AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2012. **5**: 162-221

Notes on Nonlinear Dynamics

Racicot, François-Éric

RECEIVED: 12 SEPTEMBER 2012
ACCEPTED: 24 OCTOBER 2012

Abstract

We present a selective survey of modern nonlinear modeling techniques relevant to the field of applied financial econometrics. We first established the usefulness of nonlinear modeling of financial time series and its relevance for forecasting by means of Sims's (1984) definition. Then, we describe specific univariate and multivariate nonlinear models that can be classified either as stochastic or as deterministically chaotic. We also provide several novel numerical applications of these models along with their estimation techniques and tests. We conclude this literature review by presenting an application which compares the UHF-GARCH model with the parsimonious model-free realized volatility approach. Additionally, we present an extension to the multivariate case, referred as the realized covariance. This model-free measure of dependence might be useful in order to evaluate the *volatility feedback,* which is an alternative explanation to the leverage effect theory.

Keywords:

Nonlinear models; BDS test; Chaos; UHF-GARCH models; Realized volatility; Realized correlation; MGARCH; Leverage effect; Volatility feedback; Markov switching regime model; VIX.

JEL classification:

C1; G11; G17.

Racicot, F.E. ✉ *Telfer School of Management, Member of the CGA-Canada Accounting and Governance Research Centre (CGA-AGRC), University of Ottawa, Canada.*
Address: 55 Laurier Avenue East, Ottawa Ontario, K1N 6N5, Canada.
✆ 613-562-5800 (4757).
E-mail: Racicot@telfer.uottawa.ca

AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2012. **5**: 162-221

Notas sobre **dinámica no lineal**

Racicot, François-Éric

Resumen

Este artículo consiste en un estudio selectivo de las modernas técnicas de modelización no lineal de interés en el campo de la econometría financiera aplicada. En primer lugar, se pone de manifiesto la utilidad de la modelización no lineal de series temporales de carácter financiero y su importancia en las labores predictivas por medio de la definición de Sims (1984). A continuación, se describen los modelos específicos no lineales, tanto univariantes como multivariantes, que pueden ser clasificados ya sea como estocásticos o deterministas caóticos. También se presenta una serie de nuevas aplicaciones numéricas de estos modelos junto con sus técnicas de estimación y contraste. Esta revisión de la literatura concluye con la presentación de una aplicación que compara el modelo UHF-GARCH con la metodología parsimoniosa "model-free" de volatilidad realizada. Además, se presenta una extensión al caso multivariante, a la que nos referimos como covarianza realizada. Esta medida "model-free" de la dependencia podría resultar de gran utilidad a la hora de evaluar la *retroalimentación de la volatilidad,* que es una explicación alternativa a la teoría del efecto apalancamiento.

Palabras clave:

Modelos no lineales, Test BDS test; Caos, Modelos UHF-GARCH, Volatilidad realizada, Correlación realizada, MGARCH; Efecto apalancamiento, Retroalimentación de la volatilidad, Modelos con régimen cambiante de Markov, VIX.

1. Introduction[1]

Since the seminal paper of Bachelier (1900), there has been a considerable development in nonlinear modeling of financial assets. The fact that most financial models rests on the Martingale hypothesis, including the empirical facts that financial time series are not normally distributed (Mandelbrot, 1963), triggered the developments of a plethora of nonlinear modeling techniques. This is the main reason behind the ARCH process developed by Engle (1982) and its basic extension, the GARCH model of Bollerslev (1986).

These models assume that one of the key ingredients of modern econometric models of asset pricing is a sharp focus on the difference between conditional and unconditional moments. Conditional mean forecasts, which use recent information, are known to be more efficient than the unconditional ones. Similarly, the ARCH model rests on the presumption that forecasts of variance can also be improved by using recent information, at some point in the future. In particular, volatility clustering implies that big surprises of either sign will increase the probability of future volatility. Forecasts of volatility that recognize this fact will likely be more accurate than those that do not. Since all modern theories of asset relate the first moment (risk premia) to the second one (a measure of risk), forecasted volatilities are not only indispensable ingredients for asset pricing theories, but also for strategies of portfolio management.

Black (1976) observed that the distribution of stock returns was leptokurtic. He also noticed a negative correlation between current returns and future volatility. Black (1976) and Christie (1982) suggested a plausible economic explanation called the *leverage effect,* which was later modeled by Nelson (1990). According to the leverage effect, a reduction in equity value would raise the debt-to-equity ratio (leverage); hence, raising the riskiness of the firm as manifested by an increase in future volatility. As a result, the future volatility will be negatively related to current return on the stock. The linear GARCH (p, q) specification is potent for modeling returns volatility clustering. However, due to the fact that the conditional variance is linked to past conditional variances and squared innovations, it is not able to capture this kind of dynamic pattern. These considerations triggered the development of Nelson's (1990) EGARCH (p, q) model. In this specification, the volatility depends not only on the magnitude of the past surprises in return, but also on their corresponding signs.

Recent literature (Figlewski and Wang, 2000; Bollerslev *et al.*, 2006; Bollerslev *et al.*, 2009; Sun and Wu, 2010; Manda, 2010; and Russi, 2012) provides an alternative explanation of the existence of the negative correlation between implied volatility and

[1] *See Racicot (2000).*

stock returns. According to Figlewski and Wang (2000), there is an alternative explanation comparable to the leverage theory, referred as the *volatility feedback*. In this theory, the causality between stock returns and volatility is now reversed, meaning that changes in expected volatility alter stock market prices.

In order to quantify the feedback relationship between expected volatility and stock returns, some recent studies used UHF (ultra high frequency) data (e.g. Russi, 2012). These studies generalized previous research that was developed in a univariate setting to model volatility using UHF data. For instance, Engle (2000) developed a UHF-GARCH model and Bollerslev and Wright (2001) proposed the concept of integrated volatility that was eventually called the *realized volatility*[2]. This concept can be generalized to a bivariate setting, which is known as the *realized correlation*. It is computed by means of *realized Kernels* (Barndorff-Neilsen *et al.*, 2008b; Gatheral and Oomen, 2010). In this paper, we consider the leverage effect theory in its original treatment (Black, 1976; Christie, 1982; Nelson, 1990), i.e. the univariate case of the simple realized volatility in comparison with the UHF-GARCH. Therefore, we leave the subject of the bivariate case to future research.

This article is organized as follows. Section 2 is concerned with the usefulness of nonlinear models in empirical finance. In section 3, we discuss various classical univariate and multivariate models of volatility, their estimation process, stationarity tests and empirical applications. Section 4 considers nonlinear stochastic models of the means. Section 5 deals also with the mean nonlinear models, but particularly with the deterministic ones. Section 6 is devoted to tests of nonlinearity. In section 7, we propose ways to use the models presented in the previous sections for the purpose of forecasting. Section 8 also presents forecasting methods with a specialized application based on irregularly spaced high frequency data. Finally, section 9 concludes.

2. On Establishing the Usefulness of Nonlinear Models in Empirical Finance

It is widely agreed that many time series of asset returns, while approximately uncorrelated, are not temporally independent (Mandelbrot, 1963)[3]. The dependence arises through persistence in conditional variance or perhaps in other conditional moments, as well. A number of recent theoretical developments are beginning to show that, in a general-equilibrium context, economic theory cannot discard the possibility of nonlinear dependence in the conditional mean as well as dependence in higher-order conditional moments in asset returns.

[2] *For another application using high frequency data, see Bollerslev et al. (2007).*

[3] *For more on the subject, see Haug (2007).*

Before developing these subjects, we present some definitions and concepts that will be used in the following discussion. Sims (1984) shows that general-equilibrium asset-pricing models imply martingale asset-price behavior only at arbitrarily short horizon.

2.1. Definition (Sims 1984). A process $\{P_t\}$ is instantaneously unpredictable if and only if, as $v \rightarrow 0$

$$E_t[(P_{t+v} - E_t[(P_{t+v}])^2] / E_t[(P_{t+v} - P_t)^2] \rightarrow 1, \text{ a.s.}^4 \quad (1)$$

E_t is the mathematical expectation conditional on the information set I_t. I_t includes past and present information on P_t and other related variables. For an instantaneously unpredictable process, prediction error is the dominant component of changes over small intervals. For example, if $\{P_t\}$ is a martingale[5], which is defined as

$$E_t[P_{t+v}] = P_t \text{ for all } v > 0 \ , \quad (2)$$

and $\{P_t\}$ has finite second moments, the ratio given by equation (1) is exactly 1.

Sims observes that, under (1), regressions of $P_{t+v} - P_t$ on any variable in I_t have an $R^2 \rightarrow 0$ as $v \rightarrow 0$. Under (2), $R^2 = 0$. He also shows that (1) does not rule out predictability of first moments over longer time periods. For instance, a period from one day to a week is considered as a short one for Sims. Moreover, (1) does not rule out predictability of higher order moments, such as conditional variance, even in short time periods. Before the days of nonlinear dynamics, many financial time series were believed to follow a random walk. This means that no linear dependence can be found (no autocorrelation). Now, we know that the lack of linear dependence does not exclude nonlinear dependence, which if present would contradict the random walk model. Two reasons explain why some financial time series, like stock returns[6], should deviate from the random walk model. Firstly, the variance of stock returns changes over time. This phenomenon was observed by Mandelbrot (1963), who noted that although stock returns appeared uncorrelated, large changes tended to be followed by large changes and small changes by small changes (this is called *volatility clustering*). This fact has led to the development of ARCH and GARCH models. These models attempt to capture the changing variance in time series (detailed description of these models is presented

[4] *This stands for almost sure convergence (a.s.). Sequences that converge almost surely can be manipulated almost in the same ways as non-random ones. The interest typically centers on averages such as* $b_n(\omega) = \sum_{t=1}^{n} Z_t / n$ *. We write that* $b_n \rightarrow b$ *a.s. if and only if* $P[\omega: lim_{n \rightarrow \infty} b_n(\omega) = b] = 1$*, where* ω *represents the entire random sequence* $\{Z_t\}$*. Other modes of convergence are also used in the literature. These are : converge in* r^{th} *mean* (m)*, convergence in probability* (p) *and convergence in distribution* (d) *(logical relationships among the four modes of convergence :* $m \Rightarrow p \Rightarrow d$ *and* $a.s. \Rightarrow p$ *). They are defined as follows.* $lim_{n \rightarrow \infty} E(|b_n(\omega) - b|^r) = 0$*, for some* $r > 0$*,* $lim_{n \rightarrow \infty} E[\omega : /b_n(\omega) - b/ < \varepsilon] = 1$*. Finally, a sequence* b_n *converges to* b *in distribution if the distribution function* F_n *of* b_n *converges to the distribution function* F *of* b*, at every continuity point of* F*. See Amemiya (1985, 1994), Davidson and MacKinnon (1993) and White (2001).*

[5] *For example, we can get a martingale by taking the conditional expectation of an AR (1) process containing a unit root, known as a random walk (i.e.* $p_t = \alpha p_{t-1} + u_t$ *where* $\alpha = 1$ *and* $u_t \sim WN(0, \sigma^2)$: $\mathrm{E}_{t-1}(p_t) = p_{t-1}$*) .*

[6] *Stock return* i *over a time period* $(t, t+h)$ *is defined as* $r_{i,t,t+h} = p_{i,t,t+h} = [(P_{i,t+h} + D_{i,t+h}) / P_{i,t}] - 1$ *, where* P_i *is the price of stock* i *and* D_i *is the dividend of stock* i*. An approximation of this formula used in most empirical work (which we will use in our work) is* $r_{i,t,t+h} = ln(P_{i,t+h} + D_{i,t+h}) - ln\ P_{i,t}$ *.*

below). Secondly, there are several calendar anomalies[7]. The returns differ by small amounts during different periods. It is appropriate now to give a definition of linearity because it will be used extensively in the following discussion.

2.2. Definition (Priestley 1981). A stationary process $\{P_t\}$ is a linear process if it has a Wold representation[8] like $P_t=A(L)u_t$ where u_t is required to be i.i.d.

The i.i.d. condition in the above definition plays a central role. It implies that the best MSE predictor is a linear predictor using past information. It means that the past contains no information on the future; therefore, the best predictor is simply the unconditional mean. The definition rules out prediction made of nonlinear combination of past information. For example, assuming a process $h_t=a_t+\beta a_{t-1}a_{t-2}$, where $a_t\sim WN(0,\sigma^2)$, the unconditional expectation is $E(h_{t+1})=0$ and the autocovariance is $E(h_t h_{t-k})=0$. However, the conditional expectation is $E(h_{t+1}|h_t,h_{t-1},\ldots)=\beta a_t a_{t-1}$, which is the MMSE forecast (best MSE) of a future observation. This definition of linearity is also called P-linearity. Clearly, if P-linearity is rejected, then nonlinearity prevails (deterministic or stochastic nonlinearity). The BDS test presented below is a good way to test P-linearity. Another convenient definition of linearity (Lee *et al.*, 1993)[9] states that a process $\{P_t\}$ is linear in mean conditional to I_t if

$$Pr[E(P_t|I_t)]=I_t'\,\theta^*]=1 \text{ for some } \theta^*\in\Re^k.$$

A process exhibiting ARCH (ARCH process is presented below) may nevertheless exhibit linearity of this sort, because ARCH does not refer to the conditional mean. This definition is appropriate whenever one is concerned with the adequacy of linear models for forecasting. Alternatively, $\{P_t\}$ may not be linear in mean conditional to I_t, so

$$Pr[E(P_t|I_t)]=I_t'\,\theta]<1 \text{ for all } \theta^*\in\Re^k.$$

When the alternative is true, a linear model is affected by *neglected nonlinearity*. Tests for linearity usually make the assumption of a model (e.g. an AR (p) model). Then, a test is performed on the residuals and if the null is rejected, the alternative model may provide forecasts superior to those from the linear model. However, the BDS, Bispectrum[10] and McLeod-Li tests do not require models that imply such forecasts (BDS and other tests are presented below).

---

[7] *For more information on the subject, see Racicot (2011).*

[8] *The Wold's decomposition (Wold, 1938), a fundamental theorem in time series analysis, states that every weakly stationary purely nondeterministic stochastic process $\{r_t\}$ can be represented as a linear combination (or linear filter) of a sequence of uncorrelated random variables. By stochastic process {rt} we imply a family of random variables $\{r_t, t\in T\}$ defined on a probability space (Ω,F,P). This linear filter representation ($MA(\infty)$) is given by $r_t=u_t+\psi_1 u_{t-1}+\psi_2 u_{t-2}+\ldots=\sum_{j=0}^{\infty}\psi_j u_{t-j}$, $\psi_0=1$. The sequence u_t is a white-noise process (Mills, 1990).*

[9] *I_t is a partition of a greater process (a set) $Z_t'=(P_t,I_t')$, where P_t is a scalar and I_t, is a $k\times 1$vector. I_t may (but not necessarily) contain a constant and lagged values of P_t.*

[10] *See Hinich (1982) and Ashley, Patterson and Hinich (1986).*

3. Variance-nonlinear Stochastic Models

One of the key ingredients of modern econometric models of asset pricing is a sharp focus on the difference between conditional and unconditional moments. The ability of time series analysis to forecast means well rests on the observation that forecasts conditional on recent information are more efficient than forecast that do not use this information. Similarly, the ARCH model rests on the presumption that forecast of variance at some point in the future can also be improved by using recent information. In particular, volatility clustering implies that big surprises of either sign will increase the probability of future volatility. Forecasts of volatility which recognize this fact will generally be more accurate than those which do not. Since all modern theories of asset relate first moments (risk premia) to second moments (measures of risk), forecasted volatilities are indispensable ingredients of asset pricing theories or strategies of portfolio management.

In the following discussion, we present the details of the ARCH model and its extensions. We also propose an application to financial practitioners.

