CAEPR Working Paper
#003-2009

PERSISTENCE IN NONLINEAR TIME SERIES:
A NONPARAMETRIC APPROACH

Juan Carlos Escanciano, Javier Hualde
Indiana University and Universidad Pública de Navarra

February 6, 2009

This paper can be downloaded without charge from the Social Science Research Network electronic library at: http://ssrn.com/abstract=1346052.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: http://www.indiana.edu/~caepr. CAEPR can be reached via email at caepr@indiana.edu or via phone at 812-855-4050.

©2008 by NAME. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
Abstract

The purpose of the present paper is to relate two important concepts of time series analysis, namely, nonlinearity and persistence. Traditional measures of persistence are based on correlations or periodograms, which may be inappropriate under nonlinearity and/or non-Gaussianity. This article proves that nonlinear persistence can be characterized by cumulative measures of dependence. The new cumulative measures are nonparametric, simple to estimate and do not require the use of any smoothing user-chosen parameters. In addition, we propose nonparametric estimates of our measures and establish their limiting properties. Finally, we employ our measures to analyze the nonlinear persistence properties of some international stock market indices, where we find an ubiquitous nonlinear persistence in conditional variance that is not accounted for by popular parametric models or by classical linear measures of persistence. This finding has important economic implications in, e.g., asset pricing and hedging. Conditional variance persistence in bull and bear markets is also analyzed and compared.

Keywords and Phrases: Conditional Mean; Nonlinear time series; Nonlinear Persistence; Nonlinear correlograms; Persistence in variance; Bull and bear markets.

JEL Classification: C12, C14.

We thank seminar participants at Universidad Carlos III de Madrid, CESG 2008 and University of Montreal for useful comments. We thank Lin Zhu for excellent research assistance. We also thank Marine Carrasco, Silvia Gonçalves and Bruce Hansen for helpful conversations. Research funded by the Spanish Ministry of Education and Science reference number SEJ2005-07657/ECON.

†Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

‡Research funded by the Spanish Ministry of Education and Science through a “Ramón y Cajal” contract.
1. Introduction

The concept of persistence plays a major role in macroeconomics, finance and econometrics. However, while in the linear set-up there exists a complete machinery of tools (mainly based on autocorrelations and periodograms) for studying the dynamic properties of economic series, surprisingly, there are still today relatively few useful analytical tools capable of assessing appropriately the dependence and persistence behavior of nonlinear time series (cf. Granger and Terasvirta, 1993). This problem is even more accentuated by the fact that traditional measures of dependence may be inappropriate under nonlinearity and/or non-Gaussianity of the underlying time series process.

The persistence properties of nonlinear time series have received increasing attention in recent years. Gallant, Rossi and Tauchen (1993) and Koop, Pesaran and Potter (1996) have extended the impulse-response methodology, initiated by Sims (1980), to a nonlinear framework, see also Gourieroux and Jasiak (1999, 2000), Clarida and Taylor (2003) and Kapetanios (2007) for related approaches. These extensions, however, require either a fully parametric model or nonparametric smoothed estimations of the truncated conditional mean for the time series of interest. The resulting measures of persistence are not operative when the conditioning set considered is high dimensional or involves the infinite past of the underlying time series, and, in addition, they might depend on user-chosen parameters, such as bandwidths, which affect inference in finite samples. In general, as noted in Gourieroux and Jasiak (1999), the implementation of these approaches requires a fairly large number of observations to give reasonable results, which could rule out the application of these methods to e.g. macroeconomic series, for which such large samples are sometimes unavailable.

Alternatively, one of the most ambitious programs to relate nonlinearity and persistence is that of Granger (1995). He proposed the definitions of Short Memory in Mean (SMM) and Extended Memory in Mean (EMM) as new concepts of nonlinear non-persistence and persistence in mean, respectively, see also Granger and Hallman (1991). These ideas have the potential to be fundamental in extending the concepts of persistence and cointegration to nonlinear models, although, unfortunately, they are not operative in a nonparametric framework. Even Granger (1995) noted that, though conceptually simple, these definitions are difficult to check in practice, because, in general, the dependence measures on which they are based on cannot be estimated, see also Domowitz and El-Gamal (2001) for related discussions.

In many situations the object of interest is the whole joint distribution and not just the conditional mean. Granger (2003) defines a persistent process in distribution using

Within this setting, the main purpose of this paper is to define new operative measures of nonlinear dependence and persistence. Essentially, we transform the ill-posed problem of evaluating the nonlinear dependence by local measures, like the conditional mean or some distance between the bivariate and the product of marginal densities, to a well-posed problem defined in terms of cumulative measures of dependence, which, incidentally, take the form of unconditional moments. Thus, even if our measures are of nonparametric nature, they are very easy to estimate, and avoid the choice of user-chosen parameters which arises in local estimation. Our concept of dependence is relatively inspired by that of Granger (1995), which, as shown below, is closely connected to a generalized version of the predictive dependence measures of Wu (2005). But, unlike these authors, we consider a non-adapted multivariate framework, which is particularly useful to examine various aspects of the dependence among different processes, as seems to be relevant in economic and financial settings. In this sense, our approach shares some features with the generalized impulse response functions of Gallant et al. (1993) or that of Koop et al. (1996), although our focus is not detecting sensitivity to shocks, but measuring the strength of the dependence (i.e. persistence). As a by-product of our analysis, we establish an important link between the persistence measures proposed by Granger (1995), the nonlinear impulse response literature, as discussed in Koop et al. (1996), and the dependence measures used by Wu (2005) (see the variance decomposition in (3) below).

Additionally, we provide nonparametric estimates of our measures and analyze their asymptotic properties. The paper discusses both conditional mean and distributional persistence under weak dependence conditions (mixingale and strong mixing, respectively). Our results under mixingales are new in the literature, and are of independent interest. We establish the asymptotic distribution of our new measures and confidence intervals computed with the assistance of bootstrap methods. We consider an application of our methodology to study the nonlinear persistence in the main daily stock market indices, where we find an ubiquitous nonlinear persistence in conditional variance that cannot be explained by popular parametric models or by classical linear measures of persistence. A novel nonparametric way of measuring leverage effects and its persistence is also provided and used to illustrate the difference in persistence between bull and bear stock markets.
The layout of the article is as follows. In Section 2 we review some particularly interesting measures of nonlinear dependence given in the literature, define similar pairwise definitions to those considered in Granger (1995, 2002) and establish connections among them. In Section 3 we introduce our new integrated measures of dependence. Section 4 provides formal asymptotic theory for the estimators of the new measures, Section 5 discusses the empirical application, and, finally, in Section 6 we conclude and describe further research. All proofs are gathered in an Appendix.

To conclude this section, a word on notation. Throughout, \( A' \) and \( |A| \) denote the matrix transpose and the Euclidean norm of \( A \), respectively. \( \|A\|_p \) denotes the usual \( L_p \) norm for a random variable (r.v.), i.e. \( \|A\|_p = (E|A|^p)^{1/p} \), and with an abuse of notation we denote by \( L_p \) the set of r.v.’s with \( \|A\|_p < \infty \). \( 1(A) \) denotes the indicator function of the event \( A \). For two sequences of real numbers \( (a_n)_{n \in \mathbb{N}} \) and \( (b_n)_{n \in \mathbb{N}} \) write \( a_n \sim b_n \) if \( \lim_{n \to \infty} a_n/b_n = 1 \).

2. Nonlinear measures of dependence

Let \((Y_t)_{t \in \mathbb{Z}}\) and \((X_t)_{t \in \mathbb{Z}}\) with \( Y_t \in \mathbb{R} \), \( X_t \in \mathbb{R}^p \), be processes defined on the probability space \((\Omega, \mathcal{F}, P)\) and consider the information set at time \( t \) given by \( I_t = (X'_t, X'_{t-1}, \ldots)' \in \mathbb{R}^{p \cdot \infty} := \mathbb{R}^p \times \mathbb{R}^p \times \cdots \). Here \( X_t \) may or may not contain lagged values of \( Y_t \). Extensions of our theory to multivariate \( Y_t \) are trivial, and hence they are not considered here. It is well known that under integrability of \( Y_t \) we can define the regression function \( M_{t,h}(x) = E(Y_{t+h} | I_t = x), x \in \mathbb{R}^{p \cdot \infty}, h \geq 0 \), almost surely (a.s.). The regression function \( M_{t,h} \) is \( \mathcal{F}_t \)-measurable, where \( \mathcal{F}_t \) is the \( \sigma \)-field generated by \( I_t \), i.e., \( \mathcal{F}_t = \sigma(I_t) \), and represents the “best” prediction, in a mean square sense, of \( Y_{t+h} \) given \( I_t \). We first concentrate on the concept of persistence in conditional mean. As in Granger (1995), we define the concepts of SMM and EMM as follows.

Let \( E(Y_t) = \mu_t \) for all \( t \). Throughout this paper, \( c_{h,t} \) denotes a sequence of positive numbers that tends to zero as \( h \) increases to infinity, for all \( t \in \mathbb{Z} \).

**Definition 1.** \((Y_t)_{t \in \mathbb{Z}}\) is called SMM if for all \( t \in \mathbb{Z} \)

\[
E \left( M_{t,h}(I_t) - \mu_{t+h} \right)^2 < c_{h,t}.
\]

If \((Y_t)_{t \in \mathbb{Z}}\) does not satisfy the previous condition is called EMM.

To be mathematically precise we should specify in Definition 1 that \( Y_t \) is SMM or EMM relative to \( \mathcal{F}_t \). We omit this here (and also in subsequent definitions) for the sake of brevity, but note that the definition does not only refer to \( Y_t \), but also to the \( \sigma \)-field on which we are conditioning on. Relative to this definition, Granger (1995)
just considered the adapted univariate case with $X_t = Y_t$. Granger and Hallman (1991) replaced the name of EMM by long memory in mean, which could be slightly misleading because, in general, a stationary linear long memory process is SMM. Gourieroux and Jasiak (1999) referred to SMM and EMM as nonlinear integrated and nonlinear integrated of order zero, respectively. As noted by Granger (1995) the concepts of SMM and EMM are related to a kind of “mixing in mean” property, or, more precisely, to the concept of mixingale (cf. McLeish, 1974); see Davidson (1994, Chapter 16) and Section 4 here. Similarly, Bollerslev and Engle (1993) defined the concept of nonintegration in mean by saying that $Y_t$ is not integrated in mean if

$$\lim_{h \to \infty} M_{t,h}(I_t) = \text{constant},$$

for all $t$, with probability one, the process being integrated in mean if this condition fails (they also introduced the concept of integration in variance, which depends on the limiting behavior of $E(Var(Y_{t+h} | I_{t+h-1}) | I_t)$). The use of almost sure convergence can be useful in cases where second moments may not exist.

We stress that no condition for adaptation is necessary in our definition of persistence. This is so because even without adaptation, i.e., even if $Y_t$ is not $\mathcal{F}_t$-measurable, the classical physical interpretation of persistence defined as the extent to which events today have an effect on the future history of a stochastic process is still valid, see equation (4) below. Some (asymptotic) measurability condition is, however, usually necessary to obtain central limit theorems for non-persistent series.

