour maps

For the first experiment, ξ_2 is specified as the JOHANSEN statistic as described in Section 2.1, without augmentation and assuming a bivariate first-order VAR, just as the one that actually generates the data. The unit-root statistic ξ_1 is specified as the DF statistic with a constant term included (DICKEY-FULLER's 'μ–test statistic'), with $p-1$ augmenting lags. The lag order p was found by an AIC search with the upper bound of 5. Due to the bivariate generating model, univariate autoregressions are imperfect specifications, hence lag augmentation was considered to make up for this defect. The DF statistic *without* augmentation was found to perform poorly. Its implied null distribution in the experiment (averaged over Θ_B) differs strongly from the one tabulated in the literature. With an augmentation determined from the sample, the correspondence improves. Also, the trend-augmented DF–τ statistic yields a relatively poor performance that cannot keep pace with the JOHANSEN test. This is probably due to the fact that the JOHANSEN statistic uses information on the deterministic part of the generating model while the DF–τ statistic would not. The implied JOHANSEN null distribution (averaged over Θ_A) comes quite close to the theoretical χ^2_1, with a small negative size bias.[3] Sample size was varied over $T = 50$, 100, 200. The number of replications was always 2×10^6, i.e., 10^6 for each of the hypotheses Θ_A and Θ_B.

For first-order autoregressions (see Section 3.2), Figures 1–4 show the simulated contour maps. The interpretation of the fractiles is slightly different for the two statistics. The DICKEY-FULLER statistic ξ_1 *rejects* for large negative values (the lower tail), which therefore conform to Θ_A where the unit vector cointegrates. The JOHANSEN statistic ξ_2 *rejects* for large positive values (the upper tail), which therefore conform to Θ_B where the cointegrating vector differs from the unit vector. A classical decision would prefer Θ_A in the vertical

[3] A referee has pointed out that the deviation of the small-sample distribution of ξ_2 from χ^2_1 can be substantial for certain parameter values in Θ_A. Averaging by the prior distribution over Θ_A apparently removes a large portion of the discrepancy.

band $[0.0, 0.05] \times [0.0, 0.95]$ and Θ_B in the horizontal band $[0.05, 1] \times [0.95, 1]$.[4] In the small north-west area $[0, 0.05] \times [0.95, 1]$ and in the large south-east area, classical testing yields no clear decision. A typical *ad-hoc* approach would be to compare the implied fractiles or p–values directly and to divide preference areas by a diagonal passing through $(1, 0)$ and $(0, 1)$ at an angle of $-45°$.[5]

The maps reveal that the principal information for the decision is provided by the JOHANSEN statistic ξ_2 that becomes critical if it falls into the upper decile of the assumed χ^2_1 null distribution. The main contour is sloping only gently until the DF statistic ξ_1 reaches some upper fractile. The position of this fractile point varies with the sample size and may be at 0.8 for $T = 100$. To the east of this value, Θ_B is generally preferred. Such large values of ξ_1 are usually taken as evidence on instability, locally 'explosive' behavior, or mis-specification. A traditional application of the DICKEY-FULLER test would locate the critical values in the *lower* fractiles.

As a kind of sensitivity analysis, the prior distribution was modified in Figure 3 such that the off-diagonal elements of the matrices $\mathbf{Z}$ were set to Cauchy random numbers instead of normal random numbers, as in the other experiments. The larger variation of behavior caused by the Cauchy distribution requires stronger smoothing. The vertical boundary seen in Figure 2 disappears, such that the suggested decision now relies on the Johansen statistic only. Otherwise, the structure of the decision map is similar to the basic experiment. From this and other unreported sensitivity experiments, we conclude that the choice of the prior distribution does not affect the main results too much, as long as the modification hits both hypotheses symmetrically. In further unreported experiments, similar effects were encountered by increasing $\sigma_z^2/\sigma_\varepsilon^2$ while maintaining normality.

For the important design of Figure 2, i.e., the DF test and $T = 100$, the

[4] Throughout we assume that classical testing proceeds with a significance level of 0.05. Different significance levels affect the described boundaries in a trivial way.

