Kunst, Robert M.

Working Paper

Decision maps for bivariate time series with potential threshold cointegration

Reihe Ökonomie / Economics Series, No. 121

Provided in Cooperation with:
Institute for Advanced Studies (IHS), Vienna

Suggested Citation: Kunst, Robert M. (2002) : Decision maps for bivariate time series with potential threshold cointegration, Reihe Ökonomie / Economics Series, No. 121, Institute for Advanced Studies (IHS), Vienna

This Version is available at:
https://hdl.handle.net/10419/71243

121

Reihe Ökonomie

Economics Series

Decision Maps for Bivariate Time Series with Potential Threshold Cointegration

Robert M. Kunst

121

Reihe Ökonomie

Economics Series

Decision Maps for Bivariate Time Series with Potential Threshold Cointegration

Robert M. Kunst

September 2002

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Robert M. Kunst
University of Vienna
and
Institute for Advanced Studies
Department of Economics
Stumpergasse 56, A-1060 Vienna, Austria
☎: +43/1/599 91-255
email: kunst@ihs.ac.at

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share "work in progress" in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschafts-wissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

Bivariate time series data often show strong relationships between the two components, while both individual variables can be approximated by random walks in the short run and are obviously bounded in the long run. Three model classes are considered for a time-series model selection problem: stable vector autoregressions, cointegrated models, and globally stable threshold models. It is demonstrated how simulated decision maps help in classifying observed time series. The maps process the joint evidence of two test statistics: a canonical root and an LR--type specification statistic for threshold effects.

Keywords

Model selection, Bayes testing, nonlinear time series models

JEL Classifications

C11, C15, C32

Comments
The data on U.S. interest rates are taken from the International Financial Statistics database. The author thanks Manfred Deistler, Elizaveta Krylova, and Sylvia Kaufmann for helpful comments. The usual proviso applies.

Contents

1 Introduction

Bivariate time series data often exhibit features like the pair of retail interest rates in Figure 1. The two component series appear to be closely linked such that deviations from each other are relatively small. For long time spans, the variables move up or down the positive diagonal, with very little memory with respect to this motion. Individually, the series appear to be well described as random walks or at least as first-order integrated processes with low autocorrelation in the differences. Eventually, however, the upward or downward motion seemingly hits upon some outer boundary and is reversed, such that all observations are contained in a bounded interval, in the long run. The bounds of the interval are assumed as unknown. A good example for such pairs of time series data are interest rates, such as saving and loan rates or bill and bond rates, though a deep economic analysis of interest rates is outside the scope of this paper. We just observe that they attain their maximum in phases of high inflation and that their minimum is usually strictly positive.

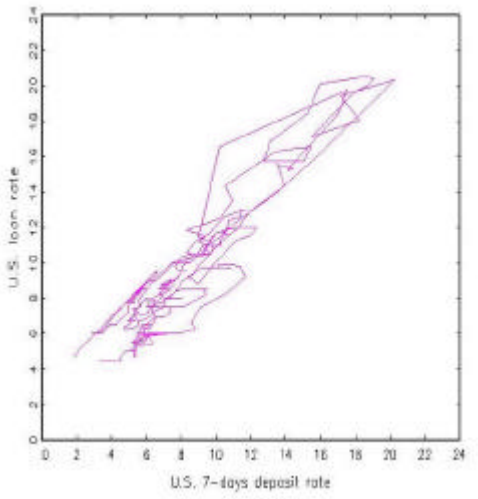

Figure 1: Time-series scatter plot of monthly data on U.S. interest rates on loans and 7-days deposits, 1963–2002.

This paper is concerned with selecting an appropriate time series model for data such as the depicted series, if the set of available classes is given by

the three following ideas. Firstly, the observed reversion to some distributional center may suggest a linear stable vector autoregression. This model class has the drawback that the autoregression is unable to reflect the random movement in the short run. Secondly, this movement and the obvious link between the components may suggest a cointegrated vector autoregression. Particularly for interest rates of different maturity, this is an idea that appears in the econometric literature, see for example CAMPBELL AND SHILLER (1987), HALL *et al.* (1992), or JOHANSEN AND JUSELIUS (1992). ENDERS AND SIKLOS (2001) even state that "it is generally agreed that interest-rate series are I(1) variables that should be cointegrated". The drawback of this model class is that it is unable to match the long-run boundedness condition. The model contains a unit root, is non-ergodic and inappropriate from a longer-run perspective (see WEIDMANN, 1999, for a similar argument in economics). Thirdly, one may consider a mixture of the two models, with cointegration prevailing in a 'normal' regime and global mean reversion in an 'outer' regime. This idea yields a threshold cointegration specification, as it was used by JUMAH AND KUNST (2002), again for interest rates. The drawback of the model is that it is nonlinear and that it contains some poorly identified parameters.

The concept of threshold cointegration is due to BALKE AND FOMBY (1997, henceforth BF) who assumed a version with cointegration in the outer regime and an integrated process without cointegration in an inner regime. JUMAH AND KUNST (2002) suggested a modification of BF's model that is in focus here. Threshold cointegration models were also considered by ENDERS AND GRANGER (1998), ENDERS AND FALK (1998), and ENDERS AND SIKLOS (2001). These contributions focus on asymmetric adjustment to disequilibrium and hence they mainly use two-regime or single-threshold models. HANSEN AND SEO (2002) analyze hypothesis testing for single-threshold cointegration models and assume cointegration in both regimes, though with different cointegrating vectors. Like BF, LO AND ZIVOT (2001) consider the case of three regimes, with cointegration in the lower and upper regimes and no cointegration in the central regime. While these authors allow for asymmetric reaction or different cointegrating structures across regimes, only symmetric reaction outside the band will be considered here, due to the limited information that is provided by the few observations in the outer region of our models. WEIDMANN (1999) analyzed bivariate time series of an inflation index and an interest rates and suggested a three-regimes model that is close to the one used here. Whereas his model assumes different

cointegration structures across regimes and achieves global stability by the choice of cointegrating vectors in the outer regimes, we impose local stability in the outer regimes. TSAY (1998) considered related models in a more general framework. His example of flow data for two Icelandic rivers may also conform to the pattern of Figure 1, as the flow is bounded from below by zero and from above by some natural maximum.

Threshold cointegration models are particular threshold vector autoregressions of the SETAR (*self-exciting threshold autoregression*) type that was considered by CHAN *et al.* (1985), TONG (1990), CHAN (1993), and CHAN AND TSAY (1998). Particularly CHAN *et al.* (1985) found that stability in the outer regimes is sufficient, though not necessary, for global stability and geometric ergodicity. It follows that the models suggested here are globally stable and geometrically ergodic. To the author's knowledge, the proof of this important property has not been given explicitly in the literature. It has been added to this paper as an appendix.

Each of the three outlined model classes deserves attention as a possible data-generating mechanism. It is therefore interesting to study methods that allow selecting among the classes on the basis of observed data. The decision set consists of three elements: the stable vector autoregression, the linear cointegration model, and the threshold cointegration model. Candidates for discriminatory statistics are likelihood ratios for any two of these model classes or approximations thereof. The theory of likelihood ratios between the first and the second class has been developed by JOHANSEN (1988), hence it is convenient to include this ratio in the vector statistic. As another statistic, we add an approximation to the likelihood ratio of the first and the third class.

