IHS Economics Series
Working Paper 16
October 1995

Estimating the Number of Unit Roots: A Multiple Decision Approach

Robert M. Kunst

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

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

Reihe Ökonomie / Economics Series No. 16

ESTIMATING THE NUMBER OF UNIT ROOTS

A MULTIPLE DECISION APPROACH

Robert M. Kunst

Estimating the Number of Unit Roots

A Multiple Decision Approach

Robert M. Kunst

Reihe Ökonomie / Economics Series No. 16

October 1995

Robert M. Kunst
Institut für Höhere Studien
Stumpergasse 56, A-1060 Wien
Phone: +43/1/599 91-255
Fax: +43/1/599 91-163
e-mail: kunst@ihssv.wsr.ac.at

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

Abstract

The problem of detecting unit roots in univariate and multivariate time series data is treated as a problem of multiple decisions instead of a testing problem, as is otherwise common in the econometric and statistical literature. Four examples for such multiple decision designs are considered: first- and second-order integrated univariate processes; cointegration in a bivariate model; seasonal integration for semester data; seasonal integration for quarterly data. In all cases, restrictedly optimum decision rules are found for finite samples based on Monte Carlo simulation.

Zusammenfassung

Das Problem, Einheitswurzeln in univariaten und multivariaten Zeitreihen aufzufinden, wird als Problem multipler Entscheidungen behandelt anstatt als Testproblem, wie es sonst in der ökonometrischen und statistischen Literatur üblich ist. Vier Beispiele solcher multipler Entscheidungsproblem werden berücksichtigt: univariate Prozesse mit Integration erster und zweiter Ordnung; Kointegration in einem bivariaten Modell; saisonale Integration für Semesterdaten; saisonale Integration für Quartalsdaten. In allen Fällen werden beschränkt optimale Entscheidungsregeln für endliche Stichproben durch Monte-Carlo-Simulationen ermittelt.

Keywords

Multiple decisions; unit roots; autoregressive processes.

JEL-Classifications

C22,C32.C49

1. Introduction

Much of the recent literature on the analysis of macroeconomic time series focuses on the problem of making decisions on their degree of non-stationarity (for a good survey of the literature, see Banerjee et al., 1993). Within this framework, researchers are particularly interested in whether the time series at hand has to be differenced once or twice or probably not at all to satisfy the usual assumption of covariance stationarity for the filtered series. Series requiring differencing once are usually called *difference-stationary* or *first-order integrated.* Additionally, in the joint analysis of two or more time series, researchers are interested in whether linear combinations of difference-stationary series may already be stationary. In this case, the linearly combined series are called *cointegrated* (for details, see the seminal paper by Engle and Granger, 1987). Interest in cointegration has been instigated by technical problems as well as by economic theory. With regard to technical matters, it can be shown easily that differencing of cointegrated series leads to extremely inefficient estimation due to loss of information on low frequencies, even though individual series are difference-stationary. With regard to economic theory, evidence on long-run features is often taken as reflecting theoretical considerations on long-run equilibrium relations. Frequently quoted economic examples of this type are the long-run income elasticity of consumption, purchasing power parity, and the joint movement of interest rates with different terms to maturity.

Until the later 1980s, decisions on whether data sets suggest differencing, double differencing, or cointegrating relations were mainly based on the univariate testing procedure developed by Fuller (1976) and elaborated by Dickey and Fuller (1979). Also in multivariate problems, these decisions tended to be based on a primary "cointegrating" regression and secondary residual analysis (see, e.g., Engle and Granger, 1987, and Phillips and Ouliaris, 1990). Johansen (1988) presented an efficient alternative framework for making such decisions. He suggested to determine the number of cointegrating relations by testing sequences and to proceed by conducting conditionally efficient estimation. Pantula (1989) took up the idea of sequential testing for deciding upon the number of unit roots in univariate series.

In consequence, current integration/cointegration analysis is dominated by two main strands of statistical techniques. The first class of methods is characterized by easy handling and inefficiency caused by univariate residual analysis under limited information. The second class of methods relies on full system estimation but is faced with the usual problems of making decisions by sequential hypothesis testing. Alternatively, some researchers have used Bayesian methods, currently still with less impact on economic users. In the tradition of objectivist Bayesian statistics, most of

them relied on continuous prior distributions designed to capture the researcher's lack of information before conducting the experiment. For a survey, see Uhlig (1994).

Here, a comprehensive framework for the problem of estimating the number of unit roots in univariate and multivariate situations is presented. In contrast to the bulk of the literature, this is not seen as a testing but as an estimation problem in the tradition of multiple decisions. In contrast to the Bayesian contributions to the literature, a uniform prior is assumed on the decision parameters leading to mixtures of discrete and continuous distributions on the primary model parameters. Section 2 outlines the formal background. Section 3 presents examples and some evidence on corresponding decision bounds generated by Monte Carlo simulations. Section 4 concludes.

2. Estimating discrete parameters

Wherever possible, we would like to enhance the formal correspondence between the estimation of continuous and of discrete parameters. The reluctance by many researchers to call discrete parameter estimation problems by that name has probably lead to occasional confusion. Often, discrete problems are called "model selection" or "sequential testing". In concordance with, e.g., the work of Hannan and Deistler (1988), we will view all these problems as problems of estimation. In contrast, *testing* problems appear whenever one out of two hypotheses is given the preferred position of a "null hypothesis" and the researcher's loss is asymmetric because of subject matter considerations or of any reasons that permit a formal equivalence to quality control problems.

2.1 The nested problem

We consider the situation that observations are being generated from an unknown element from a collection of distributions characterized by a parameter θ taken from a parameter space Θ. The parameter θ will also be called the *primary parameter*. The observer would like to make decisions on whether $\theta \in \Theta_i, i=0,\ldots,p$. Having made this decision, the user could also be interested in estimating θ within Θ_i. This problem, however, will be set aside within this study.

Although the problem may be considered under more general assumptions, we will focus on two specific situations. In the first case, the model classes (or parameter sets) are ordered by an inclusion sequence

$$\Theta_0 \subset \overline{\Theta}_1, \Theta_1 \subset \overline{\Theta}_2, \ldots, \Theta_{p-1} \subset \overline{\Theta}_p = \Theta \qquad (2.1)$$

(2.1) implies that Θ_j is contained in the topological boundary of Θ_{j+1}, which is occasionally denoted as $\Theta_j \subset \partial\Theta_{j+1}$. Hence, any set Θ_i is "small" relative to all Θ_j with $j>i$. (2.1) will be referred to as the *nested problem*. In all applications considered in Section 3, Θ will be a convex subset of the multidimensional Euclidean space and the closure operator is therefore clearly defined. Note that closure refers to the topology within Θ and not within the Euclidean space and, typically, neither Θ nor any Θ_i will be closed within the Euclidean space. Anyway, there is a natural metric d such that (Θ,d) is a metric space.

In many applications, $\Theta = \Theta_{(1)} \times \Theta_{(2)}$ such that any parameter vector will consist of two parts $\theta = (\theta_1,\theta_2)$ such that θ_1 is restricted to a bounded convex set and θ_2 is not restricted within a multidimensional Euclidean subspace. Cross restrictions are conceivable but the separation is important as a uniform distribution on the subspace $\Theta_{(1)}$ will be constructed. Sometimes, the partition can be attained by a continuous transformation of the parameter space, starting from a given primary parameterization. Then, we consider the transformed parameterization as the "natural" one, assuming that classification of any θ into the Θ_i is independent of θ_2. This convention does not define the parameterization uniquely. Selection of a coordinate system is determined by the practitioner's concern rather than by formal properties. For example, autoregressive models of fixed order have a convenient and natural parameterization if θ_1 consists of the coefficients and θ_2 of a possible mean. Decisions on the number of unit roots can be made based on θ_1, and $\Theta_{(1)}$ is bounded and convex. We could adopt the re-parameterization due to Dickey and Fuller (1979) and thus minimize the dimension of $\Theta_{(1)}$ but this coordinate space is probably less natural. In contrast, the coefficient coordinates of vector autoregressions are not bounded within the Euclidean space, hence for decisions on cointegration a re-parameterization, as e.g. suggested by Johansen (1988), is inevitable.

