INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0227

FORECASTING ECONOMIC TIME SERIES USING MODELS OF NONSTATIONARITY

S. VAN BELLEGEM and R. VON SACHS

http://www.stat.ucl.ac.be

Forecasting economic time series using models of nonstationarity

Sébastien Van Bellegem[1] Rainer von Sachs[2]

September 3, 2002

Abstract

The classical forecasting theory of stationary time series exploits the second-order structure (variance, autocovariance and spectral density) of an observed process in order to construct some prediction intervals. However, some economic time series show a time-varying unconditional second-order structure. This article focus on a simple and meaningful model allowing this nonstationary behaviour. We show that this model satisfactory explains the nonstationary behaviour of several economic data sets, among which are the U.S. stock returns and exchange rates. The question how to forecast these processes is addressed and evaluated on the data sets.

Keywords: Covariance nonstationarity, rescaled time, time-modulated process, nonparametric estimation, forecasting.

Running head: Forecasting nonstationary processes

[1] *Corresponding author.* Research Fellow of the National Fund for Scientific Research (FNRS). Institut de statistique, Université catholique de Louvain, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Telephone: + 32 10 47.41.60. Fax: + 32 10 47.30.32. E-mail: `vanbellegem@stat.ucl.ac.be`

[2] Institut de statistique, Université catholique de Louvain, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Telephone: +32 10 47.88.06. Fax: + 32 10 47.30.32. E-mail: `vonsachs@stat.ucl.ac.be`

1 Introduction

To forecast economic time series, many analyses are based on the assumption that the probabilistic properties of the underlying process are time-invariant. Even if this assumption is very useful in order to construct simple predictors, it seems not to be the best strategy in practice. Indeed, taking into account structural changes of parameters may lead to better forecasting performance because it is sometimes more adequate to capture some aspects of the real world (Clements and Hendry, 1998, Section 7.4). More surprisingly, some stylised facts, as the long range dependence of the absolute returns of financial time series, can even be explained by a parameter change of the model (Mikosch and Stărică, 1999). In terms of forecasting, Swanson and White (1997) conclude that models allowing parameter variability can show better accuracy and adaptability than constant models on macroeconomic data. These empirical and theoretical works are in favour of a model with evolutionary parameters, reflecting the evolution of the economy over time. An important consequence of this observation is the lack of stationarity of the data, which means that the *unconditional* moments of the time series *can vary during time*.

This evolution of unconditional moments has been statistically tested on various economic processes. For instance, Pagan and Schwert (1990a, 1990b) develop statistical tests rejecting the hypothesis that the monthly U.S. stock returns are covariance stationary, as long as the period of Great Depression (1929–1939) is included in the series. Loretan and Phillips (1994) confirm this conclusion for stock returns, but also for a set of exchange rate data. Similar results have been found by Los (2000) on short series of weekly price indeces of Asian stock markets.

In this article, we go one step further and study how this nonstationary behaviour can be modeled. We focus on a very simple model of nonstationarity with an unconditional variance evolving with time. This model includes the so-called "time-modulated processes" studied in Fryźlewicz, Van Bellegem and von Sachs (2002), and we generalize their definition in Section 2 of this article. In this section, we also derive some basic properties of these processes. Estimation and forecasting procedures are described in Section 3.

Section 4 is devoted to the practical evaluation of these nonstationary processes on several financial time series among which are the U.S. stock returns and exchange rates. Analyses focus on various stock returns and exchange rates and we show that our simple and meaningful model of nonstationarity provides a satisfactory explanation power of the nonstationary behaviour of the observed data.

2 Modelling evolutionary variance

The most simple nonstationary model consists of a second-order stationary process modulated by a deterministic time-varying variance. If $Y_t, t = 0, 1, 2, \ldots$, is a zero mean stationary process, a very simple nonstationary model is given by $X_t = \sigma_t Y_t$, where σ_t is a deterministic time-varying function which is strictly positive. As we suppose the process Y_t to be zero mean and stationary, the nonstationarity of X_t is only explained by its evolving unconditional variance.

With this model of variance nonstationarity, it may seem contradictory to construct a forecasting theory, since a predictor exploits generally an invariance structure in the unconditional moments of the process. This problem is overcome if we add regularity assumptions on the deterministic function σ_t. For instance, we can impose that σ_t is a piecewise constant function. More generally, we can assume that σ_t is nearly constant along intervals of a certain length τ. Using this regularity assumption, we can estimate and extrapolate the deterministic variance to build a predictor.

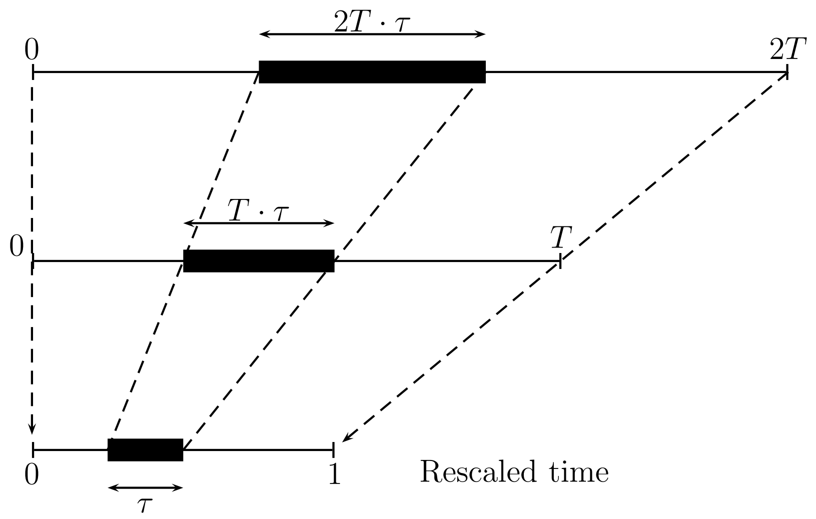
However, this approach is not satisfactory since it implicitely imposes that the function σ_t is estimable only using τ observations. In this framework, when the length of the data set increases, no improvement is possible in the estimation of σ_t over this interval of length τ. This implies that asymptotic considerations can not be used in the statistical inference of such process. This is a substantial drawback, because the usual statistical properties of estimators, as consistency, efficiency or central limit theorems cannot be used to measure and to compare the quality of different estimators.

To overcome this (theoretical) problem, Dahlhaus (1997) introduced a concept of "local stationarity" in a general context of covariance evolution.