Definition 1 offers other interesting insights. Defining the projector operator $P_lW = E(W | I_t) - E(W | I_{t-1})$, $W \in L_2$, for any $k \geq 1$, we can write

$$E(Y_{t+h} | I_t) - \mu_{t+h} = \sum_{j=0}^{k-1} P_{t-j}Y_{t+h} + E(Y_{t+h} | I_{t-k}) - \mu_{t+h}. \quad (1)$$

First, note that the r.v.’s $P_{t-j}Y_{t+h}$ relate very closely to the generalized impulse response functions of Koop et al. (1996), which represent the difference between the average response of $Y_{t+h}$ given present and past with respect to the average response given only history. As Koop et al. (1996) suggest, measures of persistence can be defined by focusing on the dispersion of these functions, so a simple way of capturing this idea is to analyze (for fixed $j$) the evolution of the variance of $P_{t-j}Y_{t+h}$ as $h$ increases. Noting that the variables in the right hand side of (1) are uncorrelated, then, for any $k \geq 1$,

$$E\left(E(Y_{t+h} | I_t) - \mu_{t+h}\right)^2 = \sum_{j=0}^{k-1} E\left(P_{t-j}Y_{t+h}\right)^2 + E\left(E(Y_{t+h} | I_{t-k}) - \mu_{t+h}\right)^2. \quad (2)$$
implying that Granger’s (1995) measure of persistence can be represented as the cumulative contribution of the variances of generalized impulse response functions at different horizons (plus a remaining term). In the case where \( Y_t \) is SMM, since the left hand side in (2) does not depend on \( k \), we can take the limit as \( k \to \infty \) and conclude
\[
E \left( E (Y_{t+h} | I_t) - \mu_{t+h} \right)^2 = \sum_{j=0}^{\infty} E (P_{t-j} Y_{t+h})^2 .
\] (3)

Similarly, it can be proved that any SMM process \( \{Y_t\} \) has the representation
\[
Y_{t+h} = \mu_{t+h} + \varepsilon_{t+h} + \sum_{j=1}^{h-1} P_{t-j+h} Y_{t+h} + \sum_{j=h}^{\infty} P_{t-j+h} Y_{t+h},
\] (4)

where \( \varepsilon_{t+h} = Y_{t+h} - E (Y_{t+h} | I_{t+h-1}) \). The representation (4) is due to Gordin (1969) in the adapted stationary univariate case. See Hansen (1982) for a multivariate extension and Hayashi (2000, p. 403) for a textbook treatment, still in the stationary and ergodic adapted framework. We stress that (4) is valid for non-stationary and non-adapted sequences. In particular, we note that the concept of persistence only involves the \( F_{t+h} \)-measurable “innovations” \( \{P_{t-j+h} Y_{t+h}\}_{j=h}^{\infty} \) and not the possibly non \( F_{t+h} \)-measurable innovation \( \varepsilon_{t+h} \).

Note that under stationarity
\[
E \left( E (Y_{t+h} | I_t) - \mu_{t+h} \right)^2 = \sum_{k=h}^{\infty} \delta(k),
\]
where \( \delta(k) \equiv E (P_{t-k} Y_t)^2 \). Hence, an alternative definition of SMM in the stationary case could be a process satisfying
\[
\sum_{k=0}^{\infty} \delta(k) < \infty.
\] (5)

Two remarks are worth to be mentioned. First, it should be noted that \( \delta(k) \) is a generalized version of the predictive dependence measures based on the idea of coupling proposed by Wu (2005) (see his Theorem 1)\(^1\). Second, condition (5) alone does not guarantee a central limit theorem; see Dedecker, Merlevede and Volny (2007)\(^2\).

\(^1\)Wu (2005) defines his measures with \( X_t \) being independent and identically distributed (iid) unobservable r.v.’s. His results depend crucially on the iid assumption. In particular, if \( Y_t \) is a linear process, and \( X_t \) is the iid innovation in the linear process, this SMM definition coincides with the classical short memory condition of linear processes; see Wu (2005).

\(^2\)A sufficient condition for a CLT of stationary SMM processes is that an asymptotic measurability
Although theoretically appealing, the measures $\delta(k)$, and hence Granger’s (1995) measure of persistence, are not operative in nonparametric settings without further restrictions on the data generating process (DGP). To solve this problem, we define the pairwise equivalent of the concept introduced by Granger (1995), which, although weaker, it is more operative because it only involves finite-dimensional r.v.’s.

**Definition 2.** $(Y_t)_{t \in \mathbb{Z}}$ is called pairwise SMM (PSMM) if for all $t \in \mathbb{Z}$,

$$E(m_{t,h} - \mu_{t+h})^2 < c_{h,t},$$

where $m_{t,h} = E[Y_{t+h} | X_t]$. If $(Y_t)_{t \in \mathbb{Z}}$ does not satisfy previous condition is called pairwise EMM (PEMM).

From the previous definitions, the law of iterated expectations and Jensen’s inequality, we easily observe that a SMM process is also PSMM. The reciprocal is false. There exist processes which are PSMM but not SMM, although they are rare in practice. An example is given by the following process, adapted from Granger (1995).

**Example 1.** Let $Y_t$ be the process generated by

$$Y_t = \sum_{k=0}^{\infty} \prod_{j=0}^{k} \varepsilon_{t-k-j},$$

(6)

where $\varepsilon_t$ is a sequence of iid r.v.’s with zero mean and unit variance, in short $\varepsilon_t \sim iid(0,1)$. With $X_t = \varepsilon_t$, this is an EMM which is PSMM because $m_{t,h} = 0$ a.s., for all $t, h \geq 1$.

Regarding distributional dependence, we formalize a definition given in Granger (2003). Let $f_{t,h}(y,x)$, $k_{t+h}(y)$, and $g_t(x)$ be, respectively, the joint and marginal densities of $Y_{t+h}$ and $X_t$. Define

$$s_{t,h}(y, x) = f_{t,h}^{1/2}(y, x) - k_{t+h}^{1/2}(y)g_t^{1/2}(x) \quad h \geq 1.$$  

(7)

**Definition 3.** $(Y_t)_{t \in \mathbb{Z}}$ is called pairwise short memory in distribution (PSMD) if for all $t \in \mathbb{Z}$,

$$H_{t,h} := \int_{\mathbb{R} \times \mathbb{R}^d} |s_{t,h}(y, x)|^2 dy dx < c_{h,t}.$$  

If $(Y_t)_{t \in \mathbb{Z}}$ does not satisfy previous condition is called pairwise extended memory in assumption holds and that $\sum_{k=-\infty}^{\infty} \delta^{1/2}(k) < \infty$. 

Alternative definitions can be given in terms of other divergence measures or distances, see Granger, Maasoumi and Racine (2004) and references therein. Notice that by simple algebra, see e.g. Strasser (1985), uniformly in

\[ H_{t,h} \leq \int_{\mathbb{R} \times \mathbb{R}} |f_{t,h}(y,x) - k_{t+h}(y)g_t(x)| \, dy \, dx \leq 2\alpha_h, \quad (8) \]

where \( \alpha_h \) is the \( \alpha \)-mixing coefficient defined as

\[ \alpha_h = \sup_{t \in \mathbb{Z}} \sup_{B \in \mathcal{F}_t, A \in \mathcal{P}_{t+h}} |P(A \cap B) - P(A)P(B)|, \quad h \geq 1, \]

where the \( \sigma \)-fields \( \mathcal{F}_t \) and \( \mathcal{P}_t \) are \( \mathcal{F}_t = \sigma(X_s, s \leq t) \) and \( \mathcal{P}_t = \sigma(Y_s, s \geq t) \), respectively. See Doukhan (1994) for a comprehensive study of the concept of \( \alpha \)-mixing and its relation with other weak dependence concepts. Thus the concept of PSMD can be understood as a kind of (pairwise) mixing condition. In fact, if in the definition of \( \alpha_h \) we use \( \mathcal{F}_t = \sigma(X_t) \) and \( \mathcal{P}_{t+h} = \sigma(Y_{t+h}) \), then by Lemma 2.15 in Strasser (1985),

\[ \alpha_h \leq \sup_{t \in \mathbb{Z}} \left[ \frac{1}{2} H_{t,h}(4 - H_{t,h}) \right]^{1/2}. \]

3. **Integrated measures of Nonlinear dependence**

In this section we propose generalizations of the usual autocovariances and cross-covariances to a nonlinear framework. It is well-known that in the presence of nonlinearity (or non-Gaussianity) these measures do not characterize the dependence, and the practitioner needs more reliable measures such as the pairwise regression functions \( E[Y_t \mid X_{t-j}] \). Robinson (1983) has studied the large sample properties of kernel estimators of lagged conditional means \( E[Y_t \mid Y_{t-j}] \) for various lags \( j \), see also Auestad and Tjøstheim (1990). Inference on these functions, however, requires smoothing estimation involving bandwidth choices, hampering their application to practical situations.

Here we consider integrated measures of the regression functions and densities. Assume that the r.v. \( Y \) is integrable, so the regression function

\[ m(x) = E[Y - \mu \mid X = x] \]
is well defined (up to a null set), with $X$ the explanatory variable and $E[Y] = \mu$. Define the integrated regression function $\gamma(\cdot)$ as

$$
\gamma(x) = E[(Y - \mu)1(X \leq x)] = \int_{-\infty}^{x} E[Y - \mu | X = z]F(dz),
$$

where the second equality follows by the law of iterated expectations, $F$ being the cumulative distribution function (cdf) of $X$. In a time series context it is particularly interesting the case where $Y = Y_{t+h} \text{ and } X = X_t$, $h \geq 0$. In this case, the measures

$$
\gamma_{t,h,\text{ind}}(x) := Cov(Y_{t+h}, 1(X_t \leq x)) = E[(Y_{t+h} - \mu_{t+h})1(X_t \leq x)],
$$

where $E(Y_t) = \mu_t$ for all $t$, are called the Integrated Pairwise Regression Functions (IPRF) in general, and the Integrated Pairwise Autoregression Functions (IPAF) when $X_t = Y_t$, see Escanciano and Velasco (2006a). As previously noted in the literature (cf. Stute, 1997)

$$
m_{t,h} - \mu_{t+h} = 0 \text{ a.s. } \iff \gamma_{t,h,\text{ind}} = 0, \text{ almost everywhere (a.e.) in } \mathbb{R}^p.
$$

More generally, instead of the parametric indicator function family $\{1(X_t \leq x) : x \in \mathbb{R}^p\}$ we could consider a general parametric family $\{w(X_t, x) : x \in \Pi\}$ satisfying

$$
m_{t,h} - \mu_{t+h} = 0 \text{ a.s. } \iff \gamma_{t,h,w}(\cdot) = E[(Y_{t+h} - \mu_{t+h})w(X_t, \cdot)] = 0 \text{ a.e. in } \Pi \subset \mathbb{R}^p. \quad (9)
$$

Additional examples of weighting functions satisfying (9) are $\exp(i x_0' X)$, with $i = \sqrt{-1}$, $\sin(x' X)$ or $1/(1+\exp(c-x' X))$. See Bierens and Ploberger (1997), Stinchcombe and White (1998) and Escanciano (2006) for sufficient conditions on $w$ and $\Pi$ to satisfy (9).

The integrated measures of dependence $\gamma_{t,h,w}$ have been shown to be useful for testing interesting hypotheses in a nonlinear time series framework. It is well known that these measures are able to capture both linear and nonlinear dependence in the conditional mean. What is not known, however, is that these measures are also capable of characterizing nonlinear persistence. Loosely speaking, we prove that equivalence (9) also holds in the limit as $h \to \infty$; see Theorem 1 below. Hence, this paper shows that cumulative measures such as $\gamma_{t,h,w}$ are appropriate nonparametric tools for studying nonlinear persistence.