The choice function defined by

$$\kappa: \begin{cases} \Theta \to \{0,\ldots,p\} \\ \theta \mapsto \kappa(\theta) \text{ if } \theta \in \Theta_{\kappa(\theta)} \end{cases} \tag{2.2}$$

provides us with a discrete parameter $\kappa(\theta)$ that summarizes all interesting information and will be called the *secondary parameter*. All other information is viewed as nuisance information. The metric generated by (Θ,d) and the function $\kappa(.)$ is not useful for the set $\{0,\ldots,p\}$ as it would be trivial due to (2.1). Alternatively, we adopt the logical position of viewing e.g. 3 to be "closer" to 2 than to 1 and we will expressly use the squared distance measure

$$d_K(i,j) = (i-j)^2 \tag{2.3}$$

corresponding to the square of a metric on $\{0,\ldots,p\}$. The researcher is interested in minimizing the distance d_K between his/her estimate of $\kappa(\theta)$ and the true value.

Maybe unfortunately, the observer typically estimates the discrete secondary parameter only indirectly by first estimating the usually continuous primary parameter θ. The estimate for θ is a random variable

$$\hat{\theta}:\begin{cases} (\Omega,\mathcal{A},P)\to\Theta \\ x_n=(X_1,\ldots,X_n)\mapsto\hat{\theta}(x_n)\end{cases} \tag{2.4}$$

as the observations x_n are realizations of a random variable on the indicated probability space. The sample size is n. We assume that the estimator (2.4) is consistent. A good estimator of the primary parameter is certainly crucial for what follows. In most applications, the estimator is some approximation to the maximum likelihood estimator under the information that $\theta\in\Theta$. After making the decision on the secondary parameter, the observer may return to this problem and replace (2.4) by a more efficient estimator under the information that $\theta\in\Theta_i$ for fixed i.

A naive suggestion for constructing an estimator for the secondary parameter would be

$$\hat{\kappa}_N=\kappa(\hat{\theta}) \tag{2.5}$$

This is, however, unattractive because of (2.1). In finite samples, $\hat{\theta}$ usually has a continuous probability density, hence the topological smallness of $\Theta\backslash\Theta_p$ is reflected by the probability measure and $P(\hat{\theta}\in\Theta_p)=1$. Thus, the estimator (2.5) is p with probability 1. This property holds for every finite n, and the estimator is inconsistent.

The inclusion sequence (2.1) has incited many researchers to solve the estimation problem via hypothesis testing. Hypothesis tests are constructed with the null hypothesis H_0:$\theta\in\Theta_i$ and the alternative Θ_{i+1} or $\Theta\backslash\Theta_i$. An estimate for the secondary parameter is obtained by a certain stopping rule in such a testing sequence. There are four methods of this type in current usage:

(i) Test $\Theta_0\cup\ldots\cup\Theta_{p-1}=\overline{\Theta}_{p-1}$ against Θ_p; if rejected stop and $\hat{\kappa}=p$; test $\overline{\Theta}_{p-2}$ against Θ_{p-1}; if rejected stop and $\hat{\kappa}=p-1$; .... ; if no rejection $\hat{\kappa}=0$.

(ii) As in (i) but always test $\overline{\Theta}_i$ against $\Theta\setminus\overline{\Theta}_i$.

(iii) Test Θ_0 against Θ_1; if accepted stop and $\hat{\kappa}=0$; test $\overline{\Theta}_1$ against Θ_2; if accepted stop and $\hat{\kappa}=1$; ... ; if everything rejected $\hat{\kappa}=p$.

(iv) As in (iii) but always test $\overline{\Theta}_i$ against $\Theta\setminus\overline{\Theta}_i$.

The testing sequences (i) and (ii) correspond to the currently favored general-to-specific tests (see, e.g., Yap and Reinsel, 1995). (iii) and (iv) are specific-to-general. (iii) only works if rejection of Θ_j against Θ_{j+1} is guaranteed if Θ_{j+k} holds with $k\geq 2$. Asymptotically these properties are guaranteed by (2.1) and therefore all four testing sequences are consistent in the sense of test consistency. Viewed as estimators for κ, they define *inconsistent* procedures unless $\kappa=p$, provided the significance level for testing is kept fixed. Reducing the significance level to zero asymptotically, one can

define consistent estimators of κ. [1] In samples of typical economic size, (i) and (ii) yield a tendency toward small-sample upward biases and (iii) and (iv) toward downward biases. These tendencies can be triggered by modifying significance levels. All of these popular estimators will be summarily called *testing estimators*.

The distance measure (2.3) can be used to generate a different type of estimators. In analogy to e.g. least-squares estimation, let us assume that the investigator endeavors to minimize the loss function

$$l(\hat{\kappa},x)=\left(\hat{\kappa}\left(x(\theta,\omega)\right)-\kappa(\theta)\right)^2=d_{\kappa}(\hat{\kappa},\kappa) \tag{2.6}$$

The arguments are random variables and the right-hand side in (2.6) is unobserved. However, one could try to minimize expected loss given fixed θ:

$$E_{\theta}l(\hat{\kappa},x)=\int_{\Omega}\left(\hat{\kappa}\left(x(\theta,\omega)\right)-\kappa(\theta)\right)^2 dP_{\theta}(\omega) \tag{2.7}$$

In statistical decision theory, this function is called the *risk function* (see, e.g., Ferguson, 1967). (2.7) is definitely not constant in κ and usually not constant in θ for given κ. Unlike in some classical problems, it is also not possible in general to solve the minimization problem analytically as this would require some knowledge about the small sample distribution of the primary parameter estimate. This turns out to be intractable in most applications. In order to make (2.7) operable in principle, one could try to finally define an estimator

$$\hat{\kappa} \text{ minimizes } EE_{\theta}l(\hat{\kappa},x)=\int_{\Theta}\int_{\Omega}\left(\hat{\kappa}\left(x(\theta,\omega)\right)-\theta\right)^2 dP_{\theta}(\omega)dQ(\theta) \tag{2.8}$$

However, (2.8) requires a definition of a probability measure Q on the parameter space Θ to define a weighting scheme. This will be done here.

Though (2.8) may look like a Bayesian problem, Q is not to be interpreted as a prior distribution reflecting prior beliefs about the parameter θ. It is simply used as a weighting for a decision problem. For such a decision problem among the discrete secondary parameters, it appears logical to attribute the same weight to each of these parameters. In the Bayesian interpretation, this amounts to a non-informative prior over the secondary parameters. In contrast to formal Bayesian analysis, we will not focus on posterior distributions but rather stick to the classical and probably more user-relevant problem of making discrete point decisions.

The uniform distribution across the secondary parameters does not completely specify the distribution over the primary parameters. As was stated before, typically the expected loss depends on θ not only on κ. It remains to define a probability

[1] Many researchers are aware of this problem but deem it to be unimportant for the practitioner (see e.g. Johansen (1995)). Some Bayesians point out the complete consistency of their tests (see Phillips and Ploberger (1994)) achieved by asymptotic reduction of significance levels but rarely view them in a unified framework with their estimation procedures.

distribution or weighting on each $\Theta_i=\kappa^{-1}(\{i\})$. In concordance with the uniform weighting for the secondary parameters, one may consider to define Q as uniform on Θ_i. This seems to be reasonable if Θ_i is bounded and convex [2]. In other cases, the uniform law is not properly defined on Θ_i. Bayesians sometimes use diffuse improperly defined priors but in most practical applications the extreme weight given to unusual "far-away" parameters is unacceptable.

Now consider the case that Q is uniform on $\Theta_{i(1)}$ where $\Theta_i=\Theta_{i(1)}\times\Theta_{i(2)}$ which is bounded and convex. Sometimes a continuous one-to-one transformation is required to achieve such parameterization. This is a good chance to separate θ into the interesting part θ_1 and the uninteresting nuisance remainder θ_2 [3]. If θ_2 is defined on some higher-dimensional product of the real line and does not influence the decision on the secondary parameters in population once θ_1 is known and does not influence the decision in large samples, one could e.g. impose standard normal distributions as weighting schemes for these nuisance parameters. In many applications, we may permit a certain degree of dependence of the expected loss in (2.7) on θ_2 in finite samples.