In our situation, this concept is as follows: Suppose we observe the series from time 0 up to $T-1$ (T observations). The local stationarity assumption postulates the existence of a deterministic function $\sigma(z)$ defined for $z \in [0,1)$ such that the approximation $\sigma_t \approx \sigma(t/T)$ holds in an appropriate way that we will define below. In this approach, two scales of time are defined: The *observed time*, which is the usual scale of time $0,\ldots,T-1$, and the *rescaled time* defined on the interval $[0,1)$. There exists a mapping between these two scales, and since this mapping depends on the sample size T, the resulting nonstationary process is doubly indexed:

$$X_{t,T} = \sigma\left(\frac{t}{T}\right) Y_t, \tag{2.1}$$

where Y_t is a zero-mean stationary process with unit variance. The regularity assumptions are now made on the function $\sigma(z)$ defined on $[0,1)$. Due to the mapping between $0,\ldots,T-1$ and $[0,1)$, the estimation of $\sigma(z)$ becomes a standard statistical problem: For instance if $\sigma(z)$ is constant on an interval of length $\tau < 1$ in the rescaled time, then it may be estimated using $\tau \cdot T$ observed data in the real time:



Consequently, if T increases, we get more observations to estimate $\sigma(z)$ on τ. Estimation of $\sigma(z)$ is then parallel to the estimation of a regression function in a nonparametric regression problem.

An important consequence of the rescaled time is the new interpretation of "asymptotics". When T tends to infinity, we get more information on the local structure of $\sigma^2(z)$ in the rescaled time, because the mapping defines a finer grid in the recaled time. However, it does *not* mean that we look into the future, because the rescaled time has a fixed bounded support $[0, 1)$. To understand this point, an analogy with the spectral analysis of stationary time series may help. Indeed, the spectrum of a stationary process is defined on the interval $[-\pi, \pi]$. When we observe T data, T finite, the spectrum is not identifiable on $[-\pi, \pi]$ because we can only estimate the covariance of the process up to lag $T-1$, and then the lowest frequencies of the spectrum are not identifiable. When T increases, we get more information on the spectrum on $[-\pi, \pi]$, and this spectrum is uniquely defined only *asymptotically* (Dahlhaus, 1996). The rescaled time framework is analogous, in the sense that "T tends to infinity" now means that we observe finer details in the rescaled time.

In this article, we study model (2.1) in some particular cases, corresponding to the specification of Y_t and $\sigma(z)$. We consider

[tm-WN] Time-modulated White Noise processes, which are processes (2.1) where Y_t is a White Noise process;

[tm-ARMA] Time-modulated ARMA processes, which are processes (2.1) where Y_t is a general ARMA process.

An other important specification is the regularity assumption made on the deterministic function $\sigma(z)$. In this article, we study two cases of regularity:

[PC] is the case where $\sigma^2(z)$ is modeled by a piecewise constant function, with a finite number of jumps,

[Lip] is the case where $\sigma^2(z)$ is a continuous function without jump. In this case, $\sigma^2(z)$ is modeled following Fryźlewicz *et al.* (2002) as a Lipschitz continuous function, i.e. $|\sigma^2(z)-\sigma^2(z_0)| \leqslant C\cdot|z-z_0|$ for all $0 < z_0, z < 1$.

In addition, we assume in the two cases that the variance is bounded away from zero, i.e. there exists $\delta > 0$ such that $\sigma(z) \geqslant \delta$ uniformly in z.

Let us now derive some properties of these processes. We first consider the case of a tm-WN process $\{X_{t,T}\}$. Our definition implies that the expectation and the conditional expectation of $X_{t,T}$ is zero, which we denote by

$$\mathrm{E}\left(X_{t,T}\right) = \mathrm{E}\left(X_{t,T} | \mathcal{F}_{t-1,T}\right) = 0$$

where $\mathcal{F}_{t-1,T}$ stands for the set of the observed values $X_{0,T}, \ldots, X_{t-1,T}$. Concerning the (conditional) variance, we get

$$\mathrm{Var}\left(X_{t,T}\right) = \mathrm{Var}\left(X_{t,T} | \mathcal{F}_{t-1,T}\right) = \sigma^2\left(\frac{t}{T}\right)$$

(recall that the variance of the stationary process Y_t in equation (2.1) is assumed to be one). The covariance and conditional covariance of a tm-WN are trivially zero. A more interesting property concerns the kurtosis κ_2:

$$\kappa_2 := \frac{\mathrm{E}X_{t,T}^4}{\left(\mathrm{E}X_{t,T}^2\right)^2} = \frac{\mathrm{E}Y_{t,T}^4}{\left(\mathrm{E}Y_{t,T}^2\right)^2} \tag{2.2}$$

and it follows that *the kurtosis of a tm-WN model is the kurtosis of its White Noise*. This simple property has a particular importance in the modeling of economic series, which are known to be leptokurtic. This stylised fact and equation (2.2) indicate that the distribution of Y_t in equation (2.1) is potentially leptokurtic. Consequently, we allow the white-noise process Y_t to be non Gaussian and to follow a possible very general distribution.

These results are easily extended to the case of tm-ARMA processes. Consider equation (2.1) where Y_t is the stationary and invertible ARMA(p,q) process

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

where $\{Z_t\}$ is a zero-mean white noise process such that $\mathrm{Var}\, Y_t = 1$. Direct considerations yield

$$\mathrm{E}X_{t,T} = 0 \tag{2.3}$$

$$\mathrm{E}\left(X_{t,T} | \mathcal{F}_{t-1,T}\right) = \sigma\left(\frac{t}{T}\right)\{\phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}\} \tag{2.4}$$

and

$$\operatorname{Var} X_{t,T} = \sigma^2\left(\frac{t}{T}\right) \qquad \operatorname{Var}\left(X_{t,T}|\mathcal{F}_{t-1,T}\right) = \sigma^2\left(\frac{t}{T}\right). \tag{2.5}$$

A similar property holds for the kurtosis: The kurtosis of a time-modulated ARMA process is the kurtosis of the constituting ARMA process, and then the distribution of the process $\{Z_t\}$ is potentially non Gaussian. This allows the tm-ARMA process to be leptokurtic, in accordance with one of the stylised facts of economic processes.

3 Estimation and forecasting

In the scope of the present article, we want to study how time-modulated processes explain the nonstationary behaviour of some data sets. This section introduces all the tools we need for this goal. First, we recall two tests of stationarity studied in the literature. Then, we propose three different estimators of the local variance function from data. Finally, we address the problem how to forecast time-modulated processes, and propose a way to construct prediction intervals in practice.

3.1 Testing for covariance stationarity

In what follows, we will use tests of covariance stationarity. Let us briefly recall two of the tests presented in Pagan and Schwert (1990b).