Assuming that we have a time series of stock returns where the returns are defined as $r_t = ln(P_t/P_{t-1})$ and where P_t is the price of a stock at time t from a time series $\{P_t\}$ (see footnote 5). Then, the AutoRegressive Conditional Heteroskedasticity (ARCH(q) Model of Engle,1982[11]) is defined as

$$r_t = \sigma_t u_t ,$$
$$V(r_t | I_t) = \sigma_t^2 = \alpha + \Sigma_{i=1}^{q} \varphi_i r_{t-i}^2 = \alpha + \Phi(L) r_t^2 \qquad (3)$$

where u_t is i.i.d. (0,1). As an example, an ARCH(1) model is given by

$$\sigma_t^2 = \alpha + \phi\, r_{t-1}^2, \qquad (4)$$

Note that if $0<\phi<1$ then r_t has an unconditional stationary[12] distribution that is non-normal with variance $\alpha / 1-\phi$. An extension to an ARCH model that allows for change in the mean is ARCH in mean (ARCH-M) and is given by

[11] *For an introduction to ARCH models and their application to finance see Alexander (2001), Bollerslev (1992), Brooks (2008), Engle (1995), Gouriéroux (1992), Mills (1999), Pagan (1996), Rachev et al. (2007) and Tsay (2005).*

[12] *There are two types: the second-order (weak) stationarity and the strict one. A stochastic process $\{r_t, t \in Z\}$, with index set $Z=\{0,\pm1,\pm2,...\}$, is strictly stationary if the joint distributions of $\{r_{t_1}, r_{t_2},...,r_{t_m}\}$ and $\{r_{t_1+k}, r_{t_2+k},...,r_{t_m+k}\}$ are the same for all positive integer m and for all $t_1,t_2,...,t_m$, $k \in Z$ (Brockwell and Davis, 1991). Thus, to be strictly stationary the joint p.d.f. of any set of observations (discrete time series) must stay unchanged by shifting all the times of observations forward or backward by any integer amount k. The second-order (weak) stationarity requires that the first and second moments stay constant trough time: 1) $E(r_t)=c$, $\forall t \in Z$ (independent of t); 2) $E|r_t^2|<\infty$, $\forall t \in Z$; 3) $\gamma(h,s)=\gamma(h+t,s+t)$ for all $h, s, t \in Z$ (also independent of time t, but depend of the distance in time). The second order stationarity and the assumption of normality are sufficient to produce strict stationarity. Any stationary process can be inverted in a (convergent) infinite MA, which can be well approximated by a low order ARMA process. Stationary is also required if one wants to do regression using time series. Spurious regression will result if one does not respect this condition. One test of spurious regression is large R^2, large t-test and very low Durbin-Watson coefficient. To be sure that the time series at hand are stationary, one must do some unit root testing before using the data. See Enders (2004), Gouriéroux (1990, 1992), Hamilton (1994) and Mills (1990).*

$$r_t = \theta_0 + \theta_1 \sqrt{\sigma_t^2} + \varepsilon_t$$
$$\mu_t = E(r_t | I_t) = \theta_0 + \theta_1 \sqrt{\sigma_t^2} \quad (5)$$

where ε_t is an ARCH(q) with $\varepsilon_t | I_t \sim N(0, \sigma_t^2)$.

Note that this model is nonlinear in the mean and the variance. A more complex form of this model may be found in the literature. The extension of the ARCH model that allows for lags in the conditional variance was first presented by Bollerslev (1986). This model, called the *Generalized AutoRegressive Conditional Heteroskedasticity* GARCH (p, q), is written as

$$\sigma_t^2 = \alpha + \Sigma_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \Sigma_{i=1}^{q} \phi_i r_{t-i}^2 \,. \quad (6)$$

As an example, a simple GARCH (1,1) model is given by

$$\sigma_t^2 = \alpha + \phi r_{t-1}^2 + \beta \sigma_{t-1}^2 . \quad (7)$$

Note that if $0 < \beta + \phi < 1$ then p_t has an unconditional stationary distribution that is non normal with variance $\alpha / [1 - \beta - \phi]$. As showed above, the GARCH model can be extended to the case where the mean is no longer assumed constant: GARCH-M and is defined as

$$\sigma_t^2 = \alpha + \Sigma_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \Sigma_{i=1}^{q} \phi_i r_{t-i}^2 \,,$$
$$\mu_t = E(r_t | I_t) = \theta_0 + \theta_1 \sqrt{\sigma_t^2} \quad (8)$$

where the process of r_t is given by equation (5). More complex forms of this model may be found in the literature. Note that if in GARCH (1, 1), for example, the parameter $\beta + \phi = 1$, then the resulting model is called the IGARCH (1, 1). The integrated GARCH model is strictly stationary, but not generally covariance stationary (see ref. 18). All the ARCH-GARCH models presented above are linear in the second moment and univariate. Nonlinear and multivariate forms of these models also exist. In the GARCH (p, q) model (6), the variance depends only on the magnitude and not on the sign of p_t. This is not consistent with the empirical findings that stock market prices are subject to the leverage effects[13]. The Exponential GARCH (p, q) developed by Nelson (1990), also known as EGARCH

[13] *Black (1976) observed that the distribution of stock returns was leptokurtic. He also noted a negative correlation between current returns and future volatility. Black (1976) and Christie (1982) suggested a plausible economic explanation, known as the leverage effect. According to the leverage effect, a reduction in equity value would raise the debt-to-equity ratio (leverage); hence, raising the riskiness of the firm as manifested by an increase in future volatility. As a result, the future volatility will be negatively related to current returns on the stock. The linear GARCH(p, q) model is not able to capture this kind of dynamic pattern, because the conditional variance is only linked to past conditional variances and squared innovations. The development of the EGARCH(p, q) model by Nelson (1990) have been motivated in light of these considerations. In this model, the volatility depends not only on the magnitude of the past surprises in returns, but also on their corresponding signs (Nelson, 1990; Bollerslev, 1992 and Pagan, 1996). As mentioned in our introduction, there are new developments regarding the leverage effect theory. An alternative explanation referred to the "volatility feedback" argues that the causality between volatility and stocks returns is reversed, which means that the expected volatility alters stock market prices. For more information on this subject, see Bollerslev et al. (2006), Manda (2010), Sun and Wu (2010).*

(p, q), seems to have a superior fit on the data compared with GARCH (p, q). The EGARCH (p, q) model, which is a nonlinear GARCH (p, q) model, is given by

$$ln\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_1(\phi u_{t-i} + \theta_1(|u_{t-i}| - E|u_{t-i}|) + \sum_{i=1}^{p} \gamma_i \, ln\sigma_{t-i}^{\ 2} \ . \tag{9}$$

Unlike the linear GARCH (p, q) model in (6), β_i (here γ_i) and ϕ_i are not restricted. This is to ensure non-negativity of the conditional variances. Note that (9) looks like an unrestricted ARMA (p, q) model for log σ_t^2. If $\phi_i\theta<0$, the variance tends to rise (fall) when p_t is negative (positive) in accordance with empirical evidence for stock market returns. If u_t is assumed to be i.i.d. normal, then p_t is a covariance stationary conditional that all the roots of the autoregressive polynomial $\beta(\lambda)$ lie outside the unit circle. The EGARCH model is related to the Multiplicative ARCH model developed by Milhøj (1987) that is defined as

$$log\ \sigma_t^2 = \alpha + \sum_{i=1}^{q} \theta \, log\ u_{t-i}^2 + \sum_{i=1}^{p} \beta_i \, (log\ u_{t-i}^2 - log\ \sigma_{t-i}^2). \tag{10}$$

Many other ARCH formulations are proposed in the literature (e.g. Pagan, 1996). We present the Hentschel's (1995) model. Hentschel applied a transformation similar to the Box-Cox on a generalization of the absolute GARCH model. The resulting model is

$$\frac{\sigma_t^{\lambda} - 1}{\lambda} = \alpha_0 + \beta \left(\frac{\sigma_{t-1}^{\lambda} - 1}{\lambda} \right) + \alpha \sigma_{t-1}{}^{\lambda} \, (f(u_t))^{v}, \tag{11}$$

where

$$f(u_t) = |u_t - b| - c(u_t - b).$$

This model encompasses most of the models presented previously. For example, when $\lambda = v = 2$ and $b = c = 0$, we have the standard GARCH model. The absolute value GARCH model sets $\lambda = v = 1$ with b and c free. Finally, the exponential GARCH or EGARCH model of Nelson (1990) is obtained by setting $\lambda = 0$, $v = 1$ and $b = 0$ to get

$$log\ (\sigma_t) = \alpha_0 + \beta \, log(\sigma_{t-1}) + \alpha(|u_t| - cu_t). \tag{12}$$

This model is appealing because it does not require any parameter restrictions to ensure that conditional variance of returns are always positive.

The ANN-GARCH Model

In this section, we discuss an extension of the artificial neural network model (ANN). Following Donaldson and Kamstra (1997), we present an extension of the standard GARCH model that includes an ANN model to further capture the nonlinearities in the financial time-series.

This can be described as follows. We first assume an AR (1) process for the returns of the financial time series

$$r_t = \alpha_1 + \alpha_2 r_{t-1} + e_t \tag{13}$$

where $r_t = ln(P_t|P_{t-1})$, $e_t \sim (0, \sigma_t^2)$ and,

$$\sigma_t^2 = c + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j e_{t-j}^2 + \sum_{k=1}^{T} \varphi_k D_{t-k} e_{t-k}^2 + \sum_{h=1}^{S} \varepsilon_h \Psi(z_t \lambda_h) \ . \tag{14}$$

Equation (14) is the ANN-GARCH model. We can see that the first two components forms a standard GARCH (p, q) model. Adding the third component, we get the sign-ARCH model developed by Glosten, Jaganathan and Runkle's (1993), called the GJR model. The last term is the ANN. More precisely,

$$D_{t-k} = \begin{cases} 1 & \text{if } e_{t-k} < 0 \\ 0 & \text{if } e_{t-k} \leq 0 \end{cases} \tag{15}$$

$$\Psi(z_t \lambda_h) = \left[1 + \exp\left(\lambda_{h,0,0} + \sum_{d=1}^{v} \left[\sum_{w=1}^{m} \left(\lambda_{h,d,w} z_{t-d}^{w} \right) \right] \right) \right]^{-1} \tag{16}$$

$$z_{t-d} = \left[e_{t-d} - E(e) \right] / \sqrt{E(e^2)} \tag{17}$$

$$1/2\lambda_{h,d,w} \sim U(-1, +1) \tag{18}$$

Equation (15) is part of GJR model and it is simply a dummy variable. Equation (16) describes the logistic ANN nodes. This equation is called the transfer function. A popular model for the transfer function is the nonlinear logistic function[14]

$$y = a + \left((1 + \exp\left[-(c + bx) \right] \right)^{-1} .$$

Equation (17) shows how the data must be transformed when $E(e)$ and $E(e^2)$ are the mean and variance of the innovations. To estimate the parameters α, β, γ, φ and ε in equation (14), a value of λ, in equation (18), is chosen using a uniform random number generator allowed to vary between –1 and +1 and then, estimation of the parameters is done by maximum likelihood. The ANN-GARCH is considered as a seminonparametric model, because we have to select values for the λ's that are the scaling factors used to identifies the ε's.

Application

The ANN-GARCH was applied to four financial time series, namely S&P500, NiKKEI, FTSE, and TSEC. Donaldson and Kamstra (1997) used an AR (1)-ANN (1)-GARCH

[14] For an introduction to artificial neural network, see Campbell et al. (1997), Franses and van Dijk (2000) and Alexander (2001).

(1,1), and estimate the parameters α_1, α_2, c, β_1, γ_1, φ_1 and ε_1. They report that in comparison with other model like the standard GARCH, the EGARCH, the GJR model and the ANN-GARCH seems to be the best performer. In particular, this model is able to capture both the symmetric and the asymmetric vo latility effects not captured by the standard ARCH-type models for most of the financial time series.

The GARCH option pricing model

We consider the ARCH models as DGP's[15] of a particular financial time series exhibiting certain characteristics, which is mainly the volatility clustering. However, these models have been also generalized to the pricing of options (Duan, 1995, 1996 and Duan *et al.*, 1999)[16]. In this section, we show that the standard GARCH (p, q) can be extended for option pricing models.

Supposing that P_t is the price of an asset observed at discrete time t. Transforming this variable into return and assuming that it is conditionally lognormally distributed, we obtain

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right) = r + \lambda\sqrt{\sigma_t^2} - \frac{1}{2}\sigma_t^2 + e_t \tag{19}$$

where $e_t|I_{t-1} \sim N(0,\sigma_t^2)$; r is the risk free rate and λ, the risk premium. We assume that

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i e_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 \tag{20}$$

is a standard GARCH model. To ensure that the Black and Scholes (1973) model is a special case of our model, assume that p=0 and q=0, (19) and (20) are reduced to a standard homoskedastic lognormal process. To obtain the GARCH option-pricing model, the conventional risk-neutral valuation has to be generalized for heteroskedasticity of the asset return process. This generalization is called the locally risk-neutral valuation relationship. It implies, under a pricing measure denoted by (*) that

$$r_t = r - \frac{1}{2}\sigma_t^2 + \varepsilon_t \tag{21}$$

where $\varepsilon_t|I_{t-1} \sim N(0,\sigma_t^2)$ and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i\left(\varepsilon_{t-i} - \lambda\sqrt{\sigma_{t-i}^2}\right)^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2$. As a corollary, we get

$$P_T = P_t e^{\left[(T-t)r - \frac{1}{2}\sum_{s=t+1}^{T}\sigma_s^2 + \sum_{s=t+1}^{T}\varepsilon_s\right]}. \tag{22}$$

[15] *DGP stands for Data Generating Process.*

[16] *Another model that uses a continuous GARCH process to price European options is the Heston and Nandi (2000) stochastic volatility model (Heston, 1993). It proposed an analytical formula for the pricing of options. For an application of this model and a VBA code to estimate a GARCH (1,1) process in Excel, see Rouah and Vainberg (2007).*

Discounting the asset price using the risk free interest rate, we obtain the martingale property. Then, assuming a GARCH (p, q) process, a European call option with exercise price X and maturity T has a value equal to

$$C_t^{GARCH} = e^{-(T-t)r} E^* \left[Max(P_t - X, 0) | I_t \right] \tag{23}$$

where E*[.] is the expectation computed in a risk neutral world, conditional on information set $I_t : (e_t, \ldots, e_{t-q+1}, \sigma_t^2, \ldots, \sigma_{t-p+1}^2, X)$. It should be noted that there is no analytic solution for equation (23). This is due to the fact that the conditional distribution for more than one period cannot be derived analytically. To solve that problem, we can use Monte Carlo simulation to compute a value of (23).

The Delta of (23), i.e. a measure of the sensitivity of the call premium to the underlying asset, is given by

$$\Delta_t^{GARCH} = e^{-(T-t)r} E^* \left[\frac{P_T}{P_t} 1_{(P_T \geq X)} | I_t \right] \tag{24}$$

where $1_{(P_T \geq X)}$ is an indicator function taking the value 1 if $P_t \geq X$ and 0 otherwise. As for equation (23), this measure is also computed by Monte Carlo simulation. The European put GARCH option price can be derived by using the put-call parity relationship.

The Black and Scholes model may be considered as a special case of the GARCH process. More precisely, for p=0 and q=0, we obtain the homoskedastic lognormal process, namely the Black and Scholes model.

Finally, the GARCH option-pricing model may be used for extracting the implied volatility instead of using the variance of the underlying asset return. The *smile* is then obtained, for the GARCH option price, by plotting its implied volatility as a function of its strike price (X)[17].

Modelling correlation

The GARCH model can be generalized for modelling the conditional covariance[18]. Then we can build a covariance matrix, where each element is assumed to have the following process

$$\sigma_{i,j,t} = \omega_{i,j} + \alpha r_{i,j,t-1} r_{ij,t-1} + \beta \sigma_{i,j,t-1} \tag{25}$$

where α and β are estimated by maximum likelihood. $\omega_{i,j}$ is the log-run covariance, $\sigma_{i,j,t-1} = 1/T \sum_{k=1}^{T} r_{i,t-k} \, r_{j,t-k}$ and $r_{i,t}$ is the return of asset i at time t.

[17] *For an introduction on this subject, see Hull (2012).*

[18] *See Bhansali (1998), Tsay (2005) and Hull (2012).*

To forecast the covariance at period $t+k$, we have to compute the conditional expectation based on information set I_t

$$\begin{aligned}\sigma_{i,j,t+k/t} &= E_t[r_{i,t+k}\, r_{j,t+k}] = E_t[\sigma_{i,j,t+k}] \\ &= \omega_{i,j} + (\alpha+\beta)\, E_t[\sigma_{i,j,t+k-1}]\end{aligned} \tag{26}$$

where $E_t\,[\omega_{i,j}]=\omega_{i,j}$ and $E_t\,[r_{i,\,t+k-1}\, r_{j,\,t+k-1}]=\sigma_{i,\,j,\,t+k-1}$. This implies by repeated substitution that

$$\begin{aligned}E_t[\sigma_{i,j,t+k}] &= \omega_{i,j} + (\alpha+\beta)\left(\omega_{i,j} + (\alpha+\beta)E_t[\sigma_{i,j,t+k-2}]\right) \\ &= \omega_{i,j} + (1+\alpha+\beta) + (\alpha+\beta)^2\, E_t[\sigma_{i,j,t+k-2}]\,.\end{aligned} \tag{27}$$

Solving $E_t\,[\sigma_{i,j,t+k-2}]$ by recursive substitutions and renaming variables, we get

$$E_t[\sigma_{i,j,t+k}] = \bar{\sigma}_{i,j,} + (\alpha+\beta)^{k-1}\left(\sigma_{i,j,t+1} - \bar{\sigma}_{i,j}\right) \tag{28}$$

where $\bar{\sigma}_{i,j,}$ is the mean value of the covariance between of asset i and j.