[5] The coordinates of the graphs are implied fractiles rather than p–values. The two concepts only coincide for one-sided tests with critical regions in the lower tail. Thus, the coordinates are p–values for the ξ_1 but not for ξ_2.

effect of deterministic terms was also studied. Figure 5 shows the effects of a randomized added drift with $\sigma_\mu^2 = 1$. This modification is apparently beneficial for the discriminatory power of the DF statistic ξ_1, as the boundary now runs vertically at the upper decile of its null distribution, unless ξ_2 reaches large values. Again, the suggested critical value is in the *upper* tails and not in the *lower* tails, which would be the traditional approach of unit-root testers. The variant confirms that the decision should be based on assigning a much larger weight to the statistic ξ_2 and to re-consider this decision only when the statistic ξ_1 points to a data-generating process in the explosive region. We also note that the north-east corner now yields a high posterior probability for Θ_B, recalling that such pairs correspond to rejection of the self-cointegration null jointly with large and positive DF statistics. Relative to the maps without drift, there are increased areas where the probability of one of the two hypotheses exceeds 0.75. The added trends permit more accurate decisions, therefore decreasing global risk.

For second-order autoregressions (see Section 3.3), the Figures 6-8 are obtained. These are generally similar to the first-order cases. The area with strong preference for Θ_A increases, which points to a steeper reaction along the decision contour and hence to a decrease in risk. With increased sample size, the evidence provided by DF statistics in the upper tails of their null distribution becomes more reliable, hence the Θ_B area stretches southward from its north-eastern habitat. Notwithstanding these minor differences between first-order and second-order autoregressions, the main message is the same. The optimal decision should rely mainly on the JOHANSEN test statistic. Large ξ_2 values indicate the validity of Θ_B, with 'large' defined as the upper decile of the χ^2 null distribution. For conspicuously low or high values of ξ_1, this decision should be adapted. These basic simulations are contrasted with a variant with randomized drift in Figure 9. The effect of the added trends are similar to those for the first-order autoregression. Global risk decreases and preference for Θ_B is restricted to the north-east corner.

Figures 10–12 study the CADF test according to HANSEN (1995). Because

both tests now use multivariate information, one may expect a relative improvement of performance relative to the DF test. For $T = 50$, the diagram shown in Figure 10 was obtained. The separating contour is almost horizontal and again gives preference to the JOHANSEN statistic. However, the secondary contour with a posterior probability of 0.75 for Θ_A is now vertical at a p–value that comes close to those used in traditional statistical analysis. In other words, a decision based on ξ_2 alone implies a substantial risk of an incorrect choice of Θ_A, although the data actually stems from Θ_B. Nevertheless, the risk of an incorrect decision for Θ_B in the large south-east region is larger, hence the 'non-rejection' according to the CADF test should be ignored. For $T = 100$, Figure 11 is obtained, which is similar and just shows a slight gain of the preference area for Θ_B in the east, i.e., for uncommonly large CADF statistics. The expansion of the preference area for Θ_B continues as the sample is increased to $T = 200$. This experiment is shown as Figure 12. The increase in sample information has only small benefits for the targeted decision, as the 0.75 support area for Θ_A hardly changes. Apparently, most of the additional information is exploited by the CADF statistic that now helps in discriminating among the 'pseudo-explosive' trajectories that yield a large DF statistic. A similar feature was observed for the univariate DF statistic (see Figures 1–4), hence the effect of incorporating multivariate information in the CADF test fails to convince.

Finally, it is of interest to study whether classical testing is supported in the sense that those areas that are clearly allotted are fully contained in the optimal preference areas. The figures reveal that this is not necessarily the case. The band $[0, 0.05] \times [0, 0.95]$ is always covered in the Θ_A preference area, which indicates that a rejection of the DF (or CADF) test in conjunction with an acceptance of the JOHANSEN test is to be interpreted as good evidence on self-cointegration and stationarity. On the other hand, the band $[0.05, 1] \times [0.95, 1]$ is not always fully covered in the Θ_B area, indicating that even a mild rejection of the ξ_2 test in conjunction with near-rejection of the DF test still supports stationarity of the component variable. For the north-west corner of conflict, the experimental evidence is fragile. Some maps support a decision for Θ_A, while

other maps point to a comparison of implied fractiles as a good approximation to optimal decision rules. Conflict outcomes in the south-east region will often be resolved by giving preference to the acceptance by ξ_2 in current econometric work (see NEUSSER, 1991). The maps, however, show that the DF (CADF) statistics should not be ignored. The further these statistics move into the upper tail of their asymptotic distribution, the more does the probability of non-stationarity of the investigated variable increase, even if ξ_2 formally supports its stationarity by non-rejection.

