Costationarity of locally stationary time series

Alessandro Cardinali†
University of Bristol, U.K.
Guy P. Nason
University of Bristol, UK

1. Introduction

Loosely speaking, a stationary time series is one whose statistical properties remain constant over time, whereas the statistical properties of *locally* stationary (LS) time series change slowly over time. As a consequence, LS series can appear stationary when examined close up, but appear non-stationary when examined on a larger scale. Priestley (1983) and Nason and von Sachs (1999) review locally stationary (LS) time series.

Recently, Dahlhaus and Polonik (2006) introduced a general infinite order time-varying moving average (MA) representation for LS processes:

$$X_{t,n} = \sum_{r=-\infty}^{\infty} a_{t,n}(r)\epsilon_{t-r}, \tag{1}$$

where ϵ_i are assumed to be i.i.d. with zero mean and unit variance, where the $\{a_{t,n}\}$ sequence satisfies a number of technical conditions, see also Dahlhaus and Polonik (2009), and $t = 1, \ldots, T$. The LS processes introduced by Dahlhaus (1997), referred to below as locally stationary Fourier (LSF) processes, and the LS wavelet (LSW) processes of Nason et al. (2000) (NvSK) are examples of infinite order time-varying MA processes.

The (second-order) statistical content of many kinds of LS time series is quantified by a time-varying spectrum. For example, the LS processes of Dahlhaus and Polonik (2006) and LSF processes have a spectrum denoted by $f(z, \lambda)$, where z is time and λ is frequency. LSW processes have a spectrum denoted by $S_j(z)$, where z is time and j indicates scale or frequency band. Here, we denote the generic time-varying spectrum by $p(t, \nu)$. Here t is time and ν is another parameter, or set of parameters. For LSF processes $p(t, \nu) = f(t, \nu)$ and ν would be the usual frequency $\nu \in (-\pi, \pi)$. For LSW processes $p(t, \nu) = S_\nu(t)$, where $\nu = j \in \mathbb{N}$ would be the usual scale index. For stationary processes the spectrum $p(t, \nu) = p(\nu)$ does not depend on time. So far, we are not prescriptive about the domain of the indices t, ν, as these are prescribed according to the precise modelling context. For example, in the LSW and LSF cases above, the time $z \in (0, 1)$ is rescaled time $z = t/T$, for the oscillatory processes of Priestley (1983) time could be any $t \in \mathbb{R}$. (In the following we use z and t interchangeably).

What is common to all these models is that when the (potentially) time-varying spectrum, $p(t, \nu)$, is a constant function of t then, and only then, is the process second-order stationary. For brevity, we henceforth assume stationary means second-order stationary.

†*Address for correspondence*: Alessandro Cardinali, CTEU, Bristol Heart Institute, Bristol Royal Infirmary, Marlborough Street, Bristol, BS2 8HW. Email: A.Cardinali@bristol.ac.uk

2. Costationarity

2.1. Introduction

The main question we pose here is: given two LS time series, X_t, Y_t is it possible to find a linear combination, $Z_t = \alpha X_t + \beta Y_t$, such that Z_t is stationary, where α, β are deterministic? (Obviously, if X_t or Y_t are stationary then the problem is trivial.) Practically, it turns out that this concept is too restrictive because the statistical evolution of X_t, Y_t is such that, for real data, there is usually no *constant* (α, β) for which Z_t is stationary. Consequently, we seek deterministic *time-varying* linear combinations which result in Z_t being stationary. This leads us to the following definition:

DEFINITION 1. Let $\{X_t\}$, $\{Y_t\}$ be LS with time-varying spectra $p_X(t, \nu)$, $p_Y(t, \nu)$ of bounded variation respectively. Then $\{X_t\}$ and $\{Y_t\}$ are said to be *costationary* if there exist deterministic complexity constrained sequences (with constraint C) (α_t, β_t) such that

$$Z_t = \alpha_t X_t + \beta_t Y_t, \tag{2}$$

where $\{Z_t\}$ is a stationary process. The (α_t, β_t) sequences are called costationary vectors.

REMARK 1. Definition 1 refers to (α_t, β_t) as being complexity constrained with constraint C. The reason for the constraint is that given *any* observed time series X_t and Y_t, you can choose any stationary process Z_t and any deterministic α_t you like, then just define $\beta_t = (Z_t - \alpha_t X_t)/Y_t$ (assume wlog $Y_t \neq 0$). In other words, costationarity can always be made to 'work' in purely formulaic terms with no constraint on the (α_t, β_t), but those solutions might be pathological and not interpretable.

On the other hand, the most interpretable solutions would be $\alpha_t = \alpha, \beta_t = \beta$, but, as indicated just above Definition 1, the constancy of the α, β here is too restrictive. Hence, we look for (α_t, β_t) that are, in some sense, 'the least complex but still achieve costationarity'. In practice, there are various ways in which we can achieve this. We might constrain (α_t, β_t) to be samples from a smooth function with smoothness constraint, e.g. integrated squared second derivative, C, as in Section 4.2. Alternatively, we might constrain (α_t, β_t) to be piecewise constant functions and impose an upper limit, C, on the number of breaks for both functions. Below, in Section 4.1, we admit piecewise constant functions with breaks laid out on a dyadic grid and C controls the fineness of grid. However, the precise choice is up to the user and influenced by the particular characteristics of the problem at hand.

The complexity of the resultant solutions (α_t, β_t) is an indicator of the character of the relationship between X_t and Y_t. A low complexity solution indicates that the two time series have "universal" relationship spread widely over time, a high complexity solution (where α_t, β_t have to vary a lot to achieve costationarity) indicates a more frequently changing relationship.

Finally, the particular costationary solution mentioned above is not unique because ϕZ_t is stationary if Z_t is stationary and ϕ is a constant, and hence $(\phi\alpha_t, \phi\beta_t)$ would also be a costationary solution. We shall say more on more non-trivial forms of uniqueness in Section 3.1.

REMARK 2. In some *applications*, for example the wind power application in Section 4.2, it can be useful to find (α_t, β_t) that make Z_t *closer* to stationary than either of the original X_t, Y_t series, although Z_t itself might *not be* stationary. Some of the mathematical conditions on the parameters of the systems will implicitly be tightened up for specific examples later. However, the definition is concrete and can be deployed: either theoretically, with mathematically defined LS processes and measure of stationarity as in Examples 3 and 4, or practically, with real data and a statistical hypothesis test of stationarity as in Section 4. The next section motivates the costationarity concept, provides more explanation and gives some examples.

2.2. Motivation, elaboration and examples

There are four main reasons why discovering costationarity is important: (i) learning/discovery of any costationary relationship itself, (ii) estimating the strength of any such relationship, (iii) using the derived stationary series, in some applications in preference to either of the original series and (iv) using the relationship to learn about X_t from data on Y_t or vice versa. For the last point, rearranging (2) yields $X_t = \alpha_t^{-1}(Z_t - \beta_t Y_t)$. Later, we present two applications, portfolio selection in finance and wind energy volatility mitigation, which both show examples of how it can be advantageous to use the stationary derived series, Z_t, in preference to the original series X_t or Y_t.

Additionally, costationary relationships induce an "error-" or "variance-correction" formula which can give further valuable information on the nature of the relationship between the series, more details are given in Section 2.3.

REMARK 3. To find costationarity, given arbitrary LS X_t and Y_t, one could adopt a model-independent approach to search for costationary vectors by applying the covariance operator to Z_t in (2) yielding

$$\gamma_Z(\tau) = \mathrm{cov}(Z_t, Z_{t+\tau}) = \mathrm{cov}(\alpha_t X_t + \beta_t Y_t, \alpha_{t+\tau} X_{t+\tau} + \beta_{t+\tau} Y_{t+\tau}), \tag{3}$$

for all t, τ, where $\gamma_Z(\tau)$ is the autocovariance function of some unknown stationary process. Even if Z_t were known *a priori*, solving this large set of nonlinear simultaneous equations for suitable (α_t, β_t) would be difficult both analytically or numerically. The actual situation is even more tricky since Z_t is, in general, unknown. Hence, we do not pursue this avenue further here. Our approach is via models, e.g. LSF, LSW and their associated spectral quantities which turns out to be computationally feasible as described next.

REMARK 4. Our overall approach evaluates whether the spectrum of Z_t, $p_Z(t, \nu)$, is a constant function of time for a *given* set of vectors (α_t, β_t). If it is, then X_t and Y_t are costationary. Given actual data realizations of X_t, Y_t for $t = 1, \ldots, T$ we adopt a 'projection pursuit' approach to discovering flat combination spectra. Our method is fully described in Section 3, but see Friedman and Tukey (1974), Jones and Sibson (1987) or Nason (2001) for general projection pursuit references. Our "projection index" measures the stationarity of Z_t, via the constancy of its spectrum, $p_Z(t, \nu)$, and we optimize the index over the vectors (α_t, β_t) using numerical methods. As with projection pursuit, our method shares the advantage of overcoming a "curse of dimensionality" in that rather than solve the multiple nonlinear sets of simultaneous equations mentioned in Remark 3, assessment of the stationarity of Z_t is always a univariate problem irrespective of the number of LS time series one wishes to interrogate for costationarity. To emphasize, we could extend the concept in Definition 1 to more than one time series and the statistical problem would still be evaluation of the time-constancy, or not, of the single univariate spectrum $p_Z(t, \nu)$. Hence, finding the solution might be more computationally intensive, but not conceptually more difficult.

Given analytical process specifications for X_t and Y_t, as in the examples below, one can investigate theoretically whether a costationary solution exists, or failing that, investigate the problem using numerical methods. Again, the problem of searching for a flat spectrum is considerably easier than solving the multiple nonlinear equations mentioned above.

REMARK 5. Costationarity is analogous to and inspired by, but not the same as, the cointegration of Engle and Granger (1987), where a linear combination of integrated series is sought that makes the combination stationary. These works fall into the general category of *comovements*, neatly summarized by Croux et al. (2001) who also summarize the key works at that time, Engle and Granger, the codependence of Gourieroux and Peaucelle (1992) and the common features

of Engle and Kozicki (1993). A recent review of common features can be found in Urga (2007). Other related works are cointegration relationships with structural breaks, a form of change, see, for example, Johansen et al. (2000).