This model can be used to construct the term structure of correlation, which is determined in term of forward variances and covariances. Assuming that there is no serial correlation (no autocorrelation) in the returns, we can write the term structure of correlation as

$$R^{(k)}_{i,j,t} = \frac{\sum_{l=1}^{k} \sigma_{i,j,t+l/t}}{\sqrt{\sum_{l=1}^{k} \sigma_{i,t+l/t} \sum_{l=1}^{k} \sigma_{i,t+l/t}}} \tag{29}$$

where $\sum_{l=1}^{k} \sigma_{i,j,t+l/t} = E_t\left[\sum_{l=1}^{k}\left(r_{i,t+l}\, r_{j,t+l}\right)\right] = \sum_{l=1}^{k}\left(E_t\left(r_{i,t+l}\, r_{j,t+l}\right)\right)$ and $E_t\left(r_{i,t+l}\, r_{j,t+l}\right) = \sigma_{i,j,t+l/t}\,.$

We can observe the similarities between the GARCH model and the exponential smoothing model,

$$\sigma_{i,j,t} = (1-\lambda)\sigma_{i,j,t-1} + \lambda r_{i,t-1}\, r_{j,t-1}\,. \tag{30}$$

In this model, we must find the best λ that matches the financial time series. Since $\Delta\sigma_{i,j,t} = \lambda\left(r_{i,t-1}\, r_{j,t-1} - \sigma_{i,j,t-1}\right)$ and $E_t\,[\Delta\sigma_{i,j,t}] = \lambda\left(E_t\left(r_{i,t-1} r_{j,t-1} - \sigma_{i,j,t-1}\right)\right) = 0$, the exponential smoothing model has intrinsically no forecasting power.

Multivariate ARCH models (MGARCH)

Multivariate linear formulation exists for the ARCH models. Many issues in asset pricing and portfolio allocation decisions can be analyzed in a multivariate context. Let r_t be a $N\times 1$ vector stochastic process, then any process that permits the representation

$$r_t = u_t \Omega_t^{1/2} \tag{31}$$

where u_t is assumed to be i.i.d. with $E(u_t)=0$, $V(u_t)=I$, and Ω_t is a time-varying $N\times N$ positive definite covariance matrix (and measurable conditional to information I_t), which is referred as a multivariate linear ARCH model. Kraft and Engle (1983) defined Ω_t, in their multivariate linear ARCH model, as a linear function of the contemporaneous cross-products in the past squared errors: $vech(p_{t-1}p_{t-1}')$,..., $vech(p_{t-q}p_{t-q}')$; where $vech(.)$ is the operator that stacks the lower portion of an $N\times N$ matrix as an $(N(N+1)/2)\times 1$ vector. Bollerslev, Engle and Wooldridge (1988) generalized this model to a multivariate linear GARCH (p, q). Ω_t is transformed as follows

$$vech(\Omega_t)=W+\sum_{i=1}^{q}A_i vech(p_{t-1}p_{t-1}')+\sum_{i=1}^{p}B_i vech(\Omega_{t-i}) \tag{32}$$

where (Ω_t) is a $(N(N+1)/2)\times 1$ vector, and A_i and B_i are $(N(N+1)/2)\times(N(N+1)/2)$ matrices. Only three of these models are presented here[19], which are the Bollerslev's (1990) MGARCH[20], the BEKK model of Engle and Kroner (1995), and the Koutmos (1996) extensions of Bollerslev's model that includes a multivariate EGARCH for the innovations. These are presented in section 3.1.4.

3.1. Estimating and testing ARCH models

3.1.1. Testing for ARCH effect

Testing for the presence of ARCH effect in the error term of a model like

$$y_t=\beta' x_t+e_t, \tag{33}$$

is done by using the Lagrange multiplier (LM) test firstly proposed by Engle (1982). The procedure is as follows. Firstly, by applying ordinary least-squares on

$$y_t=\beta' x_t+e_t,$$

where x_t may include lagged variables: $x_t=[1 \;\; y_{t-1} \;\; y_{t-2} \ldots y_{t-p}]$, we get

$$\begin{aligned}\hat{e}_t &= y_t-\hat{\beta}' x_t\\ &= y_t-\hat{\beta}_0-\hat{\beta}_1 y_{t-1}-\hat{\beta}_2 y_{t-2}-\ldots-\hat{\beta}_{t-p}\, y_{t-p}.\end{aligned} \tag{34}$$

---

[19] When there are many time series to be jointly modeled, the number of parameters to be estimated becomes very large. Another model that takes into account this problem is the orthogonal GARCH developed by Alexander and Chibumba (1997). The implementation of the orthogonal GARCH model involves using principal component analysis and univarite GARCH methods. The orthogonal GARCH method has other qualities, namely the lack of dimensional restrictions and the fact that the matrices are positive definite. The problem of non positive definite matrices is often encountered in the MGARCH framework.

[20] It should be noted that Engle (2000) has developed a generalized version of that model called the dynamic conditional correlation (DCC) model. As we will see in section 3.1.4, the MGARCH of Bollerslev (1990) involves computing a conditional covariance matrix $\Omega_t=D_t\Gamma D_t$, where Γ is supposed to be a time invariant conditional correlation matrix. In his generalization, Engle (2000) lets Γ to be time dependent and denotes it Γ_t, which contains dynamic conditional correlations that are more in accordance with the stylized facts observed in the financial time series.

Secondly, we apply OLS on

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 \hat{e}_{t-1}^2 + \ldots + \alpha_q \hat{e}_{t-q}^2 , \tag{35}$$

This test is based on the resulting R^2. Knowing that

$$T \times R^2 \xrightarrow{d} \chi_q^2 , \tag{36}$$

Therefore, the test rejects the null hypothesis, which is $H0$: $\alpha_0 = \alpha_1 = \ldots = \alpha_q = 0$, if

$$T \times R^2 > \chi_q^2 ,$$

with probability of type I error of $\alpha = 5\%$, for example.

3.1.2. Estimating univariate ARCH models

Estimation of ARCH models is generally done by maximum likelihood (ML). Assuming that the disturbances of the following model

$$y_t = \beta' x_t + e_t , \tag{37}$$

are normally distributed

$$e_t \sim N(0, \sigma_t^2), \tag{38}$$

where σ_t^2 follows any of the ARCH models presented previously. If σ_t^2 follows an ARCH (1) model, the likelihood function is constructed by using the following steps.

i) The joint p.d.f. of the errors is written as

$$f(e_1, e_2, \ldots, e_T | \beta', \alpha_0, \alpha_1) = f(e_1 | \beta', \alpha_0, \alpha_1) \times f(e_2 | \beta', \alpha_0, \alpha_1) \times \ldots \times f(e_T | \beta', \alpha_0, \alpha_1)$$

$$= \Pi_{t=1}^{T} f(e_1 | \beta', \alpha_0, \alpha_1) , \tag{39}$$

where $f(e_t | \beta', \alpha_0, \alpha_1) = \dfrac{1}{\sqrt{2\pi\sigma_t^2}} \times \exp\left\{ -\dfrac{1}{2} \left(\dfrac{e_t - E(e_t)}{\sigma_t} \right)^2 \right\}$.

ii) By applying the Jacobian transformation $f(y) = (e) \times \left|\frac{\partial e}{\partial y}\right|$ where $\left|\frac{\partial e}{\partial y}\right| = \begin{vmatrix} \frac{\partial e_1}{\partial y_1} & \frac{\partial e_1}{\partial y_2} & \cdots & \frac{\partial e_1}{\partial y_T} \\ \vdots & & & \vdots \\ \frac{\partial e_T}{\partial y_1} & \frac{\partial e_T}{\partial y_2} & \cdots & \frac{\partial e_T}{\partial y_T} \end{vmatrix}$, we

we get the joint p.d.f. of the y_t's

$$f(y_1, y_2, \ldots, y_T) = \Pi_{t=1}^{T} f(y_t | \beta', \alpha_0, \alpha_1) . \tag{40}$$

With this result, the likelihood function is written as

$$L(\alpha_0, \alpha_1, \beta'|y_1, y_2, \ldots, y_T) = \prod_{t=1}^{T} f(y_t|\beta', \alpha_0, \alpha_1)$$

$$= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \times \exp\left\{-\frac{1}{2}\left(\frac{e_t - E(e_t)}{\sigma_t}\right)^2\right\}$$

$$=(2\pi)^{-T/2}\sum_{t=1}^{T}(\sigma_t^2)^{-1/2} \times \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{y_t - \beta' x_t}{\sigma_t}\right)^2\right). \qquad (41)$$

iii) Finally, in order to find the maximum of this function and to simplified the numerical calculations, the logarithm transformation is applied to the likelihood function. The result is

$$lnL = -T/2 ln(2\pi) - 1/2\ ln \sum_{t=1}^{T}\sigma_t^2 - 1/2\left(\sum_{t=1}^{T}\left(\frac{y_t - \beta' x_t}{\sigma_t}\right)^2\right). \qquad (42)$$

Application

We provide a numerical application of an ARCH model, namely the popular EGARCH model, using a sample of montly S&P500 data ranging from January 1982 to February 2012. A plot of these returns is provided at Figure 1.

Figure 1. Monthly SP&500 returns – January 1982 to February 2012

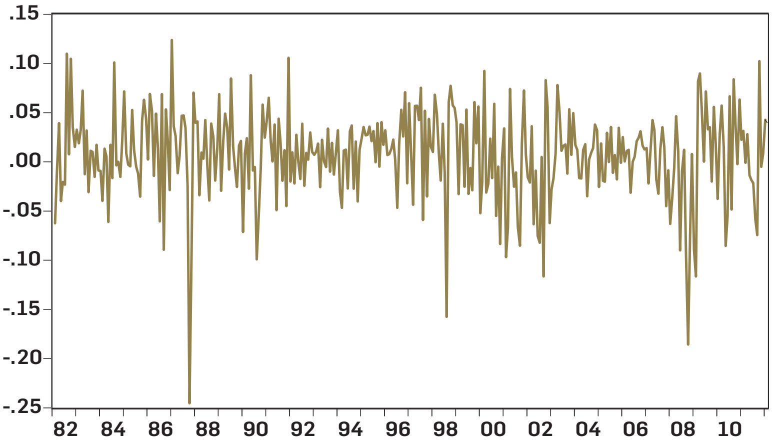


Figure 1 shows that there are evidences of changing variance so an ARCH-type model should perform well in modeling this time series. Table 1 provides the EViews estimation of the popular EGARCH process on the time series shown in Figure 1.

Table 1. EViews estimation an AR(1)-EGARCH model for the S&P500 returns (January 1995 to March 2012)

Dependent Variable: RSP500
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 04/18/12 Time: 10:11
Sample (adjusted): 1995:03 2012:03
Included observations: 205 after adjustments
Convergence achieved after 25 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)
*RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	*z*-Statistic	Prob.
AR(1)	0.148347	0.078698	1.885014	0.0594
Variance Equation				
C(2)	-1.277	0.572366	-2.23108	0.0257
C(3)	0.256524	0.138581	1.851075	0.0642
C(4)	-0.28038	0.095627	-2.93202	0.0034
C(5)	0.825574	0.08531	9.677368	0.0000
R-sq.	0.002522	Mean dependent var		0.0052
Adjusted R-sq	0.002522	S.D. dependent var		0.0466
S.E. of reg.	0.046512	Akaike info criterion		-3.438
Sum sq. resid	0.441334	Schwarz criterion		-3.357
Log likelihood	357.3583	Hannan-Quinn criter.		-3.405
DW stat	2.042729			
Inv. AR Roots	0.15			

As shown in this table, the coefficients of the EGARCH[21] process are all quite significant in terms of p-values. This implies that there would be some leverage effect in our sample. As it is well-known, the goodness-fit measures used to evaluate the overall performance of these models are the information criterions like the Akaike's one. The R^2 criterion is known to be unreliable in the context of time-series analysis.

Some authors consider modeling stock prices directly (e.g. Mun, 2006)[22]. We consider a similar approach, assuming an ARMA (1, 1), for the prices plus a mean reversion factor. Thus, the suggested model for the mean of the process is given by

$$P_t = \beta_0 + \beta_1 P_{t-1} + \beta_2 \varepsilon_{t-1} + \beta_3 (P_t - P_{t-1}) + \varepsilon_t .$$

[21] *EViews provides several other useful ARCH-type models.*

[22] *Stock prices can also be modeled as a function of dividends. For instance, Foerster and Sapp (2006) use the following model: $P_t - P_{t-1} = \alpha_0 + \alpha_1(D_{t-1} - D_{t-2}) + \alpha_2(D_{t-2} - D_{t-3}) + \varepsilon_t$ to explain the changes in stock prices. The model is used to determine how investors react to changes in dividends. A positive increase in the dividends might be considered by investors as a positive indication on the future value of the company. Therefore, its stock prices increase. Foerster and Sapp (2006) found that the model has good fit in general, but α_2 is not significant. Note that the specification of this model is similar to a standard model of dividends which is given by: $D_t - D_{t-1} = \beta_0 + \beta_1(D_{t-1} - D_{t-2}) + \beta_2(D_{t-2} - D_{t-3}) + \varepsilon_t$. This specification is designed to determine whether there is a persistence in the changes in dividends, where the persistence is captured by the parameters β_1 and β_2.*

Notes on Nonlinear Dynamics. *Racicot, F.E.*
AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2012. **5**: 162-221

We also assume that the residuals of this model follow an EGARCH (1,1) process. The EViews results appear at Table 2.

Table 2. EViews estimation an ARMA (1, 1)-EGARCH model with a mean reversion factor for the S&P500 prices January 1995 to March 2012

Dependent Variable: SP500
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 04/18/12 Time: 11:36
Sample (adjusted): 1995:03 2012:03
Included observat ions: 205 after adjustments
Convergence achieved after 49 iterations
MA Backcast : 1995:02
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(5) + C(6)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(7)
*RESID(-1)/@SQRT(GARCH(-1)) + C(8)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	*z*-Statistic	Prob.
C	1573.573	458.4589	3.43231	0.0006
AR(1)	0.987622	0.007647	129.1536	0.0000
MA(1)	0.989238	0.003379	292.7858	0.0000
D(SP500)	0.505848	0.004722	107.1205	0.0000
Variance Equation				
C(5)	0.248612	0.097187	2.558091	0.0105
C(6)	0.279824	0.137714	2.031927	0.0422
C(7)	-0.11442	0.069825	-1.63859	0.1013
C(8)	0.926551	0.023902	38.76513	0.0000
R-squared	0.98954	Mean dependent var		1115.2
Adjusted R-sq.	0.989384	S.D. dependent var		249.68
S.E. of reg.	25.7255	Akaike info criterion		9.0513
Sum sq. resid	133022.1	Schwarz criterion		9.181
Log likelihood	-919.762	Hannan-Quinn criter.		9.1038
DW stat	1.832442			
Inv. AR Roots	0.99			
Inv. MA Roots	-0.99			

As shown at Table 2, the results are good for most of the parameters of the EGARCH model. The mean model is significant in terms of p-values and the overall fit is also quite high as shown by the adjusted R^2 (and the DW is sufficiently close to 2). Thus, our conclusion is that this data set shows significant leverage effect. We arrived to a similar conclusion in the model presented at Table 1.

3.1.3. Testing for unit roots

Up to now, we have presented models that where stationary, because of the assumption that the process $\{p_t\}$ followed: $p_t = \sigma_t u_t$, where u_t was supposed to be a white noise. However, in this section, we consider a generalization of this model that allows for autocorrelations in the process $\{p_t\}$ and lags in the innovations. This means that the process follows an ARMA (*p, q*) process. Thus, the generalized model might be written as

$$\theta(L)p_t = c + \varphi(L)u_t \ , \tag{43}$$

where $u_t \sim WN(0,\sigma_t^2)$. That means that u_t might take the value of

$$u_t = \left(\sqrt{\gamma + \beta\sigma_{t-1}^2 + \varphi u_{t-1}^2}\right)\varepsilon_t \tag{44}$$

if σ_t^2 is $E_t(u_t^2)$ assumed to follow a GARCH (1, 1) process and ε_t is an i.i.d. (0,1) process. For example, an AR (1)-GARCH (1, 1) can be written as

$$\begin{aligned} p_t &= \alpha + \theta p_{t-1} + u_t \ , \\ V(p_t / I_t) &= \sigma_t^2 = \gamma + \beta\sigma_{t-1}^2 + \varphi u_{t-1}^2 \ . \end{aligned} \tag{45}$$

One condition that this model must respect is stationarity. To check if the stationarity[23] condition is respected, we present two classical tests, namely the Dickey-Fuller test and the augmented Dichey-Fuller test[24].

