

IHS Economics Series
Working Paper 51
November 1997

Decision Bounds for Data-Admissible Seasonal Models

Robert M. Kunst





INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):

Robert M. Kunst

Title:

Decision Bounds for Data-Admissible Seasonal Models

ISSN: Unspecified

1997 Institut für Höhere Studien - Institute for Advanced Studies (IHS)

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

All IHS Working Papers are available online: http://irihs.ihs.ac.at/view/ihs_series/

This paper is available for download without charge at:

<https://irihs.ihs.ac.at/id/eprint/1029/>

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

Reihe Ökonomie / Economics Series

No. 51

**Decision Bounds for Data-Admissible
Seasonal Models**

Robert M. Kunst

Decision Bounds for Data-Admissible Seasonal Models

Robert M. Kunst

Reihe Ökonomie / Economics Series No. 51

November 1997

Robert 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@ihs.ac.at
and
Johannes Kepler University Linz

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

The Institute for Advanced Studies in Vienna is an independent center of postgraduate training and research in the social sciences. The **Economics Series** presents research done at the Economics Department of the Institute for Advanced Studies. Department members, guests, visitors, and other researchers are invited to contribute and to submit manuscripts to the editors. All papers are subjected to an internal refereeing process.

Editorial

Main Editor:

Robert M. Kunst (Econometrics)

Associate Editors:

Walter Fisher (Macroeconomics)

Arno Riedl (Microeconomics)

Abstract

The selection problem among models for the seasonal behavior in time series is considered. The central decision of interest is between models with seasonal unit roots and with deterministic cycles. In multivariate models, also the number of stochastic seasonal factors is a discrete parameter of interest. To enable restricting attention to data-admissible models, a new attempt is made at defining data admissibility. Among data-admissible model classes, statistical decision rules are constructed on the basis of weighting priors and decision-bounds analysis. The procedure is applied to some exemplary economics series. Many univariate series select models without seasonal unit roots but the bivariate experiments enhance the importance of seasonal unit roots with restricted influence of seasonal constants. The framework of decision-bounds analysis offers a convenient alternative to sequences of classical hypothesis tests.

Keywords

Unit roots, seasonal cointegration, model selection

JEL-Classifications

C32

1 Introduction

For a long time, seasonal characteristics of economic series were viewed as being residual and without inherent economic interest. The usual practice was to focus on de-seasonalized data, at least in the field of macroeconomics. Typically, de-seasonalization was achieved by the application of seasonal adjustment filters. However, recently the drawbacks of this practice and the possible economic relevance of seasonality have been fully recognized. Many economists have realized, firstly, that the supposedly seasonal noise part of the series contains important information about the non-seasonal component and, secondly, that the seasonal component of one series contains information about the seasonal and non-seasonal components of other series. In short, any attempt to decompose the world of economics into two mutually independent worlds, one being non-seasonal and of economic interest and the other one being seasonal and uninteresting, is misguided. A good survey of the historical developments that led to the modern econometrics of seasonality can be found in HYLLEBERG (1992).

The need to view the whole of the economic variable of interest within a full model that incorporates seasonal features as well as trend and business cycle characteristics has instigated the development of various seasonal time series models. Building on the traditional low-order autoregressive (AR) model, which is a convenient first-order approximation to the dynamic properties of most time series in the absence of a theoretical model, three ways have been used repeatedly in order to capture seasonal characteristics. Firstly, seasonal constants can be added to the deterministic part of the model. These seasonal-dummy models are characterized by inflexible and repetitive cycles and by predictions that are mainly projections of the average in-sample cycle into the future. Secondly, stable conjugate complex roots in the AR polynomial may reflect non-persistent seasonality. Thirdly, seasonal unit roots have been suggested by HYLLEBERG ET AL. (1990) and others to allow permanent shifts in the seasonal structure. These models with non-stationary stochastic seasonality are characterized by flexible and shape-changing cycles and by predictions that mainly project the present shape of the seasonal cycle into the future, with a low degree of predictive precision due to the non-stationarity. Amalgams of these models have also been used but we will point to some of their drawbacks in this paper.

This apparent variety of available models calls for an efficient model selection strategy. In this paper a framework for a multiple decision strategy with regard to model selection is developed. The basic choice set of seasonal models is determined by arguments of data admissibility, as we feel that one should be suspicious of models that are not data-admissible. The concept of data admissibility permits to impose plausibility restrictions on process trajectories and to discard models that have a high probability

of generating implausible trajectories. Following the exclusion of non-admissible models from consideration, the choice among the remaining candidates is often guided by sequences of hypothesis tests, whereas other researchers use Bayesian methods. This paper introduces a multiple decision (MD) strategy as a third alternative that builds on Bayesian ideas but assigns a central position to a loss function that is designed to penalize certain mismatches between selected and true model harder than others. The elicitation of weighting priors within each model class is guided by the idea of uniform distributions over bounded regions of parameter values. A similar approach was followed by FRANCES ET AL. (1997) for certain seasonal processes in a fully Bayesian setting. However, the present work differs in two main aspects. Firstly, for our bivariate experiments we use uniform weighting on eigenvalues and not on the coefficient space. Secondly, we form decision rules by loss functions.

The remainder of this paper is organized as follows. A step toward a more rigorous definition of data admissibility is taken in Section 2. In Section 3 existing seasonal models are reviewed under the aspect of data admissibility. Section 4 concentrates on constructing multiple decision bounds in order to permit an efficient selection among competing models. In Section 5, empirical applications to some macroeconomic series are reported. We robustify the results by also reporting some sensitivity experiments. Section 6 concludes. An appendix expounds the basics of the multiple decisions technique used in this paper.

2 The concept of data admissibility

The reference work by HENDRY (1995, p.364) gives the following verbal definition of data admissibility:

”A model is data admissible if its predictions automatically satisfy all known data constraints.”

In this definition, a correct interpretation of the word ‘prediction’ appears to be crucial. If it stands for point prediction from the realized sample — or possibly subsamples thereof — by means of conditional expectation, the definition may appear slightly too liberal for empirical applications. For example, consider three models: (a) white noise plus a constant a ; (b) a random walk started from the value a ; (c) some nonlinear but symmetric generating law that also starts from a but is stationary only in a local neighborhood of a and is explosive otherwise. All three models result in the same conditional-expectation point forecast a . However, we feel that for a naturally bounded economic variable only model (a) is admissible, whereas for an unbounded variable we may also accept model (b) but not (c).

On the other hand, if ‘prediction’ means the predictor process in its interpretation as a random variable, though possibly conditioned on some starting values in the data sample, the definition is very restrictive. We formalize the two extreme interpretations in two tentative definitions. We remark that, as we are concerned with models for time series, we view ‘models’ as parametric collections of time-series processes.

Definition 1. A process is called *data-admissible in mean prediction* if none of its point forecasts at any step size defined by conditional expectations from observed samples or subsamples thereof violates logical constraints. A model parameterized by a parameter $\theta \in \Theta$ is called *data-admissible in mean prediction* if all $\theta \in \Theta$ define admissible processes.

The defined property is sample-dependent. It is conceivable that we construct an artificial sample that is close but not identical to the observed one, apply the model under investigation with a certain fixed value of θ , and violate the logical constraints for a certain step size. The following alternative definition marks the other extreme.

Definition 2. A process is called *strictly data-admissible* if it is conceivable as a data-generating mechanism for the observed n -dimensional economic variable for all $t \in I$, where I is the index time range for the considered class of time-series processes. Typically, $I = \mathbf{N}$. A parametric model is called *strictly data-admissible* if all $\theta \in \Theta$ define data-admissible processes.

Verbal elements such as ‘logical constraints’ and ‘conceivable’ are intended to leave room for expert evaluation by economic theorists. Some logical constraints are certain or almost sure, such as definitional identities, but for others their violation is just extremely unlikely. Under the latter category come e.g. trajectories for unemployment rates that remain at a level beyond 90% for several decades, under the former category come trajectories with unemployment rates outside the range $[0,1]$. Practical usage of the definition requires the assumption of a set of conditions \mathcal{A} which conceivable trajectories are not allowed to violate. \mathcal{A} could exclude impossible values but also other features, such as disproportionate growth, explosive cycles, or sudden jumps. In the following, we will use as sets \mathcal{A} interval restrictions to avoid impossible values or cone-type restrictions to avoid excessive expansion from starting values.