The first test is called *post-sample prediction test*. Suppose we observe the zero-mean process $X_{0,T},\ldots,X_{T-1,T}$ and split the time axis in $T = T_1 + T_2$ with $T_1 = T_2$. If we want to test the hypothesis that the variance on $X_{0,T},\ldots,X_{T_1-1,T}$ is equal to the variance on $X_{T_1,T},\ldots,X_{T_1,T-1}$, a suitable test statistic is

$$\hat{\tau} = \hat{\sigma}_1^2 - \hat{\sigma}_2^2$$

where $\hat{\sigma}_i^2$ is the sample variance on the ith segment. Under the null hypothesis, the distribution of $T_1^{1/2}\hat{\tau}$ is asymptotically normal if $X_{t,T}^2$ is a stationary process with autocovariance γ_j (Hoffman and Pagan, 1989):

$$T_1^{1/2}\hat{\tau} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, 2(\gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j)\right) \tag{3.1}$$

as T tends to infinity. The asymptotic variance of this distribution is estimated using the kernel-based estimate

$$\hat{\nu}_\ell = \hat{\gamma}_0 + 2 \sum_{j=1}^{\ell} \left(1 - \frac{j}{\ell + 1} \right) \hat{\gamma}_j$$

where $\hat{\gamma}_j$ is the jth serial covariance of X_t^2 and ℓ is a truncation number. A discussion on this estimator can be found in Newey and West (1987), where a consistency result is established when $\ell = \ell(T)$ tends to infinity with T and is such that $\ell(T) = O(T^{-1/4})$. Discussions about the choice of ℓ may be found in Phillips (1987) and White and Domowitz (1984).

A second test for covariance stationarity is the *CUSUM test*. Define

$$\psi(r) = \frac{1}{\sqrt{T\hat{\nu}}} \sum_{t=1}^{[Tr]} \left(X_{t,T}^2 - \hat{\sigma}_T^2 \right) \tag{3.2}$$

where $0 < r < 1$ and $\hat{\sigma}_T^2$ is the classical variance estimate over the whole segment of length T. This test compares the global variance estimate with the partial sum of the squared process (recall that we assume the process to be zero-mean). If the $X_{t,T}$ obey the moment and mixing conditions in Phillips (1987), then Lo (1991) proves that, under the null, $\psi(r)$ converges in distribution to a Brownian bridge. This means that the probability $\mathrm{P}(\psi(r) < c)$ is equal to the probability that an $\mathcal{N}(0, r(1-r))$ random variable is less than c.

3.2 Estimation of the unconditional variance

The unconditional variance function $\sigma^2(z)$, defined in the rescaled time $z \in (0, 1)$, is a function which is piecewise constant [PC] or continuous [Lip].

In the [Lip] case, the unconditional variance is estimated as follows. Consider the squared process $X_{t,T}^2$. This process is clearly an unbiased estimate of $\sigma(t/T)$ but it suffers from a large variability (see Proposition 1 in Appendix A). Then we smooth it using a kernel estimator.

In the [PC] case, we proceed as follows. Suppose the variance function has no jump in the interval $I = [z_0, z_1]$, $0 < z_0 < z_1 < 1$. A natural estimate of $\sigma(z)$ for $z \in I$ is the empirical variance

$$\hat{s}_T := \frac{1}{|z_1 - z_0|T} \sum_{t=[z_0 T]}^{[z_1 T]} X_{t;T}^2, \tag{3.3}$$

where $[\,\cdot\,]$ denotes the integer part of a real number. The resulting estimate is an unbiased and consistent estimator of $\sigma^2(z)$ on I (see Proposition 2 in Appendix A). However, we have to detect the jump points, and for this we propose to follow one of these two procedures:

1. Consider the random variable

$$s(t) = \left|\hat{\sigma}_{t+1}^2 - \hat{\sigma}_t^2\right| \qquad t = 0, \ldots, T-2, \tag{3.4}$$

   where $\hat{\sigma}_t^2$ denotes the sample stationary variance computed on $\{X_{0,T}, \ldots, X_{t,T}\}$. We decide that there is a jump in the variance at $t = t_0$ if $s(t_0)$ is larger than a certain threshold λ. In practice, we consider a data-dependant threshold: We choose λ such that 3 % of the realisations of $s(t)$ are larger than λ. Thus, on average, 3 % of the data are considered as jumps,

2. We can also use the CUSUM test to decide if there is a jump in the unconditional variance. Fixing the level of the CUSUM test at α, we select the first jump by searching the time point $t_0 = r_0 T$, $0 < r_0 < 1$, such that $\psi(r_0)$ in (3.2) is outside the confidence interval of the test. Then, we reproduce the procedure on segment $X_{[r_0 T],T}, \ldots, X_{T-1,T}$. In practice, we set $\alpha = 0.10$.

3.3 Forecasting procedure

The local stationarity framework imposes to write the nonstationary process with a double index, see equation (2.1). If we want to forecast h values of an observed process, we arrange the indices as follows: We denote by $X_{0,T}, \ldots, X_{T-h-1,T}$ the observed values with the convention that the ratio h/T tends to zero as T tends to infinity. This implies that the forecasting horizon h may depend on T, but does not increase faster than T.

The last observed realisation is denoted by $X_{T-h-1;T}$, which implies that, in the rescaled time, the local variance function $\sigma^2(z)$ can be estimated only on the interval

$$\left[0, 1 - \frac{h+1}{T}\right]$$

which asymptotically tends to $[0, 1)$. If we denote $\zeta_h = 1 - (h+1)/T$, then the value of $\sigma^2(z)$ outside $[0, \zeta_h]$ are the *predicted values of the unconditional variance in the rescaled time* (see Figure 1).