Dickey-Fuller test (DF test)

Let:

The TS (Trend Stationary) model be

$$p_t = \gamma_0 + \gamma_1 t + u_t \ , \tag{46}$$

The DS (Difference Stationary) model be

$$p_t = \gamma_1 + p_{t-1} + u_t \ . \tag{47}$$

Combining (46) and (47) as suggested by Bhargava (1986) gives

$$\begin{aligned} p_t &= \gamma_0 + \gamma_1 t + v_t; \qquad v_t = \alpha v_{t-1} + u_t \\ &= \gamma_0 + \gamma_1 t + \alpha(p_{t-1} - \gamma_0 - \gamma_1(t-1)) + u_t \ . \end{aligned} \tag{48}$$

Since (48) is nonlinear in the parameters, it is convenient to reparametrize it as

$$p_t = b_0 + b_1 t + \alpha p_{t-1} + u_t, \tag{49}$$

---

[23] *For a test of stationarity against the alternative of a unit root, see Brooks (2008) or Kwaitkowski, D. et al. (1992). This test is known as the KPSS test. It has been developed for alleviating the criticisms of DF and Phillips-Perron-type tests of having low power when the process is stationary, but with a root close to the non-stationarity boundary. We provide an EViews application of this test in the application subsection presented below.*

[24] *In the presence of structural breaks and misspecification of the short run component like their possible nonlinearity, the tests presented below might be inappropriate. Breitung (2002) suggests nonparametric tests that are robust to such misspecifications.*

where b_0 and b_1 are obtained by manipulating (49) as

$$\begin{aligned} p_t &= \gamma_0+\gamma_1 t+\alpha p_{t-1}-\alpha\gamma_1 t+\alpha\gamma_1 +u_t \\ &= \gamma_0(1-\alpha)+\alpha\gamma_1+\gamma_1(1-\alpha)t-\alpha p_{t-1}+u_t\ , \end{aligned} \tag{50}$$

and by setting $b_0=\gamma_0(1-\alpha)+\gamma_1\alpha$ and $b_1=\gamma_1(1-\alpha)$. (50) hides the fact that $b_1=0$ when $\alpha=1$. Subtracting p_t by p_{t-1} both side gives

$$\begin{aligned} p_t-p_{t-1} &= b_0+b_1 t+\alpha p_{t-1}-p_{t-1}+u_t \\ \nabla p_t &= b_0+b_1 t+(1-\alpha)p_{t-1}+u_t\ .\text{[25]} \end{aligned} \tag{51}$$

As we can see, if $\alpha<1$, (51) is equivalent to model TS (46); and if $\alpha=1$, (51) is equivalent to model DS (47). Equation (51) can be consistently estimated ($p\,lim(\hat{b})=b$) by using OLS. As showed by Dickey and Fuller (1979), the resulting t-statistic for testing the null hypothesis $\alpha-1=0$ must be compared to the corresponding τ values. Otherwise, type I error might be committed. The critical values (for regression (51), which includes a constant and trend) of the τ statistics are respectively for α =1%, 2.5%, 5%, 10%; $\tau_{ct}=-3.96, -3.66, -3.41, -3.13$ (Davidson and Mackinnon 1993, chap. 20).

Augmented Dickey-Fuller test (ADF test)[26]

From (51) we can write

$$\nabla p_t = x_t\beta+(\alpha-1)p_{t-1}+u_t\ , \tag{52}$$

where $x_t\beta=\beta_0+\beta_1 t$. x_t may also include any set of non stochastic regressors that we might want to include in the test regression; namely a constant, a linear trend, a polynomial trend (e.g. the quadratic trend $x_t\beta=\beta_0+\beta_1 t+\beta_2 t^2$). Supposing that u_t follows the stationary AR (1) process $u_t=\rho u_{t-1}+e_t$, (52) becomes

$$\begin{aligned} \nabla p_t &= x_t\beta+(\alpha-1)p_{t-1}+\rho u_{t-1}+e_t \\ &= x_t\beta+(\alpha-1)p_{t-1}+\rho(p_{t-1}-p_{t-2}-x_{t-1}\beta-(\alpha-1)p_{t-2})+e_t \\ &= x_t\beta-x_{t-1}\rho\beta+(\alpha+\rho-1)p_{t-1}+p_{t-2}(\rho-\rho-\rho\alpha)+e_t \\ &= x_t\beta^*+(\alpha+\rho-1)p_{t-1}-\rho\alpha p_{t-2}+e_t \\ &= x_t\beta^*+(\alpha+\rho-1-\alpha\rho)p_{t-1}+\alpha\rho(p_{t-1}-p_{t-2})+e_t \\ &= x_t\beta^*+\overbrace{[(\alpha-1)-\rho(\alpha-1)]}^{=(\alpha-1)(1-\rho)}p_{t-1}+\alpha\rho(\Delta p_{t-1})+e_t \\ &= x_t\beta^*+\beta_1' p_{t-1}+\beta_2'\Delta p_{t-1}+e_t\ , \end{aligned} \tag{53}$$

where $\beta_1'=(\alpha-1)(1-\rho)$ and $\beta_2'=\alpha\rho$.

[25] *∇ is the difference operator with $d=1$. ∇^d means differencing d times and it is equal to $(1-L)^d$, where L is the lag operator. For example, $\nabla^2=(1-L)^2=1-2L+L^2$ and if multiplied by p_t, it gives: $p_t-2p_{t-1}+p_{t-2}$.*

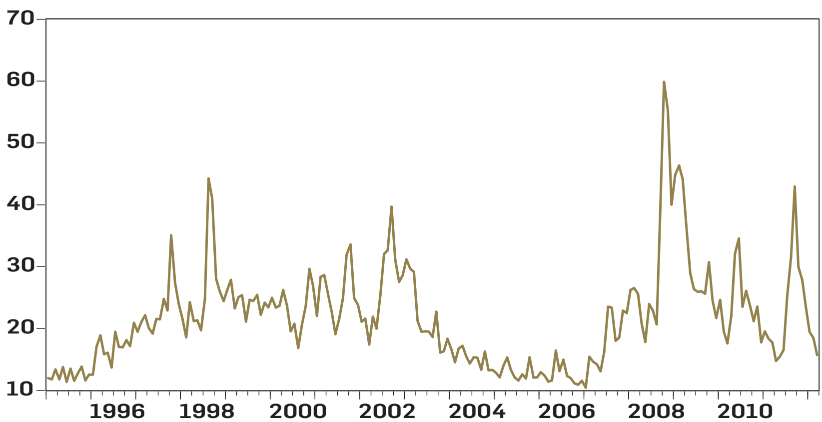
[26] *See Dickey and Fuller (1981).*

Thus, one performs an ADF test simply by running OLS on (53) and by comparing the resulting t statistic, called τ, of parameter β_1^{*} to the τ_{ct} asymptotic critical values. The null hypothesis is $\beta_1^{*}=0$, which is the equivalent of testing $\alpha=1$.

An application of standard unit root tests on VIX data

In this section, we discuss a numerical application of the standard unit root tests using a sample on of S&P500 Volatility Index (VIX) ranging from January 1995 to March 2012. Figure 2 shows this time series.

Figure 2. **Monthly S&P500 Volatility Index (VIX) – January 1995 to March 2012**



Some words should be said about the CBOE S&P500 Volatility Index (VIX). This index is computed using OTM puts and calls as[27] $\hat{\sigma}^2=\frac{2e^{rT}}{T}\left[\int_0^{F_{0,T}}\frac{1}{K^2}P(k)dK+\int_{F_{0,T}}^{\infty}\frac{1}{K^2}C(k)dk\right]$; where $P(K)$ and $C(K)$ are puts and calls, which are functions of the strike price K, and $F_{0,T}=S_0e^{rT}$ is the forward price. In practice, the CBOE use a discretized approximation of this equation, which takes the form $\sigma^2=\frac{2}{T}\sum_{K_i\leq K_0}\frac{\Delta K_i}{K_i^2}e^{rT}Put(K_i)+\frac{2}{T}\sum_{K_i>K_0}\frac{\Delta K_i}{K_i^2}e^{rT}Call(K_i)-\frac{1}{T}\left[\frac{F_{0,T}}{K_0}-1\right]^2$. It should be noted that the *expected realized variance* can be estimated using this formula. This means that by creating a portfolio of OTM puts and calls weighted by the inversed squared strike price, the variance estimate can be replicated, so trading options is another way to proceed. For instance, Russi (2012) provides some details about how one should use S&P500 futures and VIX futures when going to high-frequency analysis. We present further discussion on this subject in section 8.

Table 3 provides several estimations of unit tests.

[27] *See MacDonald (2006), Rouah and Vainberg (2007) and Hull (2012). Racicot (2009) briefly discuss this formula.*

Table 3. Several EViews unit root tests

Panel A.
The DF test

Null Hypothesis: VIX has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.948487
Test critical valu	1% level	-2.576236
	5% level	-1.942376
	10% level	-1.615674

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 03/30/12 Time: 11:25
Sample (adjusted): 1995:02 2012:03
Included observations: 206 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.08135	0.02759	-2.948487	0.0036
R^2	0.040667	Mean dependent var		0.018058
Adj. R^2	0.040667	S.D. dependent var		4.573883
S.E. of reg.	4.479913	Akaike info criterion		5.841927
Sum sq. resid	4114.273	Schwarz criterion		5.858082
Log likelihood	-600.7185	Hannan-Quinn criter.		5.848461
DW stat.	1.857193			

Panel B.
The ADF test

Null Hypothesis: VIX has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.204206	0.0008
Test critical valu	1% level	-3.462095	
	5% level	-2.875398	
	10% level	-2.574234	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(VIX)
Method: Least Squares
Date: 03/30/12 Time: 11:23
Sample (adjusted): 1995:02 2012:03
Included observations: 206 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VIX(-1)	-0.15731	0.037417	-4.204206	0
C	3.436128	0.868852	3.954789	0.0001
R-squared	0.079735	Mean dependent var		0.018058
Adj. R^2	0.075224	S.D. dependent var		4.573883
S.E. of reg.	4.398486	Akaike info criterion		5.810059
Sum sq. resid	3946.723	Schwarz criterion		5.842369
Log likeli.	-596.4361	Hannan-Quinn criter.		5.823126
F-statistic	17.67535	Durbin-Watson stat		1.796368
Prob(F-stat)	0.000039			

Panel C.
The PP test

Null Hypothesis: VIX has a unit root
Exogenous: Constant
Bandwidth: 5 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-4.130745	0.0011
Test critical valu	1% level	-3.462095	
	5% level	-2.875398	
	10% level	-2.574234	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	19.15885
HAC corrected variance (Bartlett kernel)	18.35354

Phillips-Perron Test Equation
Dependent Variable: D(VIX)
Method: Least Squares
Date: 03/30/12 Time: 11:33
Sample (adjusted): 1995:02 2012:03
Included observations: 206 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VIX(-1)	-0.15731	0.037417	-4.204206	0
C	3.436128	0.868852	3.954789	0.0001
R^2	0.079735	Mean dependent var		0.018058
Adj. R^2	0.075224	S.D. dependent var		4.573883
S.E. of reg.	4.398486	Akaike info criterion		5.810059
Sum squared resid	3946.723	Schwarz criterion		5.842369
Log likeli.	-596.4361	Hannan-Quinn criter.		5.823126
F-statistic	17.67535	Durbin-Watson stat		1.796368
Prob(F-stat.)	0.000039			

Panel D.
The Ng-Perron test

Null Hypothesis: VIX has a unit root
Exogenous: Constant
Lag length: 0 (Spectral GLS-detrended AR based on SIC, maxlag=14)
Sample: 1995:01 2012:03
Included observations: 207

		MZa	MZt	MSB	MPT
Ng-Perron test statistics		-16.0614	-2.83279	0.17637	1.52947
Asymptotic criti	1%	-13.8	-2.58	0.174	1.78
	5%	-8.1	-1.98	0.233	3.17
	10%	-5.7	-1.62	0.275	4.45

*Ng-Perron (2001, Table 1)

HAC corrected variance (Spectral GLS-detrended AR)	19.9722

Panel E.

The KPSS test

Null Hypothesis: VIX is stationary				
Exogenous: Constant				
Bandwidth: 10 (Newey-West automatic) using Bartlett kernel				
				LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin(KPSS) test stat.				0.16958
Asymptotic critical values*:	1% level			0.739
	5% level			0.463
	10% level			0.347
*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)				
Residual variance (no correction)				66.93209
HAC corrected variance (Bartlett kernel)				439.7135
KPSS Test Equation				
Dependent Variable: VIX				
Method: Least Squares				
Date: 03/30/12 Time: 11:24				
Sample: 1995:01 2012:03				
Included observations: 207				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21.69903	0.570011	38.06771	0
R^2	0	Mean dependent var		21.69903
Adj. R^2	0	S.D. dependent var		8.201037
S.E. of reg.	8.201037	Akaike info criterion		7.051217
Sum squared res.	13854.94	Schwarz criterion		7.067318
Log likeli.	-728.801	Hannan-Quinn criter.		7.057728
DW stat.	0.309547			

Panel F.

The Elliot-Rotenberg-Stock test

Null Hypothesis: VIX has a unit root		
Exogenous: Constant		
Lag length: 0 (Spectral OLS AR based on SIC, maxlag=14)		
Sample: 1995:01 2012:03		
Included observations: 207		
		P-Statistic
Elliott-Rothenberg-Stock test statistic		1.773652
Test critical valu	1% level	1.9128
	5% level	3.17315
	10% level	4.33525
*Elliott-Rothenberg-Stock (1996, Table 1)		
HAC corrected variance (Spectral OLS autoregression)		19.15885

At Table 3, we can note that all the unit root tests are consistent because there are no unit roots in this time series, which means that the VIX is stationary and does not necessitate any differentiating. Consequently, one could use the VIX as a regressor in a financial regression without any sort of preliminary transformation[28].

3.1.4. Estimating multivariate ARCH models (MGARCH)[29]

In this section, we present the methodology for estimating Bollerslev's (1990) MGARCH and the BEKK (1995) models. Bollerslev's model is easier to estimate than other MGARCH models, because of the assumption of constant conditional correlation. By imposing constant correlation, the number of parameter to estimate reduces greatly. In fact, the research in this field deals mainly with two things: reducing the inflation of parameters to estimate and the problem that the covariance matrix might not be positive definite. The BEKK (1995) model, which has the advantage of being parsimonious, suggests some ways to handle this matter.

[28] *For an application, see Racicot and Théoret (2009).*

[29] *See also Alexander and Chibumba (1997).This method provides an alternative that is probably easier to implement than the MGARCH models of Bollerslev (1990) and Engle (2002). For an application of the orthogonal GARCH on covariance matrix forecasting in a stress scenario, see Byström, H.NE. (2000).*

The Bollerslev's (1990) MGARCH model

The Bollerslev's (1990) model is as follows.

Let y_t denote the $N\times1$ time-series vector of interest (y_t might be considered as a vector of p_{it}) with time varying conditional covariance matrix Ω_t,

$$y_t = E(y_t | \Psi_{t-1}) + \varepsilon_t \qquad (54)$$
$$V(\varepsilon_t | \Psi_{t-1}) = \Omega_t \ ,$$

where Ψ_t is the information set of all the available information up through time t–1, and Ω_t is almost surely positive definite for all t. The formulation in (54) allows for both conditional and/or unconditional heteroskedasticity. For example, note that $E(y_t | \Psi_{t-1})$ can be set to be equal to $x_t'\beta$, which is our standard multivariate regression model.

Also, let σ_{ijt} denote the ij^{th} element in Ω_t and y_{it} and ε_{it} the i^{th} element in y_t and ε_t, respectively. The assumption that the conditional correlation is constant is written as

$$\sigma_{ijt} = \rho_{ij}(\sigma_{iit}\sigma_{jjt})^{1/2}, \quad j = 1,\ldots,N, \quad i = j+1,\ldots,N. \qquad (55)$$

The appealing feature of this model relates directly to the simplified estimation and inference procedures. To that end, rewrite each of the conditional variances as

$$\sigma_{iit} = \omega_i \sigma_{it}^2 \ , \ i = 1,\ldots,N, \qquad (56)$$

with ω_i a positive time invariant scalar and $\sigma_{it}^2 > 0$ almost surely for all t. Given (55) and (56), the full conditional covariance matrix, Ω_t may be partitioned as

$$\Omega_t = D_t \Gamma D_t, \qquad (57)$$

where D_t denotes the $N\times N$ stochastic diagonal matrix with elements $\sigma_{1t},\ldots,\sigma_{Nt}$ and Γ is an $N\times N$ time invariant matrix with typical element $\rho_{ij}\sqrt{\omega_i\omega_j}$. Ω_t could be positive definite for all t if and only if each of the N conditional variances are well defined and Γ is positive definite. Assuming conditional normality, the log likelihood function for the general heteroskedasticity model in (54) becomes,

$$L(\theta) = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}(log|\Omega_t| + \varepsilon_t'\Omega_t^{-1}\varepsilon_t) \ , \qquad (58)$$

where θ denotes all the unknown parameters in ε_t and Ω_t. Under standard regularity conditions, ML estimate for θ is asymptotically normal and the traditional inference procedures are immediately available[30]. However, since the evaluation of the likelihood function in (58) requires the inversion of an $N \times N$ matrix for each time period, the maximization of $L(\theta)$ by iterative methods can be very costly even for small sized T and N. The assumption in (55) reduces this computation. By direct substitution,

$$L(\theta) = -\frac{TN}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log|D_t \Gamma D_t| - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' (D_t \Gamma D_t)^{-1} \varepsilon_t$$

$$= -\frac{TN}{2} \log 2\pi - \frac{T}{2} \sum_{t=1}^{T} \log|\Gamma| - \frac{T}{2} \sum_{t=1}^{T} \log|D_t| - \frac{1}{2} \sum_{t=1}^{T} \tilde{\varepsilon}_t' \Gamma^{-1} \tilde{\varepsilon}_t \tag{59}$$

where $\tilde{\varepsilon}_t = D_t^{-1} \varepsilon_t$ denotes the $N \times 1$ vector of standardized residuals. Except for the third term that is a Jacobian term arising from the transformation from ε_t to $\tilde{\varepsilon}_t$, the likelihood function in (59) is equivalent to the likelihood function for $\tilde{\varepsilon}_t$; which is conditionally normal with time invariant covariance matrix Γ. The likelihood function in (59) is still highly nonlinear in the parameters, thus an iterative maximization procedure is required. Nevertheless, comparing (59) to (58), the former is much easier to evaluate and requires only one $N \times N$ matrix inversion as opposed to T inversions in (58). Note also that $\log|D_t|$ is equal to the sum of $\log\sigma_{1t}, \ldots, \log\sigma_{Nt}$. The suggestion for maximizing (59) is to use the Berndt, Hall, Hall and Hausman (BHHH, 1974) algorithm along with numerical first order derivatives[31].