We note that, e.g., the model of Brownian motion is not strictly data-admissible for the unemployment rate as almost every trajectory of its processes crosses the boundaries of 0 and 1, even if started from real-life values. The same is probably true for many economic and econometric models used in practice. However, an economist may feel quite comfortable with such models for a certain time span and may be willing to use them to explain the local behavior of economic variables that are strictly bounded. One may consider to replace Definition 2 by a ‘local’ definition requiring the admissibility

constraints to just hold for a limited part of the time range I . However, then the plausibility of the analysis depends on the life span of trajectories that may be quite low. It makes more sense to exploit the LIL (law of the iterated logarithm) property of the random walk which guarantees that generated trajectories violate certain prescribed boundaries at most finitely often. We introduce the following definition for weak-sense data admissibility, which is stronger than Definition 1 but weaker than Definition 2 and may be able to capture the properties that are interesting in practice.

Definition 3. A process is called (*weak-sense*) *data-admissible* if its trajectories violate bounds constructed from plausibility arguments at most finitely often. A model is called (*weak-sense*) *data-admissible* if all $\theta \in \Theta$ define data-admissible processes.

Clearly, a strictly data-admissible model is data-admissible. A data-admissible model is admissible in mean prediction unless the observed sample has been created from the finite number of constraint violations. The characteristic properties of Definition 3 are seen as follows. Suppose the admissibility bounds are defined by some function $g(t)$ such that $\mathcal{A} = \{X_t \in [X_0 + at - g(t), X_0 + at + g(t)]\}$. We assume that (a) $g(0) > 0$, (b) $g(t)$ increases monotonously in t , (c) $\lim_{t \rightarrow \infty} g(t)/\log t = \infty$ and hence also $\lim_{t \rightarrow \infty} g(t) = \infty$, (d) $\lim_{t \rightarrow \infty} g(t)/t = 0$. In short, $g(t)$ is sublinear but grows faster than $\log t$. We note that, assuming $a > 0$ and $X_0 > 0$, this admissibility condition is stricter than $\mathcal{A}' = \{X_t > 0\}$, i.e. positivity of trajectories. The ‘trend-stationary’ model $X_t = X_0 + at + \varepsilon_t$ with Gaussian white noise ε_t is certainly data-admissible in mean prediction. However, it is also weak-sense admissible due to the extremal properties of the Gaussian distribution, though it can never attain strict admissibility due to the unboundedness of the support of the Gaussian law. Due to the LIL, the drifting random walk started from X_0 and with drift constant a is also a weak-sense data-admissible model. If the drift is unknown, widening \mathcal{A} to e.g. $\mathcal{A} = \{X_t \in [X_0 + a_1 t - g(t), X_0 + a_2 t + g(t)]\}$ with $0 < a_1 < a_2$ will result in data-admissible drifting random walks over a useful range of drift parameters, such that we can hope that the sample estimate falls into the prescribed range $[a_1, a_2]$. In contrast, all models with disproportionate growth, such as $X_t = X_{t-1} + a + bt + \varepsilon_t$, are clearly inadmissible.

3 Seasonal time series models

In the following, we will use the conventional time-series abbreviations, such as B for the lag operator, i.e. $X_{t-1} = BX_t$, Δ for first differences $1 - B$, and $\Delta_4 = 1 - B^4$ for seasonal differences. Without undue lack of generality, we will concentrate on the case of quarterly data throughout. We will use ε_t to denote a white-noise series. Wherever we need stronger properties, we will tacitly also assume that ε_t is Gaussian white noise.

For exposition, we first consider the seasonal time series model

$$\Delta_4 X_t = \sum_{i=1}^4 \delta_i D_{it} + \varepsilon_t \quad (1)$$

D_{it} are seasonal constants that 1 in the i -th quarter and 0 in the other quarters. The deterministic part of the right-hand side admits some alternative equivalent representations:

$$\sum_{i=1}^4 \delta_i D_{it} = \mu + \sum_{i=1}^3 \delta_i^* D_{it} = \mu + a \cos(\pi t) + b \cos\left(\frac{\pi t}{2}\right) + c \cos\left(\frac{\pi(t-1)}{2}\right)$$

There is an obvious one-one mapping between the three representations in the parameters $(\delta_1, \dots, \delta_4)$, $(\mu, \delta_1^*, \delta_2^*, \delta_3^*)$, and (μ, a, b, c) .

The model (1) is an amalgam of two popular time series models, the seasonal random walk with drift

$$\Delta_4 X_t = a + \varepsilon_t \quad , \quad (2)$$

and the random walk with seasonally varying drift

$$\Delta X_t = \sum_{i=1}^4 \delta_i D_{it} + \varepsilon_t \quad . \quad (3)$$

These models have been repeatedly used in the econometric literature to describe the behavior of seasonal data. To capture the serial correlation in the errors, they are usually ‘augmented’ with autoregressive lags of the left-hand-side variable, such as $\sum_{i=1}^{p-4} \varphi_i \Delta_4 X_{t-i}$ for cases (1) and (2). Our point is that the two simpler models (2) and (3) are data-admissible with respect to a plausible set \mathcal{A} but that the amalgam model (1) is not data-admissible. In particular we define the admissibility set \mathcal{A} by

$$\mathcal{A} = \left\{ \max_{i \leq 4} |X_t - X_{t-i}| < g(t) \right\} \quad (4)$$

with the function $g(t)$ obeying the restrictions $\lim_{t \rightarrow \infty} g(t)/\log t = \infty$ and $\lim_{t \rightarrow \infty} g(t)/t = 0$ as motivated in the last section. If m_d denotes the maximum difference $|\delta_i - \delta_j|$, additional conditions such as $g(0) > m_d$ and $g(t)$ differentiable with $g'(t) > 0$ exclude uninteresting cases. Note that \mathcal{A} is designed to contain secular expansion ($i = 4$) as well as intra-annual seasonal expansion ($i < 4$).

The process (1) is composed of four interspersed random walks with different drift constants. The difference of two of these four random walks with drift difference m_d is again a random walk Z_t with drift proportional to m_d . Because of the law of the iterated logarithm (LIL)

$$\limsup_{t \rightarrow \infty} \frac{Z_t - m_d t}{\sigma(2t \log \log t)^{1/2}} = 1 \quad \text{a.s.} \quad (5)$$

(see, e.g., DAVIDSON, 1994, p.408), the set \mathcal{A} is violated by the process with probability one. Note that even the deterministic skeleton of the model violates \mathcal{A} for t large enough. The average life span strictly decreases with m_d and can be quite low for empirically relevant parameter combinations.

In contrast, for (2) the difference $X_t - X_{t-i}$ is a random walk with zero drift. It follows from the LIL that \mathcal{A} will only be violated for finitely many t as $g(t)$ was assumed to grow faster than the denominator in (5). Hence, the model (2) is weak-sense data-admissible. For (3), the differences $Z_t = X_t - X_{t-i}$ are bounded in probability. The maximum of the Z_t is essentially governed by $c \log t$ and the Gumbel distribution (cf. JOHNSON AND KOTZ, 1970, p.276). It follows that (3) is not strictly admissible due to the unbounded support of the increments but that the model is weak-sense admissible if $g(t)$ grows faster than $\log t$. The problem evolves how then to reconcile the two ideas of deterministic and stochastic seasonality in (2) and (3) in one comprehensive model and retain the property of weak-sense data admissibility. The answer is that this is not possible for univariate X . If one really wants to include deterministic seasonality within the framework of seasonal unit roots, this can only be done by allowing for a seasonal starting pattern or by considering different, more complicated, structures. However, for multivariate X , such a reconciliation is possible.

A class of multivariate seasonal models with univariate marginal models of the admissible type (2) was suggested recently by FRANSES AND KUNST (1996) who consider special restrictions on seasonally cointegrated models. We use the following additional notational conventions. Δ_2 denotes the second-order differencing operator $1 - B^2$, $S(B)$ is the seasonal moving average $1 + B + B^2 + B^3$, $A(B)$ is the moving average with alternating signs $1 - B + B^2 - B^3$. Note that these three operators are factors of the seasonal differencing operator Δ_4 . In this notation, the n -dimensional seasonal model considered by FRANSES AND KUNST reads

$$\begin{aligned} \Delta_4 X_t &= \alpha_1 \beta_1' S(B) X_{t-1} + \alpha_2 (\beta_2' A(B) X_{t-1} + a^* \cos \pi(t-1)) \\ &\quad + \alpha_3 \{ \beta_3' \Delta_2 X_{t-2} + (b^*, c^*) (\cos \frac{\pi}{2}(t-1), \cos \frac{\pi}{2}(t-2))' \} \\ &\quad + \mu + \sum_{i=1}^{p-4} \Phi_i \Delta_4 X_{t-i} + \varepsilon_t \end{aligned} \tag{6}$$

The dimensionalities are $n \times r_i$ for α_i and β_i , $i = 1, 2, 3$, $r_2 \times 1$ for a^* , $r_3 \times 1$ for b^* and c^* , and $n \times n$ for the Φ_i matrices. The model allows for a non-zero general drift μ and also for deterministic seasonal influences $\alpha_2 a^*$ and $\alpha_3 (b^*, c^*)$ at the frequencies π and $\pi/2$. These are proportional to the loading vectors of the seasonal error-correcting structures. Hence, the coefficients a, b, c of the regressors $\cos(t-1)\pi, \cos(t-1)\pi/2, \cos(t-2)\pi/2$

are restricted by $a = \alpha_2 a^*$, $b = \alpha_3 b^*$, $c = \alpha_3 c^*$. These parameter restrictions cannot be expressed conveniently in the representation with coefficients δ_i , $i = 1, \dots, 4$ and the regressors D_{ti} or in the parameterization $(\mu, \delta_1^*, \delta_2^*, \delta_3^*)$.