Model selection is a finite-action problem and requires procedures beyond the dichotomy of the standard Neyman-Pearson framework of null and alternative hypotheses. Here, the three competing model classes are modeled as three alternative Bayes data measures. Each measure can be conditioned on the values of a vector statistic, such that the probability for each hypothesis can be evaluated conditional on the given or observed test statistic. The model or hypothesis with maximum probability is then the suggested choice for the observed value of the statistic. In the space of null fractiles of the test statistics, one obtains 'decision maps' that show distinct regions of preference for each model class. This approach builds on a suggestion by HATANAKA (1996) and was used by KUNST AND REUTTER (2002), among others. For the present problem, non-informative prior distributions are elicited on the

basis of Jordan distributions (see also KUNST, 2002). Discrete uniform priors over the three hypotheses are allotted implicitly. All calculations of decision maps have been conducted by means of simulation. Decision maps are particularly well suited for model selection based on the joint evaluation of two test statistics.

The remainder of this paper is structured as follows. Section 2 explains the decision maps approach and details the properties of the entertained models, including the elicitation of non-informative prior distributions. Section 3 reports the simulation results, including the maps and their tentative interpretation. Section 4 concludes.

2 Designing the simulation

2.1 Decision maps

The decision maps approach can be applied to all parameterized problems $(f_\theta, \theta \in \Theta)$ for finite-dimensional Θ, where a decision is searched among a finite number of rival hypotheses (a partition of Θ), preferably on the basis of a bivariate vector statistic. For a univariate statistic, the maps collapse to intervals on the real line. For vector statistics with a higher dimension, the visual representation of the maps encounters technical difficulties.

Assume a decision concerns the indexed set of hypotheses $\{H_j = \{\theta \in \Theta_j\}$, $j = 1, \ldots, h\}$. Usually, model selection utilizes $h-1$ likelihood-ratio statistics $S_{1(j),2(j)}$ or approximations thereof, for 'null' hypothesis $H_{1(j)}$ and alternative $H_{2(j)}$, with $1 \leq j < h$, $1(j) \neq 2(j)$, and $1(j)$, $2(j) \in \{1, \ldots, h\}$. For ease of notation, let these statistics be collected in an $h-1$–dimensional vector statistic $S = (S_1, \ldots, S_{h-1})'$, such that S_j and $S_{1(j)2(j)}$ can be used equivalently. The classical approach requires nested hypotheses, such that $\Theta_{1(j)} \subset \bar{\Theta}_{2(j)}$, where bars denote topological closure. If parameter sets can be completely ordered, one may write $\Theta_j \subset \bar{\Theta}_{j+1}$ for $1 \leq j < h$. Then, two typical choices of vector likelihood-ratio statistics are $S = (S_{1,2}, S_{2,3}, \ldots S_{h-1,h})'$ and $S = (S_{1,h}, S_{2,h}, \ldots, S_{h-1,h})'$.

Let weighting *prior distributions* be defined on each Θ_j, $1 \leq j \leq h$ by their densities π_j. For any pair $(1(j), 2(j))$, $1 \leq j < h$, the collection of distributions $\left(f_\theta, \theta \in \Theta_{1(j)}\right)$ defines a *null distribution* $f_{1(j)}$ of the statistic S_j

via the implied *p.d.f.* of this statistic under θ, denoted as $f_{j;\theta}$ by

$$\int_{\Theta_{j_1}} f_{j;\theta}(x)\pi_j(\theta)d\theta = f_{1(j)}(x) \quad . \tag{1}$$

Note that it is equally possible to define a null distribution $f_{2(j)}$ of the very same statistic. Let $F_{1(j)}$ denote the *c.d.f.* that corresponds to $f_{1(j)}$. Then, the *preference area* for Θ_j is defined by

$$\begin{aligned} PA_j &= \{z = (z_1, \ldots, z_{h-1})' \in (0,1)^{h-1} \mid j = \arg\max_j P\left(H_j|S\right), \\ &\quad z_k = F_{1(k)}\left(S_k\right)\}. \end{aligned} \tag{2}$$

The transformation $F_{1(k)}$ serves to represent the preference areas conveniently in a simplex instead of some possibly unbounded subspace of $\mathbb{R}^{h-1}$. A graphical representation of preference areas is called a *decision map*.

The meaning of (2) can be highlighted by considering its counterpart in classical statistics. Suppose $h = 3$, and two statistics are evaluated. In a $(0,1)^2$–diagram for the fractiles of the null distributions for S_{12} and S_{23}, a given sample of observations defines a point of realized statistics. Classical statistics would base its decision on rejections in a test sequence, for example starting from the more general decision on S_{23}. If one-sided tests are used that reject for the upper tails of their null distributions, the *classical preference area* for hypothesis Θ_3 or H_3 is the rectangle $A_3 = (0,1) \times (0.95,1)$. If Θ_2 is not rejected, test S_{12} will be applied and separates the preference areas for Θ_1, $A_1 = (0,0.95) \times (0,0.95)$, and for Θ_2, $A_2 = (0.95,1) \times (0,0.95)$. Classical statistics may face difficulties in uniquely determining the null distributions, as fractiles usually vary within each collection. The $(0,1)^2$–chart split into the three rectangles constitutes a *classical decision map*.

Bayesian decision maps are more complex than classical decision maps, as the boundaries among the preference areas may be general curves. Excepting few simple decision problems, it is difficult to calculate the value of a conditional probability for a given point of the $(0,1)^2$–square in the fractile space. It is more manageable to generate, by numerical simulation, a large number of statistics from the priors with uniform weights across the hypotheses and to collect the observed statistics in a bivariate grid over the fractile space. Within each 'bin', the maximum of the observed values can be evaluated easily. In computer time, the simulation can be time-consuming but requires little more storage than hg^2, where h is the number of considered hypotheses and g is the inverse resolution of the grid, i.e., there are g^2 bins.

In summary, numerical calculation of a decision map consists of the following steps: first, a conveniently large number of replications for each of the two statistics under their respective null distributions are generated; second, a grid of fractiles are calculated from the sorted simulated data; third, both statistics are generated from any of the competing hypotheses, i.e., from the prior distributions π_j. These statistics are allotted into the bins as defined by the fractiles grid. Finally, each bin is marked as 'belonging' to that hypothesis from which it has collected the maximum number of entries.

The simulation of the null fractiles requires a number of replications that is large enough to ensure a useful precision of the fractiles. For the purpose of the graphical map, a high precision may not be required. If the map is intended for usage in later applications, it may also be convenient to replace the exact null distribution by an approximation, particularly if that approximation is a standard distribution that allows an evaluation in a closed form. If the null fractiles indeed have to be simulated, a large number of replications slows down the simulation considerably due to the sorting algorithm that is applied in order to determine empirical fractiles.

By contrast, a fairly large number of replications can be attained for the simulation of statistics conditional on the respective hypotheses. Computer time is limited by the calculation time of the statistics only, which may take time if some iteration or non-linear estimation is required, while only a $g \times g$ matrix of bins for each hypothesis is stored during the simulation. For $g = 100$, 10^6 replications yield acceptable maps. For h hypotheses, this gives $10^6 \cdot h$ replications. It was found that kernel smoothing of the bin entries improves the visual impression of the map more than considerably increasing the number of replications (see Section 2.4).

2.2 The model hypotheses

For model selection, prior distributions with point mass on lower-dimensional parameter sets are required. Then, flat or Gaussian distributions are used on the higher-dimensional sets, with respect to a convenient parameterization. The specific requirements of decision map simulations rule out improper or Jeffreys priors, hence the priors do not coincide with the suggestions for unit-root test priors in the literature (see Bauwens *et al.*, 1999). They are maybe closest to the reference priors of Berger and Yang (1994), that peak close to the lower-dimensional sets and are relatively flat elsewhere.