Now, we define the analogous concepts to those of Section 2 but relative to the measures $\gamma_{t,h,w}$, which we assume belong to an appropriate metric space, say $S = $
with \( d \) a metric on \( M \). For instance, under mild conditions on \( w \), \( \gamma_{t,h,w} \) will belong to \( \ell^\infty(\Pi) \), the space of all complex-valued functions that are uniformly bounded on \( \Pi \), or to \( L_2(\Pi,W) \), the space of \( W \)-square integrable functions on \( \Pi \), for an absolutely continuous measure \( W \). The choice of the metric space \( M \) depends on the choice of the metric \( d \) used. The choice of \( d \) in turn is up-to the practitioner. If the metric is the uniform \( (\sup-) \) metric \( d_{\infty}(f,g) = \sup_{x \in \Pi} |f(x) - g(x)| \), then a space like \( \ell^1() \) can be considered, whereas if the interest is in the \( L_2 \) metric \( d_2^2(f,g) = \int |f(x) - g(x)|^2 W(dx) \), \( L_2(\Pi,W) \) will be convenient to work with. In any case, we denote by \( ||f||_{\mathcal{d}} \) the norm associated to the metric \( d \).

**Definition 4.** \((Y_t)_{t \in \mathbb{Z}}\) is called \((w,d)\)-pairwise SMM ((\(w,d\))-PSMM) if for all \( t \in \mathbb{Z} \),

\[
\|\gamma_{t,h,w}\|_{\mathcal{d}} < c_{h,t}.
\]

If \((Y_t)_{t \in \mathbb{Z}}\) does not satisfy previous condition is called \((w,d)\)-pairwise EMM ((\(w,d\))-PEMM).

The dependence of our measures on the particular choice \((w,d)\) is examined in the next theorem. We say that \( ||\cdot||_d \) (or \( d \)) is nondecreasing if for all \( f, g \in M \), such that \( |f(x)| \leq |g(x)| \), for all \( x \in \Pi \), then \( ||f||_d^2 \leq ||g||_d^2 \). Examples of such norms are the norms associated to \( d_{\infty} \) and \( d_2^2 \), which we shall denote henceforth by \( ||\cdot||_{\infty} \) and \( ||\cdot||_{2,W} \), respectively.

**Theorem 1.** If the metric \( d \) is nondecreasing, \( \sup_{t} EY_t^2 < C \), \( w \) satisfies (9) and \( \sup_{x \in \Pi} E |w(X_t,x)|^2 \leq C \), then \((Y_t)_{t \in \mathbb{Z}}\) is PSMM if and only if \((Y_t)_{t \in \mathbb{Z}}\) is \((w,d)\)-PSMM.

Theorem 1 tells us that as long as \( w \) and \( d \) satisfy some mild conditions, the \((w,d)\)-PSMM concept is invariant to the choice of the pair \((w,d)\). This allows us to transform the ill-posed problem of using the pairwise regression functions to a well-posed problem based on the new measures \( \gamma_{t,h,w} \). Two remarks are in order. First, Theorem 1 holds true if we replace PSMM by SMM and \( X_t \) by \( I_t \), so the dimension of \( X_t \) does not play a role here. Second, although convergence to zero is invariant to \((w,d)\), the actual rate of convergence to zero of \( ||\gamma_{t,h,w}||_{\mathcal{d}} \) might depend on the choice of \((w,d)\). This feature is intrinsic to the infinite-dimensional character of our problem; see Nelson (1990) for a related discussion. In short, two metrics in an infinite dimensional space are not necessarily equivalent. The choice of \((w,d)\) determines a particular metric, and hence a particular measure of nonlinear persistence. This feature of nonlinear persistence is in stark contrast to that of linear persistence.
Our measures of dependence can be related in a simple manner to those of Granger (1995) (and hence to those of Koop et al., 1996, and Wu, 2005). Assuming $\|\cdot\|_d$ is a nondecreasing norm and letting $F_t$ be the cdf of $X_t$,

$$\|\gamma_{t,h,w}\|_d^2 = \left\| \int_{-\infty}^{\infty} E(Y_{t+h} - \mu_{t+h}| X_t = z) w(z, \cdot) dF_t(z) \right\|_d^2$$

$$\leq \int_{-\infty}^{\infty} E^2(Y_{t+h} - \mu_{t+h}| X_t = z) dF_t(z) \left\| \left( \int_{-\infty}^{\infty} w^2(z, \cdot) dF_t(z) \right) \right\|_d^{\frac{1}{2}}^2$$

$$\leq E(M_{t,h}(I_t) - \mu_{t+h})^2 \left\| \left( \int_{-\infty}^{\infty} w^2(z, \cdot) dF_t(z) \right) \right\|_d^{\frac{1}{2}}^2, \quad (10)$$

where the first inequality follows from the Cauchy-Schwarz inequality, whereas the second follows from

$$\left\| E(Y_{t+h} - E(Y_{t+h}| X_t)) \right\|_d^2 \leq \left\| E(Y_{t+h} - E(Y_{t+h}| I_t)) \right\|_d^2.$$

Thus, assuming $\sup_t \|E^{1/2}(w^2(X_t, \cdot))\|_d^2 < \infty$, if $Y_t$ is SMM (relative to $F_t$), then $Y_t$ is $(w, d)$-PSMM, ensuring that the rate at which $\|\gamma_{t,h,w}\|_d^2$ decreases in $h$ is at most that of Granger’s (1995) measure.

We illustrate the behavior of our measures by means of several examples.

**Example 2.** Let $Y_t$ be a sequence of strictly stationary, mean-zero, Gaussian r.v.’s with $r(h) = E[Y_tY_{t+h}], h \geq 1, r(0) = 1$. It can be easily shown that, choosing $X_t = Y_t$,

$$\gamma_{t,h,w}(x) = r(h) \int_{-\infty}^{\infty} z\omega(z, x)(2\pi)^{-1/2}\exp(-0.5z^2)dz, \quad h \geq 1, x \in \mathbb{R}^p.$$

Therefore, whenever $0 < \|E[Y_tw(Y_t, \cdot)]\|_d < \infty$ we have $\|\gamma_{t,h,w}\|_d \sim C_{t,w,r}(h)$, for a constant $C_{t,w,d}$ which in this particular case is not $t$-dependent. Thus, under Gaussianity and mild regularity conditions, our concept of dependence reduces, as expected, to the traditional linear one. Similarly, defining $Z_t = Y_t^2$, then

$$Cov(Z_{t+h}, Z_t) = 2r^2(h),$$

for all $h \geq 1$, and, again, our measures of conditional mean dependence replicate this
behavior. Noting that by the Mehler’s formula

\[
f(y_{t+h}, y_t) = \phi(y_{t+h}) \phi(y_t) \left( 1 + \sum_{j=1}^{\infty} \frac{r^j(h)}{j!} H_j(y_{t+h}) H_j(y_t) \right),
\]

where \( f \) is the bivariate density of \( Y_{t+h}, Y_t \), \( \phi \) is the standard normal density and \( H_p(\cdot) \) is the \( p \)-th order Hermite polynomial, the first few being

\[
H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \ldots,
\]

it can be easily shown that

\[
E((Z_{t+h} - 1) w(Z_t; x)) = \sum_{j=1}^{\infty} \frac{r^j(h)}{j!} E(H_2(y_{t+h}) H_j(y_{t+h})) E(w(Z_t; x) H_j(y_t))
\]

\[
= r^2(h) E(w(Z_t; x) H_2(y_t)),
\]

where by the properties of the Hermite polynomials,

\[
E(H_j(y_t) H_k(y_t)) = k!, \quad j = k, \quad 0, \text{ otherwise},
\]

so that whenever \( 0 < \|E[H_2(Y_t)w(Y_t^2, \cdot)]\|_d < \infty \), choosing \( X_t = Z_t \), then \( \|\gamma_{t,h,w}\|_d \sim D_{t,w,d} r^2(h) \), for a constant \( D_{t,w,d} \) (not \( t \)-dependent in this case).

**Example 3.** Let

\[
Y_t = c + \phi Y_{t-1} + \varepsilon_t,
\]

where \(-1 < \phi < 1\), and \( \varepsilon_t \) is conditionally (on the past) Gaussian with zero mean and variance \( \sigma^2_t \), where

\[
\sigma^2_t = a + \alpha \varepsilon^2_{t-1},
\]

with \( a > 0, 0 < \alpha < 1 \), so \( Y_t \) follows an AR(1) with ARCH(1) errors model. Gallant et al. (1993) analyzed the persistence properties of this model by means of its conditional mean and volatility profiles for the case \( c = 0 \). Based on our measures, the persistence in conditional mean depends on

\[
\gamma_{t,h,w}(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [E(Y_{t+h} - E(Y_{t+h})|X_t = z)] w(z, x) F(dz),
\]

where \( X_t = (Y_t, Y_{t-1})' \), \( z = (z_1, z_2)' \), and \( F \) is the cdf of \( X_t \). It can be easily shown
that
\[ \gamma_{t,h,w}(x) = \phi^h \text{Cov} \left( Y_t, w \left( X_t, x \right) \right), \]
so if \( \| \text{Cov} \left( Y_t, w \left( X_t, \cdot \right) \right) \|_d \) is bounded and bounded away from zero, \( \| \gamma_{t,h,w} \|_d \) converges to zero at the exact rate \( \phi^h \). In the particular case \( w(z,x) = \mathbf{1}(z \leq x), x = (x_1,x_2)', \)
\[ \text{Cov} \left( Y_t, w \left( X_t, x \right) \right) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} z_1 F \left( dz \right) - \frac{c}{1-\phi} F \left( x \right), \]
which is uniformly bounded because \( \sup_x |\text{Cov} \left( Y_t, w \left( X_t, x \right) \right)| \leq E|Y_t| \).

The persistence in conditional second moments can be captured by
\[ \gamma_{t,h,w}(x) = E \left( (Y_{t+h}^2 - E(Y_{t+h}^2)) \cdot w \left( X_t, x \right) \right). \]
First, it can be shown that
\[ E \left( Y_{t+h}^2 \left| X_t = z \right. \right) = \frac{a}{1-\alpha} \sum_{j=1}^{h} (1-\alpha^j) \phi^{2(h-j)} + \frac{c^2 (1-\phi^h)^2}{(1-\phi)^2} + \frac{2c\phi^h (1-\phi^h)}{1-\phi} z_1 + (z_1 - c - \phi z_2)^2 \sum_{j=1}^{h} \alpha^j \phi^{2(h-j)} + \phi^{2h} z_1^2, \]
which, noting
\[ E \left( Y_{t+h}^2 \right) = \frac{a}{(1-\alpha) (1-\phi^2)} + \frac{c^2}{(1-\phi)^2}, \]
implies that
\[ \gamma_{t,h,w}(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( -\rho_1(h) - \rho_2(h) + \rho_3(h) z_1 + \rho_4(h) (z_1 - c - \phi z_2)^2 + \phi^{2h} z_1^2 \right) \cdot w(z,x) F \left( dz \right), \]
where
\[ \rho_1(h) \sim K_1 (a,\alpha,\phi) \left( \phi^{2h} > \alpha \right) + h\phi^{2h} \left( \phi^2 = \alpha \right) + \alpha^h \left( \phi^2 < \alpha \right), \]
\[ \rho_i(h) \sim K_i (c,\phi) \phi^h, \quad i = 2,3, \]
\[ \rho_4(h) \sim K_4 (\alpha,\phi) \left( \phi^{2h} > \alpha \right) + h\phi^{2h} \left( \phi^2 = \alpha \right) + \alpha^h \left( \phi^2 < \alpha \right) \quad (12) \]
where \( K_i \) in (12) are constants depending on the specified parameters. Assuming
\[ \| E \left( w \left( X_t, \cdot \right) \right) \|_d, \| E \left( Y_t w \left( X_t, \cdot \right) \right) \|_d, \| E \left( \varepsilon_t^2 w \left( X_t, \cdot \right) \right) \|_d, \| E \left( Y_t^2 w \left( X_t, \cdot \right) \right) \|_d, \]
are bounded
and bounded away from zero, the results in (12) determine the exact rates at which \(\|\gamma_{t,h,w}\|_d\) converges to zero as \(h\) increases, which basically depend on the relation between \(\alpha\) and \(\phi\). If \(\phi < \alpha\), \(\|\gamma_{t,h,w}\|_d\) decreases at the exact rate \(\alpha^h\); if \(\phi^2 < \alpha < \phi\), this rate is \(\phi^h\), but only due to the presence of \(c \neq 0\) in (11), which induces persistence in the squares, the rate being \(\alpha^h\) if \(c = 0\); if \(\alpha < \phi^2\) (assuming \(c = 0\)), the rate is \(\phi^{2h}\); finally, if \(\alpha = \phi^2\) (with \(c = 0\)), \(Y_t\) is more persistent in volatility than if \(\alpha \neq \phi^2\), with a rate of decrease given by \(h\alpha^h\). Essentially, apart from the effect of the constant, the AR structure of the model generates persistence driven by \(\phi^2\), whereas that implied by the ARCH structure depends on \(\alpha\), so our measure captures which one of these effects dominates. In contrast with the volatility impulse response function of Gallant et al. (1993), which was specifically designed to analyze the second moment properties of the process (separately from first moment properties), our measure of persistence of volatility captures first and second moment effects. In our view, this is not a drawback of our approach, because the parameter \(\phi\) has influence in the persistence of the squares (as it is the case when there are not ARCH effects).