The multivariate model (6) is a variant of the seasonal cointegration model introduced by HYLLEBERG ET AL. (1990) and LEE (1992). These articles should be consulted for all details. We just recall for convenience that the first three expressions on the right hand side correspond to cointegration at the long-run frequency $\omega = 0$ and at the two seasonal frequencies $\omega = \pi$ and $\omega = \pi/2$, i.e., the semi-annual and the annual frequency. The respective ranks r_i , $i = 1, 2, 3$, are the cointegrating ranks or the number of cointegrating relationships at the three frequencies and are usually identified by sequences of hypothesis tests guided by tables of significance points as presented by LEE (1992) and, for the modified version (6), by FRANCES AND KUNST (1996). Note that (6) allows for a deterministic seasonal influence only in the presence of seasonal cointegration. If $r_2 = r_3 = 0$, there cannot be any seasonal deterministics. FRANCES AND KUNST (1996) show that, in (6), the expansion of seasonal cycles is contained for all individual variates. Hence, the model is weak-sense data-admissible with \mathcal{A} defined as in (5). The seasonal constants only enter in the error-correcting seasonal equilibrium structures, which are stationary except for the added cycles, and therefore the model (6) also incorporates the other admissible model type (3).

4 Decision bounds

Most decisions in present econometrics are based on the framework of hypothesis testing. In hypothesis testing, one out of two decisions is formally identified with a lower-dimensional manifold Θ_0 in a parameter space Θ and then is given the name of null hypothesis. The other decision is identified with the generic remainder $\Theta \setminus \Theta_0$ and is called alternative hypothesis. Typically, the hypothesis test is conducted in the following steps. A test statistic is calculated from the observations. A significance level α is fixed. The distribution of the test statistic under Θ_0 is evaluated. The alternative is preferred if the value of the test statistic is in the α tail region of the null distribution and the null is preferred otherwise.

The problems of this approach are well known. Firstly, the labels of ‘null’ and ‘alternative’ are arbitrary and occasionally they can be interchanged by adopting a different parameterization. Secondly, the decision presupposes an asymmetric loss function with respect to incorrect decisions. In small samples, the null hypothesis appears to be preferred whereas in large samples the alternative is always preferred if it is correct whereas the null is still rejected with probability α . Thirdly, the distribution under the null is

typically not constant over Θ_0 but depends on the position of $\theta \in \Theta_0$ on the manifold, which is usually called ‘nuisance’. Fourthly, it seems difficult to generalize the approach to decision problems that are not binary. Usually, this difficulty is resolved by a sequence of (binary) hypothesis tests, which brings in a variety of further problems, such as the order of sequential decisions, the distinction of nested and non-nested situations, and the meaning of significance levels. Adopting an alternative framework in the spirit of a Bayesian version of discriminant analysis, also called multiple decisions (MD) approach, KUNST (1996) presents a different solution for the problem of optimal selection of parameter subspaces. Each parameter subset is given a discrete prior distribution of $1/k$ with k the number of ‘hypotheses’ or model classes. Within each of the k subsets, some continuous heuristic prior is defined. A loss function is defined on the set of classes Ξ , and the expectation of this loss function is then minimized. The loss-function approach assigns symmetric loss to incorrect decisions and imposes a more severe penalty on decisions that are more incorrect than others. The calculation of optimum decision bounds even for a restricted set of decision rules imposes a heavy computational burden. However, once such bounds have been established, the decision rules are readily applicable to the real world, the same way that significance tables are applicable in hypothesis testing. One advantage of the MD approach is that it necessarily yields asymptotically correct selection of hypotheses, given that such a decision is possible in the considered problem. For further details, see the Appendix.

4.1 The univariate model

With respect to seasonal time series, KUNST (1996) considers as Example 4 the following decision problem. A univariate time series is generated from a fourth-order autoregression. The autoregressive polynomial $\Phi(\cdot)$ is allowed to have at most one unit root at any of the frequencies 0 , π , and $\pi/2$. All non-unit roots are assumed to be stable. The occurrence of all unit roots is coded as $(1,1,1)$, of just one unit root at $\omega = 0$ as $(1,0,0)$ etc., and decisions among all $2^3 = 8$ cases are considered. Assuming a uniform prior on $\Xi = \{(i_1, i_2, i_3), i_j \in \{0, 1\}, j = 1, 2, 3\}$ and some reasonable prior within the classes, a double-squared loss function is minimized and decision bounds are tabulated for given sample sizes. In the following we will denote the discrete parameter space by Ξ and a typical discrete parameter by κ . Hence, $\hat{\kappa}$ denotes an estimate of a discrete parameter or equivalently a selection of a certain model class based on observed data.

Here, we consider another problem of seasonal model selection, as we want to discriminate between stochastic and deterministic conceptions of seasonality, as expressed in (2) and (3). To simplify the basic decision problem, we assume that there is a unit root at $\omega = 0$. Furthermore we exclude cases that are not data-admissible such as (1). This

latter assumption prevents the application of any hypothesis tests that are designed for nested situations, as we deliberately exclude the ‘closure’ model from consideration. We are left with the following possibilities:

1. Model (2) holds. There are unit roots at $\omega = \pi$ and $\omega = \pi/2$ but there is no deterministic seasonal pattern. We code this event as (1,0).
2. Model (3) holds. There are no seasonal unit roots but there is a deterministic seasonal pattern. We code this event as (0,1).
3. There is no seasonality in the process, neither deterministic nor stochastic. Any visible indications of seasonality may be rooted in stationary (non-persistent) cycles at frequencies close to the seasonal frequencies $\omega = \pi$ or $\pi/2$. We code this event as (0,0).

In order to keep the decision design reasonably simple, we do not separate between the roots at $\omega = \pi$ and $\omega = \pi/2$ although we are aware of the fact that partial occurrence of one of these roots has been reported in empirical studies.

In analogy to similar problems considered by KUNST (1996) we use a double-squared loss function

$$d_k((i_1, i_2), (j_1, j_2)) = \{(j_1 - i_1)^2 + (j_2 - i_2)^2\}^2 \quad (7)$$

which imposes a large penalty of 4 on misspecifying a deterministic seasonal model as a stochastic seasonal model and vice versa and a lesser penalty of 1 on misclassifying any of these two models as non-seasonal and vice versa.

The non-seasonal model has to be equipped with a weighting prior distribution. It reads

$$\Delta X_t = \mu + \sum_{1 \leq i \leq 3} \tilde{\varphi}_i \Delta X_{t-i} + \varepsilon_t \quad (8)$$

and we assume a continuous uniform distribution on the area $S_3 \subset \mathbf{R}^3$ that is determined by the stability of the roots of $\tilde{\Phi}(z) = 1 - \tilde{\varphi}_1 z - \tilde{\varphi}_2 z^2 - \tilde{\varphi}_3 z^3$. These processes were simulated via a mixture of some outer bounds of a simple geometrical shape that contains S_3 and brute-force rejection of unstable roots. For μ we assume a standard normal weighting prior, whereas $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ is fixed at 1 as we do not expect the results to depend critically on σ_ε .

The stochastic seasonal model reads

$$\Delta_4 X_t = \mu + \varepsilon_t \quad (9)$$

and we assume a standard normal prior on μ and a degenerate on the value of 1.0 for σ_ε .

The deterministic seasonal model reads

$$\Delta X_t = \sum_{i=1}^4 \delta_i D_{ti} + \varepsilon_t \quad (10)$$

and we assume a four-variate normal prior with mean $(0, \dots, 0)'$ and variance $\Sigma = \mathbf{I}_4$ for the seasonal constants $(\delta_1, \dots, \delta_4)$. σ_ε is again fixed at 1.0.