The *first* model H_1 is the *stable* vector autoregression

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \begin{pmatrix} x_{t-j} \\ y_{t-j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{3}$$

with all zeros of the polynomial $Q(z) = \det(\mathbf{I} - \sum_{j=1}^{p} \boldsymbol{\Phi}_j z^j)$ outside the unit circle. For $p = 1$, this condition is equivalent to the condition that all latent values of $\boldsymbol{\Phi} = \boldsymbol{\Phi}_1$ have modulus less than one. This again is equivalent to the property that $\boldsymbol{\Phi}$ has a Jordan decomposition

$$\boldsymbol{\Phi} = \mathbf{T}\mathbf{J}\mathbf{T}^{-1} \tag{4}$$

with non-singular transformation matrix $\mathbf{T}$ and 'small' Jordan matrix $\mathbf{J}$. If one restricts attention to real roots and to non-derogatory Jordan forms, the matrix $\mathbf{J}$ is diagonal with both diagonal elements in the interval $(-1, 1)$. Therefore, the prior distribution for this model π_1 can be simply taken from the family of Jordan distributions and is defined by

$$\begin{aligned} t_{12}, t_{21} &\sim N(0,1) \\ t_{11} = t_{22} &= 1 \\ j_{11}, j_{22} &\sim U(-1,1) \\ \boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})' &\sim N(0, \mathbf{I}_2), \end{aligned} \tag{5}$$

where the notation $\mathbf{J} = (j_{kl})$ etc. is used. The concept can be extended easily to the case $p > 1$ and to non-zero $\boldsymbol{\mu}$. For the basic experiments, $p = 1$ and $\boldsymbol{\mu} = 0$ is retained. Unless otherwise indicated, all draws are mutually independent. For example, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_s$ are independent for $s \neq t$ in all experiments, thus assuming strict white noise for error terms.

This prior distribution is not exhaustive on the space of admissible models. It excludes derogatory Jordan forms and complex roots. Derogatory Jordan forms are 'rare' in the sense that they occupy a lower-dimensional manifold. Complex roots are covered in an extension that was used for some experiments. In this variant, 50% of the $\boldsymbol{\Phi}$ matrices were drawn from $\boldsymbol{\Phi} = \mathbf{T}\mathbf{J}_c\mathbf{T}^{-1}$ instead of (4), with

$$\mathbf{J}_c = \begin{pmatrix} r\cos\phi & r\sin\phi \\ -r\sin\phi & r\cos\phi \end{pmatrix}. \tag{6}$$

It is known from matrix algebra that $\mathbf{J}_c$ is obtained from an original diagonal 2×2–matrix $\mathbf{J}$ with conjugate complex elements that is transformed by a complex matrix

$$\mathbf{T}_c = \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix} \tag{7}$$

such that $\mathbf{J}_c = \mathbf{T}_c \mathbf{J} \mathbf{T}_c^{-1}$. The prior distribution for $\mathbf{J}_c$ is constructed by drawing r from a $U(0,1)$ distribution and ϕ from a $U(0,\pi)$ distribution. The specification for the priors for $\mathbf{T}$ is unchanged. For these experiments with complex roots, also the prior distribution for H_3 was modified accordingly.

In H_1, both variables x and y have a finite expectation and revert to it geometrically. In geometric terms, the (x,y)–plane has a unique equilibrium point $(\bar{x}, \bar{y})$, which in the case $\boldsymbol{\mu} = 0$ collapses to $(0,0)$.

The *second* model class H_2 consists of *cointegrating* vector autoregressions, with the cointegrating vector defined as the difference of the two variables $(1,-1)$. For interest rates of different maturity, this difference is the 'yield spread'. For saving and loan rates, it is the mark-up of banks. For first-order vector autoregressions of this type, a prior distribution is obtained from the error-correction representation

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \boldsymbol{\mu} + \boldsymbol{\Pi} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{8}$$

with a matrix $\boldsymbol{\Pi}$ of rank one. The matrix $\boldsymbol{\Pi}$ can be represented in the form

$$\boldsymbol{\Pi} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1,-1) \quad . \tag{9}$$

The elements α_1 and α_2 are chosen in such a way that explosive modes in the system are avoided. Again, this condition is more readily imposed on the Jordan representation $\boldsymbol{\Pi} = \mathbf{T}\mathbf{J}\mathbf{T}^{-1}$ with diagonal matrix $\mathbf{J} = \mathrm{diag}\,(\lambda, 0)$ for $\lambda \in (-2, 0)$. The specification

$$\mathbf{T} = \begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix} \tag{10}$$

with $a \sim N\,(0,1)$ covers a wide variety of admissible matrices $\boldsymbol{\Pi}$. The implied form of $\boldsymbol{\Pi}$ is

$$\boldsymbol{\Pi} = \frac{1}{1-a} \begin{pmatrix} \lambda & -\lambda \\ a\lambda & -a\lambda \end{pmatrix} \tag{11}$$

and satisfies the general form (9). The class H_2 again assumes $\varepsilon_{jt} \sim N(0,1)$ for the disturbances and, in the basic specification, $\boldsymbol{\mu} = 0$. The thus defined prior π_2 is more difficult to generalize to higher p than π_1.

The *third* model class H_3 are *threshold cointegrating* models. The basic form of these models is

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \boldsymbol{\mu} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} (1,-1) \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \end{pmatrix} (1,0) \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} I\{|x_{t-1} - \xi| > c\} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (12)$$

The symbol $I\{.\}$ denotes the indicator function on the set $\{.\}$. There are two cointegrating vectors. The first one, $(1,-1)$, is always active whereas the second one, $(1,0)$, is only activated at 'large' values of the trigger variable, in our case x_{t-1}. An obvious variant is obtained by replacing the second vector by $(0,1)$ and the trigger variable by y_{t-1}. We do not focus on the choice of the trigger variable, nor do we consider a variation of the trigger lag ('delay'), as is common in the literature on threshold time series models.

The model is ergodic and both variables x and y have finite expectation (see Appendix). The typical behavior is obtained if the mean implied by the 'outer' linear regime

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \boldsymbol{\mu} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (13)$$

is contained in the set $C = \{|x_{t-1} - \xi| < c\}$. Then, the mean is targeted for 'large' values of x and, because of cointegration, also for large values of y, that imply large values of x. If the band C is reached, i.e., if x is 'small' again, the 'outer' mean is no longer interesting. Instead, the dynamic behavior of the variables resembles cointegrated processes, until the band is left and the cycle starts anew. Whenever the 'outer' mean falls outside the band, typical trajectories will remain near the 'outer' mean for long time spans. Only atypically large errors will shift them into the band, where cointegrated behavior takes over. In the first case, the intersection of C and the generic error-correction vector $\{(x,x)\,|x \in \mathbb{R}\}$ can be regarded as an 'equilibrium', with the error-correction vector possibly suitably shifted up or down by restrictions on $\boldsymbol{\mu}$. In the second case, the implied mean of the outer regime constitutes a further element of the equilibrium or attractor set.

Hence, the threshold model allows for substantial variation in behavior. In concordance with the other models, we do not elicit informative priors

but rather define non-informative reference structures with stochastic parameters. To this aim, we adhere to the following basic principle. Suppose we are given the traditional statistical problem of testing a point value against a finite interval. In that case, we would assume weights of 0.5 for each hypothesis and a uniform prior on the interval for the 'alternative'. Treating the present problem in an analogous manner, we use a normal distribution for ξ and a half-normal distribution for c. These laws are sufficiently flat around 0 to mimic the behavior of the typical uniform and normal laws. However, as a consequence of these assumptions, many processes show trajectories with little indication of threshold behavior. In fact, many trajectories closely resemble those drawn from the first model. Occasionally, 'non-revealing' trajectories occur if the threshold criterion is never activated in the assumed sample length. Statistical criteria cannot be expected to classify such cases correctly. In summary, it may be more difficult to discriminate H_3 from $H_1 \cup H_2$ than to discriminate between H_1 and H_2.