5 Summary and conclusion

In a series of Bayes-test experiments, the optimum decision that can be based on a joint application of two different unit-root tests was evaluated. The decision based on the depicted iso-odds curves is optimal in the sense that any deviation from the curve increases the probability of an incorrect decision. In its use of Bayes testing and of test statistics derived from classical statistics, the method blends frequentist with Bayesian ideas. In order to randomize the nuisance parameters, i.e., those parameters that are neither characteristic nor distinctive in defining a hypothesis, these are given a regular prior distribution. This prior distribution exhausts all possible models under any of the two hypotheses under investigation. The same number of stochastic simulations are conducted under either hypothesis. Fractiles for approximate null distributions are evaluated and the actual outcomes are compared to these null fractiles for both tests. The shape of the iso-odds curves critically depends on the prior specifications, hence care must be taken to avoid biasing the decision in favor of one of the hypotheses. While comparable priors across hypotheses serve in achieving this goal, some may feel that the shown figures are to be interpreted in a qualitative rather than exactly quantitative sense.

Further unreported plots have shown that the DICKEY-FULLER test with a trend regressor is not useful for joint testing with a JOHANSEN restriction test in the given model frame where a cointegrating rank of one has been established or

assumed. For the DICKEY-FULLER specification without a trend regressor and for the modified test that was suggested by HANSEN, the situation is different. The main features are summarized below:

1. In most contour plots, the iso-odds curve suggests basing the principal decision on the JOHANSEN statistic. Significant JOHANSEN statistics should usually be taken as evidence against self-cointegration or stationarity of the investigated component.

2. The contour shows a gentle downward slope, however, and the principal decision should therefore be re-considered in the presence of very negative unit-root statistics. The tested variable is likely to be stationary if, for example, the JOHANSEN statistic is in its upper decile but the unit-root statistic is in its lower decile. The exact position of the decision contour varies with the sample size, the intensity of short-run memory, and the deterministic part. The second effect was investigated by varying the lag order of the generating model. Adding deterministic parts usually admits a sharper decision and thus decreases decision risk.

3. The DF or CADF statistic also becomes conspicuous when it indicates explosiveness, i.e., is in its upper decile. Explosive values of unit-root statistics imply a high risk of misclassification. In smaller samples, such values suggest lowering the critical values for the JOHANSEN test, e.g., from 0.9 to 0.8. In larger samples, they indicate a unit root in the tested component variable.

The decision problem is relevant for practitioners who analyze economic data sets on the basis of widely available software that enables them to conduct the popular JOHANSEN procedure as well as traditional unit-root tests. Often, conflicting evidence is resolved on the basis of conjectures and *ad hoc* decisions, which may seem unsatisfactory. This motivates further work that is particularly directed at helping practitioners who face such empirical puzzles. The presented method can be refined in several directions, such as higher system dimension,

varying the search method for the correct lag specification, or more sophisticated deterministic structures. In all such extensions, the 'global' Bayes-test method may reveal features that remain unnoticed in the traditional 'local' frequentist framework with its fixed designs of data-generating processes. Of course, this observation should not discourage traditional power studies that focus on local or large-sample (asymptotic) features that necessarily remain out of the focus of the global approach. Hence, the two approaches are to be seen as complementary rather than as substitutes.

References

[1] Charemza, W.W., and E.M. Syczewska, 1998, Joint application of the Dickey-Fuller and KPSS tests, *Economics Letters* **61**, 17–21.

[2] Dhrymes, P. (1998) *Time Series, Unit Roots and Cointegration.* Academic Press.

[3] Dickey, D.A., and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of *the American Statistical Association* **74**, 427–431.

[4] Hansen, B.E., 1995, Rethinking the Univariate Approach to Unit Root Testing, *Econometric Theory* **11**, 1148–1171.