As mentioned before, nonlinear or non-Gaussian pairwise distributional dependence can be characterized by the difference between the joint and marginal densities, i.e.,

\[
d_{t,h}(y, x) = f_{t,h}(y, x) - k_{t+h}(y)g_t(x).
\]

The estimation of \(d_{t,h}(y, x)\) was first considered by Rosenblatt (1975), who used nonparametric kernel estimation. Arguing as before, the function \(d_{t,h}\) can be characterized by the cumulative measure

\[
\eta_{t,h,\text{ind}}(y, x) = \int d_{t,h}(u, v)1(u \leq y)1(v \leq x)dudv
\]

\[
= \text{Cov}(1(Y_{t+h} \leq y), 1(X_t \leq x))
\]

\[
= F_{t,h}(y, x) - K_{t+h}(y)G_t(x),
\]

where \(F_{t,h}(y, x)\), \(K_{t+h}(y)\), and \(G_t(x)\) are, respectively, the bivariate and marginal cdf’s of \(Y_{t+h}\) and \(X_t\). More generally \(d_{t,h}\) can be characterized by

\[
\eta_{t,h,w}(y, x) := \text{Cov}(w(Y_{t+h}, y), w(X_t, x)), \quad (y, x) \in \Upsilon \subset \mathbb{R}^2,
\]

where \(w\) is as in (9).

With an abuse of notation, assume that the measures \(\eta_{t,h,w}\) belong to an appropriate metric space \(S = (M, d)\), with \(d\) a metric on \(M\). For instance, under mild conditions on \(w\), \(\eta_{t,h,w}\) will belong to \(\ell^\infty(\Upsilon)\) for a proper set \(\Upsilon\). Then we define the
concept of distributional persistent process as follows.

**Definition 5.** \((Y_t)_{t \in \mathbb{Z}}\) is called \((w, d)\)-pairwise SMD \(((w, d)\)-PSMD) if for all \(t \in \mathbb{Z}\),

\[
\|\eta_{t,h,w}\|_d < c_{h,t}.
\]

If \((Y_t)_{t \in \mathbb{Z}}\) does not satisfy the previous condition is called \((w, d)\)-pairwise EMD \(((w, d)\)-PEMD).

Definition 5 resembles an asymptotic pairwise independence assumption and is implied by, e.g., a strong mixing condition (cf. (8)). We illustrate the behavior of these measures by means of two examples.

**Example 4.** Taking \(Y_t\) as in Example 2 and setting \(X_t = Y_t\), from Appendix 2 in Gourieroux and Jasiak (2002)

\[
\eta_{t,h,w}(y, x) = r(h)E[Y_t w(Y_t, x)]E[Y_t w(Y_t, y)] + o(r(h)), \quad h \geq 1, x \in \mathbb{R}.
\]

Thus, whenever \(0 < \|E[Y_t w(Y_t, \cdot)]E[Y_t w(Y_t, \cdot)]\|_d < \infty\), we shall have \(\|\eta_{t,h,w}\|_d \sim C_{t,w,d}r(h)\), for a constant \(C_{t,w,d}\), which is not \(t\)-dependent.

Again, under Gaussianity and additional regularity conditions, our concept of persistence in distribution captures the same dependence as linear measures. The next example illustrates the fact that while nonlinear transformations of the data could be less persistent than the original process, and linear measures applied to the transformed process do not in general capture the persistence in the original series, our persistence in distribution measures can unmask the original dependencies.

**Example 5.** Taking \(Z_t\) as in Example 2, and setting \(X_t = Z_t\), our persistence in distribution measures applied to \(Z_t\) equal

\[
\eta_{t,h,w}(y, x) = \sum_{j=1}^{\infty} \frac{r_j(h)}{j!}E[w(Z_{t+h}, y)H_j(Y_{t+h})]E[w(Z_t, x)H_j(Y_t)],
\]

so that provided \(0 < \|E[Y_t w(Z_t, \cdot)]\|_d < \infty\), the rate at which \(\|\eta_{t,h,w}\|_d\) decreases as \(h\) increases is characterized by \(r(h)\).

### 4. Estimation and Asymptotic Theory

This section provides formal asymptotic theory for the estimators of the new measures. For simplicity in the exposition, we restrict ourselves to strictly stationary and
ergodic sequences, although our results allow for extensions to some non-stationary or non-ergodic sequences, as in e.g. Andrews and Pollard (1994) and Hansen (1996). Under the strictly stationary assumption, the natural estimators for $\gamma_{h,w} \equiv \gamma_{t,h,w}$ and $\eta_{h,w} \equiv \eta_{t,h,w}$, based on a sample $\{Y_t, X_t\}_{t=1}^n$ are, respectively,

$$
\hat{\gamma}_{h,w}(x) = \frac{1}{n-h} \sum_{t=1}^{n-h} (Y_{t+h} - \bar{Y}_{n-h}) w(X_t, x),
$$

with $\bar{Y}_{n-h} = (n-h)^{-1} \sum_{t=1}^{n-h} Y_t$, and

$$
\hat{\eta}_{h,w}(y, x) = \frac{1}{n-h} \sum_{t=1}^{n-h} w_c(Y_{t+h}, y) w_c(X_t, x),
$$

where $w_c(X_t, x) = w(X_t, x) - (n-h)^{-1} \sum_{s=1}^{n-h} w(X_s, x)$.

In this section we establish the asymptotic limits of the new estimated measures $\hat{\gamma}_{h,w}$ and $\hat{\eta}_{h,w}$. Since these are random functions, to develop the asymptotic theory we consider results from empirical processes theory as discussed, for instance, in van der Vaart and Wellner (1996). Throughout the rest of the paper the lag $h \geq 1$ is fixed. For simplicity in the exposition, we only consider the indicator weighting function $w(X, x) = 1(X \leq x)$ and the $(sup-)$ metric $d_{\infty}$ (our results can be similarly proved, in fact with simpler proofs, for other weights and other metrics). To simplify notation, denote $\hat{\gamma}_{h,w}$ and $\hat{\eta}_{h,w}$ ($\gamma_{h,w}, \eta_{h,w}$) by $\hat{\gamma}_h$ and $\hat{\eta}_h$ ($\gamma_h, \eta_h$), respectively, whenever $w(X, x) = 1(X \leq x)$, and define $1_e(X_t \leq x) \equiv 1(X_t \leq x) - E[1(X_t \leq x)]$. We consider $\hat{\gamma}_h$ as a process in $\ell^\infty(\mathbb{R}^p)$, where $\mathbb{R}^p := [-\infty, \infty]^p$, by extending its definition as $\hat{\gamma}_h(\infty) = 0 = \hat{\gamma}_h(-\infty)$. We denote $\Rightarrow$, the weak convergence in $(\ell^\infty(\Pi), d_{\infty})$ in the sense of J. Hoffmann-Jørgensen (see Definition 1.3.3 in van der Vaart and Wellner, 1996).

Under strictly stationarity and ergodicity, the uniform consistency of $\hat{\gamma}_h(\cdot)$ or $\hat{\eta}_h(\cdot)$ follows from the Ergodic Theorem and a Glivenko-Cantelli’s argument, see e.g. Koul and Stute (1999). Thus, this section focusses on the weak convergence of

$$
v_{n,h}(\cdot) = \sqrt{n-h}(\hat{\gamma}_h(\cdot) - \gamma(\cdot)) \quad \text{and} \quad \omega_{n,h}(\cdot) = \sqrt{n-h}(\hat{\eta}_h(\cdot) - \eta_h(\cdot)),
$$

under weak dependence assumptions. We establish the asymptotic theory for $v_{n,h}(\cdot)$ and $\omega_{n,h}(\cdot)$ under two different dependence scenarios, namely, mixingale classes and strong mixing processes. These weak dependence concepts are specially adapted to our SMM and SMD concepts, respectively. We also discuss the special case of no
dependence (martingale difference sequences and iid processes) as corollary of our main results. With these asymptotic results, we can compute uniform confidence bands for \( \hat{\gamma}_{h,w}, \hat{\eta}_{h,w} \), and the significance of \( \gamma_{h,w}, \eta_{h,w} \), can be tested.

A well-known result in the econometrics and statistical literature is that the weak convergence of a process \( R_{n,h}(\cdot) \) (e.g. \( v_{n,h}(\cdot) \) or \( \omega_{n,h}(\cdot) \)) in \((\ell^\infty(\Pi), d_\infty)\) will follow from the following conditions (see e.g. p. 2251 in Andrews 1994):

(i) \((\Pi, \rho)\) is a totally bounded pseudometric space, for some pseudometric \( \rho \);

(ii) \((R_{n,h}(z_1), \ldots, R_{n,h}(z_m))\) converges in distribution to a Borel law as \( n \to 0 \), for all fixed points \( z_1, \ldots, z_m \) in \( \Pi \);

(iii) \( \{R_{n,h}; n \geq 1\} \) is stochastically equicontinuous, i.e., for every \( \varepsilon, \eta > 0 \) there exists a \( \delta > 0 \) such that

\[
\limsup_{n \to \infty} P^* \left[ \sup_{z_1, z_2: \rho(z_1, z_2) < \delta} |R_{n,h}(z_1) - R_{n,h}(z_2)| > \varepsilon \right] < \eta;
\]

Examples of pseudometrics in condition (i) are typically

\[
\rho_{m,h}(x_1, x_2) = \left( E \left[ (Y_{t+h} - \mu)^2 \mid 1_e(X_t \leq x_1) - 1_e(X_t \leq x_2) \right]^2 \right)^{1/2}
\]

and

\[
\rho_{d,h}(z_1, z_2) = \left( E \left[ \mid 1_e(Y_{t+h} \leq y_1)1_e(X_t \leq x_1) - 1_e(Y_{t+h} \leq y_2)1_e(X_t \leq x_2) \right]^2 \right)^{1/2},
\]

for \( v_{n,h}(\cdot) \) and \( \omega_{n,h}(\cdot) \), respectively, where \( z_i = (y_i, x_i), i = 1, 2 \).

To prove weak convergence to a Gaussian process we must verify that conditions (i-iii) hold. Condition (i) usually follows under mild continuity assumptions on the marginal and joint distribution functions. Conditions implying (ii) are abundant in the literature; see e.g. de Jong (1997) and Dedecker, Merlevede and Volny (2007) for mixingales and Ango Nze and Doukhan (2004) for strong mixing processes. Usually, the most difficult condition to verify is the stochastic equicontinuity in (iii).