One difference between cointegration and costationarity, as formulated here, is that cointegrating vectors are, classically, constant functions of time. However, Bierens and Martins (2010) have recently extended cointegration to include time-varying cointegrating vectors where, as here, such vectors have to be complexity penalized in some way. Another major difference between the two is that cointegration is concerned with first-order nonstationarity whereas costationarity is concerned with second-order.

We briefly recall the definitions of LSF and LSW processes to provide some context for the costationary examples that follow. However, costationarity works for wider classes of processes that have time-varying spectrum $p(t, \nu)$ and "$p(t, \nu)$ constant in t iff stationary". Recall also the definition of rescaled time $z = t/T$. We conjecture that costationarity could also be made to work in the context of a whole host of other non-stationary time series models, for example, piecewise stationary GARCH models, see Andreou and Ghysels (2002) or Davis et al. (2008), time-varying ARCH models, see Dahlhaus and Subba Rao (2006), Dahlhaus and Subba Rao (2007), Granger (2008), Markov-switching models, see Hamilton and Raj (2002), time-varying Hurst exponent models, see Grech and Mazur (2004) and references therein. The relations of any of these models to LSW/LSF is unclear and requires further research.

EXAMPLE 1. *(LSF processes)* A (zero mean) LSF process, $\{X_{t,T}\}$, with spectrum $f(z, \lambda)$, is a triangular array of stochastic processes with the following representation:

$$X_{t,T} = \int_{-\pi}^{\pi} A_{t,T}(\lambda) \exp(i\lambda t) d\xi(\lambda), \tag{4}$$

for $t = 1, \ldots, T > 0$, where $\xi(\lambda)$ is an orthonormal increments process, and where there exists a function $A(z, \lambda)$ such that $A_{t,T}(\lambda) \approx A(t/T, \lambda)$ and $|A(z, \lambda)|^2 \approx f(z, \lambda) = p(t, \nu)$. Local stationarity of $X_{t,T}$ is achieved by (total variation) constraints upon $A(z, \lambda)$, see Dahlhaus and Polonik (2006). For further details see Dahlhaus (1997), Van Bellegem and Dahlhaus (2006). In this model the time-varying autocovariance, $c(z, \tau) = \int_{-\pi}^{\pi} f(z, \lambda) \exp(i\lambda\tau)\, d\lambda$, see Dahlhaus (1996) formula (1.5). The local variance of $X_{t,T}$ is $v_X(z) = c(z, 0) = \int_{-\pi}^{\pi} f(z, \lambda)\, d\lambda$.

EXAMPLE 2. *(LSW processes)* A LSW process, $\{X_{t,T}\}$, with spectrum $\{S_j(z)\}_{j=1}^{\infty}$, is a triangular array of stochastic processes that admit the following representation

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{5}$$

for $t = 1, \ldots, T$, where $\{\xi_{j,k}\}$ are a collection of uncorrelated random variables ('innovations') with mean 0 and variance 1, the $\{\psi_{j,t}\}$ are a set of discrete nondecimated wavelets (compactly supported oscillatory vectors with support proportional to 2^j), and $\{w_{j,k;T}\}$ a collection of amplitudes that are 'smooth' in a particular way as a function of k. The smoothness of $w_{j,k;T}$ controls the degree of local stationarity of $X_{t,T}$. The spectrum is linked to the process by $w_{j,k;T}^2 \approx S_j(k/T) = p(k(t), j)$. A full definition of LSW processes, and comparison with LSF processes, can be found in NvSK. Further explanation, including wavelets, can be found in Nason (2008). Jumps in the spectrum are permitted by the generalized extension developed by Van Bellegem and von Sachs (2008) by extending $S = p$ to functions of bounded variation. We adopt this here.

For LSW the time-varying autocovariance $c(z,\tau) = \sum_{j=1}^{\infty} S_j(z)\Psi_j(\tau)$, where $\Psi_j(\tau)$ is the autocorrelation function of $\psi_{j,t}$, see NvSK formula (14). The local variance of $X_{t,T}$ is $v_X(z) = c(z,0) = \sum_{j=1}^{\infty} S_j(z)$, as $\Psi_j(0) = 1$ for all j due to the wavelet orthonormality.

We now give two examples of costationarity.

EXAMPLE 3. *(Uncorrelated LSW)* Let X_t, Y_t be two uncorrelated LSW processes with spectra $p_X(z,1) = S_1^X(z) = \cos^2(z)$ and $p_Y(z,1) = S_1^Y(z) = \sin^2(z)$ for $z \in (0,1)$, all other $p_X(z,j), p_Y(z,j) = 0$ for $z \in (0,1)$ and $j > 1$. Then, X_t and Y_t are costationary as the spectrum of $X_t + Y_t$ is exactly 1 for $z \in (0,1)$. The costationary vectors are $\alpha_t = \beta_t = 1$.

EXAMPLE 4. *(TVAR/TVARMA)* Let X_t be a TVAR(1) process with time-varying AR parameter of $a_1(z)$, $z \in (0,1)$ and constant variance parameter σ^2. From Dahlhaus (1997) X_t is LSF with spectrum given by $p_X(t,\nu) = f_X(z,\lambda) = \sigma^2\pi^{-1}\left|1 + a_1(z)e^{i\lambda}\right|^{-2}$ for $z \in (0,1)$ and $\lambda \in (-\pi,\pi)$. Let Y_t be a TVARMA$(1,1)$ process with spectrum $p_Y(t,\nu) = f_Y(z,\lambda) = \pi^{-1}G^2(z,\lambda)\left|1 + a_1(z)e^{i\lambda}\right|^2$ where

$$G^2(z,\lambda) = \sigma_1^2\left\{(1 + a_1^2(z)) - \sigma^2\sigma_1^{-2} + 2a_1(z)\cos(\lambda)\right\}, \tag{6}$$

with σ_1^2 large enough (as specified in the proof). Then $X_t + Y_t$ is second-order stationary with spectrum σ_1^2/π, hence X_t and Y_t are costationary with costationary vectors $\alpha_t = 1, \beta_t = 1$. See Appendix B for more details.

More generally, this article is written with "model-independent" concepts in mind. We refer to generic time-varying autocovariances and spectra. However, later, when it comes to estimation we choose particular estimators that are often derived from model-based formulae. For the ones we use, the estimators are asymptotically unbiased for the "model-independent" quantities and possess other attractive statistical properties. Of course, other estimators could be used.

2.3. Cross-covariance and error-correction formulae

Suppose X_t and Y_t are costationary. Take the variance of the general linear combination (2) to obtain:

$$\sigma_Z^2 = \alpha_t^2 v_X(t) + \beta_t^2 v_Y(t) + 2\alpha_t\beta_t\mathrm{cov}(X_t, Y_t). \tag{7}$$

where $\sigma_Z^2 = \mathrm{var}(Z_t)$ is well-defined and constant since $\{Z_t\}$ is stationary, and where $v_X(t) = \mathrm{var}(X_t)$ and $v_Y(t) = \mathrm{var}(Y_t)$ are the usual time-varying variances mentioned in Examples 1 and 2. Given data it is easy to estimate σ_Z^2 by standard methods and the local variances, $v_X(t), v_Y(t)$, can be estimated by integrating over ν an estimate of the time-varying spectra, $p_X(t,\nu)$ and $p_Y(t,\nu)$ respectively, again as mentioned in Examples 1 and 2.

In classical cointegration theory there is a strong connection between cointegration and error-correction models, see Engle and Granger (1987) for example. Roughly speaking, cointegrated series are constrained so that their cointegrated linear combination remains stationary even in response to shocks, say, on one of the series. The other series has to "respond" to maintain stationarity.

To understand error-correction in our setup see that equation (7) becomes

$$\sigma_Z^2 = \alpha_{[zT]}^2 v_X(z) + \beta_{[zT]}^2 v_Y(z) + 2\alpha_{[zT]}\beta_{[zT]}\gamma_{XY}(z,0), \tag{8}$$

where $t = [zT]$ and $[x]$ denotes the greatest integer $\leq x$. Now let $z_2 = z + \delta z$, where δz is small, and use the costationary argument again to write:

$$\sigma_Z^2 = \alpha_{[z_2T]}^2 v_X(z+\delta z) + \beta_{[z_2T]}^2 v_Y(z+\delta z) + 2\alpha_{[z_2T]}\beta_{[z_2T]}\gamma_{XY}(z+\delta z,0). \tag{9}$$

From Remark 1 recall that α_t and β_t are slowly varying. Hence, assume that $\alpha_{[z_2 T]} \approx \alpha_{[zT]} = \alpha$ and similarly for β_t. In the practical example in Section 4.1 the approximation is indeed equality as we use piecewise constant functions (assuming no jump at this location).

For the remainder of this section, let us further assume that v_X, v_Y and $\gamma_{XY}(z, 0)$ are differentiable. Now subtract (8) from (9), divide by δz and let $\delta z \to 0$. We have

$$0 = \alpha^2 v'_X(z) + \beta^2 v'_Y(z) + 2\alpha\beta\gamma'_{XY}(z, 0). \tag{10}$$

Writing $r = \alpha/\beta$ and rearranging (10) gives us our 'variance correction' formula:

$$v'_Y(z) = -r^2 v'_X(z) - 2r\gamma'_{XY}(z, 0). \tag{11}$$

The formula speaks for itself and can be interpreted in numerous useful ways. For example: since $r^2 > 0$ any increase in the variance of X must be matched by an *decrease* in the variance of Y to maintain costationarity, assuming $\gamma_{XY}(z, 0)$ remains unchanged.

REMARK 6. Further interesting relations can be derived for costationary series. For example, drop the subscript $[zT]$ in equation (8) for brevity, but remember α, β are still local, and then rearrange using the fact that variance is non-negative to obtain the following bounds

$$\gamma_{XY}(z, 0) \left\{ \begin{array}{c} \leq \\ \geq \end{array} \right\} \frac{\sigma_Z^2}{2\alpha\beta} \left\{ \begin{array}{l} \text{if } \alpha\beta > 0 \\ \text{if } \alpha\beta < 0. \end{array} \right. \tag{12}$$

The bound for the condition $\alpha\beta < 0$ implies that the correlation between X_t and Y_t is bounded below (by a negative number) and bounded above (by a positive number) if $\alpha\beta > 0$. Intuitively this makes sense. The bound implies that if α and β are of the same sign ($\alpha\beta > 0$), then X_t, Y_t cannot become *too* positively correlated as they will 'tend towards the same process', in second-order terms. This would result in a contradiction, because if X_t, Y_t were the same LS process then combination Z_t would be LS and not stationary.