Given these weighting prior distributions, we then simulate 30,000 trajectories with the empirically relevant lengths $N = 100, 150, 200$ of the general model, which results in approximately 10,000 trajectories for each of the cases (8), (9), and (10). For each trajectory, we evaluate two summary statistics that are useful for a discrimination among the cases. Then, the expected risk as defined by the loss function (7) is minimized by a grid search over potential decision bounds for the two summary statistics. We now refer to the results of this search as they are summarized in Table 1. Under the labels b_1 and b_2 , the table shows the identified optimum bounds. b_1 is the bound for stochastic seasonality and b_2 is the bound for deterministic seasonality. These bounds are calculated from an auxiliary encompassing regression

$$\begin{aligned} \Delta_4 X_t = & \mu + \delta_1^* \cos(\pi t) + \delta_2^* \cos\left(\frac{\pi t}{2}\right) + \delta_3^* \cos\left(\frac{\pi(t-1)}{2}\right) \\ & + \pi_1 A(B) X_{t-1} + \pi_2 \Delta_2 X_{t-1} + \pi_3 \Delta_2 X_{t-2} + \varepsilon_t \end{aligned} \quad (11)$$

This equation is estimated by least squares. If the norm of the estimated 3-dimensional coefficient vector $c_1 = \|(\pi_1, \pi_2, \pi_3)'\|$ exceeds b_1 and at the same time the norm of the estimated 3-dimensional coefficient vector $c_2 = \|(\delta_1^*, \delta_2^*, \delta_3^*)'\|$ does not exceed b_2 , we opt for (0,0), i.e., no seasonal unit roots and no deterministic seasonality. If $c_1 < b_1$, we decide for (1,0), i.e., seasonal unit roots. In this case we ignore the decision that would be suggested by c_2 , as the two types of seasonal features are not allowed to co-exist. If $c_1 > b_1$ and $c_2 > b_2$, we decide for (0,1), i.e., deterministic seasonality. It was also attempted to reverse the decision sequence, that is to give priority to the criterion c_2 but this resulted in higher minimum risk.

TABLE 1. Optimal multiple decision rules for the univariate seasonal problem and simulated frequencies of classification. 30,000 replications.

N	b_1	b_2	d_{\min}
100	0.2733	0.4548	0.04140
150	0.2549	0.4034	0.02380
200	0.1899	0.3502	0.01460

generated model	identified model								
	$N = 100$			$N = 150$			$N = 200$		
	(0, 0)	(1, 0)	(0, 1)	(0, 0)	(1, 0)	(0, 1)	(0, 0)	(1, 0)	(0, 1)
(0, 0)	9745	60	194	9849	44	106	9889	27	83
(1, 0)	9	9937	60	0	10003	3	1	10000	5
(0, 1)	483	64	9448	352	50	9593	239	17	9739

Under the heading d_{\min} , Table 1 shows the attained minimum value of the expected loss function d_k . For a fully consistent test (decision) procedure, this value must reach 0 for $N \rightarrow \infty$. In three separate tables we show the frequency of correct classification and of misclassification in the simulation. In a simple binary test it suffices to report the frequency of type I and type II errors. Here, there are three hypotheses, the cases of classification errors are more involved and these should be reported properly. For example, for $N = 200$ we generated 10006 processes with seasonal unit roots (1,0). Only 6 of them were classified incorrectly. In contrast, approximately 2.5% of all (0,1) processes with deterministic cycles were misclassified, most of them as (0,0) processes. In a sloppy classical interpretation, one may conclude that the ‘power’ of the decision procedure against deterministic cycles is almost 0.975 or 97.5%. On average, the frequency of misclassifications decreases monotonously if N increases but due to the different rates of convergence in the coefficients this is not always so clear and it pays to see the decision procedure as a whole.

4.2 The bivariate model

Building on the n -variate data-admissible seasonal model (6) with $n = 2$, we also consider a bivariate decision problem. For the moment we exclude the possibility of frequency-zero cointegration and also impose $p = 4$, hence we obtain the simplified model

$$\begin{aligned} \Delta_4 X_t = & \mu + \alpha_2 (\beta_2' A(B) X_{t-1} + a^* \cos \pi(t-1)) \\ & + \alpha_3 \{ \beta_3' \Delta_2 X_{t-2} + (b^*, c^*) (\cos \frac{\pi}{2}(t-1), \cos \frac{\pi}{2}(t-2))' \} + \varepsilon_t \end{aligned} \quad (12)$$

Then, we have two discrete seasonal decision parameters. In the ordered pairs (i_1, i_2) , i_1 varies in the set $\{0,1,2\}$ and denotes the number of seasonal cointegrating vectors. i_2 is 0 or 1 and reflects the absence or presence of deterministic seasonality. We do not impose the condition that the seasonal cointegrating vectors at $\omega = \pi$ and $\omega = \pi/2$ have to be the same but we focus on equal cointegrating ranks at these two frequencies. A separation of the ranks would complicate the analysis by introducing a third decision parameter that would hardly reflect the main features of interest.

In this context, the loss function (7) is unsatisfactory as the range over which the two decision coordinates vary is not the same. It appears preferable to penalize the largest loss with respect to i_2 as much as the largest loss with respect to i_1 . Hence we use

$$d((i_1, i_2), (j_1, j_2)) = (j_1 - i_1)^2 + 4(j_2 - i_2)^2 \quad (13)$$

This function may be viewed as an amalgam of the multiple binary decision problem we faced in the univariate case and of the estimation of an integer number. In the first case, a double-squared loss function is needed to sufficiently penalize errors in many binary entries, as single-squared loss would be equivalent to the sum of absolute errors. In the second case, a single-squared loss function is adequate.

Within the model classes, realizations of (12) must be generated according to a weighting prior for the continuous parameters. Just as in the univariate experiment, we used independent standard Gaussian random draws for the unbounded parameters μ, a^*, b^*, c^* . For a seasonal cointegration rank of 1, matrices $\mathbf{\Pi}_i = \alpha_i \beta_i'$, $i = 2, 3$, were constructed from the Jordan representation $\mathbf{\Pi}_i = \mathbf{T} \mathbf{D} \mathbf{T}^{-1}$. Because $\mathbf{\Pi}_i$ is singular, the diagonal matrix \mathbf{D} contains one element of 0. The other element of \mathbf{D} was drawn from a uniform distribution on $(0,2)$. The off-diagonal elements of the rotation matrix \mathbf{T} were drawn independently from a standard normal distribution and the diagonal elements were scaled at 1. If $\mathbf{\Pi}_j = 0, j \neq i$, the resulting process (12) is non-explosive. Otherwise, this is not guaranteed, and stability has to be checked by the eigenvalues of the state space transition matrix. For explosive solutions, all random numbers were re-drawn. A similar strategy was also used for the cointegration rank of 2, with \mathbf{D} containing two uniform random diagonal elements, where the cases of re-drawing because of explosive configurations increased considerably.

Minimization of the loss function (13) has to be conducted on the basis of test statistics. Due to the known optimality properties of likelihood-ratio statistics for binary decision problems, we again adopt LR-type statistics for our problem. Hence, decisions on i_2 rely on the ratio

$$c_3 = \frac{\tilde{\sigma}_U^{(1)} \tilde{\sigma}_U^{(2)}}{\tilde{\sigma}_R^{(1)} \tilde{\sigma}_R^{(2)}} \quad (14)$$

of the residual variance estimates from the unrestricted bivariate autoregression (with seasonal constants) and the restricted bivariate autoregression (without dummies) as estimated by least squares. We used the residual variance estimate rather than the errors variance estimate in order to keep c_3 in the interval $(0,1)$, which we found convenient as it permits a joint evaluation with the other correlation-type decision statistics. In our Monte Carlo design, we did not allow for correlation among the two error processes and we imposed this restriction tacitly in (14). In practice, one may want to replace c_3 by a ratio of determinants.

For the seasonal cointegration problem, it is well known that LR statistics can be constructed from squared canonical correlations (for details, see LEE, 1992, and FRANCES AND KUNST, 1996). There are two canonical correlations at each frequency, appropriately conditioned on deterministic influences at other frequencies, between $\Delta_4 X_t$ and $(A(B)X_{t-1}, \cos(t\pi))$ or $(\Delta_2 X_{t-2}, \cos(t\pi/2), \cos((t-1)\pi/2))$. If the larger root is smaller than a certain boundary value, this is commonly taken as an indication that there is no cointegration. If the larger root exceeds a significance bound but the smaller root is insignificant, one may opt for a cointegrated model. If both roots are significant, one may opt for a model without seasonal unit roots at the respective frequency. A similar MD solution for the cointegration problem was outlined in KUNST (1996). Unfortunately, this is not the LR test for a joint test for cointegration at two separate frequencies. The joint LR test happens to be quite complicated and we therefore simply use geometric averages of the smaller and larger non-zero roots at the two frequencies as our decision criteria c_1 and c_2 . If $c_2 < b_2$ for some decision bound b_2 that is determined by simulation, we conclude that the ranks at both seasonal frequencies r_2 and r_3 are 0. If $c_2 > b_2$ we rest the decision on whether the matrices Π_2 and Π_3 have full rank 2 or reduced rank 1 on a comparison of the first decision criterion c_1 and a numerically determined decision bound b_1 .