It is worth while to compare a thus constructed distribution on Θ with prior distributions used in the literature. Firstly, uniform priors are widely avoided as they may produce strange results in some cases and are not invariant to transformations of the coordinates in Θ. This is less of a problem if a "natural" parameterization exists. Secondly, mixed priors are rarely used. In the examples that will constitute our main focus of interest, i.e. autoregressive processes, previous research has given positive weight to parts of the parameter space that are non-admissible a priori, such as explosive processes. The intention of this positive weighting of the non-admissible parameter set may be to draw attention to the admissible boundary a posteriori. In this interpretation, though the zero weighting of an interesting hypothesis and mixing of continuous and discrete distributions is avoided, it may be difficult to see the equivalence between the assumed "prior" and the researcher's true prior if such a one is hypothesized to exist and to be reasonable.

In the following examples, evaluation of optimum decision bounds will be based entirely on Monte Carlo simulation. The complicated metric imposed on the primary parameter space prevents analytical derivations, excepting the simplest case of just two secondary parameters.

Without further restrictions, (2.8) can hardly be solved directly for all possible estimators $\hat{\kappa}$. However, by restricting the considered class of decision rules, conditionally optimal solutions can be found numerically. Under some regularity

---

[2] Generalizations of the property of convexity are conceivable but will not be needed in our practical applications. Usually, local convexity suffices.

[3] Formally, if there is a continuous one-to-one transformation T on Θ such that $T(\theta) = \theta^*=(\theta_1,\theta_2)$, we will consider θ^* as the "natural" parameterization θ of our problem.

conditions - e.g. monotonous likelihood ratios - it can be shown by statistical theory that decision rules based on sufficient statistics and likelihood ratios are optimal in some sense. Not in all of our problems the corresponding criterion statistics are sufficient but it will always be assumed that the practitioner is primarily interested in keeping the decision rules simple. The loss of our optima relative to the unrestricted optima is probably small. However, note that in the following, - in the notation of the nested problem - natural restrictions such as $\Theta_i \subset \tilde{\kappa}^{-1}(i)$ for $0<i<p$ and $\tilde{\kappa}$ defined by $\hat{\kappa} = \tilde{\kappa} \circ \nu \circ \hat{\theta}$ are in general violated. This is the price paid for keeping decision rules simple and well corresponds to classical analysis in the framework of the testing estimator.

2.2 *The multiple binary problem*

In the nested problem, the set of secondary parameters appears to be naturally ordered. This corresponds well to cases where, for example, the number of non-zero or unit eigenvalues in a matrix are estimated. The set of secondary parameters will always be equivalent to a finite sequence of natural numbers, such as $\{0,1,2,...,k\}$, or possibly the whole of $\mathbf{N}_0$. In the *multiple binary* problem, the secondary parameters are k-tuples of binary numbers, such as (0,1,0,1). Formally, the space of secondary parameters is some $\{0,1\}^k$. This corresponds well to problems where k interesting and mutually (logically) independent features are either absent or present in the data. We note that the space of secondary parameters is the *whole* of $\{0,1\}^k$ and all k-tuples will be given the same weight. In other words, the prior weighting will be uniform over all k-tuples. The set of decisions or secondary parameters is also reminiscent of the power set over $\{0,...,k\}$ - note that this is *not* the σ-field used to construct a probability space here but the set of elementary events - or of a Boolean algebra of order k. Therefore, we could also call it the *lattice problem*.

To handle the problem in a similar way to the nested problem, we have to convene a distance measure. Two extensions of the quadratic distance measure are conceivable. Firstly, one may use

$$d_1((a_1,\ldots,a_k),(b_1,\ldots,b_k)) = \sum_{i=1}^{k}(a_i - b_i)^2 \qquad (2.9)$$

All the entries a_i and b_i are either 0 or 1, hence d_1 weights the maximum distance just by k. This corresponds to a linear weighting of large distances and does not appear to penalize large errors sufficiently. We will therefore use

$$d_\kappa((a_1,\ldots,a_k),(b_1,\ldots,b_k)) = \left[\sum_{i=1}^{k}(a_i - b_i)^2\right]^2 \qquad (2.10)$$

Note that these distance measures are not metrics on the parameter spaces as triangle inequalities fail to hold. Simple transforms would be metrics but would be uninteresting for our purposes. However, after forming expectations, metrics on probability spaces can be defined by taking e.g. square roots of the expectation of (2.3) or (2.9) or quartic roots of the expectation of (2.10).

Most observations can be transferred directly from our handling the nested problem to the multiple binary problem. Interestingly, even the classical treatment in the literature has been equivalent. Typically, one of the two cases - the "feature" being present or being absent - is reflected in a "generic" set of primary parameters. The non-generic feature is then used as "null hypothesis" and is "tested" against the generic alternative. Obviously, the main problem is now what should be assumed about the remaining features at different entries of the k-tuple. Classical testing usually chooses the convenient way of testing the non-generic feature at entry i under the (maintained) assumption that the non-generic feature also holds at all entries $j \neq i$ [4]. This design expresses a strong a priori belief in the non-generic features at all entries and can run into severe problems when more than one feature turns out to be generic.

As a viable alternative, we view the multiple binary problem as an *estimation* problem, where the secondary parameter i is estimated such that the expected double squared loss expressed by the distance function d_K is minimized. Again, uniform weighting is assumed in the classes of primary parameters defined by $\kappa^{-1}(i)$, or, in the common presence of unbounded nuisance parameters, uniform weighting on $\Theta_{i(1)}$.

A further generalization to "multiple nested problems" is straightforward. In this case, each feature can appear with the frequency $l \in \{1,...,p_i\}$ or not at all. In some applications, p_i will be constant over all i and the set of secondary parameters will be equivalent to $\{0,...,p\}^k$. Unsurprisingly, even this complicated discrete estimation problem has been handled by sequences of binary tests in the classical literature leading to inconsistent secondary parameter estimates.

3. The examples

3.1 *Order of integration in second-order autoregressions*

We consider the second-order autoregressive model

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t \qquad (3.1)$$

with n.i.d. $(0,\sigma^2)$ errors ε_t. It is well known that all sensible combinations of the parameters (φ_1,φ_2) are situated in and on a triangle flanked by the three lines

[4] F-type restriction tests offer an example for an exception where the maintained hypothesis for $j \neq i$ is the generic feature. This thoroughly analyzed problem is not treated within this paper's framework.

$$\varphi_1+\varphi_2=1 \qquad -\varphi_1+\varphi_2=1 \qquad \varphi_2=-1 \tag{3.2}$$

See our Figure 1 for a geometric interpretation. In the following, we will refer to this triangle as the SODE triangle for "second-order difference equations" whose stability conditions are reflected in it. (see, e.g., Hamilton, 1994). All parameter combinations outside the triangle define anticipative or explosive processes and will therefore be excluded from the investigation. The set of sensible parameters consists of the inner part

$$\Theta_2 = \{(\varphi_1,\varphi_2) \in \Re^2 | \varphi_1+\varphi_2<1, -\varphi_1+\varphi_2<1, \varphi_2>-1\}$$

and the boundary of the triangle. All parameters in Θ_2 define stationary AR(2) processes. The boundary of the triangle defines homogeneous non-stationary processes that are also called *integrated processes*. The maybe best known example $(\varphi_1,\varphi_2)=(1,0)$ is the *random walk*. All parameters on the north-east boundary

$$\Theta_1 = \{(\varphi_1,\varphi_2) | \varphi_1+\varphi_2=1, 0<\varphi_1<2\}$$

define *first-order integrated processes*. These are characterized by exactly one root of +1 in their characteristic polynomial and equivalently by the fact that they become stationary after one first-differencing transformation. The south-east corner point

$$\Theta_0 = \{(2,-1)\}$$

defines a *second-order integrated process*. It is a convolution of a random walk and is the only process of its kind among the AR(2) processes. The other parts of the triangle boundary will be excluded for the moment. They are related to processes with very dominant periodicity, including the "mirror image" of the random walk $X_t=X_{t-1}+\varepsilon_t$. These will be examined more closely in example 3.