[5] Hatanaka, M., 1996, *Time-Series-Based Econometrics: Unit Roots and Co-Integration.* Oxford University Press.

[6] Horvath, M.T.K., and M.W. Watson, 1995,Testing for cointegration when some of the cointegrating vectors are prespecified, *Econometric Theory* **11**, 984–1014.

[7] Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control* **12**, 231–254.

[8] Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

[9] Johansen, S., and K. Juselius, 1995, Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP of UK, *Journal of Econometrics* **53**, 211–244.

[10] Kunst, R.M., and M. Reutter, 2002, Decisions on seasonal unit roots. *Journal of Statistical Computation and Simulation*. **72**, 403–418.

[11] Maddala, G.S., and I.M. Kim, 1998, *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.

[12] Neusser, K., 1991, Testing the long-run implications of the neoclassical growth model, *Journal of Monetary Economics* **27**, 3–37.

[13] Pantula, S.G., 1989, Testing for unit roots in time series data, *Econometric Theory* **5**. 256–271.

[14] Rahbek, A., and R. Mosconi, 1999, Cointegration rank inference with stationary regressors in VAR models, *Econometrics Journal* **2**, 76–91.

[15] Tanaka, K., 1996, *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley.

Figures

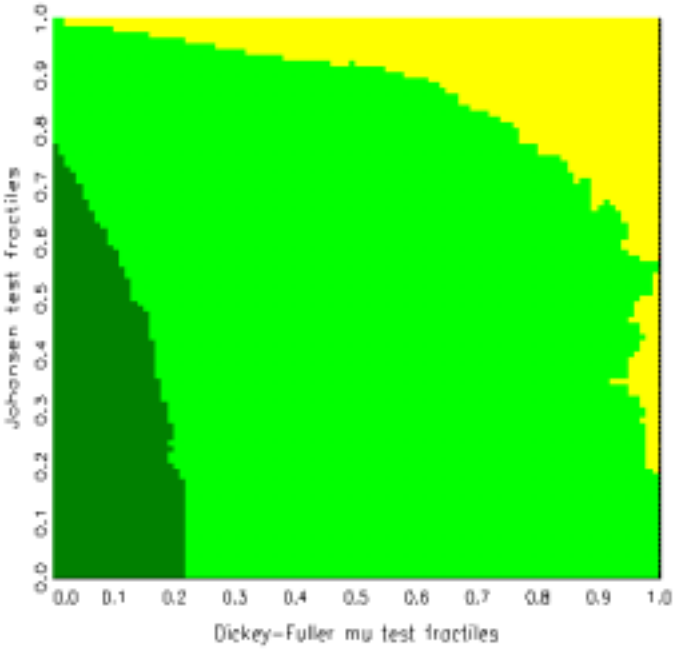

Figure 1: Boundary of optimum decision areas for $T = 50$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 31.

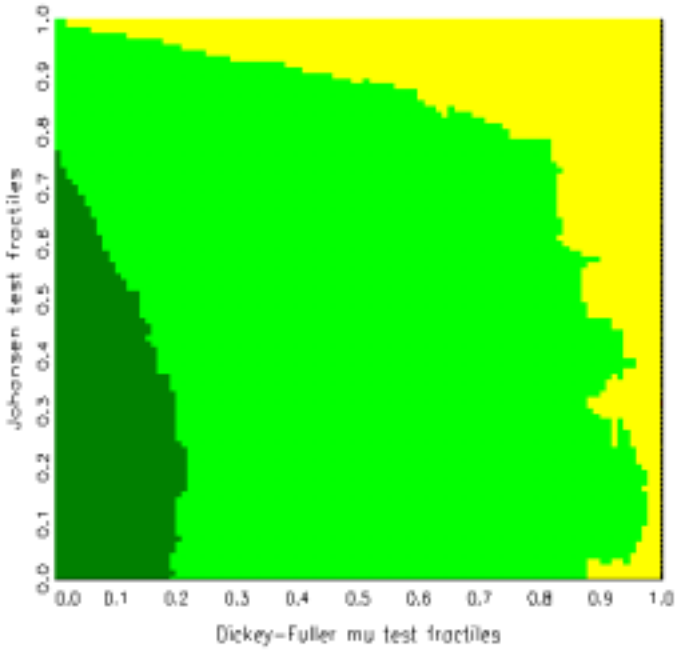

Figure 2: Boundary of optimum decision areas for $T = 100$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 33.