REMARK 7. Under costationary one can estimate $\mathrm{cov}(X_t, Y_t)$ by solving for it in (7) using the usual estimates of σ_Z^2, $v_X(z)$ and $v_Y(z)$, but we do not pursue this further here as it can be computed/estimated directly: for example, for LSF processes direct estimates can be obtained from the cross-spectral estimates, see Dahlhaus (2000), for LSW processes, see Sanderson et al. (2010).

3. Discovering costationarity in practice

3.1. General algorithm

Suppose we have two LS series $\{X_t, Y_t\}_{t=1}^T$. As indicated in Section 1 we adopt a pragmatic, computational, "project-pursuit" approach to discovering possible (α_t, β_t) combinations that cause $\{Z_t\}$ to be stationary. Examples 3 and 4 show that costationary is theoretically possible and there the spectrum, $p_Z(t, \nu)$, of the combined series is mathematically constant as a function of t. *In practice*, we form actual combinations of the series and then evaluate a *spectral estimate* of the combinations which are subject to estimation and sampling error. We then statistically assess whether a spectral estimate of a combination is constant or not as a function of time. Our precise definition of constancy, $\mathcal{T}$, is given in Section 3.2.

Another important point is that costationary solutions for two *processes* $\{X_t, Y_t\}$ need not be unique (in a mathematical sense). On top of this, when dealing with realizations, the costationary solutions need not be unique as minor perturbations in the solution might not change the practical

statistical assessment of the stationarity of Z_t. Hence, as in projection pursuit, we are interested in discovering *any* potential costationary solutions that might be of interest. Typically, to try and find costationary solutions we run the algorithm many times from different starting points. A later step examines all found solutions and tries to find the set of interesting solutions that differ by more than small numerical amounts. This is achieved by methods of multivariate analysis where all the solution vectors are treated as standard multivariate vectors and then multidimensional scaling or hierarchical cluster analysis is used to find classes of 'like' solutions.

Given realizations of (X_t, Y_t) for $t = 1, \ldots, T$, the algorithm to compute the constancy of the spectral estimate is as follows:

(1) Given input vectors (α_t, β_t) for $t = 1, \ldots, T$.

(2) Form the combination $Z_t = \alpha_t X_t + \beta_t Y_t$ for $t = 1, \ldots, T$.

(3) Compute the spectral estimate $\hat{p}_Z(t, \nu) = \hat{p}_{\alpha,\beta}(t, \nu)$ on $\{Z_t\}_{t=1}^T$.

(4) Compute the constancy of the spectral estimate using $\mathcal{T}(\hat{p}_{\alpha,\beta})$.

The α, β subscripts are used to denote that the spectral estimate depends on the combination vectors (α_t, β_t). We now define the complete costationarity discovery algorithm.

`Costat`: Given starting vectors (α_t, β_t) proceed with the penalised numerical minimization of the quantity $\mathcal{T}(\hat{p}_{\alpha,\beta})$ over the vectors (α_t, β_t) with a smoothness/complexity penalty on (α_t, β_t) as discussed in Remark 1. Then, after the minimization algorithm has finished, resulting in 'optimal' combination vectors (α_t^*, β_t^*), we then apply a statistical test of stationarity to Z_t via the final spectral estimate $p_{\alpha^*,\beta^*}(t, \nu)$. If the process Z_t is deemed stationary then the processes (X_t, Y_t) are deemed to be costationary with costationary vectors (α_t^*, β_t^*).

For implementation there are six aspects of the `Costat` that need to be specified: the choice of the starting vectors, the type of spectral estimate, the penalty, the numerical minimizer, the measure of constancy, and the test of stationarity. Appendix A provides implementation details about the first four, the next two sections provide more details about the last two.

3.2. Measure of constancy

Our measure of constancy is a metric that quantifies the variation of the spectrum $p(z, \nu)$ as a function of $z = t/T$. Specifically,

$$\mathcal{T}(p) = K^{-1} \int \int_0^1 \{p(z, \nu) - \bar{p}(\nu)\}^2 dz d\nu, \tag{13}$$

where $K = \int d\nu$ and $\bar{p}(\nu) = \int_0^1 p(z, \nu)\, dz$. It is clear that if $p(z, \nu)$ is a constant function of z then the metric $\mathcal{T} = 0$. Also, if the underlying process is not stationary then $\mathcal{T}(p) \geq 0$ with increasing $\mathcal{T}$ for "increasingly nonstationary" processes as $\mathcal{T}$ is just the L^2 metric. Clearly, other topologically equivalent metrics could be used just as well. In practice, conditions are imposed to prevent $\mathcal{T}(p)$ being affected by changes in overall value of p as is typical in projection pursuit.

EXAMPLE 5. *(LSF processes)* We could use $\mathcal{T}\{f(z, \lambda)\} = \int_{-\pi}^{\pi} \int_0^1 \{f(z, \lambda) - \bar{f}(\lambda)\}^2 dz d\lambda$.

EXAMPLE 6. *(LSW processes)* Let $J = \log T$. We could use $\mathcal{T}\{S_j(z)\} = J^{-1} \sum_{j=1}^{J} \int_0^1 \{S_j(z) - \bar{S}_j\}^2 dz$.

It is possible to detect constancy in the time-varying autocovariance instead.

3.3. Our test and bootstrap assessment of significance

An early test of stationarity is described Priestley and Subba Rao (1969) which applies a two-factor ANOVA test to the log of their evolutionary spectrum. Ahamada and Boutahar (2002) derive the exact distribution of a similar test statistic and apply it to the analysis of Euro/US dollar exchange rate. von Sachs and Neumann (2000) introduced a stationarity test based on the "segmented periodogram" (an estimate of the periodogram based on the segment $X_K, \ldots, X_L$) which estimates the evolutionary spectrum, $f(z, \lambda)$. They then apply a two-dimensional wavelet transform which simultaneously smoothes across frequency (father wavelets) and identifies inhomogeneities using mother wavelets. Other recent stationarity tests are Stărică and Granger (2005) and Paparoditis (2009) that measure the difference between a periodogram and model spectrum on intervals.

The typical stationarity test null hypothesis is:

$$H_0 : p(z, \nu) \text{ is a constant function of } z \text{ for all } \nu \tag{14}$$

versus

$$H_A : p(z, \nu) \text{ is not constant function of } z \text{ for some } \nu. \tag{15}$$

The statistical *assessments* carried out in this paper assume Gaussian innovations enabling us to use the parametric bootstrap, see, e.g. (Davison and Hinkley, 1997, 4.2.3). However, it should be emphasized that (i) all the concepts up to this point do not rely on this assumption (ii) it would be perfectly possible to substitute other assessments that do not rely on the Gaussianity assumption. However, the Gaussian assumption is not an unusual one to make and much used for classical stationary theory, e.g. Priestley (1983) (Chapters 5 and 6). Even with Gaussian innovations both LSW and LSF processes are more general and flexible than they might first appear. This is because the processes are time-varying combinations of Gaussians and hence fat or slim tails could be modelled. We are currently investigating extension of the bootstrap to a more general distributional situation.

Our test uses the test statistic $\mathcal{T}$ defined in (13). For the fixed sample size, for testing purposes, the spectrum $\{p(\nu)\}_\nu$ completely specifies the distribution of the time series Z_t under the null hypothesis. Thus, we can use a parametric Monte Carlo test to assess the significance of $\mathcal{T}$ on the sample by using B bootstrap simulations as follows

---

`Bootstat`:

(a) Evaluate $\mathcal{T}$ as in (13) on the actual data sample, call this value $\mathcal{T}^{(1)}$.

(b) Compute $\bar{p}(\nu)$ for all ν, from the sample, as defined just after (13). This is just the classical spectral estimate which assumes the data are stationary.

(c) Repeat for i in 2 to B

  (i) Simulate Z_t from the appropriate *stationary* model with squared amplitudes given by $\bar{p}(\nu)$ using Gaussian innovations.

  (ii) Compute $\mathcal{T}$ as in (13) on the simulated data, call this value $\mathcal{T}^{(i)}$.

(d) Define the p-value of the test by $p = \{\text{Number of } \mathcal{T}^{(i)} > \mathcal{T}^{(1)}\}/B$.

Heuristically the test evaluates how unlikely is the value of $\mathcal{T}$ computed on the real data compared to other 'likely' values of $\mathcal{T}$ computed on stationary realizations of Z_t having a similar constant spectrum to the actual Z_t. Empirically, we have found, with series of the lengths in this paper,

setting $B \geq 200$ to be adequate, but further work would be necessary to investigate the performance and satisfactory values of B.

The bootstrap procedure requires (statistically) consistent estimation of the spectral parameters to work, Davison (2008). Results for consistency of spectral estimation in the LSF and LSW contexts can be found in von Sachs and Schneider (1996) and Fryzlewicz and Nason (2006) respectively. Appendix C gives more details on the actual estimator we use and establishes its consistency.

Process simulation in step (c-i) of `Bootstat` is achieved by either: *(LSF processes)* the elegant 'circulant embedding' method as described by Percival and Constantine (2006) which simulates a stationary time series given a (consistent) non-parametric spectral density estimate or *(LSW processes)* the constructive formula given by (5) using the fast average-basis 'inverse' non-decimated wavelet transform, see Coifman and Donoho (1995).

Section 5 describes a simulation study that demonstrates good empirical size and power of some of our tests under both stationary and LS conditions.

4. Examples

Up to now this article has proposed a new set of statistical methods and, naturally, there needs to be a continuing debate over the precise specifics: over the test of stationarity, the measure of constancy, the nature of the costationary vectors, and so on. However, here we present two examples: asset allocation in finance and intermittency mitigation in wind energy generation. In both cases our methods generate combination series, Z_t, that are shown to be useful by independent measures, generally accepted in their respective literature.