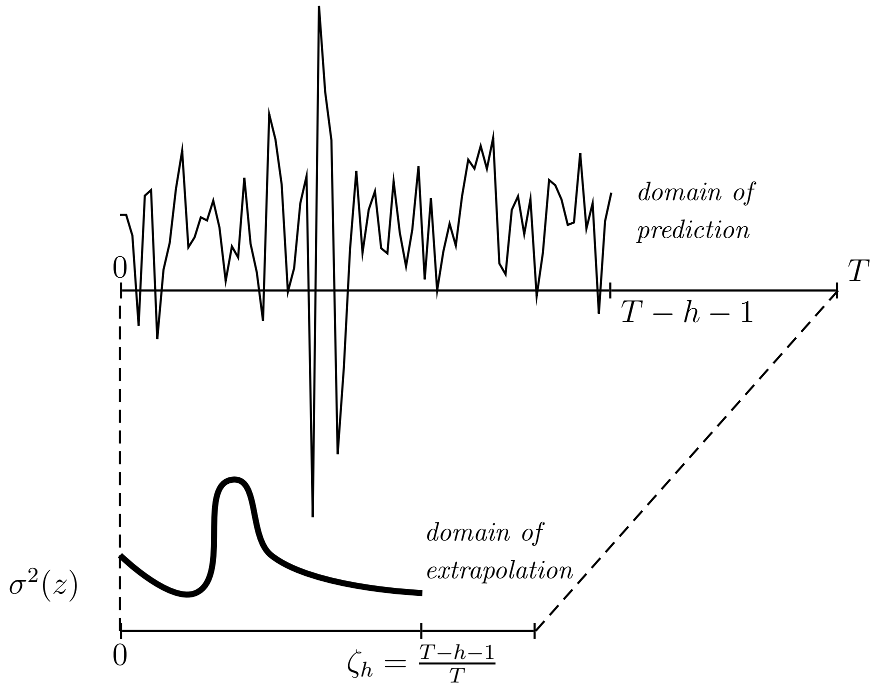


Figure 1: This picture illustrates the forecasting procedure explained in Section 3.3

We can now present our forecasting procedure for time-modulated processes.

1. Predict the local variance. Predicted values of the variance consists in an extrapolation of the estimated variance computed on $[0, \zeta_h]$. In the [PC] case, the most natural extrapolation is to prolong the last constant piece of the function by the same value. In the [Lip] case, we use a kernel extrapolator.

2. Define $\tilde{\sigma}^2(z)$ as the estimator of the local variance for $z \leqslant \zeta_h$ and its extrapolation for $\zeta_h < z < 1$.

3. Define the standardised data by
$$\tilde{X}_{t,T} = \frac{X_{t,T}}{\tilde{\sigma}\left(\frac{t}{T}\right)} \qquad t = 0, \ldots, T-h-1. \tag{3.5}$$

4. From equation (2.1), the standardised process $\tilde{X}_{t,T}$ is a stationary (zero mean) ARMA(p,q) processes. Then, use a standard linear forecaster for stationary process to forecast $\tilde{X}_{t,T}$.

5. From equation (2.3), we define the *h-step predictor* $\hat{X}_{T-1,T} = \hat{X}_{T-1,T}(h)$ in a natural way as
$$\hat{X}_{T-1,T} = \tilde{\sigma}\left(\frac{T-1}{T}\right) \mathrm{E}\left(Y_{T-1} | \mathcal{F}_{T-h-1}\right) \tag{3.6}$$
where $\mathrm{E}\left(Y_{T-1} | \mathcal{F}_{T-h-1}\right)$ denotes the classical h-step predictor of an ARMA process constructed from the standardised data (3.5).

Note that this predictor is zero for a tm-WN process.

The construction of the prediction intervals is similar. However, recall that the unconditional distribution of Y_t in equation (2.1) is not supposed to be Gaussian. It means that the (zero-mean) random variable $\tilde{X}_{t,T}$ is not Gaussian in general. Then, to construct the prediction interval at the level α, we propose to compute the interval
$$\left[\tilde{q}_{\frac{\alpha}{2}} \quad ; \quad \tilde{q}_{1-\frac{\alpha}{2}}\right]$$
where $\tilde{q}_\alpha$ is the α-quantile of the empirical distribution of the standardised process $\{\tilde{X}_{t,T}\}$, $t = 0, \ldots, n-1$. Finally the prediction interval for the forecasted value of $X_{T-1,T}$ is given by
$$\left[\hat{X}_{T-1,T} - \tilde{\sigma}\left(\frac{T-1}{T}\right)\tilde{q}_{\frac{\alpha}{2}} \quad ; \quad \hat{X}_{T-1,T} + \tilde{\sigma}\left(\frac{T-1}{T}\right)\tilde{q}_{1-\frac{\alpha}{2}}\right]. \tag{3.7}$$

Based on the empirical distribution of $\{\tilde{X}_{t,T}\}$, these prediction intervals are more suitable for the model (2.1), in which the distribution of the stationary process Y_t may be very general.

4 Experimental results

In this section, we focus on the practical evaluation of time-modulated processes. Our experiments are presented for five different economic data sets that we present in Section 4.1. In Section 4.2, we study the question of the covariance stationarity of these time series, and we apply the statistical tests recalled in section 3.1. Then, Section 4.3 studies the covariance stationarity of the standardised process (3.5), when we estimate the local variance $\tilde{\sigma}(z)$ with the estimators proposed in Section 3.2. Finally, Section 4.4 is devoted to evaluate the forecasting performance of the procedure described in Section 3.3.

4.1 The data sets

[**Nas-100**] The log-returns of the Nasdaq-100 Index. This index includes 100 of the largest non-financial domestic and international companies listed on the Nasdaq National Market tier of The Nasdaq Stock Market, Inc. from January 3, 2000 through May 31, 2002 (603 observations);

[**NasFin-100**] The log-returns of the Nasdaq Financial-100 Index. This index includes of 100 of the largest financial organizations listed on the Nasdaq National Market tier of The Nasdaq Stock Market, Inc. from January 3, 2000 through May 31, 2002 (603 observations);

[**Daily SR**] The daily stock returns to the Standard & Poor's composite portfolio from January 3, 1950 through December 29, 1962 (from Schwert (1990) – 3100 observations);

[**Ex DM-BP**] Daily percentage nominal returns for the Deutsche Mark/British Pound exchange rate from January 3, 1984 to December 31, 1991 (from Bollerslev and Ghysel (1996) – 1974 observations);

[**Ex US-BP**] Daily spot exchange rate of the US dollar to the British pound over a period of 1000 days ending on August 9, 1996 (1000 observations).