2.3 The discriminatory statistics

In order to discriminate among the three candidate models, two discriminatory statistics were employed. The statistic S_2 is designed to be powerful in discriminating H_1 and H_2. In the notation of section 2.1, it would be labelled S_{21}. S_2 is defined as the smaller canonical root for $(\Delta x_t, \Delta y_t)$ and (x_{t-1}, y_{t-1}). As JOHANSEN (1995) pointed out, this root makes part of the likelihood-ratio test for hypotheses that concern the cointegrating rank of vector autoregressions. If the larger canonical root is zero, (x, y) forms a bivariate integrated process without a stable mode. If only the smaller canonical root is zero, (x, y) is a cointegrated process with a stationary linear combination $\beta_1 x + \beta_2 y$. If also the smaller canonical root is non-zero, (x, y) is a stationary process with all modes being stable. It was outlined above why the rank-zero model is not acceptable for interest rates. Hence, the smaller root is in focus.

The statistic to appear on the x–axis, S_1, is designed to discriminate H_1 and H_3, hence in the notation of section 2.1 it would be labelled S_{13}, though it may also be useful in discriminating H_2 and H_3. S_1 is an approximate likelihood-ratio test statistic for the stationary vector autoregression H_1 versus the quite special threshold-cointegrating model with c and ξ being determined over a grid of fractiles of x. In detail, c is varied from the halved interquartile range to the halved distance between the empirical 0.05 and

0.95 fractiles, with ξ thus assumed in the center of the range. For assumed models of type (12), the error sum of squares is minimized over a grid with step size of 0.05. As $T \to \infty$, this grid should be refined. Conditional on ξ and c, estimating (12) is an ordinary least-squares problem with a corresponding residual covariance, whose log determinant can be compared to that of the unrestricted linear autoregression. The residual under the assumption of model class H_j is denoted by $\hat{\varepsilon}_t^{(j)}$. For $j = 1, 2$, this residual is calculated using the maximum-likelihood estimator. For $j = 3$, the approximate maximum likelihood estimator over the outlined grid is used. The residual covariance matrix is denoted by $\hat{\Sigma}_j = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^{(j)} \hat{\varepsilon}_t^{(j)\prime}$. With this notation, the second statistic is defined as $S_1 = \ln\left(\det \hat{\Sigma}_3\right) - \ln\left(\det \hat{\Sigma}_1\right)$.

Note that S_1 can be positive or negative. $S_1 < 0$ for strong nonlinear threshold effects, as H_3 attains a better fit. $S_1 > 0$ for stable autoregressions. The hypotheses H_1 and H_3 are not nested. The statistic S_1 uses optimizing the likelihood ratio over a limited range, which is the prevalent approach for nonlinearity testing. An alternative is the semi-parametric test by Tsay (1998) that relies on sorting the observations according to the source of nonlinearity. The fully parametric nature of our decision problem suggests the use of S_1, however.

The *null fractiles* were generated as follows. For S_2, Jordan priors were used on a cointegrating autoregression—the classical lower-dimensional 'null' hypothesis H_2. The cointegrating rank was fixed at one, whereas the cointegrating vector was not specified in constructing S_2. The null fractiles differ from those that were tabulated by Johansen (1995), as those were calculated under the hypothesis of multivariate random walks. For S_1, Jordan priors on stable vector autoregressions were used, corresponding to the classical 'null' hypothesis H_1.

2.4 Smoothing

The technique of discretizing the statistics in bins corresponds to a rectangular smoothing kernel in density estimation. Similarly, simulated boundaries of the decision areas often have a rough appearance, even for a high number of replications. It was found that smoothing the original numbers in the bins across neighboring bins is not so reliable as smoothing approximate posterior probabilities for the hypotheses. This effect is likely due to the scaling

of numbers. As a smoothing kernel, an inverse absolute function

$$w(i,j) = \frac{W}{1 + |i - i_0| + |j - j_0|}, \qquad \begin{array}{l} i_0 - n_w \leq i \leq i_0 + n_w, \\ j_0 - n_w \leq j \leq j_0 + n_w \end{array} \tag{14}$$

was used over an $(2n_w + 1) \times (2n_w + 1)$ submatrix of the complete matrix of bins. This submatrix is centered at the location (i_0, j_0), where the value is to be estimated. W is set such that the sum of kernel weights $\sum w(i,j)$ over the submatrix equals one. As a general rule, the number n_w was selected as the minimum number, at which smooth boundary curves were obtained. The maps show some deliberate variation of n_w, in order to demonstrate its influence on the results.

3 Simulation results: the maps

Figures 2 and 3 show decision maps resulting from 3×10^6 simulations of process trajectories of length $T = 50$ with stochastic parameters according to the prior distributions that were described in Section 2.2, i.e., 10^6 simulations for each model. In Figure 2, no smoothing was performed ($n_w = 0$), while $n_w = 3$ was set in Figure 3. In color coding, the preference area for the stationary autoregression H_1 is indicated in yellow (light gray), the one for the cointegrated autoregression H_2 in blue (dark gray), and for the threshold cointegrated model H_3 in red (medium gray). One sees that the main effect of smoothing is a better separation of the preference areas for H_1 and H_3, which is mainly achieved by eliminating the scattered preference specks for the rather inhomogeneous hypothesis H_3. Further increases of n_w distort the main boundaries of H_1 and H_2 and of H_1 and H_3.

The main features of the decision map are somewhat surprising. The threshold statistic S_1 appears to be valuable in discriminating cointegrated and stable models, while it was designed to point out threshold structures. The Johansen-type statistic S_2 separates linear cointegrated models from cases of threshold cointegration, while it was designed to test for potential cointegration in stable vector autoregressions. A closer look reveals that H_3 models are indeed characterized by small values of S_1, as expected, and hence are most numerous in the left part of the chart. However, H_2 models also incur small values of the statistic S_1, as a threshold structure with the critical value pushed away from the starting values will achieve a better fit

to data than an unrestricted vector autoregression for cointegrated models. In other words, linear cointegration results as a limiting case of threshold cointegration. The posterior probability for H_2 is more concentrated than that for H_3, hence H_2 dominates the left part of the diagram. Similarly, rejection of the smaller canonical root being zero may point to a stable linear model without unit roots but it may also point to a threshold model, which is stationary and ergodic though non-linear. The joint evidence of two non-zero canonical roots and a better fit by a restricted structure yields the preference area for H_3 in the north-west.

The map implies a crude empirical guideline. First, use the threshold statistic S_1. For values larger than the 0.3 fractile, a linear stable model is suggested. For smaller values, use the lesser canonical root S_2. If this root is 'significant' at 0.05, consider the threshold model, otherwise opt for linear cointegration.

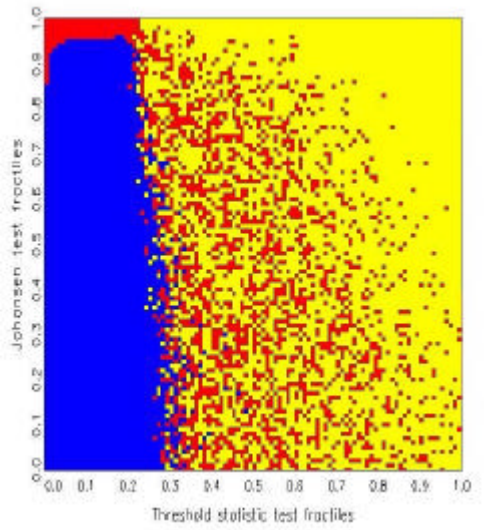

Figure 2: Decision map for processes without deterministic part. $T = 50$, $n_w = 0$. The blue (dark) area prefers H_2, the yellow (light) area prefers H_1 and the red (medium) area prefers H_3. The same color code is used for the other figures.