4.1. Costationary asset allocation

The aim of modern capital market theory is to provide a framework to explain how individuals decide to allocate their wealth. The capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966) has historically been considered the benchmark approach for a long time. According to this, investment decisions were taken by means of the Markowitz (1952) mean-variance criterion. Although this approach is still widely used, several criticisms have been raised, both on the theoretical specification of this model and the empirical validation of related studies. The CAPM is a static, single-period, model although it has often been spuriously treated as if it holds over time (intertemporally), both for the validity of its results and the information provided as model input. As an example of the latter, the mean-variance approach is used for portfolio selection even when asset returns are modeled with time-varying conditional statistical properties, e.g. using univariate or multivariate GARCH processes such as in Ledoit et al. (2003). Fama (1970) justified an intertemporal use of CAPM, by showing that, if preferences and future investment opportunities are time-invariant, then intertemporal portfolio selection can be treated as a static portfolio maximization. Merton (1971) showed that portfolio selection from an intertemporal maximizer behaves very differently when preferences and investment opportunities change over time. We use costationary solutions in such a way that links the intertemporal asset allocation to the static CAPM: this kind of optimization is missing in the single period approach.

The Markowitz (1952) mean-variance portfolio (MVP) theory takes a set of n assets, $i = 1, \ldots, n$, uses the expected future rate of return, and the covariance structure of the assets (returns), to select the optimal set of weights $\{w_i\}_{i=1}^n$ that minimizes the risk for an acceptable expected rate of return. Following recent compelling thinking, e.g., Stărică and Granger (2005), we assume that the assets are LS. For the purposes of this example, we take pairs of asset returns over time, apply our costationary methodology to those pairs and seek sets of costationary solutions. In this

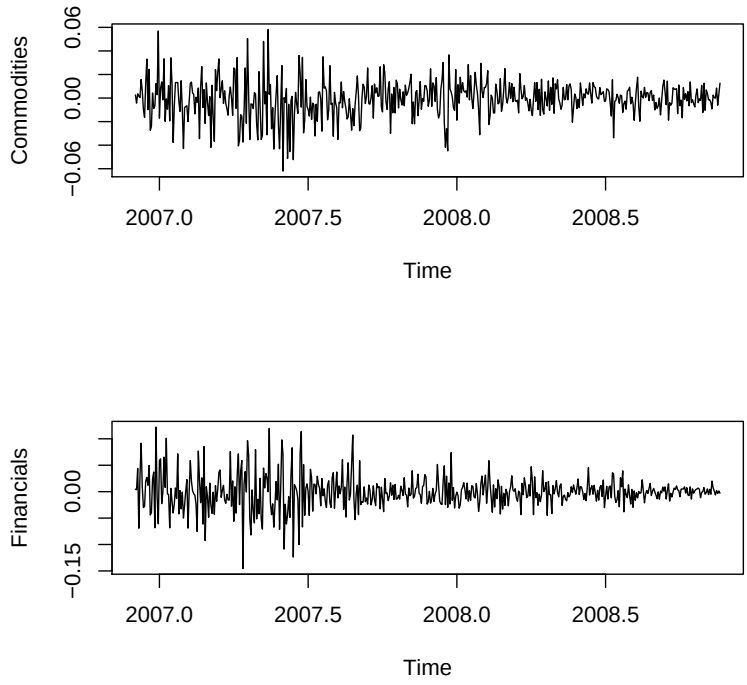

Fig. 1. Dow Jones Commodities (DJC) and NASDAQ Financials (IXF) indices log-returns.

example we choose our combination vectors to be piecewise constant. Our end point is a set of synthetic assets each containing two assets in different proportions over time. The key point is that, because of costationarity, the returns associated with each synthetic asset are stationary (or at least very close to stationarity). Hence, the covariance structure of the synthetic asset can be extremely well-estimated because the estimation is based on the *whole length* of the data series. Compare this to the vector GARCH model which bases the optimization on the conditional quantities computed at each portfolio rebalancing point.

We remarked earlier that, for a given set of time series, there might be more than one costationary solution. The other innovation we propose is permitting the use of more than one costationary solution in a portfolio that arises from the same pair of assets.

For example, Figure 1 shows a pair of stock index log returns from the Dow Jones Commodities and NASDAQ Financials indices. As described below we use 20 subsamples of size 512, rolling forward each time by five days ahead. The first sample started at 1st December 2006 and end 17th December 2008. The dates of the next 19 samples are obtained by rolling the trading dates (without weekends and holidays) by five days. Both series are assessed as being highly non-stationary with small p-values $p < 0.001$ arising from our `Bootstat` test. We form three kinds of portfolios using the stock data. For each rolling window we (i) use the classic MVP portfolio selection, (ii) refit a constant conditional correlation multivariate VECH GARCH(1,1) model with BEKK, see Bollerslev et al. (1988); Bollerslev (1990) and Engle and Kroner (1995). This is followed by repeated application of the single-period MVP portfolio selection considering a single budget constraint and setting the risk appetite parameter equal to 1, see Best and Grauer (1991). Careful attention was paid to check the adequacy of model fit. Our method, (iii), form between 4 and 8 costationary synthetic asset solutions, followed by MVP to select the overall portfolio mix. These

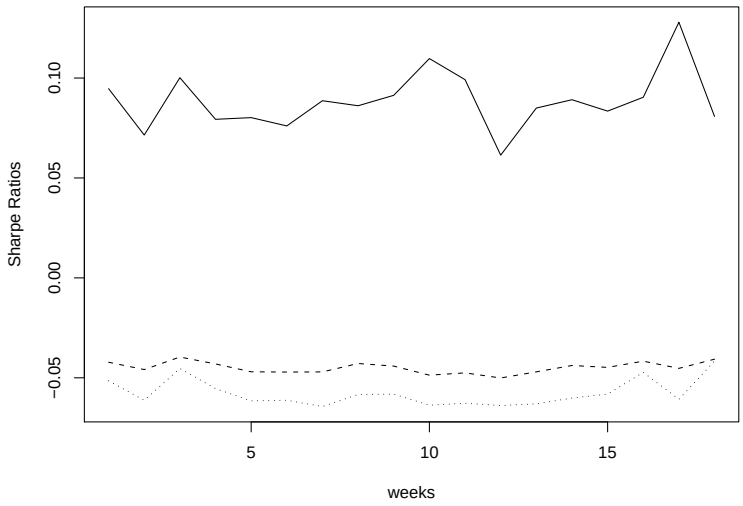

Fig. 2. Sharpe ratios for MV Costationary portfolio (solid line), MV Garch portfolio (dashed line) and class MV portfolios (dotted line). Higher ratios mean better performance of the portfolio.

analyses were applied to consecutive subsets of length 512 trading days with a rolling window stepping ahead five days a total of 20 times.

We compare the outcome of the three portfolio selection methods by means of Sharpe (1966) ratios which are a commonly used measure of how well an asset return compensates an investor for the given amount of risk. (Let r be the asset return, r_b the return on a benchmark asset and σ the standard deviation or volatility of an asset. Then the Sharpe ratio is $(r - r_b)/\sigma$ and is a reward per risk ratio. We use this ratio for comparative purposes, so we set $r_b = 0$. Clearly, investors like assets that reward well with little risk and hence have a large Sharpe ratio. Of course, it is vital to remember that the Sharpe ratio is one, very blunt, tool in the investors' bag). Figure 2 shows the series of 20 Sharpe ratios for the period 2007–8 over the 20 week rolling window. The figure shows that the costationary portfolio performs much better than the GARCH-based method, which is itself a little better than the classic method. Indeed, the GARCH and classic methods return negative Sharpe ratios which indicate that they are not better than riskless asset that the portfolios are being compared to (possibly because the market experienced severe downward pressure during this period). We have conducted similar analyses on other assets during earlier time periods where the Sharpe ratios of GARCH and classic methods are positive, but again not as large as the costationary portfolio. It is clear from these examples that costationary portfolio selection has a great deal of potential, and, as a method, is fundamentally different from existing methods. The synthetic costationary asset idea begs further investigation both theoretically and practically. For example, extension to include more than two assets at a time.

We calculated Sharpe ratios on the same rolling windows for all the three approaches that we compared. For the GARCH and classical approaches the portfolio means and variances were calculated based on the actual assets. For costationary portfolios, these are calculated on costationary solutions Z_t, that we use here as factors. These factors are not completely independent. In fact, we estimate their covariance matrix, which is not diagonal. Using this matrix we then calculate the portfolio variance. Although we are using multiple solutions, there is no dependence (in a mathematical sense) and the covariance matrix is not singular either. There is no surprise in having negative

Sharpe ratios for both the classical and GARCH approaches. We have selected the dataset to account for a down-market tendency. In the costationary approach, ratios are calculated on stationary quantities derived from the time-varying asset combinations, obtaining positive portfolio means. In this case we typically use more than four costationary solutions, obtaining a more flexible portfolio specification than the direct two-assets models.

4.2. Mitigating wind power intermittency

Wind is a key renewable energy source that has great potential by providing a partial alternative to fossil fuel carbon emissions, does not require fuel in the conventional sense and enhances energy security by both domestic siting and enhancing supply diversity. However, wind energy suffers from two main drawbacks: (i) it is hard to predict, intermittent and volatile and hence cannot always produce a reliable supply even in the face of constant demand and (ii) it is not controllable and hence cannot respond to changes in demand, see Hirst (2002), Lewis (2008) or Skea et al. (2008). Wind intermittency is a real cost as generators have to establish conventional generation that can be started or stopped at short notice, such as that provided by the UK National Grid Fast Reserve or Hot Standby contracts. Another interesting factor is that as wind energy market penetration increases, it becomes increasingly likely that not all possible wind energy is taken, and some is dumped so as to achieve stability and regulation in the system. For example, for 20% penetration it is estimated that about 10% of wind power will be lost through this mechanism, see Holtinnen (2005).

The key to dealing with intermittency is 'dumping' and aggregation, Hirst (2002), Holtinnen (2005). Aggregation can mean pooling wind energy from multiple turbines at a single site, or at multiple sites across a geographical region with varying wind characteristics and possibly mixing in other energy sources also. The goal is to provide a stable non-volatile energy source that is more easily absorbed into the grid. Our new methods provide a way of combining turbine energy outputs to mitigate intermittency and reduce volatility. Of course, aggregation is just another kind of portfolio: combining wind resources to increase return and decrease overall risk.