The BEKK model of Engle and Kroner (1995)

Another popular specification, the BEKK model of Engle and Kroner (1995), named after an earlier working paper by Bollerslev, Engle, Kraft and Kroner, guarantees positive definiteness by working with quadratic forms rather than with the individual elements of Ω_t. The model is

$$\Omega_t = C'C + B'\Omega_{t-1} B + A' \varepsilon_t \varepsilon_t' A \ , \tag{60}$$

where C is a lower triangular matrix with $N(N+1)/2$ parameters, and B and A are square matrices with N^2 parameters each, for a total parameter count of $(5N^2+N)/2$. Weak restrictions on B and A guarantee that Ω_t is always positive definite. The log likelihood of this model is obtained by replacing the value (60) in (58). Numerical methods are required to find the value of the parameters.

The multivariate VAR-EGARCH model

Multivariate autoregressive models have been applied mainly to account for the interrelation between time series. For example, Koutmos (1996) used a multivariate

[30] *If the model correctly specifies the first two conditional moments, even if conditional normality is violated and under suitable regularity conditions, the quasi-maximum likelihood estimates obtained from (58) will still be consistent and asymptotically normal. However, the usual standard errors will have to be modified.*

[31] *In this model, there are $n(n+2)/2$ correlation coefficients ρ_{ij} and N conditional variances to estimate. If these conditional variance were a univariate GARCH $(1, 1)$, then the total number of parameters would be equal to $3n+(n(n+1)/2)$.*

VAR-EGARCH (MVAR-EGARCH) for modeling the interaction of stock markets across four European countries: France, Germany, Italy and UK. The MEGARCH structure is used to capture the leverage effect and the linkage across the innovations. The MVAR-EGARCH model and its application to stock markets follow.

Assuming that $r_{i,t}$ (e.g. possibly an index) is the return of market i at time t, having four countries : $i = 1, 2, 3, 4$. Then the MVAR-EGARCH model can be written as follow

$$r_{i,t} = \beta_{i,0} + \sum_{j=1}^{4} \beta_{i,j} r_{j,t-1} + e_{i,t} \quad i,j = 1,2,3,4 \tag{61}$$

$$\sigma_{i,t}^2 = \exp\left[\alpha_{i,0} + \sum_{j=1}^{4} \alpha_{i,j} f_j\,(z_{j,t-1}) + \gamma_i \; ln(\sigma_{i,t-1}^2) \right] \quad i,j = 1,2,3,4 \tag{62}$$

$$f_j(z_{j,t-1}) = \left(|z_{j,t-1}| - E\left(|z_{j,t-1}|\right) + \delta_j\, z_{j,t-1} \right) \quad j = 1,2,3,4 \tag{63}$$

$$\sigma_{i,j,t} = \rho_{i,j} \sigma_{i,t} \sigma_{j,t} \quad i,j = 1,2,3,4 \quad i \neq j \tag{64}$$

where (61) is a vector autoregressive model VAR for the four markets. Note that the conditional mean of each market is a function of its own past and the cross-market past returns. The coefficient $\beta_{i,j}$ for $i \neq j$ captures the interactions across markets. For example, if $\beta_{i,j}$ is significant then market j can be used to predict future returns for market i. Equation (62) is used for modeling the conditional variance in each market.

This is the EGARCH model, which is a function of its past volatility as well as its cross-market standardized innovations: $f_j(z_{j,t-1})$ where $z_{j,t-1} = e_{i,t} / \sigma_{i,t}$ is the standardized innovation. This is shown in equation (63). It permits standardized and cross-market innovations to influence the conditional variance in each market asymmetrically, which is consistent with the *leverage effect*. Equation (64) is the conditional covariance used to capture the contemporaneous relationship between the returns of the markets. This specification assumes that the correlation of the returns of markets i and j is constant. As in the model of Bollerslev (1990), this is the same specification and it simplifies the estimation process.

Parameters estimation

If we assume normality of the innovations, we can write the log likelihood function of the MVAR-EGARCH as follows

$$l(\theta) = \ln L(\theta) = -\frac{1}{2}(NT)\, ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left(ln(\Omega_t) + e_t' \Omega_t^{-1} e_t \right) \tag{65}$$

where θ is a vector of 54 parameters, N=4 is the number of equations, T is the number of observations, $e_t' = (e_{1,t}, e_{2,t}, e_{3,t}, e_{4,t})$ is a vector of innovations at time t, Ω_t is time varying variance-covariance matrix; where the diagonal elements are given by equation (62) for $i=1, 2, 3, 4$ and, the off diagonal elements are given by equation (64) for $i,j=1, 2, 3, 4$ and $i\neq j$. As in Bollerslev (1990) the parameters vector θ is estimated using numerical maximization algorithm. Koutmos (1996) used the BHHH (1974) algorithm, which uses numerical derivatives of $l(\theta)$. This algorithm may be found in GQOPT and runs in FORTRAN programming language[32].

***Application*[33]**

In our example, the MVAR-EGARCH is used to model the interdependence of conditional mean and volatility of four European countries: France, Germany, Italy and UK. Koutmos (1996) found that the β coefficients are significant for many of the countries in his study. For example, the conditional mean of France is linked to past returns of Germany and UK. Similarly, the returns in Germany are correlated to past returns in France and the UK. Italy is influenced by Germany and UK.

When analyzing the volatility interdependence, it is found that the correlations ($\rho_{i,j}=\sigma_{i,j,t}/\sigma_{i,t}\sigma_{j,t}$) between markets are significant. The conditional variance in each market is also influenced by its own past innovations and past innovations created by other markets. Only Italy and UK are not influenced in both directions. Finally, the degree to which bad news (innovations) have increased the volatility more than good news is captured by $\alpha_{i,j}$ and δ_j . For example, the impact of a ±3% variation of the innovation in market i (at time t–1 on the conditional variance of market j at time t) not only appears in the same market, but also it has an impact on other markets.

The following application presents a numerical example of the simplest M-GARCH, which is referred as the constant conditional correlation (CCC) model. It is essentially the same model as the one in equation (25). Our application uses observed monthly data on the S&P500 and the VIX ranging from January 1995 to March 2012. Our goal here is to provide an example of the use of multivariate GARCH that would share some of the ideas developed in Sun and Wu (2010); which represents an evaluation of the leverage effect theory. We used EViews to run the multivariate GARCH models presented in Table 4.

[32] *For an application of GQOPT, see Racicot (2003), chapter 2.*

[33] *Brooks (2008) provides RATS programs for estimating a diagonal VECH model and BEKK model. He also provides an application of these models for computing a dynamic hedge ratio using daily data on the FTSE 100 stock index and futures contracts on stock indexes. For more financial applications of these models, see Tsay (2005).*

Table 4. Several EViews M-GARCH estimations

Panel A. The diagonal BEKK model

System: Diag BEKK
Estimation Method: ARCH Maximum Likelihood (Marquardt)
Covariance specification: Diagonal BEKK
Date: 04/12/12 Time: 17:38
Sample: 1995:02 2012:03
Included observations: 206
Total system (balanced) observations 412
Presample covariance: backcast (parameter =0.7)
Convergence achieved after 66 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.030731	0.008404	3.656652	0.0003
C(2)	-2.06E-05	6.84E-06	-3.018044	0.0025
C(3)	3.157546	0.687692	4.591511	0.0000
C(4)	0.841486	0.024087	34.93477	0.0000
	Variance Equation Coefficients			
C(5)	0.000356	0.000134	2.649782	0.0081
C(6)	-0.035248	0.011747	-3.000592	0.0027
C(7)	5.394818	2.33936	2.306109	0.0211
C(8)	0.539375	0.077694	6.942344	0.0000
C(9)	0.343672	0.045924	7.483447	0.0000
C(10)	0.747316	0.07848	9.522364	0.0000
C(11)	0.756668	0.111476	6.787726	0.0000
Log likelihood	-136.2264	Schwarz criterion		1.607085
Avg. log likeli.	-0.330647	Hannan-Quinn criter.		1.501252
Akaike info crit.	1.429383			

Equation: RSP500 = C(1) + C(2)*SP500(-1)

R-squared	0.032329	Mean dependent var	0.005313
Adjusted R-sq.	0.027586	S.D. dependent var	0.046505
S.E. of reg.	0.045859	Sum squared resid	0.429028
DW stat.	1.790204		

Equation: VIX = C(3) + C(4)*VIX(-1)

R-squared	0.711775	Mean dependent var	21.74631
Adjusted R-sq.	0.710362	S.D. dependent var	8.19269
S.E. of reg.	4.409145	Sum squared resid	3965.875
DW stat.	1.78559		

Covariance specification: Diagonal BEKK
GARCH = M + A1*RESID(-1)*RESID(-1)'*A1 + B1*GARCH(-1)*B1
M is an indefinite matrix
A1 is a diagonal matrix
B1 is a diagonal matrix

Transformed Variance Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
M(1,1)	0.000356	0.000134	2.649782	0.0081
M(1,2)	-0.035248	0.011747	-3.000592	0.0027
M(2,2)	5.394818	2.33936	2.306109	0.0211
A1(1,1)	0.539375	0.077694	6.942344	0.0000
A1(2,2)	0.343672	0.045924	7.483447	0.0000
B1(1,1)	0.747316	0.07848	9.522364	0.0000
B1(2,2)	0.756668	0.111476	6.787726	0.0000

Panel B. The Diagonal VECH model

System: Diag VECH
Estimation Method: ARCH Maximum Likelihood (Marquardt)
Covariance specification: Diagonal VECH
Date: 04/12/12 Time: 17:32
Sample: 1995:02 2012:03
Included observations: 206
Total system (balanced) observations 412
Presample covariance: backcast (parameter =0.7)
Convergence achieved after 32 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.034863	0.007172	4.86127	0.0000
C(2)	-2.35E-05	5.55E-06	-4.23408	0.0000
C(3)	2.644108	0.757922	3.488628	0.0005
C(4)	0.864404	0.029834	28.97345	0.0000
	Variance Equation Coefficients			
C(5)	0.000498	0.000166	3.003633	0.0027
C(6)	-0.070519	0.014482	-4.869294	0.0000
C(7)	7.233572	2.878396	2.513057	0.0120
C(8)	0.245557	0.075008	3.273758	0.0011
C(9)	0.165406	0.042894	3.856133	0.0001
C(10)	0.11613	0.050375	2.305295	0.0212
C(11)	0.502698	0.122579	4.10103	0.0000
C(12)	0.284898	0.140986	2.020755	0.0433
C(13)	0.454747	0.1993	2.281713	0.0225
Log likelihood	-133.019	Schwarz criterion		1.627672
Avg. log likeli.	-0.322862	Hannan-Quinn criter.		1.502596
Akaike info crit.	1.41766			

Equation: RSP500 = C(1) + C(2)*SP500(-1)

R-squared	0.032469	Mean dependent var	0.005313
Adjusted R-sq.	0.027726	S.D. dependent var	0.046505
S.E. of reg.	0.045856	Sum squared resid	0.428966
DW stat.	1.784953		

Equation: VIX = C(3) + C(4)*VIX(-1)

R-squared	0.711158	Mean dependent var	21.74631
Adjusted R-sq.	0.709742	S.D. dependent var	8.19269
S.E. of reg.	4.413861	Sum squared resid	3974.363
DW stat.	1.822254		

Covariance specification: Diagonal VECH
GARCH = M + A1.*RESID(-1)*RESID(-1)' + B1.*GARCH(-1)
M is an indefinite matrix*
A1 is an indefinite matrix
B1 is an indefinite matrix

Transformed Variance Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
M(1,1)	0.000498	0.000166	3.003633	0.0027
M(1,2)	-0.070519	0.014482	-4.869294	0.0000
M(2,2)	7.233572	2.878396	2.513057	0.0120
A1(1,1)	0.245557	0.075008	3.273758	0.0011
A1(1,2)	0.165406	0.042894	3.856133	0.0001
A1(2,2)	0.11613	0.050375	2.305295	0.0212
B1(1,1)	0.502698	0.122579	4.10103	0.0000
B1(1,2)	0.284898	0.140986	2.020755	0.0433
B1(2,2)	0.454747	0.1993	2.281713	0.0225

* Coefficient matrix is not PSD.

Panel C. The Constant Conditional Correlation (CCC) model

System: CCC
Estimation Method: ARCH Maximum Likelihood (Marquardt)
Covariance specification: Constant Conditional Correlation
Date: 04/12/12 Time: 17:36
Sample: 1995:02 2012:03
Included observations: 206
Total system (balanced) observations 412
Presample covariance: backcast (parameter =0.7)
Convergence achieved after 35 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.033991	0.007599	4.473075	0.0000
C(2)	-2.44E-05	6.31E-06	-3.876937	0.0001
C(3)	3.203652	0.895522	3.577411	0.0003
C(4)	0.844936	0.034595	24.42339	0.0000
	Variance Equation Coefficients			
C(5)	0.000182	9.28E-05	1.965837	0.0493
C(6)	0.182803	0.069687	2.623196	0.0087
C(7)	0.73716	0.090294	8.164042	0.0000
C(8)	6.994811	4.121647	1.697091	0.0897
C(9)	0.127292	0.072633	1.752531	0.0797
C(10)	0.467707	0.296321	1.578379	0.1145
C(11)	-0.74861	0.029957	-24.98968	0.0000
Log likelihood	-135.0562	Schwarz criterion		1.595723
Avg. log likeli.	-0.327806	Hannan-Quinn criter.		1.48989
Akaike info crit.	1.418021			

Equation: RSP500 = C(1) + C(2)*SP500(-1)

R-squared	0.037924	Mean dependent var	0.005313
Adjusted R-sq.	0.033208	S.D. dependent var	0.046505
S.E. of reg.	0.045727	Sum squared resid	0.426547
DW stat.	1.793226		

Equation: VIX = C(3) + C(4)*VIX(-1)

R-squared	0.712656	Mean dependent var	21.74631
Adjusted R-sq.	0.711248	S.D. dependent var	8.19269
S.E. of reg.	4.402396	Sum squared resid	3953.742
DW stat.	1.797123		

Covariance specification: Constant Conditional Correlation
GARCH(i) = M(i) + A1(i)*RESID(i)(-1)^2 + B1(i)*GARCH(i)(-1)
COV(i,j) = R(i,j)*@SQRT(GARCH(i)*GARCH(j))

Transformed Variance Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.000182	9.28E-05	1.965837	0.0493
A1(1)	0.182803	0.069687	2.623196	0.0087
B1(1)	0.73716	0.090294	8.164042	0.0000
M(2)	6.994811	4.121647	1.697091	0.0897
A1(2)	0.127292	0.072633	1.752531	0.0797
B1(2)	0.467707	0.296321	1.578379	0.1145
R(1,2)	-0.74861	0.029957	-24.98968	0.0000

Panels A, B and C (Table 4) show that all the MGARCH models work well. We used AR (1) equations to model the means of the processes and different MGARCH processes for comparisons. In order to verify the leverage effect theory, some authors (Barndorff-Neilsen *et al.*, 2008a) used *realized kernel* correlation. Instead, we use basic parametric MGARCH models. Looking at the Panel C, we see that the coefficient of R (1, 2) is negative and significant. This means that a negative correlation is observed between the implied volatility and stock returns. The leverage effect is validated by the *volatility feedback*[34], which implies that volatility is a priced factor (Bollerslev *et al.*, 2006).