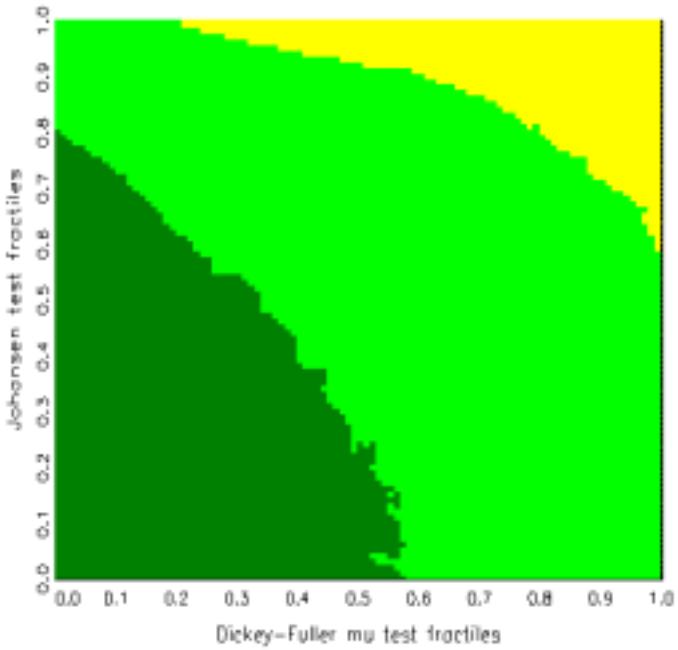

Figure 3: Boundary of optimum decision areas for $T = 100$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. First-order autoregression with Cauchy-Jordan priors. Smoothing constant is 43.

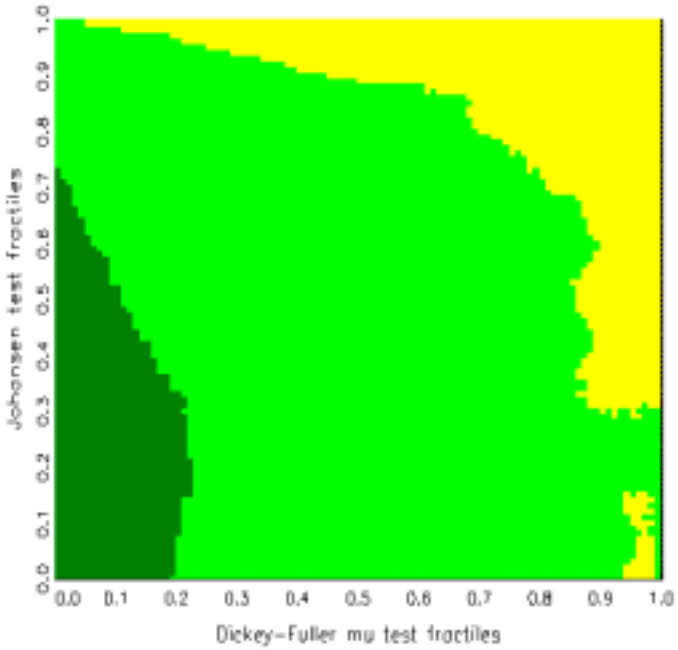

Figure 4: Boundary of optimum decision areas for $T = 200$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 39.