In Figure 4, the sample size has increased to $T = 100$, while the other simulation parameters were retained. The smoothing bandwidth was kept at

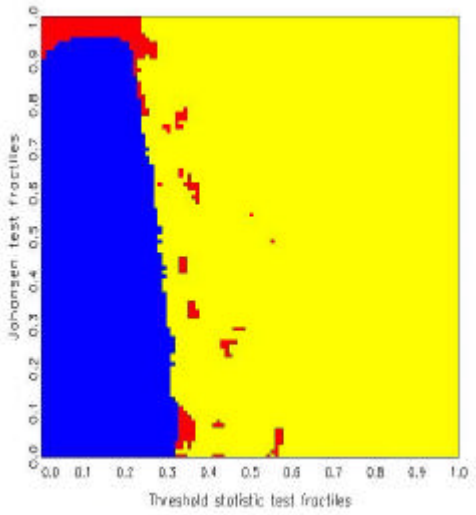

Figure 3: Decision map for processes without deterministic part. $T = 50$, $n_w = 3$.

$n_w = 3$. The nominal significance level of the canonical root has decreased to 0.02, whereas the critical fractile of the threshold statistic decreases to 0.2. These features are in line with expectations regarding large-sample performance. There is now more evidence on a preference for threshold models in a wedge between the other two models, i.e., for threshold statistics in the fractile range $(0.2, 0.3)$, particularly if S_2 is 'not too small'. The scattered appearance of the H_3 area reflects the varying shape of trajectories of length $T = 100$, which generally does not permit a safe classification. Many of these H_3 trajectories are indeed very similar or identical to H_1 trajectories from the stable model class.

Figure 5 shows the map for $T = 200$. The critical fractile of S_1 for $H_1 \cup H_3$ versus H_2 decisions shifts in to 0.12, while the implied significance level of S_2 for H_2 versus H_3 decisions has fallen to 0.01. The boundary is flanked by an unconnected preference area for H_3. Metaphorically speaking, it looks as if the dark H_2 curtain were drawn from a window that reveals a landscape that is populated by H_1 as well as H_3 models. Due to the large variation in appearance of H_3 trajectories, the typical shape of the decision map is likely to persist in even much larger samples. As a consequence, the empirical guideline for sample sizes around $T = 200$ is to first have a look at

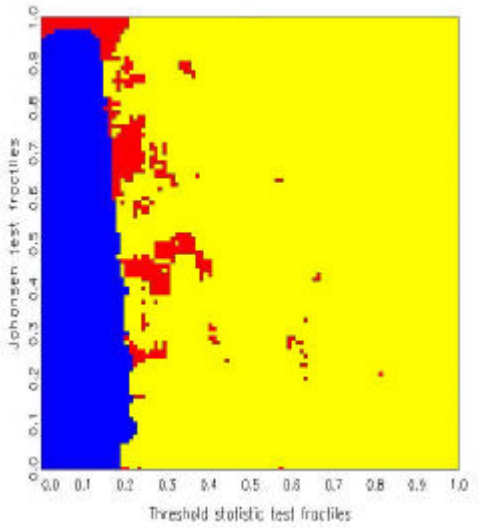

Figure 4: Decision map for processes without deterministic part. $T = 100$, $n_w = 3$.

the S_1 statistic. If it is situated between the 0.12 and 0.2 fractiles of its null distribution, H_3 or H_1 deserve consideration. Additional information from subject matter theory may help in the decision between these two classes. If S_1 is larger than its 0.2 null fractile, H_1 is recommended. If S_1 is less than its 0.12 null fractile, S_2 should be consulted. Highly significant values of S_2 imply a preference for H_3, otherwise H_2 is selected.

For the smallest sample size considered, $T = 50$, Figure 6 shows the map for a variant with standard normal constants included in the data generation mechanism. The assumption $\boldsymbol{\mu} = \mathbf{0}$ in (3), (6), (10) was replaced by $\boldsymbol{\mu} \tilde{} N(0, \mathbf{I}_2)$. The existence of a constant was also assumed in calculating the statistics S_1 and S_2. The effects of this intercept are different for each hypothesis. In H_1, only the mean is affected. In H_2, a linear trend is added. In H_3, a linear trend is generated within the inner regime. These asymmetric effects tend to simplify decisions between the model classes. Hence, the vertical boundary of the H_2 preference area is to the left of the comparable one in Figure 3. The critical fractile for this decision is around 0.1. The preference area for threshold models H_3 has grown considerably and shows a connected pattern in the upper part of the map between the 0.1 and 0.3

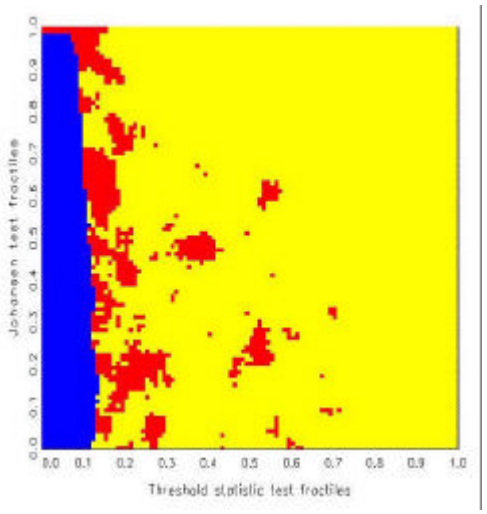

Figure 5: Decision map for processes without deterministic part. $T = 200$, $n_w = 3$.

fractiles of S_1. The influence of the canonical root S_2 on the decision has disappeared with respect to the H_2 class and is rather secondary for the H_1 versus H_3 decision. As a rough guideline, this map suggests that threshold cointegration should be considered whenever S_1 is between its lower decile and lower quartile and S_2 is not in its lower tail. It is difficult to explain the H_3 preference in the north-east corner or the S–shape in the right boundary of the main H_3 preference area. These features may be caused by specific properties of the prior distributions or may be artifacts.

If $T = 100$, the map of Figure 7 is obtained. The critical fractile of S_1 for the decision $H_1 \cup H_3$ versus H_2 has shifted in to about 0.05. By contrast, the boundary between the preference areas for H_1 and H_3 has hardly changed. The effect of the 'drawn curtain' is felt again, such that the H_3 preference area stretches down to the x–axis, i.e., to low S_2 values. The spot in the north-east persists.The asymptotic behavior suggested by Figure 7 is corroborated for $T = 200$ in Figure 8. The critical fractile for the S_1 statistic and the H_2 versus $H_1 \cup H_3$ areas decreases to 0.03, whereas the boundary between the hypotheses H_1 and H_3 remains in place. In this setup, hypothesis H_2 corresponds to the only transient model, which simplifies its detection in larger samples. Contrary to what may be an intuitive assumption, discrimi-

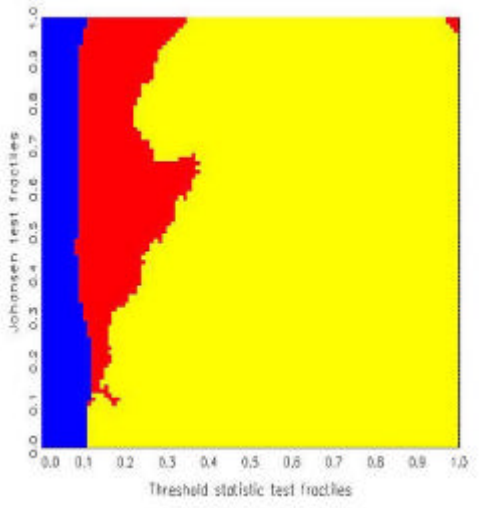

Figure 6: Decision map for processes with standard normal constant term. $T = 50$, $n_w = 3$.