### 4.1 Asymptotic theory under mixingales

This subsection introduces the concept of a mixingale class and establish the weak convergence of \( v_{n,h} \) for mixingale classes. The use of mixingales is motivated by their close relation to our concept of SMM, as can be seen from the following definition
introduced by Hansen (1996). Recall that $I_t = (X'_t, X'_{t-1}, \ldots)'$ and let $\mathcal{F}_t$ and $\mathcal{H}_t$ be the $\sigma$–fields generated by $I_t$ and $H_t = (Y_t, X'_t, Y_{t-1}, X'_{t-1}, \ldots)'$, respectively.

**Definition 6.** $\{f(\cdot), \mathcal{H}_t, \mathcal{Q}\}_{t \in \mathbb{Z}}$ is called an $L^q$–mixingale class, $q \geq 1$, if there exist sequences of nonnegative constants $\{c_t(f)\}_{t \in \mathbb{Z}}$ and $\{\zeta_m\}_{m \in \mathbb{N}}$ such that $\zeta_m \to 0$ as $m \to \infty$, and for all $n \geq 1$ and for all $f \in \mathcal{Q}$

$$\|E[f(H_{t-1}) | \mathcal{H}_{t-m-1}] - E[f(H_{t-1})]\|_q \leq c_t(f)\zeta_m$$

(14)

$$\|f(H_{t-1}) - E[f(H_{t-1}) | \mathcal{H}_{t+m-1}]\|_q \leq c_t(f)\zeta_{m+1},$$

(15)

hold for all $t$, and $m \geq 0$.

We shall apply Definition 6 to the class

$$\mathcal{B} = \{f(H_{t-1}, x) = E ((Y_{t+h} - \mu)1(X_t \leq x) | H_{t-1}) : x \in \mathbb{R}^p\}.$$ 

Note that condition (15) is redundant here since $\mathcal{Q}$ (and in particular $\mathcal{B}$) is a class of $\mathcal{H}_{t-1}$-measurable functions. According to this definition, if $\mathcal{B}$ is an $L^q$–mixingale, $q \geq 2$, taking $x = +\infty$ we conclude that $(Y_t)_{t \in \mathbb{Z}}$ is a SMM process.

Hansen (1996) established general conditions for stochastic equicontinuity for smooth (Lipschitz-continuous) mixingale classes of parametric functions. Unfortunately, $\nu_{n,h}$ involves an indicator function, which is not smooth, and hence, the results in Hansen (1996) are not applicable to the present situation. We shall generalize the stochastic equicontinuity results in Hansen (1996) to non-smooth functions using a martingale approximation and the stochastic equicontinuity theorem for martingales developed recently in Escanciano (2007). This extension is of independent interest and has applications beyond the present paper.

Define the random function $G_t(x) := E [(Y_{t+h} - \mu)^21(X_t \leq x) | H_{t-1}]$. Also, define $\sigma^2 := E[Y_t^2]$. Recall the definition of the projection operator $P_t W = E(W | I_t) - E(W | I_{t-1})$, for any $W \in L_2$. The following regularity conditions are necessary for the subsequent asymptotic analysis.

**A1:** (a) $\{Y_t, X_t\}_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process such that $\mathcal{B}$ is an $L^q$–mixingale class with $c_t(f) \equiv c_t(x) = C \|f(I_{t-1}, x)\|_s^\gamma$ and $\sum_{m=1}^\infty \zeta_m < \infty$, for some $s \geq q$ and $1/2 \leq \gamma \leq 1$; (b) $|G_t(x_1) - G_t(x_2)| \leq C_t |x_1 - x_2|^{\lambda}$, for each $(x_1, x_2) \in \mathbb{R}^2 \times \mathbb{R}^p$, some $\lambda > 0$, and where $C_t$ is a $\mathcal{H}_{t-1}$-measurable function; (c) $0 < E[|Y_t|^r] < \infty$, with $r = \max(q, 2(1 + \delta))$, for some $\delta > 0$, and $E|C_t|^q < \infty$, for some $q > \max(1, 2p/\lambda)$.

**A2:** For any $x_1, \ldots, x_m$ in $\mathbb{R}^p$ any $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^p$ with $|\lambda| = 1$ and any
Assumption A1 is a mild condition on the DGP and permits a large class of nonlinear time series, including heteroskedastic ones. For instance, if \((Y_t)_{t \in \mathbb{Z}}\) is an \(L^q\)-integrable martingale difference sequences (MDS) with respect to \(\mathcal{H}_t\), then \(B\) satisfies Definition 6. Moreover, as shown in Hansen (1996) a large class of weak dependence processes satisfy Definition 6. Assumption A2 implies the convergence of the finite dimensional distributions.

Remark: In the case of no-persistence, i.e., when \(Y_t\) is a MDS with respect to \(\mathcal{H}_t\), following the proof of our Theorem 2 one can show that the moment assumptions in Theorem 2 can be relaxed to \(E[|Y|^2(1+\delta)] < \infty\) and \(E |C_t| < \infty\).

In the MDS case, the limit process \(v_{\infty,h}\) is a Gaussian process in \(\ell^\infty(\mathbb{R}^p)\) with zero-mean, continuous sample paths, and covariance function

\[
K(x_1, x_2) = E[(Y_{t+h} - \mu)^21_e(X_t \leq x_1)1_e(X_t \leq x_2)].
\]

Under conditional homokedasticity and \(p = 1\), \(v_{\infty,h}\) is distributed as \(\sigma^2 B(F_X(\cdot))\), where \(F_X\) is the cdf of \(X_t\), and \(B\) is a standard Brownian Bridge, so the quantiles of norms of \(v_{\infty,h}\) are tabulated. For MDS with general heteroskedasticity of unknown form, the asymptotic critical values of norms of \(\hat{\gamma}_h\) for testing the significance of \(\gamma_h\) can be approximated via a wild-type bootstrap approach. The bootstrap approximation under MDS is as follows. We approximate the distribution of \(\hat{\gamma}_h\) with that of

\[
\hat{\gamma}_{h}^*(x) = \frac{1}{n-h} \sum_{t=1}^{n-h} (Y_{t+h} - \sum_{t=1}^{n-h} 1_e(X_t \leq x))V_t,
\]

where \(1_e(X_t \leq x) = 1(X_t \leq x) - (n-h)^{-1} \sum_{t=1}^{n-h} 1(X_t \leq x)\) and \(V_t\) is a sequence of independent r.v.’s with zero mean, unit variance, bounded support and also independent of the sequence \({Y_t, X_t}_{t=1}^n\). This procedure is similar to the wild bootstrap used in Wu (1986). Examples of \({V_t}\) sequences are iid Bernoulli variates with \(P(V_t = 0.5(1 - \sqrt{5})) = (1 + \sqrt{5})/2\sqrt{5}\) and \(P(V_t = 0.5(1 + \sqrt{5})) = 1 - (1 + \sqrt{5})/2\sqrt{5}\). The validity of this bootstrap approximation is proved in Theorem 4 of Escanciano.
and Velasco (2006b). With the bootstrap critical values we can compute uniform confidence bands for \( \hat{\gamma}_h \).

For the general non-adapted case described in A1, we can apply other resampling methods, such as block-bootstrap or subsampling. In particular, subsampling is a powerful resampling scheme that allows an asymptotically valid inference under very general conditions on the DGP, see the monograph by Politis, Romano and Wolf (1999). With an abuse of notation we write \( \Gamma(v_{n,h,w}) = \|v_{n,h,w}\|_d \) as a function of the data \( \{Z_t \equiv (Y_{t+h}, X^t_{t})\}_{t \in \mathbb{Z}} \), \( \Gamma(v_{n,h,w}) = \Gamma(v_{n,h,w}(Z_1, ..., Z_n)) \). Let \( G^\Gamma_n(w) \) be the cdf

\[
G^\Gamma_n(z) = P(\Gamma(v_{n,h,w}) \leq z).
\]

Let \( \Gamma(v^{b,i}_{n,h,w}) = \Gamma(v_{n,h,w}(Z_i, ..., Z_{i+b-1})) \) be the norm computed with the subsample \( (Z_i, ..., Z_{i+b-1}) \) of size \( b \). We note that each subsample of size \( b \) (taken without replacement from the original data) is indeed a sample of size \( b \) from the true DGP. Hence, it is clear that one can approximate the sampling distribution \( G^\Gamma_n(z) \) using the distribution of the values of \( \Gamma(v^{b,i}_{n,h,w}) \) computed over the \( n - b + 1 \) different subsamples of size \( b \). That is, we approximate \( G^\Gamma_n(z) \) by

\[
G^\Gamma_{n,b}(z) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1(\Gamma(v^{b,i}_{n,h,w}) \leq z), \quad z \in [0, \infty).
\]

Let \( c^\Gamma_{n,1-\tau,b} \) be the \((1 - \tau)\)-th sample quantile of \( G^\Gamma_{n,b}(z) \), i.e.,

\[
c^\Gamma_{n,1-\tau,b} = \inf \{ z : G^\Gamma_{n,b}(z) \geq 1 - \tau \}.
\]

The theoretical justification of the subsampling approximation can be grounded on Theorem 3.5.1 in Politis, Romano and Wolf (1999) under mild conditions on the strong mixing coefficients and the parameter \( b \). With these asymptotic results one can compute (uniform) confidence bands for the measures \( \gamma_h(\cdot) \).

### 4.2 Asymptotic theory under strong mixing conditions

In this subsection we establish the stochastic equicontinuity of \( \omega_{n,h} \) under strong mixing assumptions. This will follow from the general results in Andrews and Pollard (1994). We require the following assumption.

**B1:** (a) \( \{Y_t, X^t_t\}_{t \in \mathbb{Z}} \) is a strictly stationary; (b) \( |F_X(x_1) - F_X(x_2)| \leq C |x_1 - x_2|^{2\lambda_1} \), for each \( (x_1, x_2) \in \mathbb{R}^p \times \mathbb{R}^p \), some \( \lambda_1 > 0 \); (c) \( |F_Y(y_1) - F_Y(y_2)| \leq C |y_1 - y_2|^{2\lambda_2} \), for each \( (y_1, y_2) \in \mathbb{R} \times \mathbb{R} \), some \( \lambda_2 > 0 \); (c) \( \{Y_t, X^t_t\}_{t \in \mathbb{Z}} \) is a strong mixing process with
mixing coefficients such that \( \sum_{h=1}^{\infty} h^{Q-2} \alpha_{h}^{\gamma/(Q+\gamma)} < \infty \), for some even integer \( Q \geq 2 \), such that \( Q > \left( \frac{2+\gamma}{2} \right) \frac{p+1}{\lambda} \), some \( \gamma > 0 \) and \( \lambda = \min(\lambda_1, \lambda_2) \).

**Theorem 3:** Under B1, \( \omega_{n,h} \) is stochastic equicontinuous with respect to \( \rho_{d,h} \).

Different assumptions to B1 are also possible using other mixing concepts under the \( sup \)-metric; see, e.g., Rio (2000) and Andrews (1993, p. 200). For strong mixing sequences Theorem 2.17 in Bosq (2000) gives the asymptotic distribution for \( L_2 \)-norms under the mild condition \( \sum_{h=1}^{\infty} \alpha_{h} < \infty \).