In summary, we opt for $i_1 = 0$ if $c_2 < b_2$. In this case $i_2 = 0$ as $(0,1)$ is not data-admissible, hence $\hat{\kappa} = (0,0)$. If $c_2 > b_2$ and $c_1 < b_1$ then we decide $i_1 = 1$ and rest the decision on the second discrete parameter i_2 on comparing c_3 and b_3 . Finally, if $c_2 > b_2$ and $c_1 > b_1$ we decide $i_1 = 2$. $c_3 < b_3$ results in $\hat{\kappa} = (2,0)$ and $c_3 > b_3$ results in $\hat{\kappa} = (2,1)$. The effects of the possible alternative decision strategy of giving priority to the decision on i_2 at the cost of possibly ignoring indication of $i_1 = 0$ are considered in Section 5.3.

An evaluation of the loss-minimizing decision bounds $b_i, i = 1, 2, 3$ based on a Monte Carlo experiment with 50,000 replications, i.e., approximately 10,000 replications for each case, is presented in Table 2. From Table 2, we note the non-synchronous development of the bounds. In particular, b_1 does not change much as N increases from 150 to 200. Such behavior is rooted in the different rates of convergence, i.e., N for b_1, b_2

and $N^{1/2}$ for b_3 , and shows that the MD approach would be poorly substituted by any classical testing procedure with monotonously decreasing significance levels.

TABLE 2. Optimal multiple decision rules for the bivariate seasonal problem. 50,000 replications.

N	b_1	b_2	b_3	d_{\min}
100	0.110	0.167	0.772	0.1870
150	0.082	0.125	0.838	0.1292
200	0.080	0.096	0.859	0.1026

Table 3 shows what models have been identified at the decision bounds of minimum loss. Two types of processes are most vulnerable to misclassifications. Firstly, processes with neither unit-root nor deterministic seasonality (2,0) are misclassified as seasonally cointegrated processes with one stochastic seasonal component (1,0). The error frequency of this event drops from approximately 15% at $N = 100$ to about 10% for the two larger sample sizes considered. Secondly, (1,1) processes with one stochastic and one deterministic seasonal component are misclassified either as (1,0) processes — the deterministic seasonal cycle is not found — or as (2,1) processes — i.e., the stochastic seasonal cycle is ignored. The frequency of the occurrence of any of these two mistakes remains fairly constant at about 20% for $N = 100$ and $N = 150$ but drops to 16% for $N = 200$. In the first case, some of the roots are close to but not on the unit circle and hence this is equivalent to a classical ‘power’ problem. In the second case, the procedure confirms erroneous restrictions and hence the ‘optimum size’ of two partial hypothesis tests is fixed at levels of approximately 10%. Apart from these three ‘outlets’, the discrete parameter estimation is quite reliable and the frequency of some of the other possible misclassifications attains virtually zero for $N = 200$.

TABLE 3. Matching of generated and identified models at the optimum represented in Table 2. 50,000 replications, hence approximately 10,000 replications for each model class.

(a) $N = 100$

generated model	identified model				
	(0, 0)	(1, 0)	(2, 0)	(1, 1)	(2, 1)
(0, 0)	9931	16	0	49	0
(1, 0)	285	9471	12	212	11
(2, 0)	23	1502	8467	4	9
(1, 1)	89	919	16	7938	1048
(2, 1)	4	28	57	590	9319

(b) $N = 150$

generated	identified model				
model	(0, 0)	(1, 0)	(2, 0)	(1, 1)	(2, 1)
(0, 0)	9978	5	0	13	0
(1, 0)	189	9613	8	175	6
(2, 0)	6	994	8994	2	9
(1, 1)	35	591	13	7978	1393
(2, 1)	1	7	18	285	9687

(c) $N = 200$

generated	identified model				
model	(0, 0)	(1, 0)	(2, 0)	(1, 1)	(2, 1)
(0, 0)	9985	6	0	5	0
(1, 0)	89	9866	0	34	2
(2, 0)	2	1056	8947	0	0
(1, 1)	12	570	1	8409	1018
(2, 1)	0	6	17	338	9637

5 Empirical evidence

5.1 Univariate evidence

The univariate discrete estimation procedure introduced in Section 4 was applied to 18 macroeconomic time series. We used quarterly data on gross domestic product (GDP), private consumption, gross fixed investment, goods exports, wages, and a longer-term interest rate. All series are in real terms, including the interest rate which was deflated using an appropriate price index. With the exception of the interest rate, all data series are used in logarithms. Parallel data have been used for three countries: Austria (1964–1994), the Federal Republic of Germany (before unification, 1960–1988), and the United Kingdom (1957–1994). This data set coincides with the one used by KUNST and FRANCES (1996) who also provide graphical representations of the time series that show the strong seasonal effects that are present in most series. To make the procedure operable, we had to choose among two options. Firstly, the basic regression (shown above as (11))

$$\begin{aligned} \Delta_4 X_t = & \mu + \delta_1^* \cos(\pi t) + \delta_2^* \cos\left(\frac{\pi t}{2}\right) + \delta_3^* \cos\left(\frac{\pi(t-1)}{2}\right) \\ & + \pi_1 A(B) X_{t-1} + \pi_2 \Delta_2 X_{t-1} + \pi_3 \Delta_2 X_{t-2} + \varepsilon_t \end{aligned} \quad (15)$$

can be used directly, which represents a very stubborn adherence to the decision design that was also used to generate the decision bounds. Secondly, additional conditioning

may be conducted on some lags of $\Delta_4 X_t$ in order to accommodate the (typically high) autocorrelation in the error process. We report results from both variants. In the latter case we used four lags for all series and summarize evidence on serial correlation in the errors by the portmanteau statistic Q due to Ljung and Box. This standardization eases the comparison across series though it may not correspond to parsimonious time-series models for most series.

Table 4 shows the main results. We give the first decision statistic c_1 which is independent of error scales, the second decision statistic c_2 which had to be re-scaled by division through the estimated standard deviation of the errors $\hat{\sigma}$, the discrete parameter estimate following from the decision statistics and from Table 1, using $N = 100$ for Germany, $N = 150$ for the United Kingdom, and interpolating between the two values for Austria. In borderline cases, two possible estimates for the discrete parameter $\kappa = (i_1, i_2)$ are given. In the final column we display the marginal significance of Q , as stated above.

Notice that the main results differ from those of previous research based on classical methods. Most series show deterministic seasonality (0,1). In the Austrian data we find seasonal unit roots (1,0) only in consumption and interest, in both cases only in the augmented test version. In the German data we find seasonal unit roots also in the GDP and wages variables, also in the augmented version only. None of the British series is classified as having seasonal unit roots. British wages and interest are classified as (0,0). The most conspicuous results are the deterministic nature of seasonality in investment, which may be explained by the large share of the construction sector which is hit by the climatic seasonal cycle, and the contradiction in the Austrian data between total GDP and one of its main components, private consumption. The sum of a (1,0) and a (0,1) variable is certainly (1,0) but the deterministic component in the added (0,1) variable may be so strong that it disables statistical recognition of the unit roots in the aggregate series.