4. Mean-nonlinear Stochastic Models

Nonlinearity might arise either in the form of the second moment (e.g. ARCH models), or in the first moment; as presented previously. The second form of nonlinearity is considered below and a list of the most popular models is given.

The Nonlinear Moving Average (NMA) model

$$p_t = u_t + a u_{t-1} u_{t-2}\,. \tag{66}$$

The Bilinear AR Model (BAR)

$$p_t = u_t + \alpha p_{t-1} u_{t-2}. \tag{67}$$

[34] *If investors anticipate an increase in volatility, they would require a higher rate of return on their investments. Consequently, prices would have to change. The outcome would be, as in the leverage effect theory, an increase in volatility and a drop in stock returns.*

The Bilinear ARMA Model (BARMA)

$$p_t = \alpha p_{t-1} - \alpha p_{t-1} u_{t-1} + \beta u_{t-1} + u_t . \tag{68}$$

The Threshold Autoregressive Model (TAR)

$$\begin{aligned} p_t &= \alpha p_{t-1} + u_t \text{ , if } p_{t-1} < 1, \\ p_t &= \beta p_{t-1} + u_t \text{ , if } p_{t-1} \geq 1. \end{aligned} \tag{69}$$

The Nonlinear Autoregressive Model (NAR)

$$p_t = (\alpha |p_{t-1}|)/(|p_{t-1}| + 2) + u_t . \tag{70}$$

The Exponential Autoregressive Model (EAR)

$$p_t - \varphi_1 p_{t-1} - \dots - \varphi_p p_{t-p} = u_t , \quad u_t \sim WN(0, \sigma^2) \tag{71}$$

where $\varphi_i = \alpha_i + \beta_i \exp\left(-\gamma p_{t-1}^2\right)$. This model behaves in the same way as the TAR model, but its coefficients change smoothly between the time intervals. Other models using the same acronym have been developed by Lawrence and Lewis (1985). These models can also be represented in a more general form, which is called by Priestley (1980) the *state dependent* models. Other popular models like the VCM (variable coefficient model) and the STUR (stochastic unit root process) are currently used in the literature[35].

The VCM and STUR models take respectively the following forms: $p_t = \alpha_t p_{t-1} + u_t$ with $\alpha_t = 0.9 \cos(2\pi / T)$ and, $p_t = \alpha_t p_{t-1} + u_t$ where $\alpha_t = 0.1 + 0.9 \alpha_t + \eta_t$. All of these models might be estimated consistently by maximum likelihood.

The Markov Switching Model

Hamilton (1989) is known to be the first author to have proposed using a simple Markov switching AR process to model the US GDP. Following the presentation of Zivot and Wang (2006), the Markov switching AR (p) model can be written as

$$y_t = \mu_{S_t} + X_t \vartheta_{S_t} + u_t \quad \text{for } t = 1, 2, \dots, n \tag{72}$$

where $X_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p})$, ϑ_{S_t} is the AR vector of coefficients of dimension $p \; x \; 1$, $u_t \sim N(0, \sigma_{S_t}^2)$, the hidden state variable S follows a k-regime Markov chain given by

$$P(S_t = j \,|\, S_{t-1} = i) = P_{ij} \geq 0 \tag{73}$$

[35] *For an application in the frame of unit roots and cointegration testing, see Breitung (2002).*

with $i, j = 1,2,\ldots,k$, where k represents the number of possible regimes or states. As usual, the sum of the probabilities (73) must equal 1 that is

$$\sum_{j=1}^{k} P(S_t = j \mid S_{t-1} = i) = 1 \tag{74}$$

The transition probabilities can be summarized into a matrix referred as the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \cdots & P_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ P_{k1} & P_{k2} & \cdots & P_{kk} \end{pmatrix}$$

where each row of $\boldsymbol{P}$ sum to 1.

The estimation of the coefficients of (72) is usually done by means of Maximum Likelihood. Two cases must be considered when estimating this model. In the first case, the states $\boldsymbol{S} = (S_{p+1},\ldots,S_n)$ are known. In the second case, they are unknown. When knowing the states $\boldsymbol{S}$, the likelihood function is very similar to equation (42). The log likelihood function can be written as follows

$$L(\boldsymbol{\theta}|\boldsymbol{S}) = \sum_{t=p+1}^{n} \log f(y_t \mid I_{t-1}, S_t) \tag{75}$$

where $f(y_t \mid I_{t-1}, S_t) \propto \exp\left(-\frac{1}{2}\log(\sigma_{S_t}^2) - \frac{\left(y_t - \mu_{S_t} - X_t \vartheta_{S_t}\right)^2}{2\sigma_{S_t}^2}\right)$, and $\boldsymbol{\theta}$, the unknown parameters.

In the case where the states $\boldsymbol{S}$ are unknown, the likelihood function must be generalized to include the transition probabilities. It can be written as

$$L(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \log f(y_t \mid I_{t-1}, S_t) = \sum_{t=p+1}^{n} \log \left\{\sum_{j=1}^{k} f(y_t \mid I_{t-1}, S_t = j) P(S_t = j \mid I_{t-1})\right\} \tag{76}$$

This result is obtained by applying the law of total probability. Note that $f(y_t \mid I_{t-1}, S_t = j)$ is equal to $f(y_t \mid I_{t-1}, S_t)$ and that by the Bayes theorem the transition probability $P(S_t = j \mid I_{t-1})$ are $P(S_t = j \mid I_{t-1}) = \sum_{i=1}^{k} P(S_t = j \mid S_{t-1} = i, I_{t-1}) P(S_{t-1} = i \mid I_{t-1})$

$$= \sum_{i=1}^{k} P_{ij} \frac{f(y_t = j \mid y_{t-2}, S_{t-1} = i) P(S_{t-1} = i \mid I_{t-2})}{\sum_{m=1}^{k} f(y_{t-1} \mid I_{t-2}, S_{t-1} = m) P(S_{t-1} = m \mid I_{t-2})} \tag{77}$$

The log-likelihood function of the Markov switching AR (p) model can be computed iteratively using (76) and (77) given an estimate of the initial probability $P(S_{p+1} = i|I_p)$, $i = 1, 2, \ldots, k$, of each state. The unknown parameters $\boldsymbol{\theta}$ can be estimated using standard maximum likelihood.

A Numerical Application of the Markov Switching Model

Nonlinear models, like the Markov switching one, have had several applications since their inception. For instance, Billio *et al.* (2009) use it to compute volatility of their asset returns process assuming that it has different states or regimes. They have also use it, assuming a regime-switching beta model, in order to capture hedge fund exposure to market and other risk factors based on the state of the market. There are plenty of other applications of this model[36]. Our own applications follow.

We provide two applications using two standard data sets, which are the S&P500 and the VIX index. These applications were also used previously in section 3.1.3 and 3.1.4. To estimated the Markov Swtiching (MS) model, we used a computer code developed by professors Pynnönen and Knif in the EViews programming language[37]. This program assumes a mean of the process either in a low or a high volatility regime. Figures 3 and 4 present a run of this program respectively for the S&P500 returns and the VIX Index ranging from January 1995 to March 2012.

Figure 3. MS for the S&P500 returns (1995m1-2012m3)

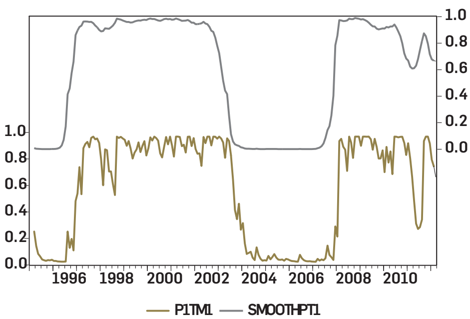
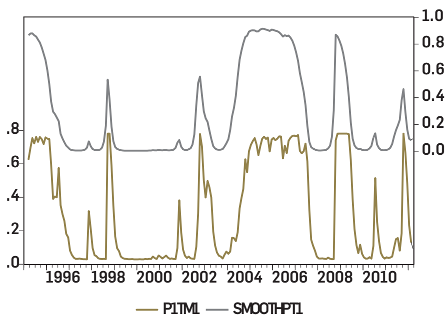


Figure 4. MS for the VIX (1995m1-2012m3)



In these Figures, P1TM1 represents the probability of regime switching. For comparison purposes, we provide the SMOOTHPT1, which is the smoothed probability of regime switching developed by Kim and Nelson (1999). In Figure 3, we can easily see that the probability of going to another regime (a high volatility regime) increased dramatically in 2007. This increase corresponds to the beginning of the financial crisis. In Figure 4, we observe the reverse phenomenon with the VIX. As we discussed previously, this would imply a negative correlation between the two series; an empirical fact justifying the *volatility feedback* or the leverage effect.

[36] *For instance, Khemiri (2012) presents an application of a Markov Switching Exponential GARCH (MSGARCH) model that provides a richer modeling of volatility dynamics. Moreover, the MSEGARCH model seems to fit the intraday data in a better way.*

[37] *The code can be found on the web site of Professors S. Pynnönen and J. Knif (Hanken), http://lipas.uwasa.fi/~sjp/Teaching/Afts/Lectures/fetsSynopsis.html.*

5. Nonlinear deterministic models[38]

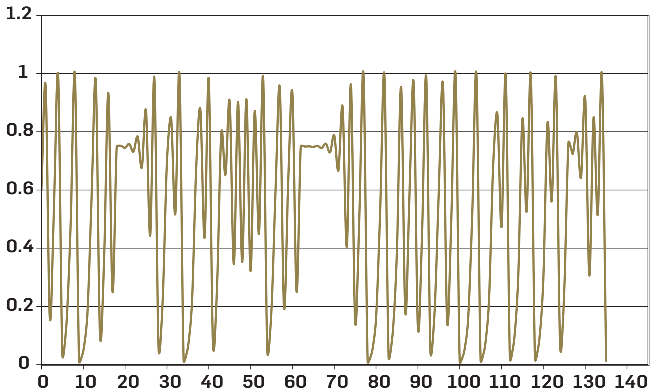
Nonlinear deterministic models might be used if the tests (e.g. BDS or R/S tests) that distinguish between a nonlinear deterministic model and a nonlinear stochastic model reject the possibility of a stochastic model. A list of these models follows.

The 'logistic map[39]' is defined as

$$p_{t+1}=\alpha p_t(1-p_t). \tag{78}$$

Figure 5 shows the erratic behavior[40] generated by this equation.

Figure 5. **The logistic map**



The 'tent map'

$$p_{t+1}=2p_t \quad \text{if} \quad p_t<.5 \;,$$
$$p_{t+1}=2-2p_t \quad \text{if} \quad p_t\geq .5 \;. \tag{79}$$

The tent map is shown in Figure 6.

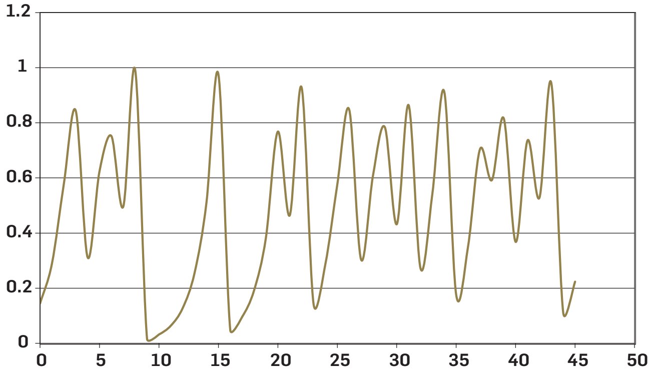
---

[38] *For an introducing to chaos theory in financial modeling, see Peters (1991, 1994).*

[39] *Only at α=3.57 that chaos can be seen. For example, at α=2.45 the logistic equation produces numbers that converge to 4 points attractor. A Lyaponuv exponent <0 means a stable system, >0 means a chaotic system (see Brock et al., 1991 and Dendney, 1991). Thus, chaos requires the largest Lyapunov exponent to be positive. The Lyapunov exponent is given by $L=lim_{t\to\infty}[ln(\|d\,F^t(x)\cdot v\|)/t]$ where $\|\ \|$ is a norm, d is the derivative and v is a direction vector. Here, '$\cdot$' denotes scalar product. $F^t(x)$ is t-th iterate of F starting at initial condition x. For a time series $\{y_t\}$ under scrutiny for temporal dependence, the definition implies that $\{y_t\}$ has a deterministic DGP. If for some state vector x_t, it's law motion can be described by $x_{t+1}=F(x_t)$ and there is some function $h(x)$ such that $y_t=h(x)$ for all t. Note also that F is assumed to have an ergodic invariant measure μ that is absolutely continuous. This means that μ is non degenerate (the series is not packed on one point, but it has a variance) and that the limiting time averages exist. These averages are independent of initial conditions (i.e. we can do time series analysis, Brock et al., 1991). Thus, ergodicity can be considered as a form of 'average asymptotic independence'. It also implies that the moments exist. For example, the Cauchy distribution is strictly stationary, but not ergodic because of the non-existence of its moments. Note that this definition involves concepts related to measure theory. For more details of these concepts, see Davidson and MacKinnon (1993), Hamilton (1994), White (2001).*

[40] *As seen in Figure 5, the logistic map seems to be an acceptable pseudo-random number generator, because it seems to generate a sufficient erratic behavior. However, this equation has two substantial problems. The first problem concerns the temporary near-periodicity or structure implicitly generated by this equation. Jäckel (2002) presents an example that shows this problem of temporary near periodicity at some values iterates. The second one concerns the non-uniformity of the invariant measure. In other words, there will be a concentration of numbers at certain iterates.*

Figure 6. **The tent map**

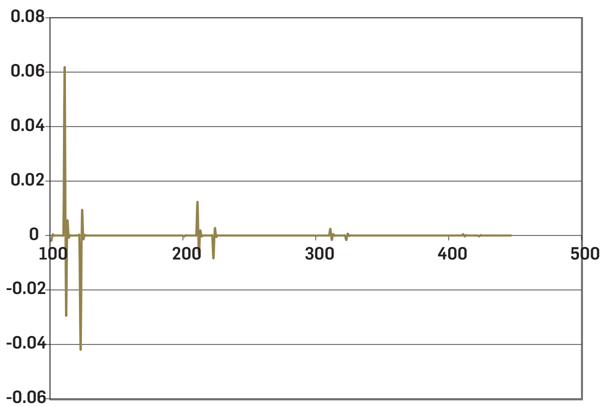


Finally, the Mackey and Glass (1977) deterministically chaotic model

$$p_t = \alpha p_{t-\tau} / [1 + p_{t-\tau}^{10}] - \beta p_t \quad . \tag{80}$$

The Mackey and Glass equation is shown in Figure 7.

Figure 7. **The Mackey and Glass (1977) equation**



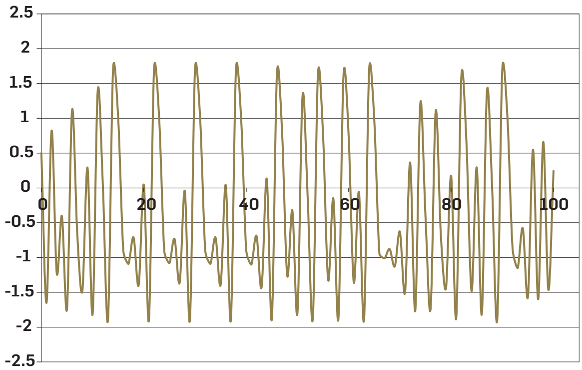
Equation (80) is formally an infinite dimensional system, but its attracting set dimension varies as the delay parameter τ is changed. For $\tau = 100$, the dimension is about 7 or larger. This model is a much higher dimensional system than the tent map. A simpler form of this model[41] is given by

$$p_t = \beta p_{t-1} - \alpha p_{t-\tau} \quad . \tag{81}$$

Figure 8 shows another chaotic model found in the literature.

[41] *The Mackey and Glass (1977) equation was developed to model red blood cell reproduction: production is based on past and current measurement. The delay τ between production and the measurement of current level produces a cycle related to that delay (see Brock et al., 1991). In Figure 7, we used equation (80) and assumed a delay $\tau = 1$.*

■ Figure 8. **Equation from the Scientific American Fractal T-Shirt**



This model is defined by the following simple recurrence equation $p_{t+1}=p_t^2+c$, where c is a given constant. We can see that, even for the simplicity of the model, it generates plausible fluctuations seen in financial time series.