For distributional dependence the least persistent case corresponds to that where \( Y_{t+h} \) is independent of \( \mathcal{F}_t \). When \( p = 1 \), the limit process of \( \omega_{n,h} \), say \( \omega_{\infty,h} \), is distributed as \( W(F_X(\cdot), F_Y(\cdot)) \) where \( W(u,v) \) is a standard bivariate Brownian Bridge, and therefore, continuous functionals of \( \tilde{\eta}_h \) are asymptotic distribution free, with critical values that can be tabulated. For \( p > 1 \), we can use a wild-bootstrap type procedure as for \( \tilde{\gamma}_h \). That is, we approximate the distribution of \( \tilde{\eta}_h \) with that of

\[
\tilde{\eta}_h(x) = \frac{1}{n-h} \sum_{t=1}^{n-h} 1_{c,Y_{t+h}}(y) 1_{c,X_t}(x) V_t,
\]

Of course, other alternative bootstrap procedures are available, e.g., the classical bootstrap of Efron (1979). For the more general case described in B1, a subsampling approximation, as discussed in Section 4.1 is, of course, possible.


In this section we apply our methodology to study the nonlinear persistence of the main stock market indices. We aim to investigate three different features of these data sets. First, we study the nonlinear persistence in conditional mean. Second, we compare nonlinear measures of persistence in conditional variance with linear ones. Finally, we employ a particular version of our results to analyze the phenomenon of bull versus bear markets in finance, which has attracted plenty of attention in recent times (see, e.g., Longin and Solnik (2001), Ang and Chen (2002), and references therein) and which involves bivariate comparisons.

The data consist in daily closed stock returns for S&P500 (SP500), Dow Jones Industrials (DJ), Frankfurt DAX Index (DAX), the London FTSE-100 Index (FTSE), Tokyo Nikkei Index (NIK) and the Hong Kong Hang Seng Index (HANG). The daily data are taken from January 1, 2003 to June 9, 2008, with a total of 1418
observations. The daily closed values for these stock indices are obtained from http://www.freelunch.com. We consider the returns of the indices obtained as the log differences of the data, which exclude dividend payments. Since our interest is in the persistence properties of these data sets, we can safely abstract from the problem of non-synchronism in national stock markets.

5.1 Martingale properties and persistence in conditional mean

The martingale properties of stocks returns have been extensively investigated in the literature, see e.g. Lo and MacKinlay (1999) and references therein. These earlier studies showed that stock price changes are not MDS, although recent evidence supporting or refuting the martingale hypothesis for some stock indices seems mixed; see e.g. Escanciano and Velasco (2006b) with the S&P500.

To gain insight in the nonlinear persistence in mean properties of this data sets we consider the behavior of Kolmogorov-Smirnov (KS) test $KS_n(h)$ given by

$$K S_n(h) := \sup_{x \in [-\infty, \infty]} \left| (n - h)^{1/2} \hat{\sigma}_{n,h}^{-1} \gamma_h(x) \right| \quad h \geq 1,$$

where

$$\hat{\sigma}_{n,h}^2 = \frac{1}{n - h} \sum_{t=1}^{n-h} (Y_{t+h} - \bar{Y}_{n-h})^2.$$

In Table 1 we show the KS tests statistics at different lags and all data sets, and the bootstrap 95% quantile computed with the wild-bootstrap approximation described in (16), for testing the significance of the measure $\gamma_h$. The function $KS_n(h)$, as a function of $h$, can be useful for detecting nonlinearities graphically and can play the same role as the usual autocorrelograms in the linear setup. Table 1 and other unreported simulations reveal two important features of these data sets. The data are uncorrelated and, moreover, seem to be MDS. This finding stands in contrast with studies documenting evidence of serial (linear or nonlinear) dependence in conditional mean. Only few isolated lags seem to be significantly different from zero at 5% nominal level.

Please, insert Table 1 here.

Thus, contrasting with earlier studies which conclude that stock indices are predictable in conditional mean given past prices, we find that the stocks considered here are unpredictable in mean. Of course, these conclusions may be period and frequency-specific and may not hold for alternative stock indices.
5.2 Persistence in conditional variances

By modern asset pricing theory the conditional variance (interpreted as a quantitative measure of risk) of future asset returns is a key component of the price of an asset. Thus, analyzing the persistence properties of conditional variances is of primary importance for a risk manager in order to take asset allocation decisions. In particular, the extent to which shocks to conditional variances have short or long-run effects is crucial to determine the risk premium of long-term contracts. These ideas have been the subject of an extensive theoretical and empirical literature, and, e.g., Poterba and Summers (1986) gave evidence of the linkage between persistence of conditional variances and stock prices.

Much of the existing literature has been restricted to linear persistence, see, for instance, the AR(1) model used in Poterba and Summers (1986) or the integrated GARCH (IGARCH) model suggested by Engle and Bollerslev (1986), but this might hide important nonlinear sources of persistence. Thus, in this subsection we study the nonlinear persistent properties in conditional variances of the stocks considered, and we also measure the extent to which linear and nonlinear persistent measures differ. To that end, we apply our new methodology and consider the estimates

$$\breve{\gamma}_{h,(2)}(x) = \frac{1}{n - h} \sum_{t=1}^{n-h} (Y_{t+h}^2 - \bar{Y}_{n-h}^2) 1(Y_t \leq x),$$

where

$$\bar{Y}_{n-h}^2 = \frac{1}{n - h} \sum_{t=1}^{n-h} Y_t^2.$$

We consider the KS functional

$$KS_n(h) := \sup_{x \in [-\infty, \infty]} \left| (n - h)^{1/2} \hat{\sigma}_{2,n,h}^{-1} \tilde{\gamma}_{h,(2)}(x) \right|, \quad h \geq 1,$$  \hspace{1cm} (19)

where $\hat{\sigma}_{2,n,h}^2$ is the sample variance of $\{Y_{t+h}^2\}_{t=1}^{n-h}$. Since we want to take into account possible asymmetric effects in conditional variance persistence we use $X_t = Y_t$ instead of the popular choice $X_t = Y_t^2$ in the definition of $\tilde{\gamma}_{h,(2)}$.

According to our previous results, these data sets seem to be MDS. Therefore, the measures $\tilde{\gamma}_{h,(2)}$ are indeed measures of nonlinear persistence in the conditional variance. There are already two well-known stylized facts that have been extensively documented in the literature of financial time series; first, conditional variances are time-varying; and second, stock returns and other financial time series are quite
persistent in conditional second moments (cf. Engle and Bollerslev, 1986, Bollerslev and Engle, 1993), especially when large sample sizes are considered. Linear measures of dependence were sufficient to uncover these two important stylized facts, but, alternatively, we ask the following questions: is nonlinear persistence different from linear persistence? If so, to what extent are they different?

We made an attempt to clarify these issues by means of the following experiment, which basically analyses whether linear components of the conditional variance persistence (captured by means of parametric models) account for all persistence (in volatility) present in the data. Thus, initially a GARCH(1,1) model with a constant mean was fitted to each data set, and we used the estimated parameters to generate 10,000 independent replications of simulated data following the parametric fitted model with Gaussian innovations. Then we estimate our nonparametric measures of dependence with the simulated data. In this way, we can assess if parametric models are able to explain the nonlinear persistence present in the data. We report in Figures 1 to 6 the nonparametric KS tests statistics at different lags and the wild bootstrap 95% quantile for $\hat{\gamma}_{h,(2)}$, jointly with mean (over the 10000 replications) KS tests from the parametric fits and the 95% confidence interval. For the sake of comparison we also plot the 95% confidence interval in the nonparametric IPAF plot, so we have a measure of the precision of the explanatory power of parametric models.

Our first finding is that we strongly reject the constant conditional variance for all stock indices and essentially all lags considered, with the exception of NIKKEY at some lags. We also observe a high linear persistence in conditional second moments. These two findings are consistent with the previously mentioned stylized facts. More important for our purposes, we find that nonlinear persistence is considerably higher than linear persistence for European and American stock returns. That is, for these data sets we find that parametric GARCH models, and hence, classical linear measures of persistence, cannot explain the nonlinear persistence present in the conditional volatilities of these stock market returns. In unreported experiments we observed that this conclusion is robust to different specifications of the volatility model, such as an EGARCH specification, different innovation’s distributions, such as a Student-t distribution, or different linear measures of persistence such as linear autocorrelograms for the squared returns.

According to our definitions, European and American indices may show an extended memory in variance behavior, which apart from having implications along the lines described before (noting that the covariance with benchmark portfolios influence the price of assets), prompts consideration of the possibility of being co-persistent in
variance (Bollerslev and Engle, 1993). This occurs when linear combinations of the stock indices display significantly less persistence than the indices themselves. Given our different measures of persistence, our concept of co-persistence is necessarily different from that in Bollerslev and Engle (1993), but could be equally framed within the common features framework (Engle and Kozicki, 1993), where some latent variables with a particular feature influence all observables and transmit to them this feature. The concept of co-persistence in variance of assets or portfolios (so that their volatility is tied together in the long run) may imply that the pricing of particular portfolios is not much affected by current shocks (unlike pricing of particular assets which form that portfolio), which, again, has important economic implications for optimal portfolio allocations and international diversification, and might lead to a better understanding of volatility spillover effects. To uncover this possible co-persistence, we carried out a simple analysis, plotting (Figure 7) the nonlinear IPRF (NIPRF) for the squared differences, that is, \( Y_t = Y_{i,t} - Y_{j,t} \), where the subindex \( i \) in \( Y_{i,t} \) indicates the \( i \)th stock index. We just considered combinations of American and European indices, noting that Asian indices show considerably less persistence. In view of our results, the S&P500 and DJ seem to be co-persistent in variance, i.e., individually they are highly persistent in variance but their difference shows little or no-persistence. Other stocks did not show this strong relationship, being noticeable that the most persistent difference of indices appear to be that between DAX and FTSE (although alternative combinations of these indices might be co-persistent). Undoubtedly, formal tests and estimation procedures along this direction deserve future research, but they are well beyond the scope of the present paper.

Please, insert Figures 1 to 7 here

5.3 Asymmetric effects in the persistence of conditional variances

In this subsection we aim to find some further insights into the high persistence properties of conditional variances. We first compare conditional variance persistence in bull and bear markets. There has been a recent interest in investigating asymmetries in contemporaneous conditional correlations between stock indices (see e.g. Hong, Tu and Zhou, 2007, and references therein). See also Mazzotta (2008) for a up-to-date review of the extensive empirical literature. Unlike this existing literature, we investigate asymmetries in nonlinear persistence of conditional variances in bull and bear markets. Our approach has several advantages over existing methods. First, it is nonparametric and hence it overcomes possible misspecification errors in related studies using parametric models. Second, our methods are able to capture nonlinear
dependence, something that is not generally possible with parametric models or linear measures of dependence. Last but not least, our approach is flexible enough so as to provide simple measures of volatility spillovers and/or leverage effects.

We consider a modification of our measures in (18) to restrict the conditioning set to the left and right tails of the distribution. Our measure of persistence in variance during bear markets is

$$\hat{\gamma}_{h,l}^{(ij)}(x) = \frac{\sum_{t=1}^{n-h} Y_{i,t+h}^2 1(Y_{j,t} \leq x)}{\sum_{t=1}^{n-h} 1(Y_{j,t} \leq x)} - \left( \frac{\sum_{t=1}^{n-h} Y_{i,t+h} 1(Y_{j,t} \leq x)}{\sum_{t=1}^{n-h} 1(Y_{j,t} \leq x)} \right)^2,$$

where the subindex $i$ in $Y_{i,t}$ indicates the $i-th$ stock index, e.g., $i =$DAX. The function $\hat{\gamma}_{h,l}^{(ij)}(x)$ is a consistent estimator of $\gamma_{h,l}^{(ij)}(x) = \text{Var}(Y_{i,t+h} | 1(Y_{j,t} \leq x))$.