TABLE 4. Estimates for discrete seasonal parameters

(a) Austrian data

	c_1	c_2	$\hat{\kappa}$	$p(Q)$
GDP	0.64171	3.68384	(0, 1)	0.00001
	0.40382	2.96462	(0, 1)	0.54605
consumption	0.26371	1.51499	(1, 0) – (0, 1)	0.00698
	0.17916	0.93987	(1, 0)	0.30315
investment	0.69244	4.92727	(0, 1)	0.80509
	0.54073	4.17534	(0, 1)	0.95806
exports	0.80326	0.75947	(0, 1)	0.98827
	0.78375	0.78911	(0, 1)	0.99871
wages	0.77809	7.44742	(0, 1)	0.00004
	0.50134	5.85293	(0, 1)	0.50428
interest	0.28200	0.57415	(0, 1)	0.00007
	0.21916	0.53456	(1, 0)	0.30913

(b) German data

	c_1	c_2	$\hat{\kappa}$	$p(Q)$
GDP	0.50964	0.66334	(0, 1)	0.00273
	0.24993	0.53925	(1, 0)	0.99217
consumption	0.64582	2.61653	(0, 1)	0.00000
	0.22901	1.36622	(1, 0)	0.78246
investment	0.58391	2.47899	(0, 1)	0.00742
	0.33720	1.43830	(0, 1)	0.99698
exports	0.76982	1.15883	(0, 1)	0.07252
	0.69788	1.01670	(0, 1)	0.38479
wages	0.57569	2.22417	(0, 1)	0.00000
	0.15816	0.96554	(1, 0)	0.79787
interest	0.19854	1.47223	(1, 0)	0.00381
	0.19931	1.13530	(1, 0)	0.01730

(c) UK data

	c_1	c_2	$\hat{\kappa}$	$p(Q)$
GDP	0.55363	0.84969	(0, 1)	0.00006
	0.31064	0.49970	(0, 1)	0.20413
consumption	0.65677	0.98788	(0, 1)	0.00000
	0.33142	0.54039	(0, 1)	0.09775
investment	0.70979	1.46300	(0, 1)	0.03056
	0.35532	0.91398	(0, 1)	0.98534
exports	0.60081	0.71460	(0, 1)	0.31281
	0.59381	0.74832	(0, 1)	0.35700
wages	0.82293	0.47569	(0, 0)	0.00408
	0.51865	0.35802	(0, 0)	0.49176
interest	0.77228	0.29415	(0, 0)	0.68022
	0.84159	0.32683	(0, 0)	0.84557

5.2 Bivariate evidence

For the bivariate examples, pairs of real wages and real private consumption series for the three countries, i.e., Austria, Germany, and the United Kingdom were used. Seasonal patterns in wages may generate similar seasonal patterns in spending, hence seasonal cointegration seems to be the most interesting in these pairs. As in the last subsection, we report two sets of values for each case, one with enough conditioning lags to eliminate residual autocorrelation and one without conditioning. Noting that univariate analysis has found no evidence for seasonal unit roots in Austrian wages and both British series, the results in Table 5 appear surprising. In the United Kingdom, all seasonality is attributed to deterministic cycles and seasonal constants if no lag augmentation is used. After accommodating for serial correlation, a (1,1) model with seasonal cointegration is selected. Note that the (1,1) model has a seasonal unit root that is not found by the univariate selection procedure. In Austria, the results also support seasonal cointegration in the stochastic part whereas the evidence on the importance of seasonal dummies is not very pronounced. Depending on the interpolation between the b_3 bounds for $N = 100$ and $N = 150$, the c_3 value of 0.779 is ambiguous. In Germany, there is no seasonal cointegration but freely developing unit-root seasonality if lags are accounted for. The statistic c_3 would point to the presence of deterministic seasonality but this decision is overruled by the requirement of data admissibility.

TABLE 5. Estimates for discrete seasonal parameters in bivariate systems consisting of real wages and real private consumption.

Country	no. of lags	c_1	c_2	c_3	$\hat{\kappa}$
Austria	1	0.057	0.180	0.779	(1,0)-(1,1)
	0	0.076	0.289	0.573	(1,1)
Germany	2	0.032	0.060	0.636	(0,0)
	0	0.066	0.252	0.196	(1,1)
United Kingdom	2	0.058	0.133	0.766	(1,1)
	0	0.158	0.257	0.442	(2,1)

In summary, the evidence is in conflict with the univariate analysis. In univariate series, a completely deterministic time series description is preferred in many variables. In the bivariate series, the importance of seasonal unit roots in Austria and Germany is underscored. This puzzle could be solved by the observation that the univariate marginal processes are unit-root processes without seasonal constants indeed but can be described satisfactorily over longer time intervals by simple seasonal constants plus a sufficiently rich stationary cyclical structure. In other words, the changes in the seasonal structure are too slow to justify the use of seasonal unit roots models for univariate series, particularly as these unit roots cannot be used simultaneously with seasonal

dummies. A model that captures 90% of the seasonal variation, say, is preferred even if it is the ‘wrong’ model. In contrast, bivariate models suggest the joint exploitation of the explanatory powers of seasonal dummies and of seasonal unit roots by restricted seasonal cointegration.

5.3 Sensitivity of the results

The empirical results presented up to here depend on the design of the MD analysis. Slight changes in that design may have strong effects on the outcome. A decision maker may feel more comfortable if the main parameter estimates prove robust toward those changes. In classical hypothesis testing, this sensitivity is checked routinely by embedding the general model as defined on the primary parameter space Θ in an even more general model by extending it to a larger primary parameter space $\Theta^* \supset \Theta$ and by considering the decision problem of whether the data still select Θ . For example, VAR models with Gaussian random errors are embedded in VAR models with non-Gaussian errors or in VAR models with first-order autocorrelated Gaussian errors. In MD analysis, this kind of sensitivity check is just one type of possible procedures and maybe not even the most amenable to its spirit. One may e.g. consider the following types of sensitivity checks:

1. Sensitivity with regard to adding or deleting a hypothesis
2. Sensitivity with regard to splitting or merging specified classes
3. Sensitivity with regard to the distributional window
4. Sensitivity with regard to opening or closing the structural part of the model window
5. Sensitivity with regard to the loss function
6. Sensitivity with regard to within-class priors or coordinate changes
7. Extending the class of decision rules with the aim of further gains in risk

For the metaphorical usage of the word ‘window’ for the basic parametric most general model considered, cf. POIRIER(1995). We adopt this metaphor in order to express our conviction that the validity of the most general model is *not testable* but rather represents a way of viewing the world of data, i.e. a ‘window’.

From this menu, only very few experiments can be conducted routinely although any item may be of particular interest. Here, we conduct three sensitivity experiments.

Firstly, we consider the effects of cointegration at $\omega = 0$ which was excluded from the basic simulation design and is often observed in macroeconomic time series. We note, however, that frequency-zero cointegration is not found by standard statistical tools in our bivariate examples.

The cointegration experiment is of type 4 according to the above list. In all classes of bivariate models, 50% of the generated processes are allowed to be cointegrated. A technical problem is that, in the presence of cointegration at the frequency 0, hard rejection must be abandoned as a principle to generate multivariate uniform distributions. The chance to hit upon a non-explosive process of the form

$$\begin{aligned} \Delta_4 X_t = & (1, \alpha_{12})' \text{diag}(\rho_{11}, 0)(1, \beta_{12}) S(B) X_{t-1} \\ & + \mathbf{\Pi}_2 A(B) X_{t-1} + \mathbf{\Pi}_3 \Delta_2 X_{t-2} + \varepsilon_t \end{aligned} \quad (16)$$

is almost negligible for non-singular seasonal cointegration matrices $\mathbf{\Pi}_2$ and $\mathbf{\Pi}_3$, i.e., in the case of no seasonal unit roots, if all matrices are still built upon their eigenvalues in the admissible ranges, i.e., $(-2, 0)$ for the frequency 0 and $(0, 2)$ for the seasonal frequencies. There are two conceivable solutions to this problem of additional cross-restrictions among the three frequencies. Firstly, one may rely on a uniform prior on the parameter ρ_{11} only and generate $\mathbf{\Pi}_2$ and $\mathbf{\Pi}_3$ by some prior distribution on the coefficient space. Secondly, one may enforce the loading vectors and cointegrating vectors to be the same across frequencies and restrict all five roots by a stability condition. In fact, this stability condition turned out to be simply

$$-\rho_{11} + \rho_{21} + \rho_{22} + \rho_{31} + \rho_{32} \leq 2$$

and a uniform distribution on this 5-variate area can be generated easily, so the latter solution was adopted. Here, ρ_{21} and ρ_{22} denote the eigenvalues of $\mathbf{\Pi}_2$ and ρ_{31} and ρ_{32} denote the eigenvalues of $\mathbf{\Pi}_3$. This has the advantage that the very same design can be used for the seasonally cointegrated classes. However, whereas the classes (2,0) and (2,1) are unrestricted, a restriction is imposed on (1,0) and (1,1), as the seasonal cointegration vectors have to be the same. This circumstance may be responsible for the observed increase in expected risk.

Table 6 summarizes the results of this cointegration sensitivity check for $N = 100$. It is obvious that the smaller non-zero seasonal root in the non-seasonal classes (2,0) and (2,1) is less likely to be found in cointegrated processes. The MD procedure lowers the corresponding decision bound in order to avoid larger losses for these misclassifications but pays with a strong increase of misclassifications of the (1,1) processes with seasonal cointegration and deterministic dummy cycles. However, the main results of the last

subsection turn out to be insensitive to the bounds changes. Apart from the experiment reported in Table 6, some more unreported variants were simulated but the main outcome was similar.