For example, consider the hourly wind speed time series measured at two Welsh Meteorological Office stations: Aberporth and Valley, previously studied by Nason and Sapatinas (2002). Aberporth is located approximately 120km south of Valley and they are mostly separated by Cardigan Bay. Wind power is proportional the cube of the wind speed, see Lewis (2008), and hence we investigate time-varying linear combinations of the cubed speed series that minimize volatility. The cubed series are shown as the grey series in the top row of Figure 3. The `Bootstat` test verifies the non-stationary nature of the original speed and power series. Since we deal with genuine combinations of physical quantities we constrain α_t, β_t to lie in $[0, 1]$. If $\alpha_t < 1$ then energy is being dumped from source X_t and similarly for β_t, Y_t. It is advisable to change proportions of power sources in a smooth manner as power systems do not respond at all well to sudden changes. Hence, we enforce smoothly changing proportions of each power series by using the following underlying Fourier representation for the Fourier interpolant of α_t namely

$$\alpha(t) = a_0/2 + \sum_{n=1}^{\infty} a_n \cos(n\pi t) + \sum_{n=1}^{\infty} b_n \sin(n\pi t), \tag{16}$$

where the $\{a_n, b_n\}$ are the usual Fourier coefficients of $\alpha(t)$, see Firth (1992). For complexity penalization we have imposed a standard $\int (\alpha_t^{''})^2 \, dt$ roughness penalty, similarly for β_t.

Figure 3 shows the results of one set of solutions that achieve stationarity. The wind power time series themselves do not look as if they fit the typical LS models. For example, they do not have zero mean. Hence, for these series, for stationarity assessment, we concentrate on the oscillations

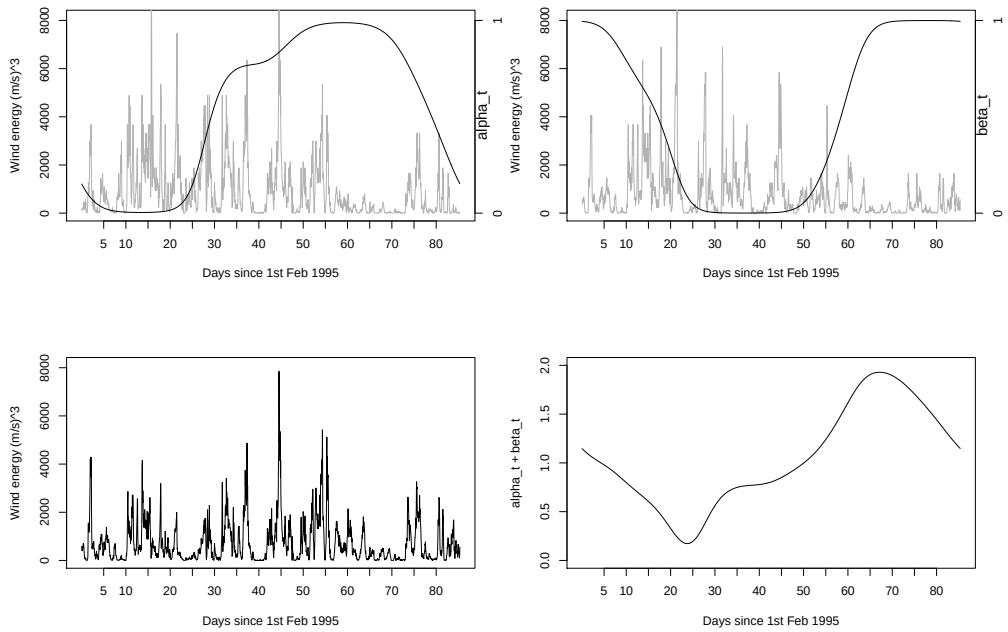

Fig. 3. *Top row:* wind power time series for Aberporth (left) and Valley (right) in grey, left axis. The solid lines correspond to the final α_t (left) and β_t (right) identified by the costationary algorithm, right axis. *Bottom left:* the 'optimal' combined series, Z_t. *Bottom right:* proportion of total resource used.

above a fixed frequency ($\approx 4.35 \times 10^{-6}$Hz, about 1 oscillation per 64 hours) by modifying the test statistic in (13) to only take account of those higher frequencies and ignore the lower ones. This makes practical sense as load balancing in a power grid is carried out on intra-hour and intra-day timescales (the latter from daily forecasting), but not typically on weekly or monthly scales, see Hirst (2002); Holtinnen (2005). In assessing the mid to high frequencies the Aberporth and Valley wind energy series are not stationary (p-value $= 0.02$) whereas the combined costationary series is (p-value $= 0.05$). Indeed, with all other criteria of stationarity we tried the combined series is more stationary than either of the individual series in that the p-value is higher. Hence, the variability in the combined power profile is lower than in either of the two individual series.

At this point the reader might be wondering what we mean by one time series being 'more stationary' than another. In mathematical terms a time series is either stationary or it is not. What we mean is that the p-value of the stationary hypothesis test, or indeed the test statistic for constancy over time can both be used to define an order relation so that (real data) series x_t is 'more stationary' than series y_t if the p-value of the stationary test for x_t is greater than that of y_t.

The bottom right plot in Figure 3 shows the total proportion of energy taken from the combined system. In practice, one would more tightly constrain the costationary vectors to reduce the amount of dumping (e.g. around day 24 less than a quarter of the total available energy is being utilized). This could be achieved by constraining $\alpha_t, \beta_t \geq d$ for some d so that the proportion of power taken from any one source would be greater than d. This might result in Z_t not being stationary, or rather assessed as such, but it would be more stable than simple aggregation.

Table 1. Empirical size (as percentage) of stationarity test based on LSW, $\mathcal{T}(\hat{S})$ and LSF, $\mathcal{T}(\hat{f})$, test statistics for the following three models: S1=white noise, S2=AR(1) process and S3=MA(15) process, see text for detailed model description.

Sample size	Model, $T(\hat{S})$			Model, $T(\hat{f})$			Sample size	Model, $T(\hat{S})$			Model, $T(\hat{f})$		
T	S1	S2	S3	S1	S2	S3	T	S1	S2	S3	S1	S2	S3
32	0	0	0	·	·	·	256	0	0	1	1	1	0
64	0	0	0	·	·	·	512	4	0	1	0	1	1
128	2	0	0	3	5	2	1024	1	0	3	1	2	4

5. Empirical assessment of our stationarity test (simulations)

5.1. Size assessment

We consider three stationary time series models described as follows. Model S1 sets $X_t \sim$ i.i.d. $N(0,1)$. Model S2 is an AR(1) model with AR parameter $\alpha = 0.9$. Model S3 is a stationary LSW process where $p(z,j) = S_j(z) = 1$ for $j = 1,2,3,4$ and $p(z,j) = S_j(z) = 0$ otherwise, which is just a particular stationary MA(15) process, see NvSK. Table 1 displays empirical size results from each of S1–S3, applying our stationarity test 100 times as described in Section 3.3 with a *nominal* size of 5% using $B = 200$ bootstrap simulations. Table 1 shows reasonable empirical size results for both LSW and LSF stationarity tests. (The LSF estimate $\hat{p}(\lambda)$ is based on a crude windowed Fourier method which only works for $T > 64$).

5.2. Power assessment

We consider the following three non-stationary models. Model P1 samples from a concatenated white noise model where the first $T/2$ observations are i.i.d. $N(0,1)$ and the following $T/2$ are i.i.d. $N(0,\sigma^2)$. Clearly, this model is nonstationary but for $\sigma^2 \approx 1$ it is not easy to detect the nonstationary nature. Model P2 samples from the LSW spectrum given by $p(z,1) = S_1(z) = 1$ for $z \in (0,1/2]$ and $p(z,1) = S_1(z) = \sigma^2$ for $z \in (1/2,1)$, other S_j are zero. Model P2's spectrum is piecewise constant with a jump increase in variability halfway through at the finest scale. Model P3 samples from the LSW spectrum given by $p(z,1) = S_1(z) = (1+\sigma^2) - 2\sigma^2|z-1/2|$ and $p(z,2) = S_2(z) = (1+4\sigma^2) - 8\sigma^2|z-1/2|$ with all other $p(z,j) = S_j(z) = 0$ all for $z \in (0,1)$. Model P3's spectrum is an example of a signal with 'small' variance at each end and a large variance in the middle controlled by σ^2. The empirical power results are shown in Tables 2 to 4. For all these the degree of non-stationarity is controlled by σ^2. If $\sigma^2 = 1$ they all are stationary. Increasing non-stationarity can be achieved by increasing σ^2 and observing increasing power. Generally, for small sample sizes the tests have low power even when the non-stationarity is reasonably large (e.g. 18% in Table 2 for $T = 64$ and $\sigma = 2.4$). Generally, for higher sample sizes the power is very good.

6. Combining locally stationary processes

Throughout this article we have assumed that time-varying linear combinations of particular LS processes themselves are in the same class of processes. That this is true for LSF processes can be established directly from the process definition in (4). The verification for LSW processes is more delicate because time-variation is achieved in LSW processes through the oscillatory basis function, the t in $\psi_{j,k-t}$ in (5), rather than through the amplitude, $A_t(\lambda)$, in LSF processes.

If we confine ourselves to constant combination functions $\alpha_t = \alpha$ and $\beta_t = \beta$ then it is easy to see that the linear combination Z_t in (2) is itself a LSW process if $\{X_t\}$ and $\{Y_t\}$ are. The next two

Table 2. Emp. power (%) for two tests for model P1. I=$\mathcal{T}(\hat{S})$=wavelet, and II=$\mathcal{T}(\hat{f})$=Fourier.

		Sample size, T												
	σ	32	64	128	256	512	1024	σ	32	64	128	256	512	1024
↑	1.5	0	1	14	52	91	100	2.1	0	13	59	99	100	100
I	1.7	0	7	27	84	99	100	2.3	2	16	68	100	100	100
↓	1.9	0	9	45	90	100	100	2.4	2	18	72	100	100	100
↑	1.5	·	·	12	40	64	98	2.1	·	·	38	96	100	100
II	1.7	·	·	14	60	97	100	2.3	·	·	47	99	100	100
↓	1.9	·	·	28	89	100	100	2.4	·	·	63	98	100	100

Table 3. Empirical power (%) of test based on $\mathcal{T}(\hat{S})$ for model P2.