Obviously, the design of this problem fulfills our assumptions for a nested problem. The set of secondary parameters is {0,1,2}. In the literature, most authors have used the testing estimator based on approximate or exact ML estimators of the coefficients φ_1 and φ_2. 5% test boundaries were fixed by simulation or numerical integration as the asymptotic distribution of the LR statistic is a known transformation of Brownian motion integrals. As was already outlined, the testing estimator with fixed significance level is inconsistent (see Johansen, 1995, and Pantula, 1989).

To evaluate the asymptotic risk of the testing estimator, one may build on the following good approximation. The exact asymptotic bias can be calculated from the formula given by Johansen (1995). If $\kappa=2$, then the estimator is consistent and the asymptotic loss is 0. If $\kappa=1$, there is a 5% chance of selecting $\kappa=2$ and a 95% chance of uncovering the true value. Asymptotic loss is 0.05/3 because of the uniform prior weights assigned to the three secondary parameters. If $\kappa=0$, there is a probability of 0.05 of incorrect asymptotic "rejections", i.e., selections of different values of κ. Assuming the two "testing" steps to be approximately independent, given $\kappa=0$, the

asymptotic loss becomes $(0.05 \cdot 0.95 + 4 \cdot 0.05^2)/3$, as the sequence of two incorrect "rejections" yields a loss of 4. The total asymptotic risk of the testing estimator is 0.03583... More efficient estimators have to be gauged against this number.

A consistent estimator with an asymptotic risk of 0 is the *Bayes-rule estimator*. In our case, its form would be:

$$\hat{\kappa} = \arg\max_k \int_{\Theta_k} \int_{\Re^n} f_\theta(x) dx d\theta$$

This is a very simple case for applying the Bayes rule as each $\Theta_k = \kappa^{-1}(\{k\})$, $k=0,1,2$, is completely expressed in the two parameters φ_1 and φ_2 and the assumed prior is uniform on Θ_k. The Bayes-rule estimator is consistent and minimizes the risk defined by the more trivial distance function

$$d_B(i,j) = \delta^i_j$$

To attain a minimum to our more complicated quadratic distance measure d_K, we took refuge to some Monte Carlo simulation. This may also help to compare the best decision bound with the optimum achieved by the Bayes-rule estimator [5] and the testing estimator. The idea is that a random walk is "closer" to a stationary process than the double unit root process $X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t$. This view naturally extends our idea that 1 is closer to 0 than 2 is. This "intensity of incorrect decision-making" should be reflected by the distance and the risk function.

Clearly, the requirement of asymptotic zero risk cannot define a decision rule uniquely. On the other hand, the theoretical optimum decision rules for a given finite sample size can be uncomfortably complex. In accordance with practitioners' needs, here *simple* and immediately operable decision rules will be preferred. A class of such simple decision rules is defined by the following design

(1) select $\hat{\kappa}=0$ if $\hat{\varphi}_1 > 2 - b_1$

(2) select $\hat{\kappa}=1$ if $\hat{\varphi}_1 + \hat{\varphi}_2 > 1 - b_2$ and $\hat{\kappa}=0$ is not selected

(3) select $\hat{\kappa}=2$ if neither $\hat{\kappa}=0$ nor $\hat{\kappa}=1$ is selected

The bounds b_1 and b_2 naturally vary with the sample size and converge to 0 for large samples. It is easily seen that the thus defined estimator $\hat{\kappa}$ is consistent if the coefficient estimators are. This means that the estimator of the secondary parameters is consistent if the estimator of the primary parameters is consistent. It is well known that exact maximum likelihood, least squares, and the method-of-moments Yule-Walker rule all define consistent estimators of the primary parameters. In this Monte Carlo study, the least squares estimator is used as it is simple to calculate and therefore much

[5] All simulations were redone with the 0-1 loss function but the differences in optimum solutions were rather small so they are not reported.

in general use. The Yule-Walker estimate is unattractive in nearly non-stationary situations.

Our choice of bounds corresponds closely to the classical solution of the testing estimator. In fact, Fuller (1976) and also some later authors used the estimated coefficients proper, rather than using likelihood-ratio test statistics, for making decisions on whether unit roots are present. Of course, the binary decision problem between Θ_0 and Θ_1 is uninteresting as the Bayes rule defines an easy-to-use estimator that, in this case, also minimizes d_K risk.

TABLE 1: *Monte Carlo bounds for estimating the number of unit roots in a univariate AR(2) model. 10000 replications were conducted*

n	b_1	b_2	risk
100	0.15	0.12	0.0564
200	0.08	0.08	0.0343
500	0.04	0.04	0.0145

Table 1 reports the results from this Monte Carlo simulation. For the smallest sample size n=100, the simulated bounds coincide well with the 5% bounds given in the literature. For larger sample sizes, their slower convergence toward 0 relative to the testing bounds becomes palpable. The bounds correspond to hypothesis tests with different size but nevertheless the achieved minimum risk may serve as a guideline in roughly suggesting that, in the absence of tables such as our Table 1, for n=500, decisions should be based on 2.5% rather than on 5% significance bounds.

3.2 *Rank of cointegrating matrix*

The first example can also be seen as estimating the rank of a certain matrix in the state-space transition form

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

$$\vec{X}_t = \mathbf{T}\vec{X}_{t-1} + (\varepsilon_t \quad 0)'$$

If the state-space transition matrix **T** has all its eigenvalues smaller than 1, the autoregressive process is stationary. If it has exactly one eigenvalue equal to 1, the process is first-order integrated. The process is second-order integrated if both eigenvalues of **T** are equal to unity. One could also think of considering the form

$$\Delta\vec{X}_t = (\mathbf{T} - \mathbf{I})\vec{X}_{t-1} + (\varepsilon_t \quad 0)'$$

Here, Θ_k corresponds to the matrix **T-I** having rank k. The estimation problem of the secondary parameter becomes equivalent to *estimating the rank* of a stochastic matrix.

A similar problem evolves in truly multivariate time series analysis. Consider the so-called error-correcting representation of a bivariate first-order vector autoregression (see Engle and Granger, 1987)

$$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = (\mathbf{\Phi} - \mathbf{I}) \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Here, the rank of the matrix $\mathbf{\Phi}$-$\mathbf{I}$ is particularly interesting. If it is 0, the processes X_t and Y_t are dynamically unrelated random walks. If it is 2, the bivariate process is stationary, provided all roots ξ of $\det(\mathbf{I}-\mathbf{\Phi}z)$ with $|\xi|<1$ or $|\xi|=1$ but $\xi \neq 1$ are excluded. If it is 1, the individual processes are first-order integrated but there is a stationary linear combination. Economists frequently interpret this stationary *cointegrating vector* as expressing a long-run equilibrium relationship in the system. Obviously, this is again a nested problem in the sense of our definition.

In order to elicit a weighting measure on a parameter space Θ, the notion of a generic event is needed. Very informally, we define a *generic event* by being true on a subset of Θ that is so large as compared to Θ that any reasonable mass distribution is likely to assign a probability mass of one to the subset. We deliberately do not give the definition based on continuous mass distributions, as we would like to allow for discontinuities where they are logically sensible. In our sense, a generic event is defined partly by its mathematical and partly by its inherent logical properties. The opposite case, which is logically assigned a mass of zero, will be called a *non-generic event*.