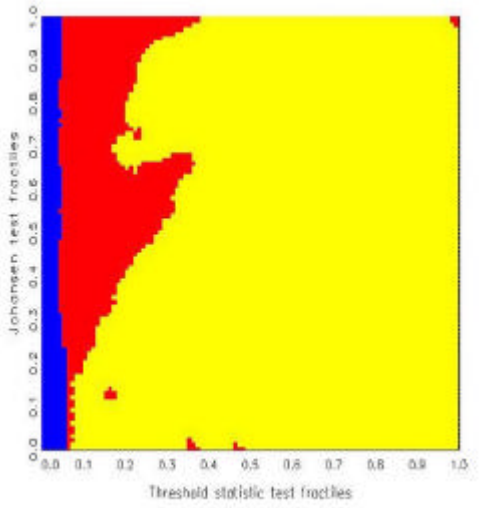

Figure 7: Decision map for processes with standard normal constant term. $T = 100$, $n_w = 3$.

nating threshold structures from linear autoregressions is not automatically simplified as $T \to \infty$. Note, however, that the resolution of the grid and the range for the grid search were kept constant. For asymptotic properties such as consistency, such parameters should depend on T.

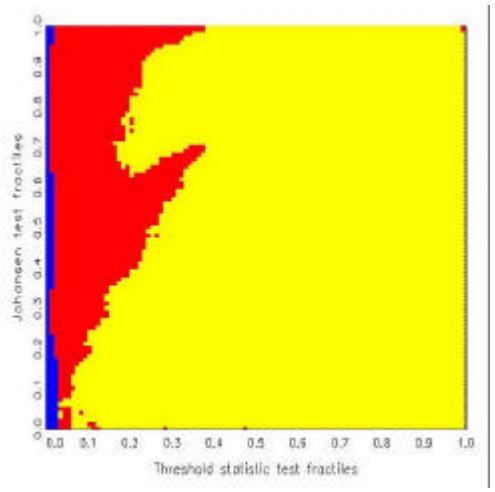

Figure 8: Decision map for processes with standard normal constant term. $T = 200$, $n_w = 3$.

These deterministic specifications were varied and further variations in the maps were obtained that are not reported here. It was found that the two reported variants—without deterministic part and with constant added to the basic dynamic model—represent important benchmark cases. In the presence of a strong 'drift', the decision according to the nonlinearity statistic S_1 should be regarded as separating $H_1 \cup H_3$ from H_2 and should not be overruled by any value of S_2. By contrast, if drifts can be ruled out *a priori*, a clear vertical boundary separates H_1 from $H_2 \cup H_3$. This classification rule is confirmed by a further experiment that is reported in Figure 9. The difference from Figure 6 is that H_2 processes were generated with a constant that is restricted to a multiple of the loading vector α. JOHANSEN (1995) showed that this restriction implies the absence of drift, thereby removing a part of the asymmetry among the unstable H_2 and the stable H_1 and H_3 in the experiments 6–8. Note that H_2 is still the only unstable case, though the divergence of its probability law as $T \to \infty$ comes at a slower pace.For restricted constants, the behavior of the decision map for $T \to \infty$ can be

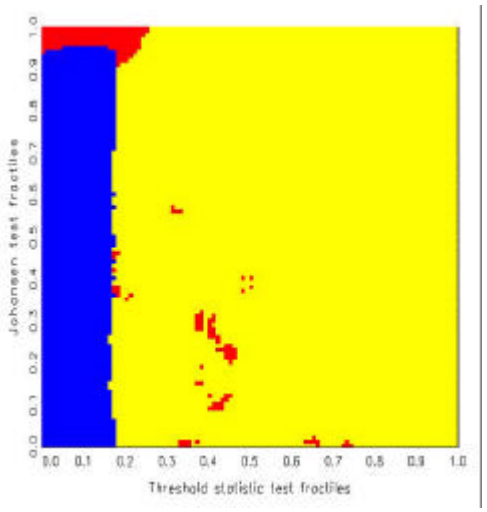

Figure 9: Decision map for processes with standard normal constant term and a no-drift restriction for linear cointegration. $T = 50$, $n_w = 4$.

guessed from the corresponding map for $T = 100$, which is shown as Figure 10. The preference area for H_3 in the northwest corner has shrunk, such that the implicit significance level of S_2 has decreased to 0.01. Scattered H_3 preference specks appear across a rather ample area.

The last variant to be reported concerns the complex-roots modification of the Jordan forms of the priors for H_1 and the outer regime of H_3, which was mentioned in Section 2.2. 50% of the coefficient matrices for hypotheses H_1 and the outer regime of H_3 were drawn from $\mathbf{T}\mathbf{J}_c\mathbf{T}^{-1}$ defined in (6) and 50% from the hitherto used $\mathbf{T}\mathbf{J}\mathbf{T}^{-1}$ with two real roots. For H_2 and the inner band of H_3, one root is fixed at unity, which implies that the other root is real. These designs remain unchanged. The deterministic constant $\boldsymbol{\mu}$ was set at zero, as in the basic experiments shown as Figures 3–5. The differences between Figure 11 and the corresponding map with an all-real design in Figure 4 are due to the cycles caused by the conjugate complex roots in H_1 and H_3. The main effect appears to be the absence of an extended support area for H_3 around the point $(0.2, 0.9)$. A tentative explanation is as follows. The complex roots in H_1 and H_3 increase the average 'distance' between the cointegrated model H_2 and the other models. The increased accuracy of decisions among H_1 and H_2 reduces the chance of the 'compromise' hypothesis H_3 to

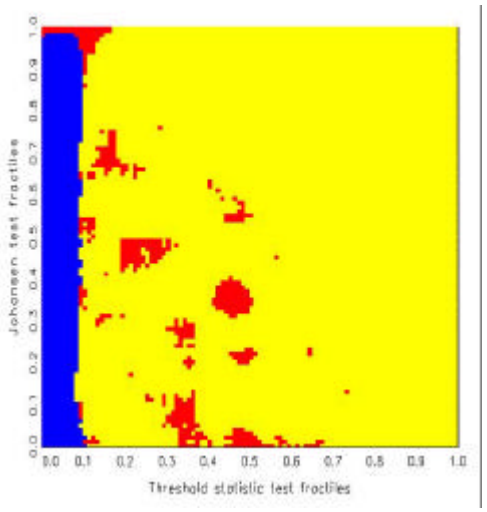

Figure 10: Decision map for processes with standard normal constant term and a no-drift restriction for linear cointegration. $T = 100$, $n_w = 4$.

rule areas of conflict. As in the other maps for models without a constant, the statistic S_1 separates the preference areas for H_1 and $H_2 \cup H_3$, while the statistic S_2 separates H_2 and H_3, conditional on the first-round decision. Several other experiments were run with minor modifications and resulted in only small differences to the reported cases.