6. Testing for neglected nonlinearity

6.1. The BDS test[42]

The BDS test uses a measure of spatial correlation[43], which is not often discussed in the literature. For this reason, we present below the details of the test. The BDS (Brock, Dechert and Scheinkman, 1987) statistic is defined as

$$W_{m,T}(\varepsilon)=T^{1/2}[C_{m,T}(\varepsilon)-C_{1,T}(\varepsilon)^m]/\sigma_{m,T}(\varepsilon)\ , \qquad (82)$$

where $C_{m,T}$ is a measure of spatial correlation (see Brock *et al.*, 1991) of scattered points in m-dimensional space, known as correlation integral. This correlation through space was defined by Grassberger and Procaccia (1983) as

$$C_{m,T}(\varepsilon)=\sum_{t<s}I_\varepsilon(p_t^m p_s^m)\times[2\,/T_m\,(T_m-1)] \qquad (83)$$

where $T_m = T-(m-1)$, $p_t^m = (p_t,\ldots,p_{t+m-1})$, T is the length of the time series $\{p_t\}$, $p_t^m = (p_t,\ldots,p_{t+m-1})$, $I_\varepsilon(p_t^m p_s^m)$ is an indicator function which equals 1 if $\|\, p_t^m-p_s^m\| < \varepsilon$ and equals 0 otherwise. $\|\ \|$ is the sup-norm. In general, this norm may be replaced by other norms like the Euclidean norm which are defined as $\|\, p_t^m-p_s^m\|=\left[\sum(p_t-p_s)^2\right]^{1/2}$. Here, to be consistent with the BDS statistic, the sup-norm (i.e. L^∞-norm) is used and defined as $\|\, p_t^m-p_s^m\|=E|p_t-p_s|$ in the Hilbert space L, which is $L(\Omega,A,P)$. The triplet (Ω,A,P) is a

[42] *This section is inspired by Brock et al. (1991). For more information on the subject, see Campbell et al. (1997) and Peters (1994). EViews has the BDS test as a standard function.*

[43] *For an introduction to spatial correlation, see Anselin (2001).*

probability space given a sample space Ω, a σ-algebra A associated with Ω, and a probability measure $P(\cdot)$ defined over A. For stochastic and deterministically chaotic systems, it can be shown that $C(\varepsilon)_{m,T} \to C_m(\varepsilon) \equiv Pr\left[\,\left\| p_t^m - p_s^m \right\| < \varepsilon\right]$. $C_m(\varepsilon)$ is the probability that the pair of points (p_t^m, p_s^m)are within the distance ε of each other $\sigma_{m,T}(\varepsilon)$ is an estimate of the standard deviation under the i.i.d. null hypothesis. Here $\{p_t\}$ is a scalar time series under investigation for randomness. We can estimate intertemporal local correlation and other dependence by means of (83). To compute this correlation, we incorporate $\{p_t\}$ in an m-dimensional space by forming m-vectors $p_t^m = (p_t, \dots, p_{t+m-1})$, starting at each date t. If $\{p_t\}$ is i.i.d., then $C_m(\varepsilon) = [C_1(\varepsilon)]^m$ and $C_1(\varepsilon) = E\{G(p+\varepsilon) - G(p-\varepsilon)\}$, where $G(p) \equiv Pr\{p_t \le p\}$ is the cumulative distribution function. The following explanation helps to understand (82). It works like a student-t test, if the series under investigation is strongly nonlinear, then the discrepancy between i.i.d. null hypothesis $C_{1,T}(\varepsilon)^m$ and the alternative $C_{m,T}(\varepsilon)$ will be large enough so that $W_{m,T}(\varepsilon)$ rejects the null hypothesis. Brock, Deckert and Sckeinkman (1987) showed that under the null hypothesis of i.i.d., (82) is asymptotically normally distributed

$$W_{m,T}(\varepsilon) \xrightarrow{d} N(0,1) \text{ as } T \to \infty. \tag{84}$$

The BDS test can be compared to the standard normal distribution tables. As mentioned above, the BDS test is designed to detect nonlinear deterministic chaos as well as nonlinear stochastic dependence. If the sample size is sufficiently large, the BDS statistic has good power against both types of nonlinear dependence. However, this test is neither specific for deterministic chaos nor for ARCH-GARCH process; rather, it is sensitive to nonlinearity. Stochastic nonlinear models described above are good examples of the alternative hypotheses that the BDS test is capable of detecting. For instance, the procedure to test the specification of an ARCH process is to use the BDS statistic on the residuals obtained by fitting a particular ARCH model. If the residual are i.i.d., then this model could be a good choice[44].

Some considerations with the use of the BDS test

The data under investigation must be stationary. Otherwise, the BDS test may commit type I errors. Unit root testing may be required to verify that this condition is satisfied. Monte Carlo experiments suggest that ε should be chosen between one-half and three halves of the standard deviation of the data. The dimension m should be chosen between 2 and 5 for small data sets (200 to 500 observations) and up to 10 for large data sets (at least 2000 observations). If there are 500 or more observation for the range of m and ε values, the asymptotic distribution provides a good approximation of the finite sample distribution. In actual application, it is recommended to do some bootstraps experiments.

[44] See Bollerslev et al. (1993).

An application of the BDS test

Table 5 presents an application of the BDS test using the data generated by the MacKey and Glass (1977) equation shown in Figure 7.

● Table 5. **The Brock, Dechert and Scheinkman (BDS) test**

BDS Test for MG1977					
Date: 03/28/12 Time: 17:50					
Sample: 1 500					
Included observations: 500					
Dimension					
2	0.154900	0.009090	17.04102	0.0000	
3	0.247785	0.014590	16.98308	0.0000	
4	0.299125	0.017578	17.01692	0.0000	
5	0.322741	0.018554	17.39515	0.0000	
6	0.327851	0.018132	18.08143	0.0000	
Raw epsilon		2.51E-06			
Pairs within epsilon		128319.0	V-Statistic	0.703777	
Triples within epsilon		45864717	V-Statistic	0.589108	
Dimension	C(m,n)	c(m,n)	C(1,n-(m-1))	c(1,n-(m-1))	c(1,n-(m-1))^k
2	58693.00	0.648362	63591.00	0.702469	0.493463
3	53990.00	0.599223	63583.00	0.705694	0.351438
4	49484.00	0.551809	63580.00	0.708997	0.252684
5	45178.00	0.506179	63580.00	0.712357	0.183438
6	41066.00	0.462294	63580.00	0.715741	0.134442

In Table 5, we used EViews to compute the values of the BDS statistics for several dimensions. These values are very significant at 1% level, which implies that our generated data cannot be considered i.i.d.[45].

6.2. Tests for nonlinear structure

To distinguish between models that are nonlinear in mean or in variance, Tsay (1986) suggests doing the following test.

Test for models that are nonlinear in the mean

The test is implemented as follows (see Campbell *et al.*, 1997).

i) Assuming that we have a time-series of observations $\{p_t\}$ then, regress p_t on its own lags and lags of its cross-products

$$g(p_{t-1}, p_{t-2}, \ldots) = \sum_{i=1}^{n} a_i p_{t-i} + \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} p_{t-i} p_{t-j} \,. \tag{85}$$

[45] *Salazar and Lambert (2010) used the BDS test to investigate if the independence assumption of their data set holds.*

Notes on Nonlinear Dynamics. *Racicot, F.E.*
AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2012. **5**: 162-221

ii) The test of the null hypothesis that all the nonlinear terms are not significant is

$$F = \frac{SCR^* - SCR/J}{SCR/N-K} \overset{a}{\sim} F(J, N-K) \quad , \tag{86}$$

where SCR^* is the restricted sum of squares residuals and SCR is the unrestricted sum. For example, for $n=m=2$ the regression to estimate is

$$\begin{aligned} p_t &= a_1 p_{t-1} + a_2 p_{t-2} + b_{11} p_{t-1}^2 + 2b_{12} p_{t-1} p_{t-2} + b_{22} p_{t-2}^2 \\ &= a_1 p_{t-1} + a_2 p_{t-2} + b_{11} p_{t-1}^2 + b_{12}^* p_{t-1} p_{t-2} + b_{22} p_{t-2}^2 \quad , \end{aligned} \tag{87}$$

where $b_{12}^* = 2b_{12}$ by assuming that $b_{12} = b_{21}$, then the test F might be constructed and compared with its critical value.

The third moment test

The third moment test is written as (see Brock *et al.*, 1991)

$$\tau = T^{1/2} r_{xxx}(i,j) / [\omega(i,j)/\sigma_x^6] \tag{88}$$

where $r_{xxx}(i,j) = [\Sigma x_t x_{t-i} x_{t-j}/T]/[\Sigma x_t^2/T]^{3/2}$. $\omega(i,j)/\sigma_x^6$ is consistently estimated by:

$$\hat{\omega}(i,j)/\hat{\sigma}_x^6 = [\Sigma x_t^2 x_{t-i}^2 x_{t-j}^2/T]/[\Sigma x_t^2/T]^3 \tag{89}$$

This test is similar to the Tsay test (1986) for nonlinearitiy.

7. Forecasting from Nonlinear Stochastic Models

In this section, we study the forecasting methods based on stochastic models that present nonlinearity in variance.

Forecasting conditional volatility from ARCH models

Forecasts from ARCH models are constructed using similar procedure as in the linear ARMA model. For example, in a GARCH (1, 1) model, an n steps ahead forecast of future conditional volatility is constructed as follows (see Campbell *et al.*, 1997 or Gouriéroux, 1992).

First, we need the one step ahead forecast and it is computed as

$$\begin{aligned} E_t(\sigma_{t+1}^2) &= \alpha_0 + \alpha_1 E_t(e_t^2) + \beta\sigma_t^2 \\ &= \alpha_0 + \alpha_1\sigma_t^2 + \beta\sigma_t^2 \\ &= \alpha_0 + \sigma_t^2(\alpha_1 + \beta) \ . \end{aligned} \tag{90}$$

As in the one step forecast, the two steps forecast is written as

$$
\begin{aligned}
E_t(\sigma^2_{t+2}) &= \alpha_0+\alpha_1 E_t(\sigma^2_{t+1})+\beta E_t(\sigma^2_{t+1}) \\
&= \alpha_0+(\alpha_1+\beta)E_t(\sigma^2_{t+1}) \ . \qquad (91)
\end{aligned}
$$

By substituting the value found in (90) in equation (91), we get

$$
E_t(\sigma^2_{t+2})=\alpha_0(1+\alpha_1+\beta)+\sigma^2_t(\alpha_1+\beta)^2 \ . \qquad (92)
$$

Moreover, note that (92) might be written as

$$
\begin{aligned}
E_t(\sigma^2_{t+2}) &= \frac{(1+\alpha_1+\beta)(1-\alpha_1-\beta)}{1-\alpha_1-\beta}\alpha_0+\sigma^2_t(\alpha_1+\beta)^2 \\
&= \frac{1-(\alpha_1+\beta)^2}{1-\alpha_1-\beta}\alpha_0+\sigma^2_t(\alpha_1+\beta)^2 \ , \qquad (93)
\end{aligned}
$$

Finally, the n steps ahead forecast is computed simply by replacing the exponent 2 in (93) by n

$$
\begin{aligned}
E_t(\sigma^2_{t+n}) &= \frac{1-(\alpha_1+\beta)^n}{1-\alpha_1-\beta}\alpha_0+\sigma^2_t(\alpha_1+\beta)^n \\
&= \frac{\alpha_0}{1-\alpha_1-\beta}-\frac{\alpha_0(\alpha_1+\beta)^n}{1-\alpha_1-\beta}+\sigma^2_t(\alpha_1+\beta)^n \\
&= \frac{\alpha_0}{1-\alpha_1-\beta}+(\alpha_1+\beta)^n\left(\sigma^2_t-\frac{\alpha_0}{1-\alpha_1-\beta}\right). \qquad (94)
\end{aligned}
$$

Note that equation (94) might be used for computing forecasts at any horizon simply by replacing n by a value of interest. When $\alpha_1+\beta=1$, the conditional expectation of volatility n periods ahead is

$$
E_t(\sigma^2_{t+n})=\sigma^2_t+n\alpha_0 \qquad (95)
$$

The GARCH (1, 1) model with $\alpha_1+\beta=1$ has a unit autoregressive root, consequently today's volatility affects forecasts of volatility into the indefinite future. As presented in section 3, this is known as an integrated GARCH or IGARCH (1, 1) model. For higher-order GARCH, e.g. GARCH (p, q), multiperiod forecasts can be constructed in a similar fashion.

Forecasting conditional covariance from ARCH models

As we explained in section 3, forecasts of the covariance of period $t+k$ can be computed as

$$
E_t\left(\sigma_{i,j,t+k}\right)=\bar{\sigma}_{i,j}+(\alpha+\beta)^{k-1}\left(\sigma_{i,j,t+1}-\bar{\sigma}_{i,j}\right) \ . \qquad (96)
$$

The procedure to obtain this formula is very similar to what we have shown (equation 94) for the forecast of conditional variance (see Bhansali, 1998).

8. Modeling ultra-high-frequency data[46]

The models and estimation methods that we presented in previous sections can also be used to describe data observed at very high frequencies[47]. However, the main problem with this is the irregularity of information arrivals. For example, when we estimate a GARCH process on the S&P500, we generally use daily, weekly or monthly returns. This means that the interval between each observation is equal: a day, a week or a month. But when analyzing intra-day observations, like the IBM stock transactions, the information arrives sometimes in clusters and at different time intervals. This problem is called time deformation, because the economic time is not the same as the calendar time. As a consequence, we might observe a loss of information that could be important when using aggregated daily, weekly or monthly data. Considering the aggregation bias, it is important to account for the increases of information not only from an econometric, but also from a market microstructure theory perspective.

Another problem arises when analyzing intraday data, there is an 'intraday seasonality' problem. More specifically, intraday durations between transactions follow a U-shaped, which should be considered when working with these observations. This intraday seasonality can be accounted for using a spline function (Engle, 2000). Recently, Huptas (2009) investigated some nonparametric methods as alternative ways to tackle this problem. Therefore, researchers working with UHF financial data have some solutions for tackling such phenomenon. In our application, we use Engle's basic spline approach.

Recently, UHF data was used has an attempted to justified the 'volatility feedback', which is an alternative explanation to the theory of the leverage effect (Bolleslev *et al.*, 2007; Bolleslev *et al.*, 2009).

To provide further evidence regarding the dynamics of the leverage effect, the concept of realized volatility was generalized. This is known as *realized correlation,* which is implemented by means of realized Kernels (Barndorff-Nielsen *et al.*, 2008a,b; Getharal and Oomen, 2010). This method was applied to analyze UHF financial data. It was used to investigate the dynamics of the leverage effect based on UHF data on the VIX and the S&P500 (Russi, 2012).

---

[46] This section in an updated version of Racicot (2003). See also Racicot et al. (2007, 2008).

[47] It should be noted that a new subfield of physics called econophysics is interested in modeling intra-day transactions or UHF data. From the research in this field has emerge the concept of random matrix theory (RMT). Essentially, the RMT theory supposes, as a null hypothesis, that we have a random matrix C constructed from mutually uncorrelated time series. Deviations of the properties of C from a random matrix would show genuine correlations. Due to the fact that RMT predictions are universal, they can be applied to a wide class of systems. The stable distributions are also studied in Random Matrix Theory and financial econometrics fields of research. For instance, the Lévy stable distribution has fat tails, which is one property of financial time series. See Stanley, Gopikrishnan, Plerou, and Amaral (2000).

Here, we intend to investigate volatility computations using UHF data and to compare the realized volatility and the UHF-GARCH models. We also provide a way of using them for forecasting purposes and we discuss an application related to the valuation of volatility swaps. We leave aside the subject of UHF realized correlations for future research.

This section is organized as follows. Firstly, we present the ACD model and the UHF-GARCH model. Secondly, we discuss the parsimonious approach of the realized volatility. Then, we show our application of these models. Finally, we present a possible use of these volatility calculations to the pricing of volatility swaps.

8.1. The autoregressive conditional duration (ACD) model

The ADC model was firstly developed by Engle and Russel (1998). This model was improved and applied, in a similar context, by Jasiak (1999), Gouriéroux, Jasiak and Le Fol (1999), Gouriéroux and Jasiak (2001) and Engle (2000). The basic formulation of the ACD model is as follows. Let $x_i = t_i - t_{i-1}$, called the duration, be the interval between two arrival times. Also, let the expectation of the i^{th} duration be

$$E\left(x_i \mid x_{i-1}, \ldots, x_1\right) = \theta\left(x_{i-1}, \ldots, x_1; \psi\right) \equiv \theta_i \tag{97}$$

Assuming that

$$x_i = \theta_i e_i \tag{98}$$

where $\{e_i\} \sim i.i.d.$, ψ is a set of parameters to be estimated. The ACD class of models are functional forms for (97). The model takes its name from the fact that the conditional expectation in (97) depends on past durations.

A general formulation of (97) that has its roots in the ARMA process is called the ACD (p, q) given by

$$\theta_i = w + \sum_{j=0}^{p} \alpha_i x_{i-j} + \sum_{j=0}^{q} \beta_j \theta_{i-j} \tag{99}$$

where p and q are the orders of the lags. It is important to notice that this model is concerned only in modeling the arrival times. It can be used for studying the marks associated with the arrival times so that hypothesis from the market microstructure theories can be tested. A generalization of this model to accommodate both the arrival times and the prices jointly have been proposed by Engle (2000).