In this problem, primary parameters are the elements of $\mathbf{\Phi}$ or $\mathbf{\Phi}$–$\mathbf{I}$ or some other parameterization enhancing the eigenvalues of $\mathbf{\Phi}$–$\mathbf{I}$. Note that admissible eigenvalues of $\mathbf{\Phi}$ lie in (-1,1), hence admissible eigenvalues of $\mathbf{\Phi}$–$\mathbf{I}$ are in (-2,0), with the borderline case 0 corresponding to unit roots. In this example, it is not so "natural" to choose a certain parameterization and therefore it is not so easy to find the correct weighting scheme with respect to Θ. Here, the following idea was selected. The matrix $\mathbf{\Phi}$–$\mathbf{I}$ can be expressed via its Jordan canonical form:

$$\mathbf{\Phi} - \mathbf{I} = L \begin{bmatrix} \lambda_1 & \delta \\ 0 & \lambda_2 \end{bmatrix} L^{-1}$$

For "most" matrices, the element δ is 0. $\delta=1$ for some matrices with $\lambda_1=\lambda_2$. For the design of the following simulation study, we assume that this case plays little role. It is probably not very costly to exclude an event such as $\lambda_1=\lambda_2$ anyway as it is non-generic. A further difficulty is much more important. The Jordan canonical form is only valid in general if complex eigenvalues are admitted. For real matrices, these must be complex conjugates. The Jordan matrix can be represented in an all-real form but we chose to exclude complex eigenvalues altogether. In examples 3 and 4, this problem will be taken up again. Some cursory simulations allowing for complex conjugate eigenvalues proved that the results are not sensitive to our all-real design.

In this bivariate model, note that complex conjugates appear to imply $|\lambda_1|=|\lambda_2|$, i.e., another non-generic event. However, this kind of argument is probably faulty. The SODE triangle (Figure 1) shows a bottom area bordered by a dashed parabola corresponding to conjugate complex eigenvalues. This lower part covers two thirds of the entire area. Note, however, that it touches upon Θ_1 only close to the corner point Θ_0. The discrimination of complex-rooted stationary processes from unit-root processes is probably a minor problem as compared to "average" real-rooted cases. In higher-dimensional models, this restriction may be more critical.

The matrix L in the Jordan decomposition can be any matrix provided it is non-singular. It is not uniquely determined. To reduce the effect of the non-uniqueness with respect to scaling, the innocuous normalization $l_{ii} = 1$, i=1,2, is imposed. The off-diagonal elements are allowed to take on any real value as long as these values do not succeed in making L singular, which again is a non-generic event but was not excluded a priori. The primary parameters l_{ij}, $i \neq j$, are treated as unbounded nuisance of the type θ_2 and are weighted according to a standard normal distribution. In later experiments, it may be interesting to vary this weighting on θ_2 and evaluate the sensitivity. We presume that the θ_2 weighting is unimportant.

To check on the rank of **Φ-I**, a decision criterion could rely on the eigenvalues λ_1, λ_2, $|\lambda_1| \leq |\lambda_2|$, of **(Φ-I)(Φ-I)'**, as the number of non-zero eigenvalues of this symmetric matrix corresponds to the rank of **Φ-I**. We preferred to use squared canonical correlations between (X,Y) and $(\Delta X,\Delta Y)$, $\rho_1 \leq \rho_2$, instead, as suggested by Johansen (1988). These are related to the likelihood ratio and can be extended easily to account for conditional influences in higher-order models or for correlation among ε_{1t}, ε_{2t}. It is shown easily that $\rho_i=0$ if and only if the rank of **Φ-I** is less than i. Also, $\rho_i=0$ if and only if $\lambda_i=0$, provided that **Φ–I** is diagonalizable. However, note that the prior weighting distribution was uniform on (-2,0) for λ_1,λ_2 but not on (0,1) for ρ_1,ρ_2.

Results from this bivariate cointegration experiment are summarized in Table 2. Actual decisions on the secondary parameters were based on sample estimates of the squared canonical correlations, in concordance with the likelihood-ratio analysis by Johansen (1988). Our decision rule was defined in the following way:

Calculate the squared canonical roots and order them $0 \leq \hat{\rho}_1 \leq \hat{\rho}_2 \leq 1$.

Choose the stationary model if $\hat{\rho}_1 \geq b_1$.

Otherwise, choose the cointegrated model if $\hat{\rho}_2 \geq b_2$.

Otherwise, choose the fully integrated model.

This decision rule is similar in spirit to the classical eigenvalue test suggested by Johansen (1988) as an alternative to the trace test whose fractiles are tabulated there. One may envisage the difference between the two decision rules - firstly, ours and Johansen's eigenvalue test and, secondly, Johansen's trace test - by plotting ρ_1 and ρ_2

in a plane where the permitted area is bounded by a triangle as the eigenvalues have been ordered by $\rho_1 \leq \rho_2$. The b_2 rule corresponds to a horizontal line whereas the trace test rule corresponds to a 45° negatively sloped line. Both "cut off" the area around the origin that indicates fully integrated processes. In all cases, the rule on the smaller eigenvalue corresponds to a vertical line. Table 2 allows a comparison between our procedure and the classical one under the caveat that the decision rules are slightly different with respect to ρ_2.

TABLE 2. *Monte Carlo bounds for estimating the cointegrating rank in a bivariate AR(1) model. 10000 replications were conducted. Approximate classical bounds and the risk of the overall rule are given in square brackets.*

n	b_1		b_2		risk	
100	.045	[.041]	.127	[.075]	.0778	[.0914]
200	.026	[.021]	.070	[.038]	.0447	[.0665]
300	.022	[.014]	.053	[.026]	.0329	[.0564]
500	.015	[.008]	.038	[.015]	.0207	[.0523]

It is not surprising that the risk of the classical decision rule substantially exceeds the optimum risk, as the classical test operates on an entirely different concept of risk that it tries to minimize. It is also not surprising that the classical test appears to settle down at risk values slightly above 5% at n=500. Its decision rule is not consistent. In contrast, our procedure attains 2% at n=500. If n=100, the usual 5% critical values roughly match those evolving from the multiple decision problem. However, one can easily calculate that, for n=500, the significance level of the *classical* tests would have to be lowered to 1% to establish this equivalence. In consequence, for n=100, b_1 corresponds roughly to Johansen's trace value whereas, for n=500, b_1 is 1.8 times as large as the classical decision bound. On the other hand, b_2 is always much larger than the classical bound, which indicates that the classical procedure tends to avoid the fully integrated model even in small samples. In summary, substantially more fully integrated and (insubstantially) more co-integrated processes will be found by the multiple decision procedure, these cases both gaining at the cost of stationary solutions. To put it conversely, the classical procedure appears to find uncomfortably many covariance-stationary processes when the true model is integrated. [6]

Note that the shape of the decision rule *per se* does not change much between the classical and our multiple decision framework and that the main difference is in the significance levels not in the decision criterion.

[6] Note that also Phillips and Ploberger (1994) find more integrated models than previously used classical tests.

3.3 *Multiple binary problems: seasonal unit roots*

Let us take up the SODE triangle again. In example 1, we had excluded the triangle boundary except for the north-east line segment, closed in the south-east by Θ_0 and open at the north corner. In particular since the publication of the article by Hylleberg et al. (1990, HEGY), econometricians have focused on cyclical and seasonal non-stationarity possibly explicable by unit roots at -1 rather than at +1 or jointly at both locations. This model with "integration", i.e., spectral singularities, at the long-run and at the Nyqvist frequency seems particularly interesting for semester (half-yearly) data, whereas additional roots at the conjugate complex pair $\pm i$ may be considered for quarterly data (see example 4). In example 3, we shall concentrate on the semester case and on the root at -1.

Second-order autoregressive processes with exactly one unit root at -1 are found on the open north-west boundary line segment. The south-west corner point has second-order integration at -1. This case appears to be of mere academic interest and is unlikely to be found in economic reality. Similar to explosive cases, the south-west corner point will be weighted with zero weight. The north pole corresponds to integration both at +1 and at -1. This is the autoregressive process

$$X_t = X_{t-2} + \varepsilon_t$$

Hylleberg et al. (1990) and other authors have found that such processes provide reasonable descriptions of trending economic variables with substantial changes in their seasonal pattern. Hence, we do want to consider this case. In summary, we now have four subsets of the overall SODE triangle parameter space:

$\Theta_{\pm} = \{(0,1)\}$ ... integrated at long run and at Nyqvist frequency

$\Theta_{+} = \{(\varphi_1,\varphi_2) | \varphi_1 \in (0,2), \varphi_1+\varphi_2=1\}$ ... integrated at long run only

$\Theta_{-} = \{(\varphi_1,\varphi_2) | \varphi_1 \in (-2,0), \varphi_1-\varphi_2=-1\}$ ... integrated at Nyqvist frequency only

$\Theta_S = \left\{(\varphi_1,\varphi_2) \in \Re^2 \middle| \varphi_1+\varphi_2<1, -\varphi_1+\varphi_2<1, \varphi_2>-1\right\}$ ... stationary

In the notation used in section 2.2, the four events can also be denoted in binary form as (1,1), (1,0), (0,1), (0,0), in this order, with the first entry corresponding to the unit root at 1 and the second entry to the unit root at -1. Note that, for simplicity, the south-east corner point was excluded. It could be rather easily be re-inserted into the multiple decision problem later on.