TABLE 6. The effects of cointegration at the zero frequency. $N = 100, 50,000$ replications. Matrix of errors and optimum decision bounds. Standard bounds from Table 2 are given in brackets.

generated model	identified model				
	(0, 0)	(1, 0)	(2, 0)	(1, 1)	(2, 1)
(0, 0)	9851	40	1	107	4
(1, 0)	251	9033	447	151	121
(2, 0)	110	3107	6726	12	41
(1, 1)	88	701	81	6022	3108
(2, 1)	18	48	119	734	9079

$$b_1 = 0.078 [0.110] \quad b_2 = 0.153 [0.167] \quad b_3 = 0.778 [0.772] \quad d_{min} = 0.2928 [0.1870]$$

A second sensitivity experiment is of type 7 according to the above rudimentary classification. Instead of giving priority to decisions on the basis of the decision bounds b_1 and b_2 and to allot the area where deterministic seasonality is found according to the bound b_3 but no seasonal cointegration to the class (0,0) with no cointegration at seasonal frequencies and no deterministic seasonality, one may consider to assign priority to the bound b_3 and to allot the doubtful area to (1,1). This strategy led to a marked deterioration in the minimum risk. This deterioration was also observed in a parallel experiment where the different decision rule was adopted in the presence of cointegration at $\omega = 0$. These results are summarized in Table 7. Although necessarily some misclassification events were reduced, it is obvious that the alternative decision rule which gives priority to the finding of deterministic seasonality is worse than the one used for generating Tables 2 and 3. Such outcomes may help to solve disputes within classical hypothesis testing about whether e.g. ‘specific-to-general’ testing is to be preferred to ‘general-to-specific’ testing. The testing sequences can simply be evaluated by MD analysis and the achieved minimum expected risks are then compared. By a similar experiment in the spirit of class 7, it can e.g. be established easily that any interchange of decisions on the bounds b_1 versus b_2 does not incur any important shifts in expected risk. We have mentioned above that a similar unreported sensitivity experiment was also conducted for the univariate problem and resulted in a similar deterioration.

TABLE 7. The effects of giving priority to decisions on the inclusion of seasonal constants. 50,000 replications. Bounds and risks from Tables 2 or 6 are shown in parentheses.

N	b_1	b_2	b_3	d_{\min}
100	0.110 (0.110)	0.154 (0.167)	0.739 (0.772)	0.2510 (0.1870)
150	0.082 (0.082)	0.116 (0.125)	0.803 (0.838)	0.1734 (0.1292)
200	0.080 (0.080)	0.087 (0.096)	0.844 (0.859)	0.1302 (0.1026)
with cointegration at $\omega = 0$				
100	0.078 (0.078)	0.144 (0.153)	0.743 (0.778)	0.3653 (0.2928)

A third sensitivity experiment was conducted to investigate the influence of potential outliers on the identified decision bounds and the decision risk. We used the arguably extreme assumption of replacing the Gaussian white-noise innovations of the basic model by independent draws from a standard Cauchy distribution. This is a sensitivity experiment of type 3 according to our tentative classification.

For Cauchy errors, none of the three discriminating statistics approximates a likelihood-ratio statistic and, additionally due to the now missing moments, one cannot expect to keep the risk at the low level of the other experiments. Nevertheless, the results summarized in Table 8 are disconcerting. The expected MD risk exceeds 1, which means that ‘on average’ the optimum decision made by the procedure is incorrect. The table of actual decisions reveals that this high risk is caused by the inability of the procedure to correctly identify deterministic cycles against the highly volatile background of Cauchy errors. In contrast, the cointegrating rank i_1 is still identified with a reliable precision.

TABLE 8. The effects of Cauchy innovations. $N = 100$ and 50,000 replications. Misclassifications and decision bounds.

generated model	identified model				
	(0, 0)	(1, 0)	(2, 0)	(1, 1)	(2, 1)
(0, 0)	9546	20	0	386	44
(1, 0)	426	5412	13	3593	547
(2, 0)	64	959	8193	95	694
(1, 1)	393	3994	2	5006	615
(2, 1)	42	674	5513	352	3417

$$b_1 = 0.088 [0.110] \quad b_2 = 0.208 [0.167] \quad b_3 = 0.888 [0.772] \quad d_{\min} = 1.380 [0.1870]$$

Just for the sake of an experiment, let us assume that the Austrian, German, and British economies are driven by Cauchy innovations and let us apply the corresponding decision bounds. Without lag augmentation, the Austrian and German wage-consumption systems are then classified as (1,1) and the British system as (2,1), i.e. there is no change

relative to the main results. With lag augmentation, however, they are classified as $(0,0)$, with two independent stochastic seasonal components but without deterministic seasonality. Although the overall MD risk is high, we note that the model class $(0,0)$ can be identified with comparative precision in the presence of Cauchy errors. Hence, the importance of seasonal unit roots is enhanced even under the implausible assumption of an extreme outliers distribution.

6 Discussion

In summary, the results of the MD analysis indicate that seasonal unit roots are a convenient way to model time-changing seasonality in macroeconomic data sets whereas the routine insertion of seasonal dummy constants into seasonal time-series models is not supported. While the univariate evidence supports purely deterministic seasonality in many cases, the bivariate structures point to the importance of seasonal unit roots. In the case of German wages and consumption, it even seems that all seasonal cycles may be caused by unit roots without deterministic influence. These stochastic cycles resemble deterministic patterns in the univariate marginals, as seasonal covariation between the variates is not taken into account and starting patterns for the seasonal cycles are very volatile. This very pronounced seasonal variation is unstable in the long run, however, and entirely different patterns will emerge eventually in the distant future.

In univariate models, it may be sufficient to use deterministic dummies, particularly as one cannot use them in conjunction with unit roots in order to sustain data admissibility. In multivariate models, the joint occurrence of strong starting patterns, which would suggest seasonal dummies, and of time-changing seasonality, which would suggest seasonal unit roots, can be accommodated by seasonal cointegration and restricted seasonal dummies.

It is tempting to identify deterministic and stochastic seasonal cycles with basic sources of seasonality, such as climate or culture. Such an economic interpretation should be conducted with care only. New technologies may alter the response to the temperature cycle over the year, and consumer preferences due to cultural traditions may experience long-run changes. Presumably, all seasonal effects in the economy, except for primary measurements of meteorological data, are changing slowly but permanently over time.

A rewarding direction for future research, which is beyond the scope of this paper, would be to include the periodic seasonal model suggested by FRANCES (1996) and GHYSELS ET AL. (1996) in the set of model classes. Periodic seasonal models have properties that bridge unit-roots and deterministic-cycles models. The number of considered classes is

certainly the most severe restriction for the MD analysis but the technique as such can be extended in a straightforward way.

The MD approach used to decide on the nature of seasonality is new and therefore may demand for some justification with respect to the details of its structure. In particular, the MD technique is Bayesian in spirit and thus its priors have to be motivated. We use two kinds of priors, those on the decision set Ξ and those on the classes Θ_j . Since Ξ is finite, a uniform discrete prior is natural and also widely used due to the principle of insufficient reason. However, it is not universally accepted and the well-known partitioning paradox indicates that care must be taken in formulating the basic classification problem, i.e., in constructing Ξ and the partitioning of Θ into the classes Θ_j . Within these classes, we chose to define uniform priors on natural parameterizations, in stark contrast to the emphasis on ‘reference priors’ in the current Bayesian literature (see, e.g., KASS AND WASSERMAN, 1996). Our choice was motivated by two observations. Firstly, the widespread skepticism against uniform priors is mainly rooted in the argument of the arbitrariness of parameterization, whereas in our examples a preference for a certain parameterization is rooted firmly in the literature. For example, it would be extremely ‘unnatural’ to parameterize autoregressive models by anything else than either coefficients or characteristic roots. We made the choice between roots and coefficients spaces as guided by convenience or by the need to construct a bounded parameter space. Secondly, whereas an updating of posteriors can be conducted on the basis of improper priors, the random numbers needed for the MD technique cannot be generated from such priors and hence priors are required to be probability distributions, thus excluding many reference priors encountered in the literature. We note that the MD priors are not the foundation of a Bayesian updating of posteriors but are weighting priors used to define an overall loss expectation.

The Bayesian approach is certainly not new. What is new, however, is the central position of the loss function in the MD problem. The loss function can be modified in such a way that certain decision errors are minutely avoided at the cost of an increase in other decision errors and hence it can be easily adjusted to the taste of the decision maker. Its shape is more important in small samples than in larger ones where most decisions made tend to be correct. For natural restrictions on the shape of the loss function, see the Appendix.

The author wishes to thank Wolfgang Polasek and the participants of a seminar in St.Oswald, Upper Austria, for helpful comments. The usual proviso applies. All computer exercises have been conducted on the basis of FORTRAN codes written by the author and some subroutines of the NAGLIB library. All computer programs are available from the author upon request.