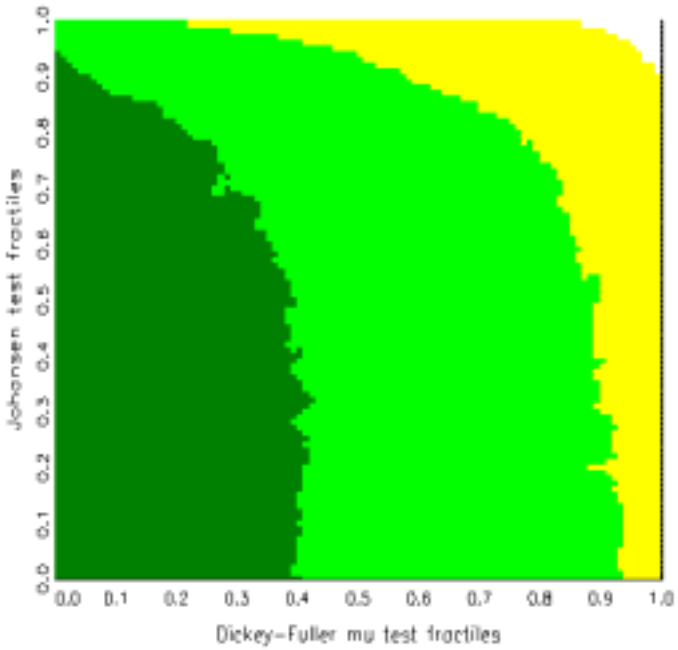

Figure 5: Boundary of optimum decision areas for $T = 100$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. First-order autoregression with random drift. Smoothing constant is 25.

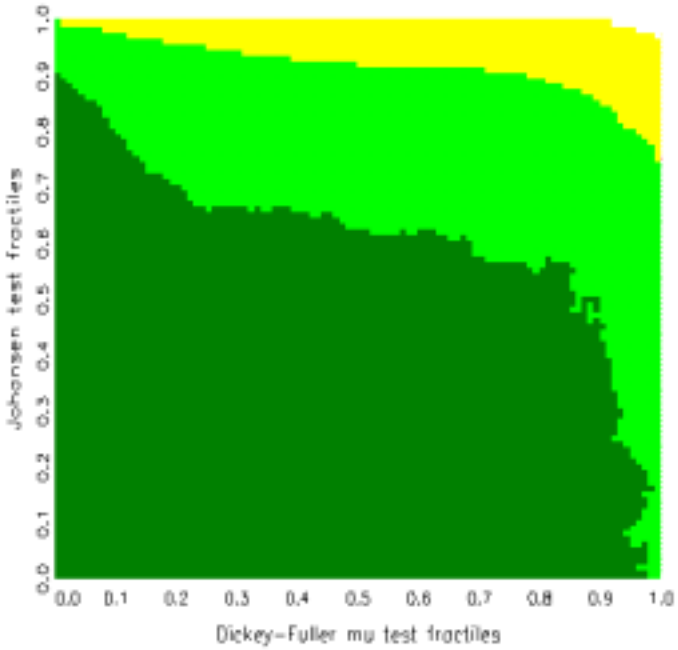

Figure 6: Boundary of optimum decision areas for $T = 50$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. Second-order autoregression. Smoothing constant is 15.

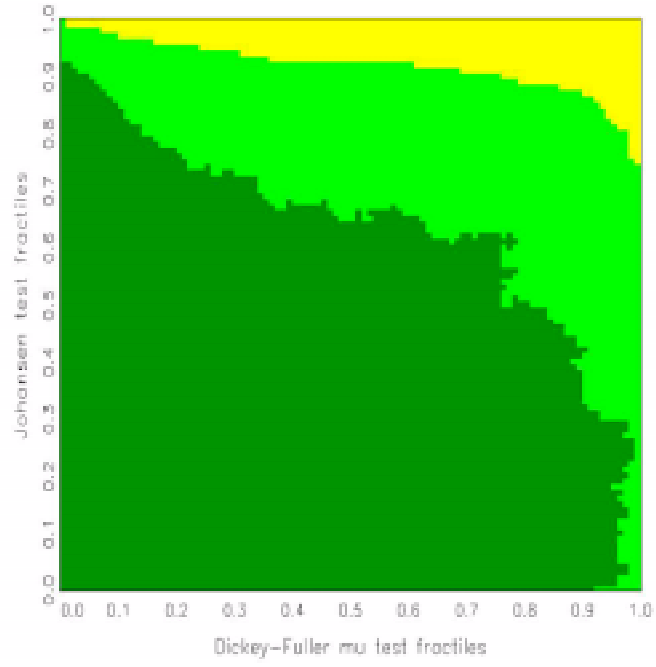

Figure 7: Boundary of optimum decision areas for $T = 100$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. Second-order autoregression. Smoothing constant is 13.