Clearly, the decision situation corresponds to the multiple binary or lattice case introduced in section 2.2. Note that there are two nested paths

$$\overline{\Theta}_{\pm} \subset \Theta_{+}, \overline{\Theta}_{+} \subset \Theta_S \quad \text{and} \quad \overline{\Theta}_{\pm} \subset \Theta_{-}, \overline{\Theta}_{-} \subset \Theta_S$$

In concordance with the distance function introduced in section 2.2, we will define it for this multiple binary problem in the following way:

$d_{\pm}$	$\Theta_{\pm}$	Θ_{+}	Θ_{-}	Θ_{S}
$\Theta_{\pm}$	0	1	1	4
Θ_{+}	1	0	4	1
Θ_{-}	1	4	0	1
Θ_{S}	4	1	1	0

This table gives a cyclical definition of distance. A unit-root process of the long-run integrated type is supposed to be "closer" to a process integrated at both frequencies than to a process integrated at -1 only. This solution to the distance definition is probably debatable. In other multiple decision problems of similar type, this distance between the case of "exactly one object A" and "exactly one object B" obviously depends on the difference between objects A and B. When estimating the number of persons in a certain room or space, in most practical situations the distinction between men/women or black/white persons matters little and would not justify our cyclical design. On the other hand, the qualification of a good econometrician - who knows about economics and statistics as well - is probably closer to that of a statistician or of an economist than the two specialists' qualifications usually are among them. In our example, the properties of processes with roots at -1 and +1 are so strikingly different that the cyclical distance measure seems justified.

Monte Carlo simulations were conducted and decisions among secondary parameters {(0,0),(0,1),(1,0),(1,1)} - now, e.g. in the ordering (0,0)-(0,1)-(1,1)-(1,0)-(0,0), algebraically reminiscent of remainder classes "modulo 4" in number theory - were based on parallels to the north-east and north-west line segment and a horizontal beneath the north pole point. As the scheme appears to be perfectly symmetric between φ_1 and $-\varphi_1$, there will be only two decision thresholds, b_1 describing the position of the horizontal and b_2 fixing the position of the skew parallels.

A technical problem derives from the fact that, as long as $b_1<b_2$, the areas pointing to $\kappa=(0,1)$ and $\kappa=(1,0)$ may overlap. To decide among $\kappa=(0,1)$ and $\kappa=(1,0)$ in this case, we convened to select $\kappa=(1,0)$ whenever $\hat{\varphi}_1 > 0$ and $\kappa=(0,1)$ otherwise. The results of some Monte Carlo simulations based on 50,000 replications are displayed as Table 3.

Since there are now four competing decisions, the risk is slightly higher than in Table 1, at corresponding sample sizes. Strikingly at odds with classical hypothesis test decisions, the optimum values for b_1 and b_2 are almost identical. It is interesting to have a closer look at these classical tests. The current recommendation seems to be to

start by choosing among the secondary parameters κ=(0,1) or (1,1) and among κ=(1,0) or (1,1) separately. These tests with identical large-sample distributions correspond to our b_1 decision. If κ=(1,1) and κ=(1,0) are selected by the two separate tests, κ=(0,1) and κ=(1,1) are discarded and κ=(1,0) vs. κ=(0,0) are subjected to a third binary decision "due to the low power of the HEGY tests relative to the more specific DF test when no seasonal unit root is present". The main conclusion to be drawn from suggesting this very complicated and hardly efficient procedure is that *a priori* confidence in the seasonal unit roots is lower than that in the more familiar cases κ=(0,1) and κ=(0,0).

TABLE 3a. *Monte Carlo bounds for deciding among long-run, seasonal, and jointly long-run and seasonal non-stationarity in AR(2) models. 50000 replications were conducted*

n	b_1	b_2	risk
100	0.133	0.136	0.0842
200	0.084	0.088	0.0488
500	0.042	0.046	0.0228

TABLE 3b: *Empirical frequencies of selecting the respective events of seasonal integration at the optimum given in Table 3a.*

	true	identified model			
		(0,0)	(0,1)	(1,0)	(1,1)
n=100					
	(0,0)	12082	680	677	53
	(0,1)	186	11610	9	770
	(1,0)	173	9	11587	763
	(1,1)	58	216	230	10897
n=200					
	(0,0)	12322	438	423	17
	(0,1)	87	11954	3	506
	(1,0)	73	11	11953	484
	(1,1)	17	112	124	11476
n=500					
	(0,0)	12451	229	223	5
	(0,1)	27	12246	2	274
	(1,0)	19	0	12244	258
	(1,1)	2	40	34	11946

3.4 *The so-called univariate HEGY model*

The possibility of the joint presence of unit roots at different locations has been shown to complicate the handling in our multiple decision framework slightly but these difficulties can be overcome. In econometric practice, quarterly or monthly data are more common than semi-annual observations. For quarterly data, it is tempting to allow for the presence of homogeneous non-stationary influences deriving from the main frequencies 0, $\pi/2$, π though other frequencies would be conceivable. In econometrics, the main reference to this problem is again HEGY (1990). There, fourth-order autoregressive structures were considered. These were transformed into the form

$$\Delta_4 X_t = a_1 S(B) X_{t-1} + a_2 A(B) X_{t-1} + (a_3\ a_4)(\Delta_2 X_{t-1}\ \Delta_2 X_{t-2})' + \varepsilon_t$$

with

$$S(B) = 1 + B + B^2 + B^3 \qquad A(B) = 1 - B + B^2 - B^3$$

HEGY (1990) then suggest to conduct *t*- and *F*-type tests on the significance of the coefficients in order to find out about the potential significance of rejecting unit roots at 1 (the coefficient a_1), at -1 (or a_2), and at $\pm i$ (a_3 and a_4 jointly). As was already stated above, we want to develop alternatives to this classical framework which is, moreover, based on the assumption of "just local tests", with the remaining unit roots assumed as being present anyway. Only asymptotically, such cross-effects among effects at different "seasonal" frequencies vanish.

To handle the HEGY problem in our framework, we have to define a distance measure on the space of secondary parameters. Secondary parameters can be coded as 3-vectors of binary numbers (m_1,m_2,m_3), convening that m_1=1 stands for the presence of a unit root at +1, m_2=1 for a unit root at -1, and m_3=1 for the complex pair $\pm i$. Then, e.g. (0,0,0) corresponding to the event of "no unit roots", and (0,1,1) to "no unit root at 1 but one each at -1 and $\pm i$. Analogously to the last example, the distance measure is defined by

$$d_K\left(\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}, \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}\right) = \left(\sum_{j=1}^{3} |m_j - n_j|\right)^2$$

noting that all m_i and n_i elements are either 0 or 1 and that the maximum distance is 9 and is e.g. reached between (1,0,0) and (0,1,1), i.e., between a process of the random-walk type and one that wholly consists of persistent seasonal cycles.

To conduct our simulation experiments, the weighting distributions over the primary parameters within the classes have to be fixed. This is trivial for (1,1,1), since there is only one fourth-order process with all three unit roots present:

(a) (1,1,1) is given the weight of 0.125 and simulated as $\Delta_4 X_t = \varepsilon_t$

(b) For (0,1,1), the process must look like $(1-\varphi B)(1+B)(1+B^2)X_t = \varepsilon_t$. We assume a uniform weighting prior on φ within the interval (-1,1). Similar rectangular priors can be chosen for (1,0,1)

(c) (1,1,0) is given a weight of 0.125 and is simulated using a uniform weighting prior over the SODE triangle to generate processes of type $(1-B^2)(1-\varphi_1 B-\varphi_2 B^2)X_t = \varepsilon_t$.