	Sample size, T												
σ	32	64	128	256	512	1024	σ	32	64	128	256	512	1024
1.5	1	6	9	11	23	49	2.1	3	8	16	49	84	98
1.7	1	6	10	24	53	81	2.3	3	9	19	58	91	100
1.9	1	6	12	35	70	97	2.4	3	11	19	61	94	100

theorems establish the result for more general combination of LSW processes.

First, we give a name to a concept that links a sequence of values, $a_{t,T}$, to a function, $a(z)$, which appeared first in Dahlhaus (1997) and later in NvSK.

DEFINITION 2. *Define a* **close pair**, $(a(t), a_{t,T})$ *with positive constant* K*, to be a function* $a : [0,1] \to \mathbb{R}$ *and a sequence* $a_{t,T}$ *such that*

$$\left| a_{t,T} - a\left(\frac{t}{T}\right) \right| < KT^{-1}. \tag{17}$$

Next, we make clear an assumption that runs through our article:

ASSUMPTION 1. *Let* $\{\xi_{j,k}\}$ *and* $\{\eta_{j,k}\}$ *be the innovations sequences associated with two LSW processes,* X_t *and* Y_t*. Throughout we assume that the two sequences are* cross-orthogonal*, that is* $\mathbb{E}(\xi_{j,k}\eta_{\ell,m}) = \rho_{j,k}\delta_{j,\ell}\delta_{k,m}$*. This is the usual assumption made in stationary process theory, see Priestley (1983) and also has been made for LS processes, see Sanderson et al. (2010).*

To show that $Z_t = \alpha_t X_t + \beta_t Y_t$ is LSW, we first establish that $\alpha_t X_t$ is LSW, and then that the sum of two LSW processes is itself LSW.

THEOREM 1. *Let* $\{X_{t,T}\}_{t=1}^{T}$ *be a LSW process with spectrum* $p_X(z,j) = |W_j(z)|^2$*, discrete nondecimated wavelets* $\{\psi_{j,k}\}$*, and innovations* $\{\xi_{j,k}\}$*, satisfying the conditions from Van Bellegem and von Sachs (2008), page 1883, Definition 1. Let* $a : [0,1] \to \mathbb{R}$ *be a function of bounded total*

Table 4. Empirical power (%) of test based on $\mathcal{T}(\hat{S})$ for model P3.

	Sample size, T												
σ	32	64	128	256	512	1024	σ	32	64	128	256	512	1024
1	0	2	4	13	37	84	6	3	7	26	65	100	100
2	0	3	10	28	74	99	7	4	7	26	68	100	100
3	2	6	16	47	90	100	8	4	7	27	71	100	100
4	2	6	21	54	96	100	9	4	7	28	73	100	100
5	2	7	23	61	99	100	9.5	4	7	29	74	100	100

variation with constant R_3, and let $(a(t), a_{t,T})$ be a close pair with constant R_2. Let the quantity $p_Y(z,j) = |a(z)|^2 p_X(z,j)$ satisfy the LSW spectrum conditions from Van Bellegem and von Sachs (2008) for all $j = 1, 2, \ldots$.

Then the array of stochastic processes $Y_{t,T} = a_{t,T} X_{t,T}$ is a LSW process with spectrum given by $p_Y(z,j)$ for $j = 1, 2, \ldots$.

The proof is in the appendix. Now we show that, under certain conditions, the sum of two LSW processes is LSW.

THEOREM 2. *Let*

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k} x_{j,k} \psi_{j,k-t} \xi_{j,k} \text{ and } Y_{t,T} = \sum_{j=1}^{\infty} \sum_{k} y_{j,k} \psi_{j,k-t} \eta_{j,k} \tag{18}$$

be LSW processes with evolutionary wavelet spectra (EWS) $p_X(z,j)$ and $p_Y(z,j)$ respectively.

Let $U_{t,T} = X_{t,T} + Y_{t,T}$. From Assumption 1 let $(\rho_{j,k}, R_j(z))$ be a close pair, where $R_j(z)$ is of bounded total variation, then $U_{t,T}$ is a LSW process with spectrum given by $p_X(z,j) + p_Y(z,j) + 2R_j(z) \{p_X(z,j) p_Y(z,j)\}^{1/2}$.

The proof is in the appendix. The quantity $R_j(z)$ for LSW processes was considered by Sanderson et al. (2010) as part of a study of the *inter*dependence of LSW processes via wavelet coherence.

7. Future directions

This article has introduced the concept of costationarity between two LS time series and proposed the benefits of relationship discovery, strength and the use of the costationary series itself. The article raises several interesting possibilities for further fruitful study. Here follows some possibilities: 1. Investigation of strategies for finding costationary solutions when there are more than two series. One could consider a large optimisation which merely enters all series into the combination and minimises the $\mathcal{T}(\cdot)$ measure of the combination as above. Alternatively, one might consider searching for sub-groups of series which can combine to make stationary series, which can then themselves be combined to form an overall stationary series. 2. Whether restricting to certain classes of LS processes could assist with theoretical approaches to finding costationary solutions, possibly in the 'model-independent' covariance domain. For example, we might expand on Example 4 to different process classes to see what the costationary partner might be in certain circumstances. 3. Investigation of the costationary-based estimate of cross-covariance that was mentioned in Remark 7. 4. Development of methods to identify confidence bands for the combination vectors. 5. An 'online' version of our method would clearly be desirable. For example, having found a particular synthetic asset in the costationary asset allocation example, or a particular way for distributing wind energy in that example, one might be interested in updating the costationary vectors progressively as more data was collected. Depending on circumstances, similar values of costationary vectors might be enough to maintain costationarity, or the vectors may need to evolve to different values, or indeed costationarity might be lost. One can envisage a continual updating by re-running the entire costationary algorithm from scratch. This might well be feasible in circumstances where the sampling rate is not too frequent. However, it would clearly be of interest to be able to develop a computationally efficient algorithm that did not have to re-run the whole analysis. 6. The `Costat` algorithm treats the 'optimal' (α_t^*, β_t^*) as deterministic functions when, of course, they are stochastic as they depend on the (X_t, Y_t) realizations. The `Costat Bootstat` method could possibly be enhanced

by recognizing this fact and include resampling from X_t, Y_t themselves. 7. More work is needed on understanding likely choices of C, the complexity constraint. Roughly speaking, the procedure we have adopted is that a value of C is chosen and if no costationarity is found then C can be relaxed, including progressively more complex functions until costationarity is detected (eventually, it must be given arbitrary functions to combine with). If costationarity is discovered straight away then C can be tightened, progressively making the function class contain less complex functions until no costationary solutions can be found (although note that this does not mean that such a solution does not exist, just that it cannot be found). Possibly, the most *interesting* value of C is the one on the boundary between no costationary solutions and one or some costationary solutions, and this says something about the degree of 'covarying' between the two series X_t and Y_t. If the series are sympathetic then costationarity can be found easily with 'simple' functions. Cross-validation could be used to discover the 'boundary' values of C.

Appendices

A. Specific implementation issues

We describe three specific implementational issues for `Costat` for the examples given in this paper.

Starting vectors. As is common in projection pursuit applications we repeatedly run `Costat` from random initial starting vectors.

Spectral estimate. We used the spectral estimate $\hat{S}_j(z)$ from NvSK, but could have used the Haar-Fisz estimate from Fryzlewicz and Nason (2006). For Fourier-inspired spectral estimation there are several time-varying spectral estimators that could be considered. For example, Neumann and von Sachs (1997), Adak (1998) or Guo et al. (2003).

Penalty. We restrict the vectors to be piecewise constant with breaks only at dyadic fractional locations. We first attempt to discover costationary vectors which are piecewise constant at $z = 1/2$ and then, if this is unsuccessful, we permit breaks at $z = 1/4, 1/2$ and $3/4$, and then successively refine the dyadic grid, at stage j the breaks are permitted at $\{2^{-j}i\}_{i=1}^{2^j-1}$. In the previous section "attempt" means run the algorithm from many, N, random starts and ascertain whether any of them indicate stationarity, if not then proceed to the next finest dyadic scale of breaks. Limitations: the penalty described here is crude, but the combination vectors have the advantage of being extremely simple to code/compute through their Haar wavelet coefficients. Alternative penalized combination vectors were considered in Remark 1 above.

Numerical minimizer The minimizer we use is `optim()` from within R with default settings.

All the issues, apart from the second, apply to any form of LS process with spectrum p. For the second, one needs to replace $\hat{S}$ by the form of spectral estimate of p related to the model in use.

B. Proofs

PROOF (OF EXAMPLE 4). To momentarily simplify notation let $\alpha = a_1(z)$. First expand $G^2(z,\lambda)$ from (6) to obtain $G^2(z,\lambda) = \sigma_1^2|1+\alpha e^{i\lambda}|^2 - \sigma^2$. Then dividing by $\pi|1+\alpha e^{i\lambda}|^2$ yields $f_Y(z,\lambda) = \sigma_1^2/\pi - f_X(z,\lambda)$. Why is Y_t a TVARMA$(1,1)$ process? The denominator of the spectrum $f_Y(z,\lambda)$ is a legitimate AR part. The numerator of the spectrum of a TVARMA$(1,1)$ is given by

$$\beta_0^2 + \beta_1^2 + 2\beta_0\beta_1\cos(\lambda) \tag{19}$$

(remember that the βs depend on z). Equating coefficients from (6) and (19) (without the σ_1^2 which goes with the π) gives

$$\beta_0^2 + \beta_1^2 = (1+\alpha^2) - \sigma^2/\sigma_1^2 \text{ and } \beta_0\beta_1 = \alpha. \tag{20}$$

We find conditions on σ_1^2 for this system to be valid, i.e. the LH equation ≥ 0. For $\beta_0^2 + \beta_1^2 \geq 0$ we can rearrange the LHS of (20) to yield $\sigma_1^2 \geq \sigma^2/(1+\alpha^2)$. Choosing $\sigma_1^2 \geq \sigma^2$ always works.