Similarly, define $\hat{\gamma}_{h,u}^{(ij)}$ as $\hat{\gamma}_{h,l}^{(ij)}$ with $1(Y_{j,t} \leq x)$ replaced by $1(Y_{j,t} \geq x)$. To make the measures scale invariant, we have standardized all data sets throughout this section. The range considered for $x$ depends on the focus on bear or bull markets. To study persistence in bear markets we consider the functional

$$K S_{n,l}^{(ij)}(h) := \sup_{x \in A^-} \left| (n-h)^{1/2} \hat{\gamma}_{h,l}^{(ij)}(x) \right| \quad h \geq 1,$$

where

$$A^- = \{ x \in \mathbb{R} : F_{n,j}^{-1}(0.05) \leq x \leq F_{n,j}^{-1}(0.20) \},$$

with $F_{n,j}^{-1}(\alpha)$ denoting the $\alpha-th$ empirical quantile of the empirical distribution function of $\{Y_{j,t}\}_{t=1}^n$. Similarly, we study persistence in bull markets by means of the functionals

$$K S_{n,u}^{(ij)}(h) := \sup_{x \in A^+} \left| (n-h)^{1/2} \hat{\gamma}_{h,u}^{(ij)}(x) \right| \quad h \geq 1,$$

where

$$A^+ = \{ x \in \mathbb{R} : F_{n,j}^{-1}(0.80) \leq x \leq F_{n,j}^{-1}(0.95) \},$$

The asymptotic theory of $K S_{n,l}^{(ij)}(h)$ and $K S_{n,u}^{(ij)}(h)$ can be established using our results in Section 4. The significance of the population analogues of $\gamma_{h,l}^{(ij)}(x)$ and $\gamma_{h,u}^{(ij)}(x)$, say $\gamma_{h,l}^{(ij)}(x)$ and $\gamma_{h,u}^{(ij)}(x)$, can be tested using subsampling methods, as described at the end of Section 4.1. In this section, we apply the subsampling approximation with $b = \lfloor kn^{2/5} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, for several values of $k$, as suggested by Sakov and Bickel (2000). For the sake of space only simulations with $k = 9$ are
reported, the results with other values are similar and omitted.

In order to compare nonlinear dependence in bull and bear markets, we can test for

\[ H_0 : \gamma_{h,l}^{(ij)}(F_j^{-1}(\alpha)) = \gamma_{h,u}^{(ij)}(F_j^{-1}(1 - \alpha)) \text{ for all } \alpha \in [0.05, 0.20], \]

against two-side or one-side alternatives, where \( F_j^{-1} \) is the quantile function associated to the cdf \( F_j \) of the \( j-th \) stock. We carried out marginal tests for \( H_0 \) and study the persistence and significance of the differences \( \gamma_{h,l}^{(ij)}(F_j^{-1}(\alpha)) - \gamma_{h,u}^{(ij)}(F_j^{-1}(1 - \alpha)) \), using the functional

\[ KS_{n,h,c}^{(ij)}(h) := \sup_{\alpha \in [0.05, 0.20]} \left| (n - h)^{1/2} \{ \gamma_{h,u}^{(ij)}(F_n^{-1}(1 - \alpha)) - \gamma_{h,l}^{(ij)}(F_n^{-1}(\alpha)) \} \right|, \quad h \geq 1. \]

We plot \( KS_{n,h,c}^{(ij)}(h) \), \( a = l, u, c \), as a function of \( h \), for all possible combinations of \( i \) and \( j \). The cases where \( i = j \) correspond to the functionals used in the previous section, when we restrict the conditioning variables to the left and right tails of the distribution.

Several conclusions can be drawn from our empirical analysis on persistence in bull and bear markets in Figures 8 to 19. First, in both situations the conditional variance seems to be highly persistent. A common feature in our applications is that the major difference between bear and bull markets is attained at the first lag for all European and American stocks. The magnitude and persistence is much higher in bear markets than in bull markets, that is, the difference between bear and bull markets variances is (statistically) significant at all lags and quite persistent. Our finding stands in contrast with some existing studies such as that of De Santis and Gerard (1997, pg. 1905) who argued that the behaviour of international returns during bear and bull markets is mostly a short-term phenomenon.

Please, insert Figures 8 to 19 here.

When we make bivariate comparisons, changing the conditioning set, so \( i \neq j \) in \( KS_{n,h,c}^{(ij)} \), we still find highly persistence variance with most indices. For the DAX the persistence in bear markets is higher than in bull markets when we condition on FTSE and American indices, but is less significant when conditioning on Asian indices. A similar behavior is found for American indices, showing a stronger link between European and American markets. For the FTSE, surprisingly, we find little difference between bull and bear markets when we condition on the DAX, although the difference at the first lag is still significant. In contrast to DAX, the FTSE

\(^3\)Available from the authors upon request.
index is more affected by HANG’s news, but it still shows little persistence when we condition on NIKKEY information. DJ and S&P500 show a remarkable similar persistence behavior. The persistence in variances of HANG and NIKKEY are rather different to American and European of indices. In particular, when conditioning on American indices, the most significant persistence is at lag two, in accordance with a time-zone effect.

In recent years there have been some interest in the persistent properties of the so-called “leverage effect”, see Gallant, Rossi and Tauchen (1993) and references therein. The leverage effect asserts that volatility is higher when past rates changes are negative. Our analysis can be certainly interpreted in terms of leverage effects. We have shown how our measures are able to gauge the different response in variances to negative and positive past values of the series and the persistence of that difference. Unlike existant measures of leverage effects which are based on specific models, e.g., the news-impact curve of Engle and Ng (1993), our new measures are nonparametric, in the sense that they do not depend on a specific model (whose correct specification is always questionable).

Summarizing, when changing the condition set and looking for spillover effects in variance, several interesting conclusions arise. In some cases we observe one-directional persistence, as between FTSE and DAX indices, or two-directional persistence as between DJ and FTSE. We also observe persistence in variance with all indices but with the NIKKEY, which seems to be more independent to bad performance in other markets. The exogeneity of NIKKEY might suggest that Japan is a good candidate for the purpose of diversification for an international investor. By the same token, the nonlinear co-persistence found among European and American stocks in bear markets suggest that the benefits of diversification between Europe and the United States may not be available when investors need them the most.

Overall, we have shown the flexibility of our measures in measuring subtle nonlinear dependence and persistence in financial data, and their ability to uncover persistence asymmetries in multivariate situations in a nonparametric way.

6. Conclusions and Summary

In this paper we have introduced new operative measures of nonlinear persistence. In contrast to related approaches, our measures are easy to estimate, do not need of large sample sizes and are nonparametric in nature, without relying on user-chosen parameters (like bandwidths). Thus, we have proposed generally applicable econometric tools for measuring nonlinear persistence in macroeconomic and financial data.
In addition, we have established an important linkage between the persistence measures proposed by Granger (1995) and the nonlinear impulse response literature, as discussed in Koop et al. (1996), and, furthermore, we have connected these results to our measures and the dependence measures used by Wu (2005). We have also investigated the asymptotic properties of the estimates of our measures under two sets of assumptions, mixingale and strong mixing conditions. For mixingale classes, we have extended a general result by Hansen (1996) to a non-smooth class involving indicator functions, a result of independent interest.

Finally, we have employed our measures to study the conditional mean and variance persistence in the main daily stock market indices. Our main finding is the ubiquitous nonlinear persistence in conditional variance that is not accounted for by classical linear measures of persistence and by popular parametric models. We also find that negative shocks contribute substantially more to the persistence in variance than positive shocks, showing an asymmetric behavior in variance persistence. We have uncovered existing links between international stock markets and their asymmetric behaviour. These findings have important economic implications in, e.g., asset pricing and hedging. In particular, our findings may show that current shocks to conditional variances may have a permanent effect on long term contracts.

As shown in the applications, linear measures can substantially understate the actual (nonlinear) persistence present in economic data. One can argue that some of the existing puzzles in economics and finance might be explained by this notorious limitation of linear methods; see for instance the so-called equity premium puzzle (inability of many asset pricing models in explaining the 6% annual equity premium and its standard deviation of 19%). In a recent paper, Bansal and Yaron (2004) have introduced an asset market pricing model with long-run risks which appears to be promising in explaining these asset pricing anomalies. This growing literature of long-run risks emphasizes the fundamental role of persistence of consumption growth on the equity premium. Our new tools can be used to assess to what extent these linear models of asset pricing are able to "fit" the nonlinear persistence of asset prices and consumption growth, as well as the relation between them.

Our work also complements, and has the potential to make operative, some of the ideas presented in Bollerslev and Engle (1993) in nonparametric situations. Our empirical findings suggest that stock market returns can be extended memory in variance, and may present co-persistent properties, as suggested in Bollerslev and Engle (1993). The economic consequences of these important findings will be investigated elsewhere.
Formal inferential procedures for testing the short and extended memory character of a given time series are also of paramount importance, and deserve further investigation. Moreover, our operative measures pave the way for a formal definition of cointegration in nonlinear set-ups that can formalize some of the ideas put forward by previous studies. We leave these and other important extensions of our methods for future research.

APPENDIX: PROOFS

Proof of Theorem 1. First, by simple application of the Cauchy-Swartz’s inequality, a PSMM process is \((w, d)\)-PSMM. Next, from \(\sup_t EY_t^2 < C\), we have that the sequence \((Y_{t+h})_{h \in \mathbb{N}}\) is uniformly tight. Also, from Prohorov’s theorem, we have that any subsequence of \((Y_{t+h})_{h \in \mathbb{N}}\) has a further subsequence \((Y_{t+s})_{s \in \mathbb{N}}\) which converges in distribution to a r.v. \(Y\), see Billingsley (1968). Because the subsequence is uniform integrable, the convergence is also in \(L_2(\mathbb{R}, F)\). From Proposition 1.2 in Chung and Williams (1990) \(m_{t,h}\) also converges to \(E(Y | X_t)\) in \(L_2(\mathbb{R}, F)\). Then

\[
\| \gamma_{t,h,s,w} \|_d^2 \to \| E[(Y - \mu)w(X_t, x)] \|_d^2 \quad \text{as } h_s \to \infty.
\]

From (9) we have that \(E(Y | X_t) = \mu\) a.s. Therefore, from the uniqueness of the limit, we have that \((m_{t,h})_{h \in \mathbb{N}}\) converges in \(L_2(\mathbb{R}, F)\) to \(\mu\) and the process is PSMM. □

In the proof of Theorem 2 we use extensively the following weak convergence theorem, which is a corollary of a general result in Escanciano (2007). Define the process

\[
R_n(x) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t 1_e(X_t \leq x), \quad x \in \mathbb{R}^p.
\]

Let \(F\) be the cumulative distribution function (cdf) of \(X_t\). We define the random function \(G_t(x) := E[Z_t^2 1(X_t \leq x) | \mathcal{H}_{t-1}]\), where \(\mathcal{H}_t\) is the \(\sigma\)-field generated by \((Z_t, X_t, Z_{t-1}, X_{t-1}, ...\). The following regularity condition is necessary for the subsequent asymptotic analysis. Define the pseudometric, with \(s \geq 2\),

\[
\rho_s(x_1, x_2) = (E[|Z_t|^s | 1_e(X_t \leq x_1) - 1_e(X_t \leq x_2)|^s])^{1/s}.
\]

C1: (a) \(\{Z_t\}_{t \in \mathbb{Z}}\) is a strictly stationary and ergodic MDS with respect to \(\mathcal{H}_t\), with \(E[|Z_t|^s] < \infty\) for some \(s \geq 2\); (b) \(|G_t(x_1) - G_t(x_2)| \leq C_t|x_1 - x_2|^\lambda\), for each \((x_1, x_2) \in \mathbb{R}^p\).
$\mathbb{R}^p \times \mathbb{R}^p$, some $\lambda > 0$, and where $C_t$ is a stationary sequence with $E|C_t| < \infty$; (c) $F$ is absolutely continuous.