(d) For (0,0,1), we use the design $(1-\varphi_1 B-\varphi_2 B^2)(1+B^2)X_t = \varepsilon_t$ again over the SODE triangle for (φ_1,φ_2).

For (1,0,0) and (0,1,0), a counterpart to the SODE triangle in the three-dimensional space would be required. However, the structure of the stationarity area for the third-order difference equation is already quite involved. It is convex but not a simplex and does not have planes at all boundaries. We decided to use "brute force" instead for any order larger than two. For three lags, noting that the coefficients in a third-order stationary difference equation have maximum absolute values of (1,3,3,1), single coefficients were drawn from uniform random variables of (-3,3), (-3,3), and (-1,1), respectively.

(e) (1,0,0) and (0,1,0) are given weights of 0.125 each and are generated from $(1-B)(1-\varphi_1 B-\varphi_2 B^2-\varphi_3 B^3)X_t = \varepsilon_t$ and $(1+B)(1-\varphi_1 B-\varphi_2 B^2-\varphi_3 B^3)X_t = \varepsilon_t$. The primary parameters $(\varphi_1,\varphi_2,\varphi_3)$ are generated by draws from three uniform distributions. Stability of the difference equation is checked and $(\varphi_1,\varphi_2,\varphi_3)$ are re-drawn if explosive roots have been found.

(f) (0,0,0) is generated from the full fourth-order design $(1-\varphi_1 B-\varphi_2 B^2-\varphi_3 B^3-\varphi_4 B^4)X_t = \varepsilon_t$. Maxima for $(|\varphi_1|,|\varphi_2|,|\varphi_3|,|\varphi_4|)$ are (4,6,4,1). Stability is checked and independent re-drawing is performed if necessary.

Table 4 gives the results from a Monte Carlo experiment according to the outlined design. For each of the sample sizes 100 and 200, 80,000 replications were simulated. This gives approximately 10,000 replications for each specific model class. Table 4 does not only show the simulated bounds but also gives the matrix of correct and incorrect decisions in the experiment. Larger sample sizes probably are not relevant in practice, due to the fact that quarterly data are rarely available for time spans of more than 50 years, maybe excepting meteorological series.

TABLE 4: *Empirical frequencies of selecting the respective events of seasonal integration if the loss function is double quadratic. Number of replications is 80,000.*

(a) n=100.

true	selected model							
	(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)
(0,0,0)	6824	1284	1424	97	314	21	22	0
(1,0,0)	89	8951	38	752	2	165	0	1
(0,1,0)	91	33	8939	769	3	0	181	4
(1,1,0)	8	444	436	9064	1	3	5	49
(0,0,1)	15	1	8	1	8868	543	563	18
(1,0,1)	1	83	0	4	207	9430	23	236
(0,1,1)	2	0	71	4	239	18	9399	254
(1,1,1)	2	8	13	156	76	757	742	8244

Bounds: b_1=0.060 b_2=0.062 b_3=0.126 Loss at minimum = 0.1444

(b) n=200.

true	selected model							
	(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)
(0,0,0)	7661	1124	983	50	152	10	6	0
(1,0,0)	40	9321	14	540	1	79	0	3
(0,1,0)	47	18	9277	601	0	0	77	0
(1,1,0)	2	270	222	9499	0	1	0	16
(0,0,1)	5	0	0	0	9194	429	378	11
(1,0,1)	0	28	0	1	77	9688	5	185
(0,1,1)	0	0	20	1	111	7	9667	181
(1,1,1)	0	3	1	65	10	450	355	9114

Bounds: b_1=0.045 b_2=0.041 b_3=0.082 Loss at minimum = 0.0876

4. Summary and conclusion

Many problems of multiple decisions are usually handled by binary sequential testing decisions. The testing framework may not correspond to the interest of the practitioner who intends to classify the data at hand into one out of a small number of categories. Two frequent cases of such problems have been treated, the *nested* and the *lattice* problem.

In the nested problem, the researcher is interested in estimating a naturally ordered discrete parameter. A related example of this type would be estimating the lag

order of an autoregressive structure but this problem has been treated so satisfactorily in the literature that it is not worth while to take it up again (see also Hannan and Deistler, 1988, Ch.5). Information criteria have been shown to provide consistent estimates of the lag order and have replaced the previous usage of less adequate sequential tests that would lead to inconsistent estimates. For the problem of estimating the order of integration in time series, a satisfactory treatment is still needed and our examples 1 and 2 have contributed to that aim.

In the lattice problem, the researcher is interested in a number of features that could be present in the data or not. A common example would be the inclusion/exclusion decision on possible regressor variables in linear regressions that is usually handled by t-tests and F-tests. However, in that case, many researchers may find themselves in an actual testing situation and their decision may closely correspond to rejecting or accepting a subject matter theory. In contrast, in estimating seasonal unit root models, such testing situation is unlikely. The researcher rather attempts to find the one model out of 4 (example 3) or 8 (example 4) structures that most closely tracks the data at hand. We have provided a new and promising framework for making such decisions.

Much work remains to be done in the future. Amalgams of the nested and lattice models appear when a variety of features can be absent/present in varying numbers *and* the number associated to each feature is interesting. Such a situation evolves, e.g., in seasonal cointegration. Another situation obtains when the absence or presence of deterministic features - such as constants, trends, fixed cycles - is investigated jointly with the unit roots. Depending on the interest in the features per se, the presence or absence of the features may define distinct decision classes or may be treated as nuisance.

To find an optimum decision rule, we assumed squared loss for the secondary parameters which are the only parameters of interest here. Squared loss is a common concept in estimating continuous parameters and we feel that multiple decisions should be treated in a joint framework. Risk typically depends on all primary parameters and we adopted uniform weighting of all primary parameters over "natural" parameterizations, conscious of the fact that uniform weighting is not invariant to re-parameterizations. We also gave equal weights to each class (secondary parameter) considered. We finally insisted on the typical researcher's goal to make binary (not quantitative) decisions on the secondary parameters, leaving Bayesian grounds with the latter viewpoint.

Viewed from a Bayesian perspective, we stressed the importance of mixed (continuous-discrete) priors in typical situations of multiple decisions. In contrast, the continuous priors used by most Bayesian researchers in unit root estimation entail two

severe problems. Firstly, they only achieve posterior mass for the non-generic classes by putting prior mass on non-admissible extensions of the parameter space, such as explosive processes. Secondly, they put probably undue emphasis on the problem of estimating primary parameters in a way that the typical researcher - at least in the examples considered - is only marginally interested.

References

Banerjee, A., J. Dolado, J.W. Galbraith, and D.F. Hendry (1993), *Co-integration, Error-correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press.

Dickey, 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.

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

Ferguson, T.S. (1967), *Mathematical Statistics - A Decision Theoretic Approach*, Academic Press.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.

Hannan, E.J. and M. Deistler (1988), *The Statistical Theory of Linear Systems*, Wiley.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo (1990) 'Seasonal Integration and Cointegration,' *Journal of Econometrics* 44, 215-238.

Johansen, S. (1988), 'Statistical analysis of cointegration vectors,' *Journal of Economic Dynamics and Control* 12,2/3,231-254.

Johansen, S. (1995), 'A Statistical Analysis of Cointegration for I(2) Variables,' *Econometric Theory* 11, 25-59.

Pantula, S.G. (1989), 'Testing for Unit Roots in Time Series Data,' *Econometric Theory* 5, 265-271.

Phillips, P.C.B. and S. Ouliaris (1990), 'Asymptotic Properties of Residual Based Tests for Cointegration,' *Econometrica* 58, 165-194.

Phillips, P.C.B. and W. Ploberger (1994), 'Posterior Odds Testing for a Unit Root with Data-Base Model Selection,' *Econometric Theory* 10, 774-808.

Uhlig, H. (1994), 'What Macroeconomists Should Know About Unit Roots - A Bayesian Perspective,' *Econometric Theory* 10, 645-671.

Yap, S.F. and G.C. Reinsel (1995), 'Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model,' *Journal of the American Statistical Association* 90, 253-267.