4.2 Testing the covariance stationarity

In a first step, we apply the two tests of covariance stationarity reviewed in Section 3.1 on the five data sets. All results are summarized in the column called "Original series" of Table 1. In this table, "PSP-test" stands for the post-sample prediction test (3.1) with $\ell = 7$ and "CS-test" refers to the CUSUM-test (3.2). For the first test, we indicate the p-value of the test. For the CUSUM-est, we present the results in a different way: We fix the level of the test at 0.01 and compare the values of $\psi(r)$ defined by equation (3.1) with the percentiles $c^{\pm}_{0.01}$ of a Brownian bridge. If the curve $\psi(r)$ is fully contained in the confidence intervals, the test does not reject the null assumption of covariance stationarity and, in this case, we indicate the symbol "–". In the contrary, if $\psi(r)$ intersects the curve $c^{+}_{0.01}$ or $c^{-}_{0.01}$, then the CUSUM test rejects the assumption of covariance stationarity. In this case, we indicate in the table the value of

$$\delta := \max_{0<r<1} \left(|\psi(r)| - c^{+}_{0.01}(r)\right) \tag{4.1}$$

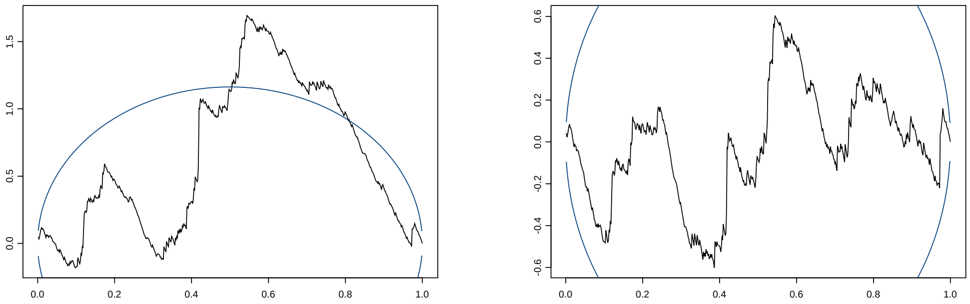
which measures the maximal deviation between $|\psi(r)|$ and $c^{+}_{0.01}(r)$ with respect to r.

To illustrate the CUSUM-test, we plot in Figure 2(a) the result of the CUSUM test appplied on the [Nas-100] series. This test concludes to a lack of homogeneity in the unconditional variance of [Nas-100], with $\delta = 0.535$, which motivates the introduction of time-modulated processes.

From Table 1, we can see that the post-sample prediction test (3.1) concludes to the lack of homogeneity in the variance of [Nas-100], [Ex DM-BP] and [Ex US-BP]. On the other hand, the CUSUM test (3.2) shows the lack of stationarity of all the series. Note that the PSP-test splits the time interval into two segments of equal length and tests the equality of variance on these two segments. This test crucially depends on the time point where we split the series into two parts. Contrary to this test, the CUSUM test controls the variance changes at each time point. Consequently, the reason why the series [Daily SR] and [NasFin-100] are not considered to be inhomogeneous in variance by the first test is simply due to the fact that the variance is not clearly broken at the middle of the series. Nevertheless, observe that there is only little doubt about the lack of homogeneity for [NasFin-100].

	Original series		Stand. (i) — [Lip]			Stand. (ii) — [PC]			Stand. (iii) — [PC]		
	PSP-test	CS-test	PSP-test	CS-test	κ_2	PSP-test	CS-test	κ_2	PSP-test	CS-test	κ_2
Nas-100	0.02	0.535	0.81	–	2.99	0.43	–	3.77	0.86	–	3.48
NasFin-100	0.20	0.053	0.74	–	3.51	0.58	–	9.64	0.71	–	6.83
Daily SR	0.37	0.291	0.78	–	3.62	0.17	0.23	9.13	0.69	0.70	5.30
Ex DM-BP	$< 10^{-2}$	1.338	0.71	–	4.26	0.22	0.15	7.80	0.69	–	8.68
Ex US-BP	$< 10^{-2}$	2.039	0.75	–	4.18	0.05	1.18	5.46	0.68	0.50	7.30

Table 1: "PSP-test" denotes the post-sample prediction test (3.1) for which we indicate the two-sided p-value. "CS-test" refers to the CUSUM test of covariance stationarity. This test compares the values of $\psi(r)$ defined in equation (3.2) and the percentiles $c^{\pm}_{0.01}(r)$ of a Brownian bridge. If $\psi(r)$ is fully contained in the confidence intervals, we indicate a "–", and in the contrary, we compute δ as in equation (4.1) — see text. The table shows results for three different standardisations (3.5) corresponding to the three different estimates (i), (ii) and (iii) of the local variance $\tilde{\sigma}^2(z)$ explained in text.



(a) The CUSUM test applied on the [Nas100] series.

(b) The CUSUM test applied on the standardised [Nas100] using the [CP] assumption.

Figure 2: In the CUSUM test, the function $\psi(r)$ is ploted for $0 < r < 1$ (see equation (3.2)). The smooth lines which are (partly) plotted are the confidence intervals of the CUSUM test at level 0.01.

4.3 Empirical properties of the standardised process

As a second step, we estimate the local variance following the three procedures mentioned in Section 3:

(i) in the [Lip] case, we smooth the squared data using a normal kernel and a bandwidth of $40/n$, where n is the total number of observations of the series;

(ii) in the [PC] case, we use the estimator (3.3) and a segmentation based on $s(t)$, see equation (3.4);

(iii) in the [PC] case the estimator (3.3) is used with the segmentation based on the CUSUM procedure.

Using our estimation, we standardise the data in the three cases, and run again the two tests of stationarity onto the standardised data. Figure 2(b) shows the result for [Nas-100] with standardisation (iii), and we observe that the standardised process is variance homogeneous. In other words, the time-varying variance $\sigma^2(z)$ modeled with [PC] fully explains the lack of stationarity of the original [Nas-100] series. The results for the other series are given in Table 1. We can observe that the first standardisation fully explains the lack of homogeneity of the series. The two other standardisations, corresponding to the [PC] case, provide less clear results. However, except for the [Daily SR] series, the third standardisation leads to better results than the second standardisation in the [PC] case.

Then, we can conclude that, in most of the cases, time-modulated processes satisfactorily explains the nonstationary behaviour observed in Section 4.2.

These first results are in favour of a smooth variance modelling. We could also model a time-modulated process (2.1) with a piecewise smooth variance. However, in this case, the estimation of the variance leads to the simultaneous problem of finding a good segmentation and choosing an appropriate smoother. Since the [Lip] and the [PC] conditions lead to satisfactory results, we will not consider this general case for the scope of this article.

In Table 1, we also mention the empirical kurtosis of the standardised process. This is to remark that the standardised process generally have not a Gaussian distribution, and this motivates the construction of prediction intervals based on the empirical quantiles, as explained in Section 3.3.

4.4 Forecasting

In the next step, we want to compare the forecasting accuracy of time-modulated models using the [PC] and [Lip] condition. In this section, we will only consider the estimators (i) and (iii). We now explain how we measure this accuracy for the one-step ahead prediction.