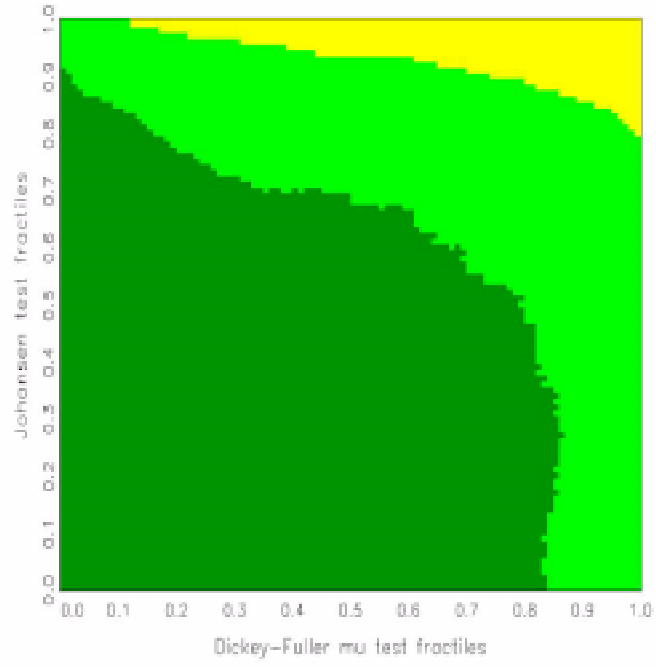

Figure 8: Boundary of optimum decision areas for $T = 200$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. Second-order autoregression. Smoothing constant is 25.

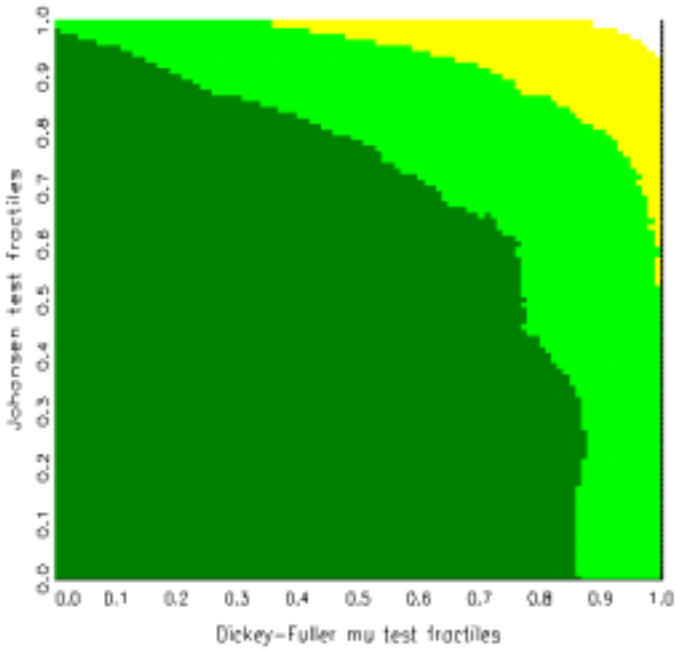

Figure 9: Boundary of optimum decision areas for $T = 100$. Dickey-Fuller μ on the x–axis, Johansen on the y–axis. Second-order autoregression with random drift. Smoothing constant is 25.