For a final note, one may return to the empirical example shown in Figure 1. The generating mechanism of this data set does not correspond exactly to the design of the maps, for two reasons. Firstly, the sample size is $T = 470$, much above the $T = 200$ that was used as a maximum in this section. Secondly, the autocorrelation function suggests a slightly longer memory in the process than a first-order autoregression, even a nonlinear one. A tentative calculation of the null fractiles according to the version with restricted drift constant and of the statistics S_1 and S_2 yields the coordinates $(0.01, 0.95)$. A simulation of the decision map for this large data set yields a leftward shift of the critical fractile for S_1 to around 0.02. Therefore, the evidence favors the model class H_2, i.e., the non-ergodic cointegrated models. If this class is regarded as unacceptable, it may be removed from the set of available alternatives. In that case, class H_3 is supported. Even with all caveats, the exemplary data set confirms the common observation that the actually

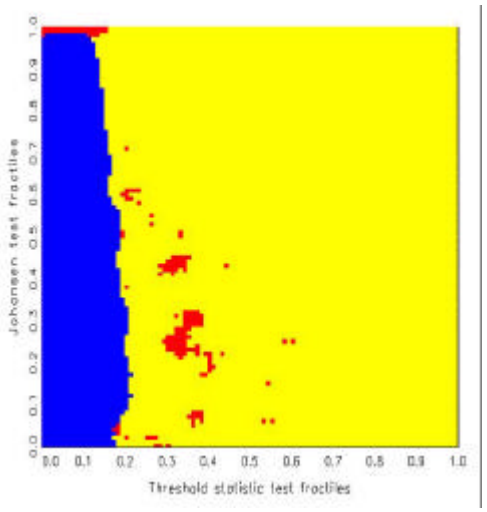

Figure 11: Decision map for processes with zero constant and a 0.5 chance of conjugate complex roots in stable autoregressive coefficient matrices. $T = 100$, $n_w = 3$.

existing boundaries are not hit frequently enough to provide clear evidence of their existence.

4 Summary and conclusion

Decision maps can be useful in their own right. Assuming a researcher analyzes a data set and his or her *a priori* plausible hypotheses correspond to the set-up of the decision map simulation. If the coordinates of the decision map are available, two statistics S_1 and S_2 can quickly be calculated from the data and can be encoded as null fractiles. Otherwise, a simulation may be used for performing the encoding. In the map, preference areas for the conflicting hypotheses are clearly separated by boundaries.

Decision maps are, however, even more important as summary guidelines for the empirical researcher. Vertical or almost vertical boundaries indicate that relying on S_1 is almost as valuable for discriminating the hypotheses as the joint evaluation of S_1 and S_2. Similarly, horizontal boundaries underscore the value of S_2 relative to S_1. In the present experiments, it was found that the main discriminatory power rests on the threshold statistic S_1, not only

among hypotheses H_1 and H_2, i.e., stationarity and cointegration. Correct identification of threshold models turned out to be extremely difficult even for relatively large samples such as $T = 200$. This difficulty is corroborated by tentatively removing H_2 from the set of available hypotheses. The implied map shows that H_2 and H_3 approximately dominate the same area, a vertical band, with H_2 more concentrated there. Formally, the decision maps suggest decisions for H_3 around the boundary between the preference areas for the other two hypotheses, i.e., for S_1 values around the 'critical values'. Another preference area for H_3 is the more 'natural' one in the northwest corner, where S_2 is in the upper tail of its null distribution. The empirical recommendation by some authors (see LO AND ZIVOT, 2001) to test for cointegration in a linear frame first and then to check for nonlinearity is not generally confirmed. Rather, the maps recommend to test for nonlinearity first. Structures with sufficiently large values of S_1 are classified as stationary H_1 models. As a second step, a cointegration test is conducted. If cointegration is rejected for a model classified as 'nonlinear', a threshold model H_3 is indicated. If cointegration cannot be rejected, a cointegrated *linear* model H_2 is supported. Traditional simulation with fixed parametric designs could never unveil this general decision pattern.

The unconnected specks of preference for H_3 reveal that threshold processes generate two species of trajectories: typical trajectories with statistics clustered in the northwest region and atypical trajectories with hardly recognizable threshold effects and statistics almost anywhere in the left part of thee $[0, 1] \times [0, 1]$–plane. Most atypical trajectories stem from designs with small values of c and therefore roughly 'look like' trajectories from stationary autoregressions. The high risk of incorrectly classifying the generating processes as H_1 may incur a relatively modest risk if one proceeds with the incorrect model, as the linear model may be a good workhorse for typical econometric tasks such as prediction. A careful evaluation of this conjecture is a promising task for future work.

Many researchers may be skeptical about the usage of decision maps, particularly when the dynamic specification of short-run nuisance for $H_j, j = 1, \ldots, 3$ is slightly simpler than time-series structures that prevail in the literature. In order to counter this argument, more sophisticated priors must be introduced, which unfortunately entails a considerable increase in computer time. For an example of higher-order autoregressions and elements of lag order search via information criteria within the decision-maps framework, see KUNST (2002). Such extensions are possible directions for future research.

Acknowledgments

The data on U.S. interest rates are taken from the International Financial Statistics database. The author thanks Manfred Deistler, Elizaveta Krylova, and Sylvia Kaufmann for helpful comments. The usual proviso applies.

Appendix: Geometric ergodicity of threshold cointegrated models

The recent econometric literature on threshold cointegrated models offers no formal proof of the stability properties of threshold cointegrated models with stable outer regimes. For the variant that is used as hypothesis H_3 in the paper, such a proof is provided here.

The 'threshold cointegration model' is defined as the nonlinear first-order autoregressive structure

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \alpha\beta' \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} x_{t-1} I\left(|x_{t-1} - x^*| > \delta\right) + \varepsilon_t \quad .$$

For simplicity, no deterministic terms are used except for the x center x^*. The model is equivalent to a stable autoregression for the outer region $\{|x_{t-1} - x^*| > \delta\}$ and to a cointegrating partial stable autoregression for the inner region. We assume the following condition:

A1: The polynomial $\det\{\mathbf{I} - (\mathbf{I} + \alpha\beta')\, z\}$ has no roots inside the unit circle or for $|z| = 1$ but $z \neq 1$. The second-order matrix condition for the Granger representation theorem (see Engle and Granger, 1987, and Johansen, 1995) $\det \alpha'_\perp \beta_\perp \neq 0$ applies, where the subscript $\perp$ denotes the orthogonal complement.

A2: The polynomial $\det\{\mathbf{I} - (\mathbf{I} + \alpha\beta' + \gamma e'_1)\}$ has all roots outside the unit circle.

For the errors ε_t we assume the regularity condition:

A3: The distribution of the errors ε_t is absolutely continuous and strictly positive on $\mathbb{R}^2$.

Conditions A1 and A2 guarantee that the model corresponds to the above concept, with A1 essentially due to Engle and Granger (1987) and to Johansen (1988) and A2 a standard assumption of time series analysis. Given A1, note that A2 restricts γ and excludes $\beta = (1, 0)'$. A3 implies

irreducibility and aperiodicity for all threshold autoregressive models. With these assumptions, the following result by TONG (1990, p. 457) can be applied:

Theorem 1 *(Drift criterion,* TONG*) Let $\{Z_t\}$ be aperiodic and irreducible. Suppose there exists a small set C, a non-negative measurable function g, and constants $r > 1$, $\gamma > 0$, and $B > 0$ such that*

$$\mathrm{E}\{r\, g\,(Z_{t+1})\,|Z_t = z\} < g(z) - \gamma, \quad z \notin C \tag{15}$$

and

$$\mathrm{E}\{g\,(Z_{t+1})\,|Z_t = z\} < B, \quad z \in C. \tag{16}$$

Then, $\{Z_t\}$ is geometrically ergodic.