FIGURE 1

The second-order difference equation

Stable solutions lie inside the triangle

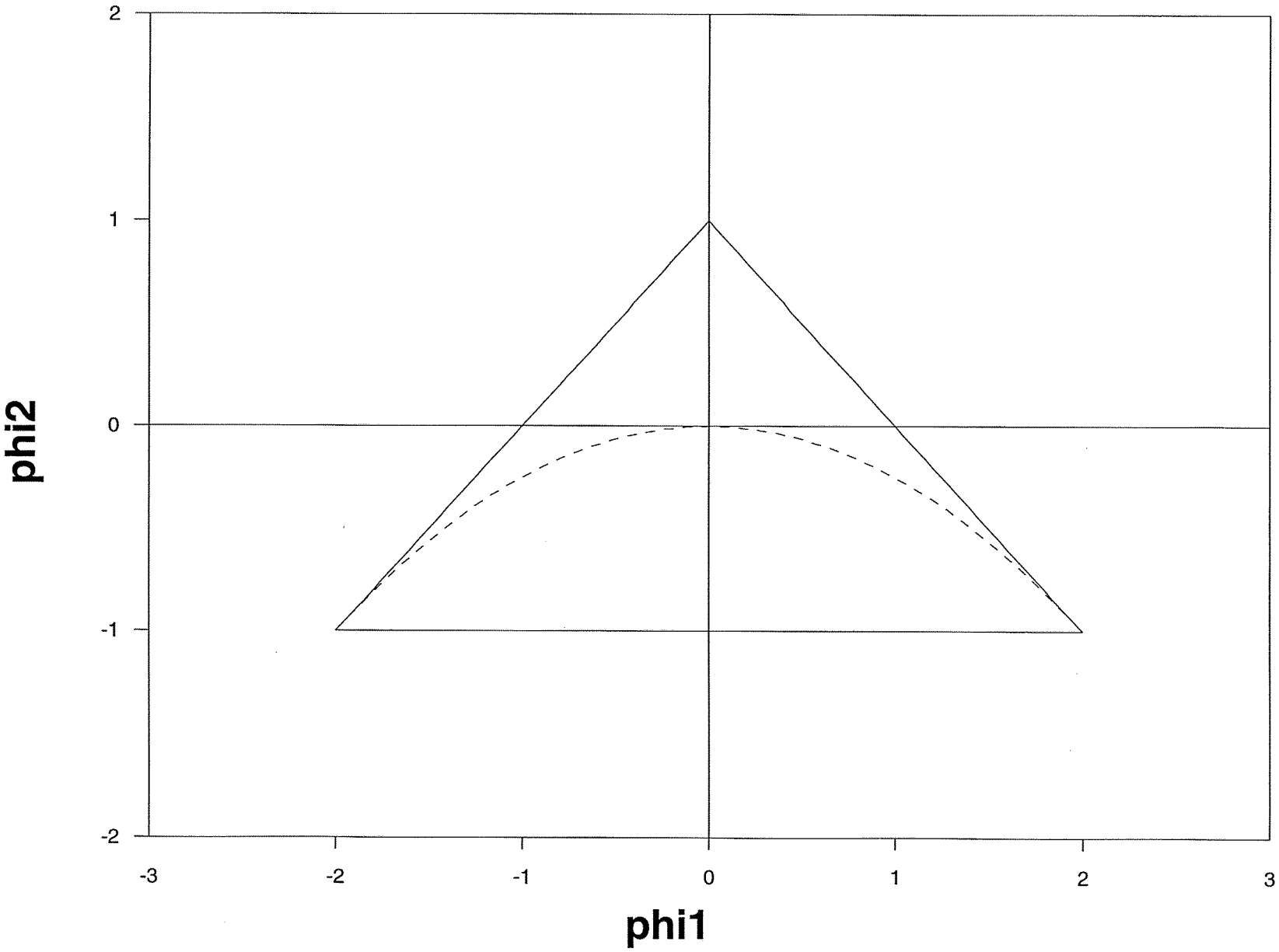
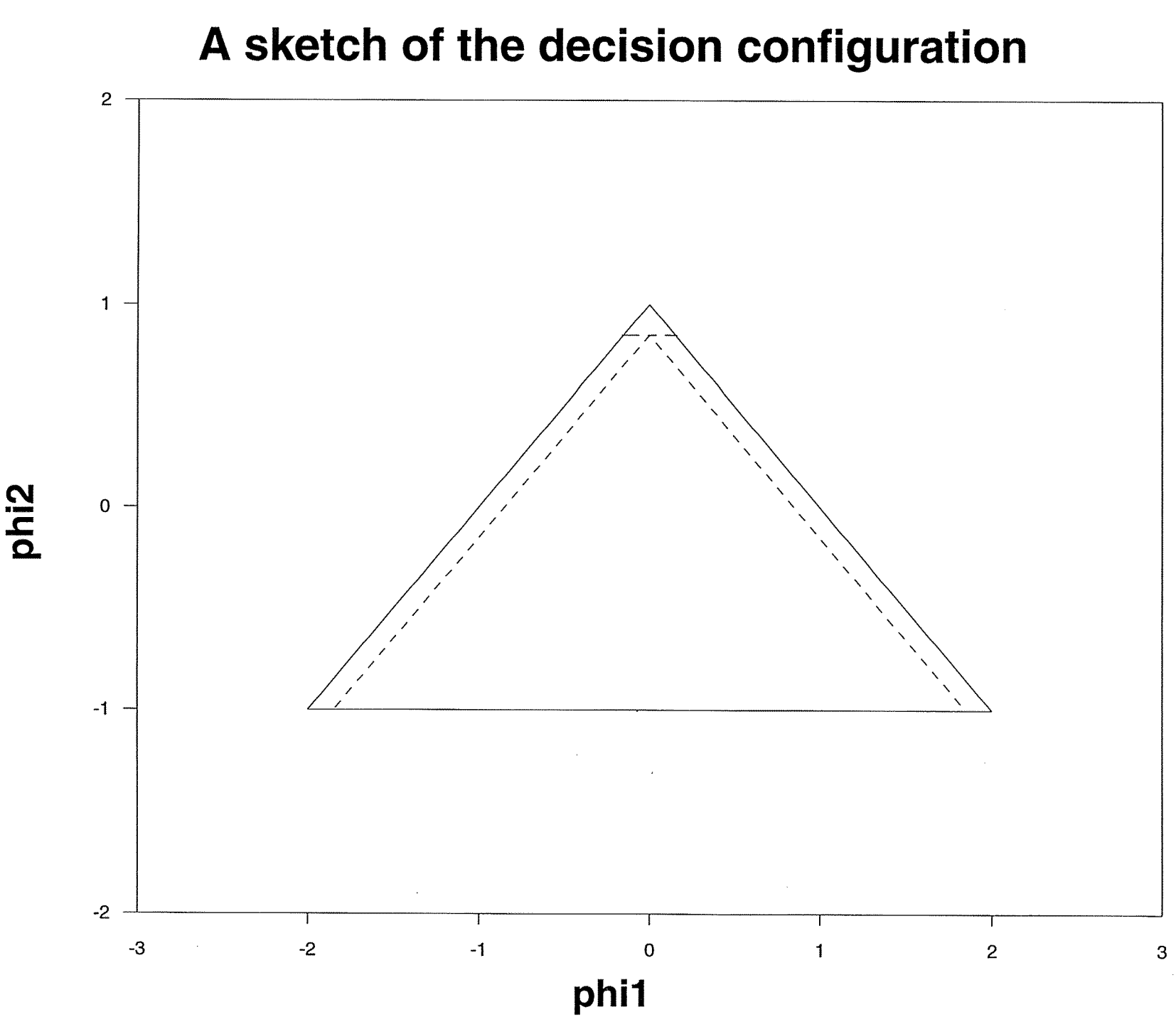


FIGURE 2


A sketch of the decision configuration
2
1
0
-1
-2
phi2
-3
-2
-1
0
1
2
3
phi1

Institut für Höhere Studien
Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria

Phone: +43-1-599 91-145
Fax: +43-1-599 91-163