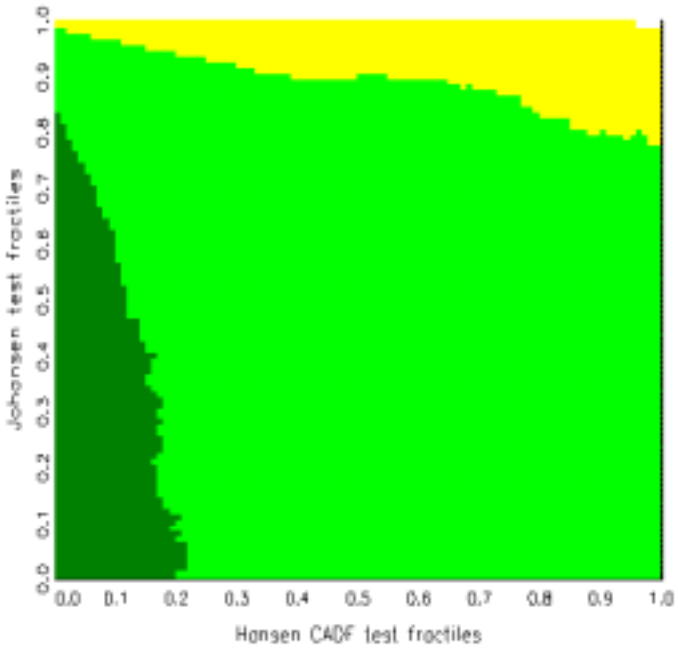

Figure 10: Boundary of optimum decision areas for $T = 50$. Hansen's CADF statistic on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 19.

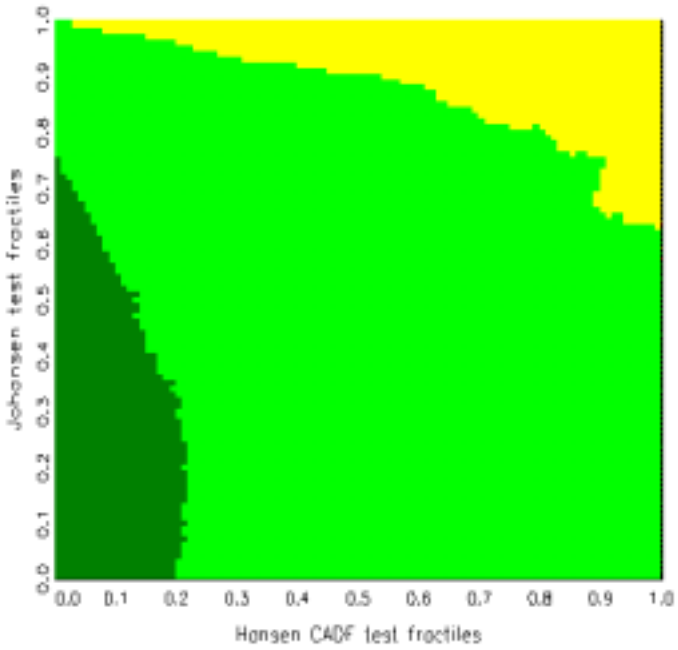

Figure 11: Boundary of optimum decision areas for $T = 100$. Hansen's CADF statistic on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 35.

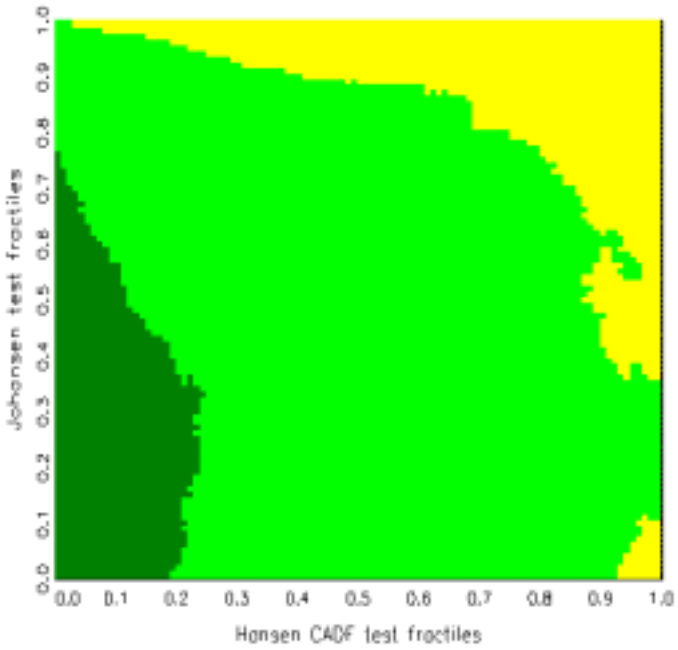

Figure 12: Boundary of optimum decision areas for $T = 200$. Hansen's CADF statistic on the x–axis, Johansen on the y–axis. First-order autoregression. Smoothing constant is 35.