First, we clip the series at time t_0 and consider the prediction at time $t_0 + 1$. With notations of Figure 1, we set $t_0 = T - h - 1 = T - 2$ because the horizon prediction is $h = 1$. Then, we can proceed to the forecast as described in Section 3.3. In our experiments, we work as follows:

1. Fit an autoregressive model to the standardised process obtained in Section 4.3. This is done through the minimisation of the Akaike's (1974) AIC criterion;

2. Compute the forecasted value at $t_0 + 1$ for the stationary process;

3. To compute the forecasted value at time $t_0 + 1$, we need to extrapolate the local variance (see equation (3.6)). This extrapolation is constant in the case of the estimator (iii). For the estimator (i), we compute $\tilde{\sigma}\left((t_0+1)/T\right)$ as the average of the local variance estimated at $t_0/T, (t_0-1)/T, \ldots, (t_0-39)/T$.

This procedure is repeated for different values of t_0. Consequently, the autoregressive process fitted at each t_0 may also change with time, in accordance with Swanson and White (1997). In our experiments, we applied this algorithm to the 100 last values of the observed process.

We follow a similar procedure for h-step ahead prediction. In this case however, the estimator (i) is computed recursively: First, we extrapolate the variance at t_0+1, then we extrapolate the variance at t_0+2 averaging the estimated values of the local variance at $(t_0+1)/T, t_0/T, \ldots, (t_0-38)/T$, and so on until the extrapolation at $(T-1)/T$. From this moving average procedure, we can see that, for a long-range prediction, the extrapolated variance is a weighted sum of the estimated variance at $t_0/T, (t_0-1)/T, \ldots, (t_0-39)/T$.

The empirical results are listed in Table 2 for a one-step ahead prediction and a four-steps ahead prediction. For each series, the three following criteria are computed (Zeng and Swanson, 1998):

- The Root-Mean-Square Error (RMSE), given by

$$\sqrt{\frac{1}{100}\sum_{t=n-101}^{n-1}\left(X_{t,T}-\hat{X}_{t,T}\right)^2},$$

  is a widely used measure of forecast accuracy. However, the RMSE suffers from some drawbacks, as it may be inconsistent with profit measure and it is not invariant to nonsingular scale-preserving linear transformations. For a detailed discussion of these two problems, we refer to Clements and Hendry (1993), Leitch and Tanner (1991) and Swanson and White (1995).

- The Mean-Absolute Deviation (MAD), defined by

$$\frac{1}{100}\sum_{t=n-101}^{n-1}\left|X_{t,T}-\hat{X}_{t,T}\right|.$$

- The Mean-Absolute Percentage Error (MAPE)

$$\frac{1}{K_{100}} \sum_{t=n-101}^{n-1} \frac{\left|X_{t,T} - \hat{X}_{t,T}\right|}{|X_{t,T}|} \, \mathbb{I}_{X_{t,T} \neq 0}$$

  where $\mathbb{I}_{X_{t,T} \neq 0}$ is 1 if $X_{t,T} \neq 0$ and 0 if not, and K_{100} is the number of nonzero terms in the sum.

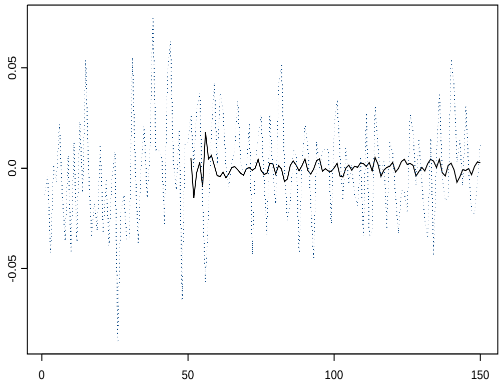
These three criteria concern the accuracy of the predicted mean. We are also interested in the accuracy of the predicted variance of the process. To measure this accuracy, we compute the two following measures:

- the percentage of observations (denoted by "Cov." for "Coverage") which fall within the corresponding prediction intervals. These intervals are constructed using formula (3.7).

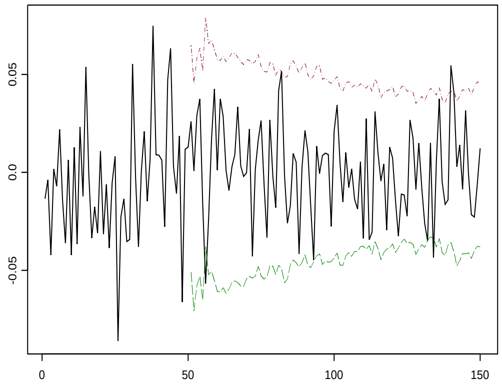
- the median of the length of these prediction intervals (MLPI).

Depending on the application we have in mind, one of these criteria will be the most important for the practitioner. The ideal predictor will have a minimal loss in the mean (RMSE, MAD or MAPE), with the best coverage and the shortest MLPI. However, it may happen that these three quantities are balancing each other. For instance, we can have a predictor with the best loss in the mean, but which has not the best coverage, etc. In this case, selecting the most appropriate predictor would be advised by the goals of the practitioner or the physical nature of the series. In other words, for an observed time series, if it is more important to get a precise forecast in the mean, then select the predictor with the best loss in the mean. On the contrary, if a precise value of the mean is less of interest than accurate predicted variability, then consider a predictor with the best MLPI, and having a sensible percentage of coverage and loss in the mean.

The results listed in Table 2 are an illustration of this balancing phenomenon. Indeed, this table indicates that the [PC] model leads to better results than the [Lip] condition as long as the accuracy in the predicted mean is considered. However, if we consider the variance prediction, the [Lip] model shows the best MLPI's. It means that at level 0.05 (i.e. if we accept 5 % of



(a) The dotted line is the actual series, and the solid line is the one-step ahead prediction.



(b) The prediction intervals of the same forecasting procedure than in Figure 3(a).

Figure 3: One-step ahead prediction at 100 time points of the series [Nas-100]. The prediction is done between January 3, 2000 and March 14, 2002 with the [Lip] model.

error in our prediction) the forecating procedure based on the [Lip] condition leads to smaller prediction intervals.

An other observation is the inherent difficulty to have a good coverage for exchange rate data. This coverage is lower than 95 % with the [Lip] model. It is better with the [PC] model, but is not satisfactory for the [Ex DM-BP] series, which leads to very large prediction intervals.

Finally, Figure 3 illustrates the forecasting procedure on a segment of the [Nas-100] series. This prediction is provided under the [Lip] assumption, and we can observe in Figure 3(b) the smooth variation of the local variance, which automatically follows the smooth evolution of the actual series.