Appendix: Consistency of the decision procedures

The multiple decision procedures used in this paper are used to discriminate among certain types of hypotheses that correspond to subsets of the parameter space Θ , which in turn is a subset of the p -dimensional Euclidean space \mathbf{R}^p . In general, the minimization of a loss function does not guarantee that asymptotically, i.e., for $N \rightarrow \infty$, the probability of a correct decision converges to 1. This important property, the consistency of the decision procedure, depends on the topological properties of the subsets Θ_i , on the form of the specified loss function, on the specified prior weighting distributions, and, of course, on the consistency in the estimation procedure of the parameter $\theta \in \Theta$. With regard to the latter three points, however, it is natural to assume the following:

- (a1) The loss function attains a value of 0 if and only if the decision is correct.
- (a2) The loss function is bounded.
- (b1) The weighting prior allots $1/k$ to each of the k subsets of Θ , among which the decision is searched.
- (b2) For every $\varepsilon > 0$ there exists a compact subset of Θ_i that contains a mass of $1 - \varepsilon$ of the conditional weighting prior on Θ_i .
- (c) For every $\varepsilon > 0$ and every $\theta \in \Theta$ there exists an integer $N(\varepsilon, \theta)$ such that an ε -neighborhood of θ contains a mass of $1 - \varepsilon$ of the distribution of $\hat{\theta}(N)$ for all $N > N(\varepsilon, \theta)$, where $\hat{\theta}(N)$ denotes the estimate from a sample of size N .

Now suppose we have a binary decision among Θ_1 and Θ_2 with $\Theta_1 \cup \Theta_2 = \Theta$. If Θ_1 is closed and Θ_2 is open in Θ , let us consider the following decision rule:

- (a) opt for $\theta \in \Theta_1$ if $\min_{\theta \in \Theta_1} |\hat{\theta} - \theta| < \delta$
- (b) opt for $\theta \in \Theta_2$ otherwise

If, for $N \rightarrow \infty$, $\delta \downarrow 0$, then the procedure is consistent for every $\theta \in \Theta_2$. Θ_2 contains an ε -neighborhood and this neighborhood contains a share of the probability mass that is arbitrarily close to 1. This just describes the ‘test consistency’ of classical tests. However, for $\theta \in \Theta_1$ one has to be more careful. The probability of incorrect decisions for Θ_1 attains a maximum on the closure of the δ -neighborhood of Θ_1 . This is the ‘size’ or ‘risk level’ of the decision procedure. If Θ_1 is bounded, then it is also compact and for sufficiently large N , its δ -neighborhood contains an arbitrarily high proportion of

probability mass. It is then possible to decrease δ with $N \rightarrow \infty$ so slowly that this proportion converges to 1 and the procedure is consistent.

If Θ_1 is unbounded, it is not compact and this construction may not be possible. Note, however, that in all problems treated in this paper Θ_1 can be represented as $\Theta_1 = \Theta_{1,1} \times \Theta_{1,2}$ with the bounded set $\Theta_{1,1}$ representing the main parameters of interest and the unbounded set $\Theta_{1,2}$ representing the ‘nuisance’ parameters. It is then necessary to restrict the influence of the nuisance in such a way that it does not prevent the sort of uniform convergence we just considered. Fortunately, this requirement is fulfilled in the examples given in the main text. Next we consider the behavior of the expected risk. The expected risk is a global concept and it may well be that the procedure is locally consistent but that it does not achieve an expected risk of 0 asymptotically. Consider

$$ER = \int_{\Theta_1} R(\theta) f_1(\theta) d\theta + \int_{\Theta_2} R(\theta) f_2(\theta) d\theta = J_1 + J_2$$

where f_1 and f_2 denote the two weighting priors on the decision subsets and $R(\theta)$ is the expected loss if θ is the true parameter and the above decision rule is used. J_1 can be made arbitrarily small if (a2) holds and if either Θ_1 is compact or other conditions guarantee the existence of the test described above. J_2 can be decomposed into two parts, the integral $J_{2,1}$ over the compactum \mathcal{C} and the integral $J_{2,2}$ over the remainder $\Theta_2 \setminus \mathcal{C}$. $J_{2,2}$ can be made arbitrarily small using (b2) and (a2). $J_{2,1}$ is calculated over the compact set \mathcal{C} and we can make the expected risk over any compact set arbitrarily small by choosing N large enough. Let us denote the maximum of the loss function by M and fix an arbitrarily small $\varepsilon > 0$. Then we can choose a compact subset $\mathcal{C}(\varepsilon)$ containing a mass of $1 - \varepsilon(3M)^{-1}$, then we fix the ‘significance level’ at $\varepsilon(3M)^{-1}$ and choose N_1 , then we choose N_2 in order to make the expected risk on the compact subset of Θ_2 smaller than $\varepsilon/3$. We finally take N as the maximum of N_1 and N_2 and $ER < \varepsilon$ as required.

Note that (b2) can also be exploited to guarantee an asymptotic risk of 0 in those cases where Θ_1 is unbounded and a test for a given significance level cannot be constructed. Then $ER \rightarrow 0$ but the decision procedure as such will not be consistent. Such extensions are possible for many problems that do not fit into our topological assumptions where locally inconsistent decision rules with an asymptotic zero risk can be found. Note that entire areas of inconsistent behavior within Θ can be avoided arbitrarily by appropriate weighting priors, hence some effort has to be made in order not to ‘cheat’ in the elicitation step. In this paper we restrict attention to priors with their support equal to Θ and uniform on a widely accepted parameterization of each Θ_j .

We have seen that by assuming compactness or uniform convergence we can construct a decision procedure that is consistent and takes expected risk to 0. Conversely, if we

minimize expected risk for finite N we minimize ER also in the limit. Since $ER > 0$, we obtain a procedure with an asymptotic risk of 0, which is also optimal in a sense as it minimizes the risk for finite samples. There is no guarantee, however, that it is optimal among all decision procedures in the sense of minimizing finite-sample risk as a different parameterization may achieve a similar procedure with lower risk. The problem of optimal parameterization is a difficult one to solve. In the binary problems, it is tantamount to the conception of a most powerful test and has led to the development of likelihood-ratio tests. Without pretending to having achieved this lowest-risk bound, it appears useful to concentrate on parameterizations that are roughly equivalent to likelihood-ratio statistics, which is what is done in the examples of the main text.

The examples of the main text, however, are not binary but multiple decision problems. The problems have very special structures that allow us to use simple extensions of the above ideas to demonstrate that consistent decision procedures also exist. It is easy to see that this is possible, for example, if all binary subproblems fall into the open/closed framework outlined above and if compactness or uniform convergence holds. In contrast to the binary problem, however, an analytic evaluation of risk minimization is usually impossible and therefore we took refuge to numerical optimization.

References

- DAVIDSON, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- FRANSES, P.H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press.
- FRANSES, P.H., HOEK, H., and PAAP, R. (1997). 'Bayesian Analysis of Seasonal Unit Roots and Seasonal Mean Shifts,' *Journal of Econometrics* 78, 359–380.
- FRANSES, P.H. and KUNST, R.M. (1996). On the Role of Seasonal Intercepts in Seasonal Cointegration. Discussion Paper TI 96-175/7, Tinbergen Institute Rotterdam.
- GHYSELS, E., HALL, A., and LEE, H.S. (1996). 'On Periodic Structures and Testing for Seasonal Unit Roots,' *Journal of the American Statistical Association* 91, 1551–1559.
- HENDRY, D.F. (1995). *Dynamic Econometrics*. Oxford University Press.
- HYLLEBERG, S. (1992), ed. *Modelling Seasonality*. Oxford University Press.
- HYLLEBERG, S., ENGLE, R.F., GRANGER, C.W.J., YOO, B.S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics* 44, 215–238.
- JOHNSON, N.L. and KOTZ, S. (1970). *Continuous Univariate Distributions 1*. Wiley.
- KASS, R.E. and WASSERMAN, L. (1996). 'The Selection of Prior Distributions by Formal Rules,' *Journal of the American Statistical Association* 91, 1343–1370.
- KUNST, R.M. (1996). Estimating Discrete Parameters: An Application to Cointegration and Unit Roots. *Oesterreichische Zeitschrift für Statistik* 25, 7–32.
- KUNST, R.M. and FRANSES, P.H. (1996). 'The Impact of Seasonal Constants on Forecasting Seasonally Cointegrated Time Series,' Report 9666/A, Erasmus University Rotterdam. *Journal of Forecasting* (forthcoming).
- LEE, H.S. (1992). 'Maximum Likelihood Inference on Cointegration and Seasonal Cointegration,' *Journal of Econometrics* 54, 1–49.
- POIRIER, D.J. (1995). *Intermediate Statistics and Econometrics. A Comparative Approach*. MIT Press.

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