The small set C is meant to contain the 'center' of the stationary distribution, assuming that such a one exists. For a first-order autoregression with stable coefficient, the condition (15) holds for any subset of $\mathbb{R}^2$ outside the mean, that is, outside a disk around zero if there are no deterministic terms. For the definition of a small set, see TONG (1990, p. 454). A technical complication is to prove that compact sets are small. For all nonlinear autoregressions of the threshold type, this can be shown as in CHAN *et al.* (1985). This result implies the following.

Theorem 2 *Let $Z_t = (x_t, y_t)'$ for $t > 0$ obey the model (1) with the conditions A1–A3 and arbitrary fixed starting conditions. Then, $\{Z_t\}$ is geometrically ergodic.*

Proof: Decompose the space $\mathbb{R}^2$ into five disjoint areas such that $\mathbb{R}^2 = \cup_{j=1}^{5} A_j$. We analyze all of them in turn.

1. $A_1 = \{x_{t-1} < x^* - \delta\}$ The process is locally geometric stable and condition (2) is fulfilled for many functions $g\,(z)$, among them all absolute values of linear functions in the arguments x and y, provided that the implied mean of the autoregression $Z_t = (\mathbf{I} + \alpha\beta' + \gamma e_1')\, Z_{t-1} + \varepsilon_t$ is outside A_1. The maximum eigenvalue λ of the regressor matrix is less than one in modulus because of A2, hence any value λ' in the open interval $(|\lambda|, 1)$ can be chosen for $1/r$. γ can be set to $(\lambda' - \lambda) \min\{g(z)|z \in A_1\}$. If the implied mean μ is inside A_1, the proof must be formulated with respect to $A_1^* = \{x_{t-1} < x^{**} - \delta\}$ for x^{**} being the x–component of μ. The area $A_1^* - A_1$ is appropriately allotted to $A_3 \cup A_4 \cup A_5$.

2. $A_2 = \{x_{t-1} > x^* + \delta\}$ Same as A_1. Again, A_2 may be replaced by A_2^* if necessary.

3. $A_3 = \{|x_{t-1} - x^*| < \delta$ and $|y| < K\}$ with K chosen large enough that the inner equilibrium line segment $\beta'(x, y)'$ is fully contained in A_3. Note that this construction assumes that the second element of β is non-zero, which is excluded by assumption A2, as the system given in A2 and defined by the 'outer regime' cannot become stable if both cointegrating vectors coincide. Clearly, condition (3) is fulfilled. The implication is unaffected by the change from A_1 to A_1^* and the implied change of A_3 to A_3^*. For the fact that A_3 is small under the assumptions A1–A3, we refer to TONG (1990) who states that, for locally linear models with continuous error distribution, all compact sets are small.

4. $A_4 = \{|x_{t-1} - x^*| < \delta$ and $y > K\}$. Defining $g(x)$ as the distance to the equilibrium line segment, for example in the Euclidean metric, yields condition (2) for this area. This function is also valid for A_1 and A_2. The implication is unaffected by the change from A_j to A_j^* for $j = 1, \ldots, 5$.

5. $A_5 = \{|x_{t-1} - x^*| < \delta$ and $y < K\}$. Same as A_4.

A_3 is small in the sense of Theorem 1, which completes the proof.■

Note that Theorem 2 gives no result for the case that the inner regime does *not* cointegrate. In fact, then 'probability mass escapes', as trajectories may wander in the direction of the unrestricted variable y. The result by CHAN *et al.* (1985) does not generalize to the multivariate case immediately, when the inner area is not completely bounded. A similar observation holds with respect to the degenerate case where the equilibrium line segment is vertical in the sense that $\beta = (1, 0)'$. Obviously, the proof is unaffected by changing the signal variable x to y or even to a linear combination of x and y that is different from $\beta' X$.

References

[1] BACKUS, D.K., AND S.E. ZIN (1993) 'Long-Memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates' *Journal of Money, Credit, and Banking* **25**, 681–700.

[2] Balke, N.S., and T.B. Fomby (1997) 'Threshold Cointegration' *International Economic Review* **38**, 627–645.

[3] Bauwens, L., Lubrano, M., and J.-F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

[4] Berger, J.O., and R.Y. Yang (1994) 'Non-informative priors and Bayesian testing for the AR(1) model' *Econometric Theory* **10,** 461–482.

[5] Campbell, J.Y., Shiller, R.J. (1987) 'Cointegration and Tests of Present Value Models' *Journal of Political Economy* **95**, 1062–1088.

[6] Chan, K.S. (1993) 'Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model' *Annals of Statistics* **21**, 520–533.

[7] Chan, K.S., Petruccelli, J.D., Tong, H., and S.W. Woolford (1985) 'A multiple threshold AR(1) model' *Annals of Applied Probability* **22**, 267–279.

[8] Chan, K.S., and R. Tsay (1998) 'Limiting properties of the least squares estimator of a continuous threshold autoregressive model' *Biometrika* **85**, 413–426.

[9] Enders, W., Falk, B. (1998) 'Threshold-autoregressive, median-unbiased, and cointegration tests of purchasing power parity' *International Journal of Forecasting* **14**, 171–186.

[10] Enders, W., and C.W.J. Granger (1998) 'Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates' *Journal of Business and Economic Statistics* **16**, 304–311.

[11] Enders, W., and P. Siklos (2001) 'Cointegration and Threshold Adjustment' *Journal of Business and Economic Statistics* **19**, 166–176.

[12] Engle, R.F. and Granger, C.W.J. (1987) 'Co-integration and Error Correction: Representation, Estimation, and Testing' *Econometrica* **55**, 251–276.

[13] Hall, A.D., Anderson, H.M., and C.W.J. Granger (1992) 'A Cointegration Analysis of Treasury Bill Yields' *Review of Economics and Statistics* **74**, 116–126.

[14] Hansen, B.E. & Seo, B. (2002) 'Testing for two-regime threshold cointegration in vector error-correction models' *Journal of Econometrics*, forthcoming.

[15] Hatanaka (1996) *Time-Series-Based Econometrics: Unit Roots and Co-Integration.* Oxford University Press.

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

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

[18] Johansen, S., and K. Juselius (1992) 'Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for the UK' *Journal of Econometrics* **53**, pages 211–244.

[19] Jumah, A., and R.M. Kunst (2002) 'On mean reversion in real interest rates: An application of threshold cointegration' *Economics Series* No. 109, Institute for Advanced Studies, Vienna.

[20] Kunst, R.M. (2002) 'Testing for stationarity in a cointegrated system' *Economics Series* No. 117, Institute for Advanced Studies, Vienna..

[21] Kunst, R.M., and M. Reutter (2002) 'Decisions on seasonal unit roots' *Journal of Statistical Computation and Simulation* **72**, 403-418.

[22] Lo, M.C. & Zivot, E. (2001) 'Threshold cointegration and nonlinear adjustment to the law of one price' *Macroeconomic Dynamics* **5**, 533–576.

[23] Tong, H. (1990) *Non-linear time series: a dynamical system approach.* Oxford University Press.

[24] Tsay, R.S. (1998) 'Testing and Modeling Multivariate Threshold Models' *Journal of the American Statistical Association* **93**, 1188–1202.

[25] WEIDMANN, J. (1999) 'Modèles à seuil et relation de Fisher: une application à l'économie allemand' *Économie et Prévision* **140/141**, 35–44.

Author: Robert M. Kunst

Title: Decision Maps for Bivariate Time Series with Potential Threshold Cointegration

Reihe Ökonomie / Economics Series 121

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996