	Series	For. hor.	RMSE	MAD	MAPE	Cov.	MLPI
[Lip]	Nas-100	$h = 1$	0.024	0.019	1.362	95 %	0.090
	NasFin-100		0.009	0.007	1.034	96 %	0.032
	Daily SR		0.004	0.003	1.022	96 %	0.020
	Ex DM-BP		0.267	0.188	0.997	93 %	0.928
	Ex US-BP		0.005	0.004	1.150	92 %	0.020
	Nas-100	$h = 4$	0.024	0.019	1.214	95 %	0.092
	NasFin-100		0.009	0.007	0.997	96 %	0.032
	Daily SR		0.005	0.003	1.004	95 %	0.020
	Ex DM-BP		0.266	0.188	1.000	91 %	0.913
	Ex US-BP		0.005	0.004	1.090	92 %	0.020
[PC]	Nas-100	$h = 1$	0.024	0.019	1.558	99 %	0.137
	NasFin-100		0.008	0.007	1.057	100 %	0.063
	Daily SR		0.004	0.003	1.107	100 %	0.026
	Ex DM-BP		0.266	0.188	1.000	99 %	2.013
	Ex US-BP		0.005	0.004	1.074	100 %	0.038
	Nas-100	$h = 4$	0.025	0.020	1.494	99 %	0.137
	NasFin-100		0.009	0.007	0.997	100 %	0.026
	Daily SR		0.004	0.003	1.160	100 %	0.026
	Ex DM-BP		0.266	0.188	1.000	99 %	2.014
	Ex US-BP		0.005	0.004	1.002	100 %	0.038

Table 2: This table summarizes the prediction accuracy of time-modulated processes. The five first rows concers time-modulated processes with [Lip] condition, i.e. using estimator (i). The five last rows concerns the [PC] variance estimated by the estimator (iii).

5 Concluding remarks

In this article, we have provided a methodology to forecast economic time series when the assumption of variance stationarity is rejected by statistical tests. This lack of homogeneity of unconditional variance has been observed on daily stock returns and exchange rate data and they confirm preliminary results published in the literature. Using the framework of rescaled time, we have shown that a simple process with a time-varying variance may be defined and explains satisfactorily the nonstationary behaviour of the observed series. This process consists in a stationary process modulated by a time-varying variance defined in the rescaled time. Two kinds of time-varying unconditional variance were considered: A smooth time-varying variance and a piecewise constant variance. The first specification gives better results in the variance forecasting, and the second specification gives better results in order to forecast the mean.

Our results confirm that the economic processes can evolve over time. Consequently, it is advisable to do a test of covariance stationarity before fitting an econometric model which assumes the constancy of the unconditional moments. This paper shows a simple and meaningful way to standardise nonstationary data in order to obtain a process which can be assumed to be stationary. It is then possible to apply the standard econometric modelling on the standardized data, and to forecast it. Forecasting the actual series then follows on the unstandardized data.

Our method of standardisation is based on the estimation of an unconditional variance. The proposed estimation procedures require the choice of parameters, like a smoothing parameter. In this article, these parameters are not automatically chosen, and it would be interessing in a future work to develop data-driven procedures to select them in an appropriate way. However, even if our parameters are chosen by hand, they provide satisfactory results, as the standardised process is stationary.

Other approaches of forecasting nonstationary signals are possible (see for instance Grillenzoni (1998) or Ramsey and Zhang (1997)). However, our approach based on the rescaled time leads to a consistent statistical theory modelling time-varying processes. Moreover, analysis with a time-modulated stationary process is related to classical statistical problems, such as nonparametric regression problem, nonparametric extrapolation and estimation of quantiles for a non-Gaussian distribution.

It is worth mentionning that our model with evolutionary variance is *not* a competitor to conditional heteroskedastic models, which are *stationary*. Indeed, one can imagine to fit for instance a (G)ARCH model to the standardised process. Our considerations on the nonstationarity are a preliminary step in the analysis of economic processes, and it may happen that evolutionary unconditional variance does not explain the behaviour of the volatility of a process.

We close by mentionning that time-modulated stationary processes are nonstationary processes with evolutionary unconditional variance only. These models can be extended to the case of a time-varying covariance (Dahlhaus (1997), Kim (1999), Fryźlewicz *et al.* (2002), Fryźlewicz (2002)).

Ackowledgements

The authors would like to thank Rohit Deo, Piotr Fryźlewicz, and Shane Whelan for very helpful comments and discussions. This research was supported by the National Fund for Scientific Research (FNRS), Walloon, Belgium, by 'Projet d'Action et de Recherche Concertes' (No. 98/03-217) from the Belgian government and by the IAP research network No. P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs). Parts of this article were presented at the 22nd International Symposium of Forecasting, Dublin, Ireland. Comments from participants at this meeting are gratefully acknowledged. The first author is grateful to the Minister-President of the Walloon-Brussels Community, H. Hasquin, and to the International Institute of Forecasters for financial supports.

A Some statistical properties of the variance estimator

In this Appendix, we focus on the statistical properties of the variance estimates. We will successively consider the case of the [Lip] assumption and the [PC] assumption.

Proposition 1. *Assume that the function σ in equation (2.1) obeys the [Lip] condition. Then,*

$$\begin{aligned}\mathrm{E}X^2_{[zT],T} &= \sigma^2\left(z\right)\\ \mathrm{Var}\, X^2_{[zT],T} &= \sigma^4\left(z\right)\left(\kappa_2 - 1\right)\end{aligned}$$

for all $0 < z < 1$, where κ_2 is the kurtosis (2.2).

The proof of this proposition is straightforward from equation (2.1), recalling that the variance of the stationary process is equal to one.

Proposition 2. *Assume that the function σ in equation (2.1) obeys the [PC] condition and has no jump between z_0 and z_1, $0 < z_0 < z_1 < 1$. Then the estimator (3.3) is such that*

$$\mathrm{E}\hat{s}_T = \int_{z_0}^{z_1} dz\; \sigma^2(z)$$

and

$$\begin{aligned}\mathrm{Var}\,\hat{s}_T = &\frac{1}{|z_1 - z_0|^2\, T}\int_{z_0}^{z_1} dz\; \sigma^4(z)\\ &+ \frac{1}{|z_1 - z_0|^2\, T}\sum_{u=1}^{[(z_1-z_0)T]-1} \gamma(u)\int_{z_0}^{z_1-\frac{u}{T}} dz\; \left[\sigma^4\left(z\right) + \sigma^4\left(z + \frac{u}{T}\right)\right]\end{aligned}$$

where $\gamma(s-t) = \mathrm{Cov}\left(Y_s^2, Y_t^2\right)$.