PROOF (OF THEOREM 1). We pick one fixed scale $j \in \{1, 2, 3, \ldots\}$. The LSW process at this scale is given by (we can omit the j)

$$X_{t,T} = \sum_{k=-\infty}^{\infty} w_{k;T}\psi_{k-t}\xi_k. \tag{21}$$

Since $X_{t,T}$ is a LSW process the quantity $w_{k;T}$ is part of a close pair with $W(z)$ with constant, R_1, say. The support of wavelet $\psi_{j,k}$ is finite on the domain $k = 0, \ldots, L_j - 1$ and so

$$X_{t,T} = \sum_{k=t}^{t+L_j-1} w_{k;T}\psi_{k-t}\xi_k. \tag{22}$$

The quantity we are interested in is

$$\begin{aligned}
a_{t,T}X_t &= \left\{a\left(\frac{t}{T}\right) + R_2T^{-1}\right\} \sum_{k=t}^{t+L_j-1} w_{k;T}\psi_{k-t}\xi_k && (23)\\
&= \left\{a\left(\frac{t}{T}\right) + R_2T^{-1}\right\} w_t\psi_0\xi_t && (24)\\
&+ \left\{a\left(\frac{t+1}{T}\right) + R_3T^{-1} + R_2T^{-1}\right\} w_{t+1}\psi_1\xi_{t+1} && (25)\\
&+ \left\{a\left(\frac{t+2}{T}\right) + 2R_3T^{-1} + R_2T^{-1}\right\} w_{t+2}\psi_2\xi_{t+2} && (26)\\
&+ \ldots && (27)\\
&+ \left\{a\left(\frac{t+L_j-1}{T}\right) + (L_j-1)R_3T^{-1} + R_2T^{-1}\right\} w_{t+L_j-1}\psi_{L_j-1}\xi_{t+L_j-1}. && (28)
\end{aligned}$$

Rearranging gives $a_{t,T}X_t = Y_{t,T} + R_2X_{t,T}/T + R_3B_{t,T}/T$ where $Y_{t,T}$ and $B_{t,T}$ are given below.

The process

$$\begin{aligned}
Y_{t,T} &= a_tw_t\psi_0\xi_t + a_{t+1}w_{t+1}\psi_1\xi_{t+1} + \cdots + a_{t+L_j-1}w_{t+L_j-1}\psi_{L_j-1}\xi_{t+L_j-1} && (29)\\
&= \sum_{k=t}^{t+L_j-1} u_{k,T}\psi_{k-t}\xi_k, && (30)
\end{aligned}$$

where $(u(t), u_{t,T})$ is a close pair where $u(z) = a(z)W(z)$, and u is of bounded variation as it is the product of two functions of bounded variation.

The process $B_{t,T}$ is given by

$$B_{t,T} = \left(w_{t+1}\psi_1\xi_{t+1} + 2w_{t+2}\psi_2\xi_{t+2} + \cdots + w_{t+L_j-1}\psi_{L_j-1}\xi_{t+L_j-1}\right) R_3/T. \tag{31}$$

The proof that $a_{t,T}X_{t,T}$ is LSW can be completed by noticing that

$$\left|(a_t + R_j/T)\, w_t - a\left(\frac{t}{T}\right) W\left(\frac{t}{T}\right)\right| \leq \left|a_t w_t - a\left(\frac{t}{T}\right) W\left(\frac{t}{T}\right)\right| + R_j|w_t|/T \to 0 \qquad (32)$$

as $T \to \infty$. This part of the proof is completely non-stochastic and essentially maps w_t to $a_t w_t$ everywhere, and due to the original 'close pair' link in the LSW defintion (formula (10) in NvSK). The same process can be applied to all scales simultaneously.

As a corollary note that both $X_{t,T}/T$ and $B_{t,T}/T$ are asymptotically zero in the mean square sense. This is because $X_{t,T}$ is obviously a LSW process with covariance of $c_T(z,\tau)$ hence

$$\mathbb{E}\left\{\left(T^{-1}X_{t,T}\right)^2\right\} = T^{-2}\left\{c\left(t/T, 0\right) + \mathcal{O}(T^{-1})\right\} \to 0,$$

as $T \to \infty$ using Proposition 1 of NvSK. A similar argument applies to $B_{t,T}/T$ which can be seen as the LSW process $X_{t,T}$ with its (square-root) spectrum multiplied by a linear function and hence, by this theorem, a LSW process.

PROOF (OF THEOREM 2). Define

$$\mu_{j,k} = x_{j,k}\xi_{j,k} + y_{j,k}\eta_{j,k}. \qquad (33)$$

Then the quantity $U_{t,T}$ can be written

$$U_{t,T} = \sum_j \sum_k \mu_{j,k}\psi_{j,k-t}. \qquad (34)$$

From the properties of $\{\xi_{j,k}\}$ and $\{\eta_{j,k}\}$ we have that $\mathbb{E}\mu_{j,k} = 0$ and

$$\begin{aligned} \mathrm{var}(\mu_{j,k}) &= \mathbb{E}(\mu_{j,k}^2) && (35)\\ &= x_{j,k}^2\mathrm{var}(\xi_{j,k}) + y_{j,k}^2\mathrm{var}(\eta_{j,k}) + 2x_{j,k}y_{j,k}\mathrm{cov}(\xi_{j,k},\eta_{j,k}) && (36)\\ &= x_{j,k}^2 + y_{j,k}^2 + 2x_{j,k}y_{j,k}\rho_{j,k} = \nu_{j,k}^2 && (37)\end{aligned}$$

Assume, for the moment, that $\nu_{j,k}^2 > 0$ for all j,k. Now let $\epsilon_{j,k} = \nu_{j,k}^{-1}\mu_{j,k}$. Clearly, $\mathbb{E}(\epsilon_{j,k}) = 0$ and $\mathrm{cov}(\epsilon_{j,k},\epsilon_{\ell,m}) = \delta_{j,\ell}\delta_{k,m}$. In other words, $\{\epsilon_{j,k}\}$ satisfies the conditions for being an LSW innovation sequence. Hence,

$$U_{t,T} = \sum_j \sum_k \nu_{j,k}\psi_{j,k-t}\epsilon_{j,k}. \qquad (38)$$

Case (a). If $\xi_{j,k}$ and $\eta_{\ell,m}$ are independent for all j,k,ℓ,m then $\rho_{j,k} = 0$ for all j,k. Since $(x_{j,k}, p_X^{1/2}(z,j))$ and $(y_{j,k}, p_Y^{1/2}(z,j))$ are close pairs we have

$$\begin{aligned} \nu_{j,k}^2 &= x_{j,k}^2 + y_{j,k}^2 && (39)\\ &= p_X(k/T,j) + \mathcal{O}(T^{-1}) + p_Y(k/T,j) + \mathcal{O}(T^{-1}). && (40)\end{aligned}$$

By Fryzlewicz (2003) p. 93 $(\nu_{j,k}^2, p_X(z,j) + p_Y(z,j))$ is a close pair and the result is proven.

Case (b.) Here we have

$$\begin{aligned} \nu_{j,k}^2 &= x_{j,k}^2 + y_{j,k}^2 + 2x_{j,k}y_{j,k}\rho_{j,k} && (41)\\ &= p_X(k/T,j) + p_Y(k/T,j) + 2p_X^{1/2}(k/T,j)p_Y^{1/2}(k/T,j)R_j(k/T) + \mathcal{O}(T^{-1}) && (42)\end{aligned}$$

and again, the result is proven (bounded variation is preserved by sums, products and square roots).

C. Actual estimator used for our LSW examples

For our LSW examples we use the simple time-average of the fast and simple $\mathbf{L}_k$ estimator of $\mathbf{S}(k/T)$ from NvSK whose consistency is demonstrated as follows.

THEOREM 3. *Let $X_{t,T}$ be a stationary Gaussian LSW process satisfying $S_j \leq D2^j$ for all j for some $D > 0$. Then $T^{-1}\sum_{k=1}^{T} \mathbf{L}_k$ is a consistent estimator of* $\mathbf{S}$.

PROOF (OF THEOREM 3). We are interested in estimation of the vector $\mathbf{S} = (S_1, \ldots, S_J)^T$ where here the S are constants wrt z because of stationarity. To begin we follow the idea of Fryzlewicz and Nason (2006) and consider estimation of the $J \times 1$ vector quantity $\beta = AS$ where A is the $J \times J$ inner product matrix of autocorrelation wavelets from NvSK. We let $\hat{\beta} = T^{-1}\sum_{k=1}^{T} \mathbf{I}_k$ be our estimator of β where $\mathbf{I}_k = (I_{1k}, \ldots, I_{Jk})^T$ is the raw wavelet periodogram from NvSK.

NvSK prove that $\mathbb{E}(\mathbf{I}_k) = A\mathbf{S}_k + \mathcal{O}(T^{-1})$. Hence in the stationary situation $\mathbb{E}(\mathbf{I}_k) = A\mathbf{S} + \mathcal{O}(T^{-1})$ for all k and thus $\hat{\beta}$ is asymptotically unbiased for β. Further, they also demonstrated that the variance of I_{jk}, call this σ_j^2, is (asymptotically) constant, i.e. this is the variance of the jth component of each term in $\hat{\beta}$. It does not depend on t.

For consistency it only remains to show that $\mathrm{var}(\hat{\beta}_j) \to 0$ as $T \to \infty$ for each j. The variance of $\hat{\beta}$ is given by:

$$\mathrm{var}(\hat{\beta}_j) = T^{-2}\sum_{k=1}^{T} \mathrm{var}(I_{jk}) + 2T^{-2}\sum_{k=1}^{T}\sum_{\ell=k+1}^{T} \mathrm{cov}(I_{jk}, I_{j\ell}). \tag{43}$$

The first term of the RHS of (43) clearly tends to zero as $T \to \infty$ as the variance is finite. Fryzlewicz and Nason (2004) show that: $(**) = \mathrm{cov}\,(I_{jk}, I_{j\ell}) = 2\left(\sum_{\tau=-\infty}^{\infty} c\left(\frac{t}{T}, \tau\right)\Psi_j(\tau + \ell - k)\right)^2 + \mathcal{O}(2^{-j}/T)$, where $c(t/T, \tau)$ is the localized autocovariance and $\Psi_j(\tau)$ is the autocorrelation wavelet from NvSK and $\{\psi_{j,k}\}$ are the discrete nondecimated wavelets as defined in their introduction.