**Theorem A1**: Under C1, $R_n$ is stochastically equicontinuous with respect to $\rho_s$.

**Proof of Theorem A1**. It follows from (a trivial) extension to multivariate regressors of Theorem 1 in Escanciano and Mayoral (2008). $\square$

**Proof of Theorem 2**. Defining the projector operator $P_l W = E(W|H_l) - E(W|H_{l-1}), W \in L_2$, we can show that, uniformly in $x$,

$$v_{n,h}(\cdot) = \sqrt{n - h(\tilde{\gamma}_h(\cdot) - \gamma_h(\cdot))}$$

$$= \sum_{j=1}^{h+1} \frac{1}{\sqrt{n - h}} \sum_{t=1}^{n} P_{t+j-1} \{(Y_{t+h} - \mu)1_e(X_t \leq x)\}$$

$$+ \frac{1}{\sqrt{n - h}} \sum_{t=1}^{n} E((Y_{t+h} - \mu)1_e(X_t \leq x)|H_{t-1}) - \gamma_h(\cdot) + o_P(1)$$

$$\equiv \sum_{j=1}^{h+1} A_{nj}(x) + B_n(x) + o_P(1).$$

Note that $A_{nj}$ are martingales, so results to establish the stochastic equicontinuity of $A_{nj}$, such as our previous Theorem A1 or Theorem 1 in Escanciano (2007) are applicable. Note that for $j \geq 2$,

$$A_{nj}(x) = \frac{1}{\sqrt{n - h}} \sum_{t=1}^{n} \{P_{t+j-1}Y_{t+h}\}1_e(X_t \leq x),$$

so Theorem A1 applies to $Z_t = P_{t+j-1}Y_{t+h}$, after noting that, for $s \geq 2$,

$$E[|Z_t|^s] \leq E[|Y_{t+h} - E(Y_{t+h}|H_{t+j-2})|^s] \leq E[|Y_{t+h}|^s] < \infty.$$  

The stochastic equicontinuity of $A_{n1}$ follows from the following arguments. Let $B_k = [x_k, y_k]$, for $k = 1, \ldots, N$, a partition of $\mathbb{R}^p$ in $\varepsilon$-brackets with respect to $\rho_{m,h}$. With an abuse of notation, we denote the previous partition associated to $\varepsilon = 2^{-q}$ simply by $B_q = \{B_{qk}; 1 \leq k \leq N_q\}$. Without loss of generality we can assume that the finite partitions described before are nested. It is clear that

$$\sum_{q=1}^{\infty} 2^{-q} \sqrt{\log N_q} < \infty,$$
which readily verifies condition (5) in Escanciano (2007). It remains to verify condition (6) in that paper. From the monotonicity of \(1(X_t \leq x)\), A1(b) and keeping the notation in Escanciano (2007), \(\alpha_n(B_q)\) is defined as

\[
\max_{1 \leq k \leq N_q} \left| \frac{1}{n} \sum_{t=1}^{n} E \left[ \sup_{x,y \in B_{qk}} \left| P_t\{(Y_{t+h} - \mu)1_e(X_t \leq x)\} - P_t\{(Y_{t+h} - \mu)1_e(X_t \leq y)\} \right|^2 \right| H_{t-1} \right|
\]

\[
\leq C \max_{1 \leq k \leq N_q} \left| n^{-1} \sum_{t=1}^{n} E[(Y_{t+h} - \mu)^2 1_e(x_k \leq X_t \leq y_k)] \right| H_{t-1} \leq O_P(2^{-2q}).
\]

This proves (6) in Escanciano (2007), and hence the stochastic equicontinuity of \(A_{n1}\).

As for \(B_n(x)\), we shall apply Theorem 1 in Hansen (1996). First note that the class of functions \(\{f(H_{t-1}, x) = E(\{(Y_{t+h} - \mu)1_e(X_t \leq x)\}) : x \in \mathbb{R}^p\}\) satisfies the Lipschitz condition

\[
|f(H_{t-1}, x) - f(H_{t-1}, x)| \leq C_t^{1/2} |x_1 - x_2|^{\lambda/2}.
\]

Therefore, the class \(\mathcal{B}\) satisfies the conditions of Theorem 1 in Hansen (1996), and the stochastic equicontinuity of \(B_n\) follows. The finite sum of stochastic equicontinuous processes is stochastic equicontinuous. Finally, the convergence of the finite dimensional distributions in (ii) follows directly from A2 and Corollary 1 in Dedecker, Merlevede and Volny (2007). □

**Proof of Theorem 3.** The theorem follows from a direct application of Theorem 2.2 in Andrews and Pollard (1994). □

**References**


Doukhan, P., Massart, P. and Rio, E., 1995, Invariance principles for absolutely regular


Granger, C. W. J. and Hallman, J., 1991, Nonlinear transformations of integrated time


### Table 1

**Persistence measures for Stock indices. Conditional Mean**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.18</td>
<td>0.70</td>
<td>1.18</td>
<td>1.19</td>
<td>0.63</td>
<td>1.02</td>
</tr>
<tr>
<td>2</td>
<td>0.83 [1.38]</td>
<td>0.78 [1.37]</td>
<td>0.62 [1.35]</td>
<td>1.24 [1.31]</td>
<td>0.55 [1.40]</td>
<td>0.73 [1.44]</td>
</tr>
<tr>
<td>3</td>
<td>0.69 [1.38]</td>
<td>0.60 [1.34]</td>
<td>0.58 [1.34]</td>
<td>0.88 [1.37]</td>
<td>0.48 [1.33]</td>
<td>0.73 [1.39]</td>
</tr>
<tr>
<td>4</td>
<td>0.60 [1.33]</td>
<td>0.72 [1.35]</td>
<td>1.46* [1.33]</td>
<td>1.06 [1.37]</td>
<td>0.74 [1.42]</td>
<td>0.82 [1.29]</td>
</tr>
<tr>
<td>5</td>
<td>0.60 [1.34]</td>
<td>0.98 [1.32]</td>
<td>0.94 [1.42]</td>
<td>0.70 [1.33]</td>
<td>0.83 [1.30]</td>
<td>0.77 [1.38]</td>
</tr>
<tr>
<td>6</td>
<td>1.26 [1.29]</td>
<td>1.25 [1.27]</td>
<td>1.09 [1.38]</td>
<td>1.18 [1.30]</td>
<td>0.74 [1.33]</td>
<td>0.79 [1.32]</td>
</tr>
<tr>
<td>7</td>
<td>0.68 [1.30]</td>
<td>0.90 [1.30]</td>
<td>0.85 [1.32]</td>
<td>0.58 [1.32]</td>
<td>0.81 [1.33]</td>
<td>0.88 [1.39]</td>
</tr>
<tr>
<td>8</td>
<td>0.60 [1.27]</td>
<td>0.61 [1.23]</td>
<td>0.78 [1.39]</td>
<td>0.50 [1.35]</td>
<td>1.58* [1.34]</td>
<td>0.79 [1.32]</td>
</tr>
<tr>
<td>9</td>
<td>0.73 [1.36]</td>
<td>0.79 [1.39]</td>
<td>0.76 [1.41]</td>
<td>0.53 [1.35]</td>
<td>1.46* [1.34]</td>
<td>0.90 [1.27]</td>
</tr>
<tr>
<td>10</td>
<td>0.51 [1.29]</td>
<td>0.65 [1.35]</td>
<td>1.00 [1.35]</td>
<td>0.70 [1.35]</td>
<td>0.78 [1.29]</td>
<td>0.59 [1.33]</td>
</tr>
<tr>
<td>20</td>
<td>1.11 [1.32]</td>
<td>1.20 [1.32]</td>
<td>0.66 [1.32]</td>
<td>0.58 [1.34]</td>
<td>0.47 [1.37]</td>
<td>1.16 [1.35]</td>
</tr>
<tr>
<td>30</td>
<td>1.41* [1.36]</td>
<td>1.19 [1.33]</td>
<td>0.88 [1.31]</td>
<td>0.72 [1.42]</td>
<td>0.98 [1.31]</td>
<td>0.67 [1.32]</td>
</tr>
<tr>
<td>40</td>
<td>0.72 [1.32]</td>
<td>0.68 [1.38]</td>
<td>0.85 [1.37]</td>
<td>0.64 [1.37]</td>
<td>0.49 [1.41]</td>
<td>0.78 [1.35]</td>
</tr>
<tr>
<td>50</td>
<td>0.79 [1.42]</td>
<td>0.96 [1.36]</td>
<td>0.93 [1.32]</td>
<td>0.79 [1.36]</td>
<td>1.16 [1.31]</td>
<td>1.10 [1.36]</td>
</tr>
<tr>
<td>60</td>
<td>0.83 [1.37]</td>
<td>0.72 [1.35]</td>
<td>1.05 [1.34]</td>
<td>1.02 [1.32]</td>
<td>0.72 [1.38]</td>
<td>1.08 [1.39]</td>
</tr>
<tr>
<td>70</td>
<td>0.71 [1.40]</td>
<td>0.62 [1.42]</td>
<td>0.47 [1.36]</td>
<td>0.98 [1.31]</td>
<td>0.66 [1.35]</td>
<td>0.80 [1.40]</td>
</tr>
<tr>
<td>80</td>
<td>0.80 [1.38]</td>
<td>0.77 [1.37]</td>
<td>0.65 [1.44]</td>
<td>0.70 [1.40]</td>
<td>0.98 [1.33]</td>
<td>0.95 [1.43]</td>
</tr>
<tr>
<td>90</td>
<td>1.08 [1.37]</td>
<td>1.14 [1.40]</td>
<td>1.71* [1.45]</td>
<td>1.71* [1.41]</td>
<td>0.71 [1.35]</td>
<td>0.91 [1.39]</td>
</tr>
<tr>
<td>100</td>
<td>0.75 [1.44]</td>
<td>0.53 [1.43]</td>
<td>0.75 [1.48]</td>
<td>0.85 [1.43]</td>
<td>1.02 [1.36]</td>
<td>1.01 [1.32]</td>
</tr>
</tbody>
</table>

Note:* Significantly different from zero at the 5% level (bootstrap test)

In brackets the 95% bootstrap critical value
Figure 1. Conditional variance persistence for DAX. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.

Figure 2. Conditional variance persistence for FTSE. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.
Figure 3. Conditional variance persistence for HANG. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.

Figure 4. Conditional variance persistence for NIKKEY. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.
Figure 5. Conditional variance persistence for DJ. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.

Figure 6. Conditional variance persistence for SP500. Nonparametric IPAF PLOT (upper) with 95% wild bootstrap critical values (bottom line, bootstrap replications=300). Mean parametric IPAF PLOT (lower) with 95% empirical quantiles from parametric IPAFs (10000 replications). Gaussian GARCH(1,1) model.
Figure 7. Nonlinear IPAF PLOT (wild bootstrap critical values, bootstrap replications=300). Conditional Variance for differences.
Figure 8. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. DAX.

Figure 9. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. DAX.
Figure 10. IPAF PLOTs (subsampling bands, $b=164$) bull vs bear markets. FTSE.

Figure 11. IPAF PLOTs (subsampling bands, $b=164$) bull vs bear markets. FTSE.
Figure 12. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. HANG.

Figure 13. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. HANG.
Figure 14. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets.

NIKKEY.

Figure 15. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets.

NIKKEY.
Figure 16. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. DJ.

Figure 17. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets. DJ.
Figure 18. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets S&P500.

Figure 19. IPAF PLOTs (subsampling bands, b=164) bull vs bear markets S&P500.