Proof. Taking the expectation of (3.3) we get, with definition (2.1):

$$\mathrm{E}\hat{s}_T = \frac{1}{|z_1 - z_0|\; T}\sum_{t=[z_0T]}^{[z_1T]-1} \sigma^2\left(\frac{t}{T}\right) = \frac{1}{|z_1 - z_0|}\sum_{t=[z_0T]}^{[z_1T]-1}\int_0^{\frac{1}{T}} dz\; \sigma^2\left(\frac{t}{T}\right)$$

Because $\sigma^2(\cdot)$ is constant over $[z_0, z_1]$, it follows

$$\mathrm{E}\hat{s}_T = \frac{1}{|z_1 - z_0|}\sum_{t=[z_0T]}^{[z_1T]-1}\int_0^{\frac{1}{T}} dz\; \sigma^2\left(z + \frac{t}{T}\right) = \frac{1}{|z_1 - z_0|}\sum_{t=[z_0T]}^{[z_1T]-1}\int_{\frac{t}{T}}^{\frac{t+1}{T}} dz\; \sigma^2\left(z\right)$$

which leads to the result for the expectation. The computation of the variance is similar: If $\gamma(u)$ denotes the autocovariance function of the process $\{Y_t^2\}$, then we can write

$$\begin{aligned}\operatorname{Var} \hat{s}_T &= \frac{1}{|z_1 - z_0|^2\, T^2} \sum_{s,t=[z_0 T]}^{[z_1 T]-1} \gamma(s-t)\, \sigma^2\left(\frac{s}{T}\right) \sigma^2\left(\frac{t}{T}\right) \\ &= \frac{1}{|z_1 - z_0|^2\, T} \sum_{s,t=[z_0 T]}^{[z_1 T]-1} \gamma(s-t) \int_{\frac{t}{T}}^{\frac{t+1}{T}} dz\, \sigma^2\left(\frac{s-t}{T} + z\right) \sigma^2\,(z)\end{aligned}$$

Because $\sigma^2(\cdot)$ is constant over $[z_0, z_1]$, it follows

$$\operatorname{Var} \hat{s}_T = \frac{1}{|z_1 - z_0|^2\, T} \sum_{s,t=[z_0 T]}^{[z_1 T]-1} \gamma(s-t) \int_{\frac{t}{T}}^{\frac{t+1}{T}} dz\, \sigma^4\,(z)$$

Defining $u := s - t$, direct computations lead to

$$\begin{aligned}\operatorname{Var} \hat{s}_T = \frac{1}{|z_1 - z_0|^2\, T} \Bigg\{ &\sum_{u=0}^{[(z_1 - z_0)T]-1} \gamma(u) \sum_{t=[z_0 T]}^{[z_1 T]-1-u} \int_{\frac{t}{T}}^{\frac{t+1}{T}} dz\, \sigma^4(z) \\ &+ \sum_{u=[(z_0 - z_1)T]+1}^{-1} \gamma(u) \sum_{t=[z_0 T]-u}^{[z_1 T]-1} \int_{\frac{t}{T}}^{\frac{t+1}{T}} dz\, \sigma^4(z) \Bigg\}\end{aligned}$$

and the result follows using the symmetry of $\gamma(u)$. □

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control*, *19*, 716–723.

Bollerslev, T. and Ghysel, E. (1996). Periodic autoregressive conditional heteroscedasticity. *J. Bus. Econ. Statist.*, *14*, 139–151.

Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean square forecast errors. *J. Forecasting*, *12*, 617–676.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting economic time series.* Cambridge University Press.

Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In P. Robinson and M. Rosenblatt (Eds.), *Athens conference on applied probability and time series analysis* (Vol. 2). Springer, New York.

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

Fryźlewicz, P. (2002). *Modelling and forecasting financial log-returns as locally stationary wavelet processes* (Research Report). Department of Mathematics, University of Bristol. (`http://www.stats.bris.ac.uk/pub/ResRept/2002.html`)

Fryźlewicz, P., Van Bellegem, S. and von Sachs, R. (2002). *Forecasting nonstationary time series by wavelet process modelling* (Discussion Paper No. 0208). Institut de statistique, UCL. (`ftp://www.stat.ucl.ac.be/pub/papers/dp/dp02/dp0208.ps`)

Grillenzoni, C. (1998). Forecasting unstable and nonstationary time series. *Int. J. Forecasting*, *14*, 469–482.

Hoffman, D. and Pagan, A. R. (1989). Post-sample prediction tests for generalized method of moment estimators. *Oxford Bull. Econ. Statist.*, *51*, 333–343.

Kim, W. (1999). *Econometric analyses of evolutionary time series.* (Manuscript)

Leitch, G. and Tanner, J. (1991). Economic forecast evaluation: Profits versus the conventional error measures. *Amer. Econ. Rev.*, *81*, 580–590.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, *59*, 1279–1313.

Loretan, M. and Phillips, P. C. (1994). Testing the covariance stationarity of heavy-tailed time series. *J. Empirical Finance*, *1*, 211–248.

Los, C. A. (2000). Nonparametric efficiency testing of Asian stock markets using weekly data. *Adv. in Econometrics*, *14*, 329–363.

Mikosch, T. and Stărică, C. (1999). *Change of structure in financial data, long-range dependence and GARCH modelling* (Technical Report). University of Groningen. (`http://www.math.ku.dk/~mikosch/preprint.html`)

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Pagan, A. R. and Schwert, G. W. (1990a). Alternative models for conditional stock volatility. *J. Econometrics*, *45*, 267–290.

Pagan, A. R. and Schwert, G. W. (1990b). Testing for covariance stationarity in stock market data. *Econ. Letters*, *33*, 165–170.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica*, *55*, 277–301.

Ramsey, J. B. and Zhang, Z. (1997). The analysis of foreign exchange data using waveform dictionaries. *J. Empirical Finance*, *4*, 341–372.

Schwert, G. W. (1990). Indexes of United States stock prices from 1802 to 1987. *J. Bus.*, *63*, 399–426.

Swanson, N. R. and White, H. (1995). A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *J. Bus. Econ. Statist.*, *13*, 265–275.

Swanson, N. R. and White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *Int. J. Forecasting*, *13*, 439–461.

White, H. and Domowitz, I. (1984). Nonlinear regression with dependant observations. *Econometrica*, *52*, 143–161.

Zeng, T. and Swanson, N. R. (1998). Predictive evaluation of econometric forecasting models in commodity futures markets. *Stud. Nonlinear Dynam. Econometrics*, *2*, 159–177.