However, $\Psi_j(\tau)$ is compactly supported with support (a, b) which depend on j (which is **fixed**) here (precise values of a and b are to be found in Eckley and Nason (2005)). Further, $\Psi_j(\tau) \leq \Psi_j(0) = 1$: the inequality because Ψ_j is an autocovariance function and the equality from NvSK.

For stationary processes the time-varying autocovariance $c(t/T, \tau) = c(\tau)$, the regular autocovariance. NvSK show that, necessarily, for stationary LSW processes $\sum_{\tau=-\infty}^{\infty} |c(\tau)| < \infty$. Using the compact support of $\Psi_j(\tau)$ means that the first term of $(**)$ can be written:
$2\left(\sum_{\tau=a-(\ell-k)}^{b-(\ell-k)} c(\tau)\Psi_j(\tau + \ell - k)\right)^2 \leq 2\left(\sum_{\tau=a}^{b} |c\{\tau - (\ell - k)\}|\right)^2$. Then from (43) we sum from $\ell = k + 1$ to T as $T \to \infty$, and since $\sum |c(\tau)| < \infty$ the sum over ℓ in (43) is finite by the comparison test.

References

Adak, S. (1998) Time-dependent spectral analysis of nonstationary time series, *J. Am. Statist. Ass.*, **93**, 1488–1501.

Ahamada, I. and Boutahar, M. (2002) Tests for covariance stationarity and white noise, with an application to Euro/US dollar exchange rate: an approach based on the evolutionary spectral density, *Econ. Lett.*, **77**, 177–186.

Andreou, E. and Ghysels, E. (2002) Detecting multiple breaks in financial market volatility dynamics, *J. Appl. Econometrics*, **17**, 579–600.

Best, M. and Grauer, R. (1991) On the sensitivity of mean-variance-efficient portfolios to changes in asset mean: some analytical and computational results, *Rev. Fin. Stud.*, **4**, 315–342.

Bierens, H. and Martins, L. (2010) Time-varying cointegration, *Econometric Theory*, **26**, (to appear).

Bollerslev, T. (1990) Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *The Review of Economics and Statistics*, **72**, 498–505.

Bollerslev, T., Engle, R., and Wooldridge, J. (1988) A capital-asset pricing model with time-varying covariances, *J. Polit. Econ.*, **96**, 116–131.

Coifman, R. R. and Donoho, D. L. (1995) Translation-invariant de-noising, in A. Antoniadis and G. Oppenheim, eds., *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pp. 125–150, Springer-Verlag, New York.

Croux, C., Forni, M., and Reichlin, L. (2001) A measure of comovement for economic variables: theory and empirics, *Rev. Econ. Stat.*, **83**, 232–241.

Dahlhaus, R. (1996) Asymptotic statistical inference for nonstationary processes with evolutionary spectra, in *Athens Conference on Applied Probability and Time Series Analysis. Lecture Notes in Statistics, Vol. 115*, pp. 145–159, Springer Verlag, New York.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes, *Ann. Statist.*, **25**, 1–37.

Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes, *Ann. Statist.*, **28**, 1762–1794.

Dahlhaus, R. and Polonik, W. (2006) Nonparametric quasi maximum likelihood estimation for gaussian locally stationary processes, *Ann. Statist.*, **34**, 2790–2824.

Dahlhaus, R. and Polonik, W. (2009) Empirical spectral processes for locally stationary time series, *Bernoulli*, **15**, 1–39.

Dahlhaus, R. and Subba Rao, S. (2006) Statistical inference for time-varying ARCH processes, *Ann. Statist.*, **34**, 1075–1114.

Dahlhaus, R. and Subba Rao, S. (2007) A recursive online algorithm for the estimation of time-varying ARCH parameters, *Bernoulli*, **13**, 389–422.

Davis, R., Lee, T., and Rodriguez-Yam, G. (2008) Break detection for a class of nonlinear time series models, *J. Time Ser. Anal.*, **29**, 834–867.

Davison, A. (2008) Personal communication.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge.

Eckley, I. A. and Nason, G. P. (2005) Efficient computation of the discrete autocorrelation wavelet inner product matrix., *Statistics and Computing*, **15**, 83–92.

Engle, R. and Kozicki, S. (1993) Testing for common features, *J. Bus. Econ. Stat.*, **11**, 369–380.

Engle, R. and Kroner, K. (1995) Multivariate simultaneous generalized ARCH, *Econ. Th.*, **11**, 122–150.

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

Fama, E. (1970) Multi-period consumption-investment decisions, *Amer. Econ. Rev.*, **60**, 163–174.

Firth, J. (1992) *Discrete Transforms*, Chapman and Hall, London.

Friedman, J. and Tukey, J. (1974) A projection pursuit algorithm for exploratory data analysis, *IEEE Trans. Comput.*, **23**, 881–890.

Fryzlewicz, P. and Nason, G. P. (2004) Smoothing the wavelet periodogram using the Haar-Fisz transform., Technical Report 03:08, Statistics Group, Department of Mathematics, University of Bristol.

Fryzlewicz, P. and Nason, G. P. (2006) Haar-Fisz estimation of evolutionary wavelet spectra, *J. R. Statist. Soc.* B, **68**, 611–634.

Fryzlewicz, P. Z. (2003) *Wavelet Techniques for Time Series and Poisson Data*, Ph.D. thesis, University of Bristol, U.K.

Gourieroux, C. and Peaucelle, I. (1992) Series Codépendantes: Application a l'Hypothèse de Parité du Pouvoir d'Achat, *l'Actualité Economique*, **68**, 283–304.

Granger, C. (2008) Non-linear models: where do we go next — time varying parameter models?, *Studies in Nonlinear Dynamics & Econometrics*, **12**, Article 1.

Grech, D. and Mazur, Z. (2004) Can one make any crash prediction in finance using the local Hurst exponent idea?, *Physica A*, **336**, 133–145.

Guo, W., Dai, M., Ombao, H., and von Sachs, R. (2003) Smoothing spline ANOVA for time-dependent spectral analysis, *J. Am. Statist. Ass.*, **98**, 643–652.

Hamilton, J. and Raj, B., eds. (2002) *Advances in Markov-Switching Models*, Physica-Verlag, Heidelberg.

Hirst, E. (2002) Integrating wind output with bulk power operations and wholesale electricity markets, *Wind Energy*, **5**, 19–36.

Holtinnen, H. (2005) Impact of hourly wind power variations on the system operation in the Nordic countries, *Wind Energy*, **8**, 197–218.

Johansen, S., Mosconi, R., and Nielsen, B. (2000) Cointegration analysis in the presence of structural breaks in the deterministic trend, *The Econometrics J.*, **3**, 216–249.

Jones, M. and Sibson, R. (1987) What is projection pursuit? (with discussion), *J. R. Statist. Soc.* B, **150**, 1–36.

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003) Flexible multivariate GARCH modeling with an application to international stock markets, *The Review of Economics and Statistics*, **85**, 735–747.

Lewis, G. M. (2008) High value wind: a method to explore the relationship between wind speed and electricity locational marginal price, *Renewable Energy*, **33**, 1843–1853.

Lintner, J. (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics*, **47**, 13–37.

Markowitz, H. (1952) Portfolio selection, *J. Finance*, **7**, 77–91.

Merton, R. (1971) Consumption and portfolio rules in a continuous-time model, *J. Econ. Th.*, **3**, 373–413.

Mossin, J. (1966) Equilibrium in a capital asset pricing model, *Econometrica*, **34**, 768–783.

Nason, G. P. (2001) Robust projection indices, *J. R. Statist. Soc.* B, **63**, 551–567.

Nason, G. P. (2008) *Wavelet Methods in Statistics with R*, Springer, New York.

Nason, G. P. and Sapatinas, T. (2002) Wavelet packet transfer function modelling of nonstationary time series, *Statistics and Computing*, **12**, 45–56.

Nason, G. P. and von Sachs, R. (1999) Wavelets in time series analysis, *Phil. Trans. R. Soc. Lond.* A, **357**, 2511–2526.

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *J. R. Statist. Soc.* B, **62**, 271–292.

Neumann, M. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *Ann. Statist.*, **25**, 38–76.

Paparoditis, E. (2009) Testing temporal constancy of the spectral structure of a time series, *Bernoulli*, **15**, 1190–1221.

Percival, D. B. and Constantine, W. L. B. (2006) Exact simulation of Gaussian time series from nonparametric spectral estimates with application to bootstrapping, *Statistics and Computing*, **16**, 25–35.

Priestley, M. B. (1983) *Spectral Analysis and Time Series*, Academic Press, London.

Priestley, M. B. and Subba Rao, T. (1969) A test for non-stationarity of time-series, *J. R. Statist. Soc.* B, **31**, 140–149.

Sanderson, J., Fryzlewicz, P., and Jones, M. (2010) Measuring dependence between non-stationary time series using the locally stationary wavelet model, *Biometrika*, **97**, 435–446.

Sharpe, W. (1964) Capital asset prices: a theory of market equilibrium under conditions of risk, *J. Finance*, **19**, 425–442.

Sharpe, W. (1966) Mutual fund performance, *J. Bus.*, **39**, 119–138.

Skea, J., Anderson, D., Green, T., Gross, R., Heptonstall, P., and Leach, M. (2008) Intermittent renewable generation and cost of maintaining power system reliability, *IET Gener. Transm. Distrib.*, **2**, 82–89.

Stărică, C. and Granger, C. (2005) Nonstationarities in stock returns, *Review of Economics and Statistics*, **87**, 503–522.

Urga, G. (2007) Common features in economics and finance: an overview of recent developments, *J. Bus. Econ. Stat.*, **25**, 2–11.

Van Bellegem, S. and Dahlhaus, R. (2006) Semiparametric estimation by model selection for locally stationary processes, *J. R. Statist. Soc.* B, **68**, 721–746.

Van Bellegem, S. and von Sachs, R. (2008) Locally adaptive estimation of evolutionary wavelet spectra, *Ann. Statist.*, **36**, 1879–1924.

von Sachs, R. and Neumann, M. H. (2000) A wavelet-based test for stationarity, *J. Time Ser. Anal.*, **21**, 597–613.

von Sachs, R. and Schneider, K. (1996) Wavelet smoothing of evolutionary spectra by nonlinear thresholding, *App. Comp. Harm. Anal.*, **3**, 268–282.