8.2. The UHF-GARCH model

Since this paper concerns nonlinear stochastic models, we conclude with an extension of a familiar model for the volatility. For the purpose of this report, we briefly discuss the ultra-high-frequency GARCH (UHF-GARCH) model.

Assuming that r_i is the return from transaction $(i-1)$ to i, the conditional variance per transaction can be defined as

$$V_{i-1}\left(r_i \mid x_i\right)=h_i \tag{100}$$

where x_i is defined as previously. The conditional variance depends on current and past returns and durations. Since volatility is always measured over a fixed time interval and frequently reported in annualized terms, the conditional volatility per unit of time is the most interesting measure to be evaluated. It is given by

$$V_{i-1}\left(\frac{r_i}{\sqrt{x_i}} \mid x_i\right)=\sigma_i^2 \tag{101}$$

which implies that the relation between (100) and (101) is

$$h_i=x_i\sigma_i^2 \tag{102}$$

Using relation (102), we can compute the forecasted conditional variance of transactions using

$$E_{i-1}\left(h_i\right)=E_{i-1}\left(x_i\sigma_i^2\right) \tag{103}$$

The familiar GARCH (1, 1) model presented in previous section can be extended to compute σ_i^2, which has the following form

$$\sigma_i^2= w+\alpha e_{i-1}^2+\beta\sigma_{i-1}^2+\gamma x_i^{-1} \tag{104}$$

where x_i^{-1} is the reciprocal of duration. According to the market microstructure model of Easley and O'Hara (1992), a fraction of the investors is informed and consequently knows if there is news concerning their assets. When it is time for the investors to do transactions, they will buy if the news is favorable, sell on bad news and they will make no transactions if there is no news. Thus, in this model, long intervals (x_i) will be interpreted as no news. This implies that in our model $\sigma_i^2$of and according to the hypothesis of Easley and O'Hara (1992), we expect a positive value for γ, because long durations indicate that there is no news and consequently a lower volatility. Note that with this formulation, long durations cannot induce the conditional variance to be negative. The usual maximum likelihood estimator might be used for estimating parameters w,α, β, γ. Other extensions of (104) can be formulated. One that seems promising is defined as follows

$$\sigma_i^2= \alpha_0+\alpha e_{i-1}^2+\beta\sigma_{i-1}^2+\gamma_1 x_i^{-1}+\gamma_2\frac{x_i}{\theta_i}+\gamma_3\xi_{i-1}+\gamma_4\theta_i^{-1} \tag{105}$$

where ξ_i is the long run volatility, θ_i is the conditional duration and might be defined by the parsimonious ACD (1, 1) model. Engle (2000) suggests computing the long run volatility by a Exponential Weighted Moving Average (EWMA) model on r^2/x as

$$\xi_i = \lambda \xi_{i-1} + (1-\lambda)\frac{r_{i-1}^2}{x_{i-1}} \tag{106}$$

In this extended model for computing volatility using high frequency data, the influences of durations on volatility have been incorporated in three parameters. These parameters measure, respectively, the effect of surprise in duration, the reciprocal duration and the expected reciprocal duration, which is the expected rate of arrivals of transactions. As in any other GARCH models, forecasting volatility can be found simply by computing the conditional expectation and it is given by

$$E_{i-1}(\sigma_i^2) = \alpha_0 + \alpha_1 e_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma_1 E_{i-1}(x_i^{-1}) + \gamma_2 + \gamma_3 \xi_{i-1} + \gamma_4 \theta_i^{-1} \tag{107}$$

This calculation reveals us that parameter γ_2 is not persistent. However, parameters γ_1 and γ_4 indicate a long run influence on future volatilities due to the persistence of the durations. These models might be estimated by QMLE (Quasi-Maximum Likelihood Estimator) without specifying the density of the disturbances. This is supported by the theorem of Bollerslev and Wooldrige (1992).

8.3. A more parsimonious approach for computing volatility using UHF data

Engle (2000) approach for modeling and computing volatility using high-frequency data seems promising on the theoretical side of the coin. However, this approach is complicated because there is lot of data manipulations, which must be done before having an estimate of volatility that might be used, for example, in daily option pricing.

The concept of *realized volatility* was firstly developed by Andersen and Bollerslev (1998) and applied for computing daily volatility forecasts of exchange rates and S&P 500 Index-Futures, respectively, by Bollerslev and Wright (2001) and Martens (2002). In other words, the realized volatility is measured by the squared value of intra-daily returns. This measure is also considered to be a more accurate measure of ex-post volatility. Assuming that the returns follow a special semimartigale process, Bollerslev and Wright (2001)[48] observe that 'the quadratic variation of this process constitutes a natural measure of ex-post realized volatility'. It also corresponds to the theoretical definition of volatility used in diffusion and stochastic volatility models[49]. A mathematical definition of realized volatility follows,

$$\sigma_I^2(m) = \frac{1}{n}\sum_{n=1}^{N} r_{m,n}^2 \tag{108}$$

where $r_{m,n}^2$ is the n^{th} squared return on day m. Due to the fact that the returns are not observed at a constant interval, the numbers of observations N will vary from day to

[48] *See also Andersen et al. (2003).*

[49] *For another application, see Barndorff-Neilson and Shephard (2001) and Hull and White (1987).*

day. Compared to the UHF-GARCH model, we can easily see the simplicity of the calculations required for obtaining an estimate of the volatility. As in the GARCH framework, it is possible to obtain a forecast of the realized volatility. The method might be described as follows. The forecasts are based on a long memory autoregressive model, where the lag p of the autoregressive process must approach infinity. The coefficients obtained from this autoregression are then used to construct a forecast function, which takes the following form

$$\sigma_I^2(m)=\frac{1}{n}\sum_{n=1}^{N}\left(\hat{\mu}+\hat{v}_{N(m-1)+n|N(m-1)}\right) \tag{109}$$

where $\hat{v}_{t+1|t}=\sum_{j=1}^{\infty}\hat{\alpha}_j v_{t-j}$ and

$$\hat{v}_{t+k|t}=\sum_{j=1}^{k-1}\hat{\alpha}_j v_{t+k-j|t}+\sum_{j=k}^{\infty}\hat{\alpha}_j v_{t+k-j} \tag{110}$$

The coefficients α_j might be estimated in the time domain by a long order autoregression[50], $v_t=log(r_t^2)-\hat{\mu}$, where $\hat{\mu}$ is the samplemean of $log(r_t^2)$. These coefficients might be also estimated in a frequency domain using a Wiener-Kolmogorov filter. The results from using either technique appear to be similar (Bollerslev and Wright, 2001). In the following application, we use the long order autoregression on the log-squared returns[51], which we assume to be a martingale difference. More precisely, $\alpha(L)\left(log(r_t^2)-\mu\right)=e_t$ where $\alpha(L)=1-\alpha_1 L-\alpha_2 L^2-\alpha_3 L^3-\ldots,\ e_t\sim WN(0,\sigma^2)$ and the lagged polynomial is assumed to converge. So to implement the forecasting formula represented by equation (103), we simply have to fit a long order autoregression to the log-squared returns and use this estimated equation to compute our forecasts. This point is made clearer in the following section. Since the log-squared returns may yield large negative numbers for returns close to zero, we applied the following transformation

$$r_t^*=log(r_t^2+\tau s^2)-\frac{\tau s^2}{r_t^2+\tau s^2} \tag{111}$$

where s^2 is the sample variance of r_t and τ is chosen to be equal to 0,02 (Fuller, 1996; Breidt and Carriquiry, 1996).

8.4. An application: comparing realized volatility to UHF-GARCH calculations using high frequency data

The purpose of this section is to give an application of the new models for high frequency data recently developed in the literature. More precisely, we will compare the UHF-GARCH model of Engle (2000) to the procedure for calculating volatility based on intra-day data developed by Bollerslev and Anderson (1998), applied by Bollerslev and Wright (2001) and Martens (2002).

[50] The time series of volatilities might be represented by an appropriate proxy such as the log-squared returns, which has an autoregressive representation.

[51] As alternative hypothesis, we might specify that the squared or absolute returns have an autoregressive representation. See Bollerslev and Wright (2001).

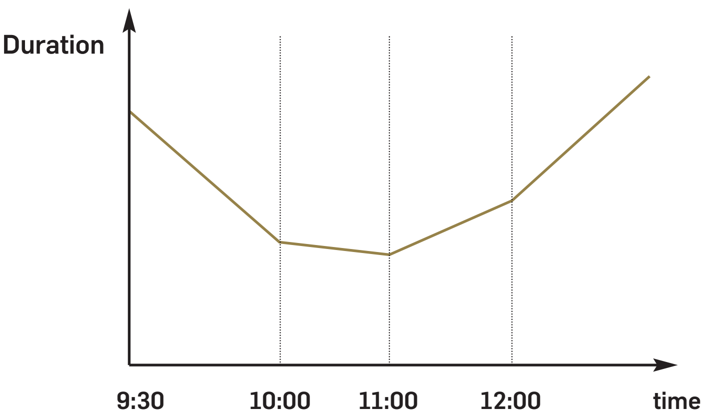
8.4.1. Data

The data set that we are using is the transactions quotes on IBM stocks[52] that are naturally irregularly spaced[53]. Two types of random variables compose the transaction data: the time of transactions and the marks at the time of transactions. In our application, a point[54] in time is the time at which a contract to trade some number of shares of IBM is traded. The marks are composed of volumes, prices and the available bid and ask prices of the contract at that time. Our data set is composed of 60,000 transactions traded on the New York Stocks Exchange on a time period, which stretches from November 1990 to January 1991. In our calculations, we proceeds as Engle (2000) and we use the trades that occur between 9:30am to 4:00pm[55]. In order to account for the calendar effect or the time of day effect, the data must be seasonally adjusted. This effect is represented by a higher frequency of transactions near open and close of the market. The adjusted duration is defined by

$$\tilde{x}_i = \frac{x_i}{\varphi(t_{i-1}; \beta)} \tag{112}$$

where $x_i = t_i - t_{i-1}$ is the duration between trades and $\varphi(.)$ is a piecewise linear spline function used to seasonally adjust the durations. Figure 9 gives an illustration of a linear spline.

Figure 9. **Linear spline of the U-shaped intra-day duration**



[52] *We used the same sample of observations as in Engle (2000), firstly, because the purpose of this section is the comparison of Engle's (2000) model has a known benchmark with the integrated volatility concept; we already know how this model behave in this context. Secondly, as explained in Dacorogna et al. (2001), it is not easy to find a reliable source of data in high-frequency finance (e.g. no errors in the variables like unintentional errors, intentional errors and system errors).*

[53] *The fact that the transactions are irregularly spaced creates heteroskedasticity.*

[54] *It refers to point process.*

[55] *The trades that occur on Thanksgiving Friday, on Christmas Eve, on New Year or overnight durations are not considered, here.*

As shown in this figure, the knots are the points where the linear pieces of the splines join together. They appeared at times 9:30, 10:00, 11:00, 12:00, 1:00, 2:00, 3:00, 3:30. Specifically, the seasonal adjustment[56] is done by regressing the durations on the time using a linear spline[57] specification, which takes the following form

$$x = c + \beta_1 t_1 + \beta_2 t_2 + \beta_3 t_3 + \beta_4 t_4 + \beta_5 t_5 + \beta_6 t_6 + \beta_7 t_7 + \beta_8 t_8 + e \tag{113}$$

where t_{i-1} for i=2,…,9 are vectors of time variables constructed from the knots. From this regression we obtain $\hat{x}_i = \varphi(t_{i-1}; \hat{\beta})$. The resulting variable $\tilde{x}_i$, which is free of the typical time of day effect, represents fractions of durations below or above normal.

8.4.2. Comparing volatility calculations

As we mentioned above, the maximum likelihood estimator is used to estimate the parameters of all of our UHF-GARCH models. After correcting the intraday seasonality, these models are easy to estimate by using standard software like EViews. For example, with a simple command in the programming language of that software, one can estimate equation (105) as follows

Exhibit
Eviews programs for the Extended ACD GARCH model

```
equation ACD.arch(c,h) sqr(duree)
acd.makegarch eduree
genr r=rends
param c(1) -.004 c(2).2 c(3) -.6 c(4) .4 c(5).3 c(6) .5 c(7).5 c(8) -.05 c(9) .1 c(10) .1
equation ACD_UHF .arch(h,s,c=.001, m=200) r ar(1) ma(1) duree @ 1/duree duree/eduree longuevola(–1)
1/duree
```

where the variable *duree* is our seasonally adjusted duration variable, *rends* is the return defined as $r_i = log\,\Delta((bid_i + ask_i)\ /\ 2)$ also adjusted for the time of day effect[58] as in the durations[59]. The *longuevola* is obtained by computing the mean of squared returns. To modify the simple GARCH, we used the option @ and then add the variable that we think might affect the calculation of volatility per seconds.

[56] *For nonparametric methods to estimate the intraday seasonality, see Huptas (2009).*

[57] *Note that we used a linear spline. We might have used a k^{th}-order spline, which is a piecewise polynomial approximation; with polynomials of degree k differentiable k-1 times everywhere. For example, a cubic spline is a spline of order three and is a piecewise polynomial differentiable twice everywhere. At each knot point the slopes must match and the curvatures from each side must also match. A cubic spline is given by the formula $s(\tau)=\sum_{i=0}^{3} a_i \tau_i + \frac{1}{3!}\sum_{p=1}^{n-1} b_p(\tau-\varepsilon_p)_+^3$ where $(\tau-\varepsilon_p)_+^3 = \max(\tau-\varepsilon_p, 0)$. As we can see, $s(.)$ is linear combination of τ_i and $(\tau-\varepsilon_p)^p$. In general, a k^{th} degree spline has $n+k$ parameters, where n is the number of knot points. A useful way of writing the function $s(.)$ is called the Bspline, noted $B(.)$. It consists in finding a set of basis function and in representing the general splines as a linear combination of them. See James and Weber (2000).*

[58] *It must also be adjusted because volatility is known to have a daily configuration (Engle, 2000).*

[59] *It is called the $(log\,\Delta)$ midquote. This is supposed to be a better measure of the $(log\,\Delta)$ price, because it reduces the econometric issue of bid-ask bounce and price discreteness (Engle, 2000).*

To make a comparison between the two methods for computing daily volatility, we have to consider that the UHF-GARCH gives volatility calculations per seconds and the realized volatility gives a volatility estimate for a day. We have to transform one of the models into comparable units. For example, to compute the values of *one-day options* on electricity[60], which requires a measure on a daily basis that uses the intra-day movements of the underlying, we have to transform the UHF-GARCH calculation on a daily basis. A way to proceed is by analogy of the realized volatility calculations. More precisely, we suggest averaging the intra-day volatilities to obtain a daily volatility calculation as

$$\sigma_d^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \qquad (114)$$

where σ_i^2 is obtained by estimating high-frequency GARCH models. At Table 6, we present a comparison between different methodologies for computing daily volatility. Using our intra-daily transactions on IBM stock for the first week of our sample, we compute the volatilities for five consecutive days, beginning on a Thursday in November 1990[61]. Thus, our assumption for comparing the volatility computations seems to work well.

Table 6. Estimations of daily volatility based on realized volatility and GARCH models

Day	Realized volatility	Simple GARCH	ACD GARCH	Extended ACD GARCH	Number of observations
Thursday	3.18	3.68	3.69	3.24	688
Friday	9.13	12.04	12.17	7.49	792
Monday	2.59	3.23	3.04	3.15	671
Tuesday	6.16	6.97	6.96	5.76	732
Wednesday	3.58	3.62	3.62	3.27	649

In fact, we can see that all the GARCH calculations follow quite closely the realized volatility methodology, which is reassuring. As explained in Bollerslev and Wright (2001) and as shown at Table 7, the simple GARCH has the worst performance compared to the realized volatility for high-frequency data.

Table 7. Average absolute percentage changes

Day	Simple GARCH	ACD GARCH	Extended ACD GARCH
Thursday	15.72%	16.04%	1.89%
Friday	31.87%	33.30%	17.96%
Monday	24.71%	17.37%	21.62%
Tuesday	13.15%	12.99%	6.49%
Wednesday	1.12%	1.12%	8.66%
Average	17.31%	16.16%	11.32%

[60] *For an introduction on this subject, see Wilmott (2000) or Pilipovic (1998).*

[61] *We have chosen this specific segment of time, simply for comparisons of calculations.*

The Extended ACD GARCH (equation 105) seems to have the best performance among the GARCH models in our comparison. To have a better idea on the performance of these volatility models, one can compute standard measures such as the R-squared of the Mincer-Zarnowitz (1969) regression as in Bollerslev and Wright (2001) or Martens (2002). We follow the same approach. In the next section, we present a comparison of forecasts based on our four models.

8.4.3. Comparing volatility forecasts

Our objective is to make forecasts based on our four models and then to compare them with the resulting Mincer-Zarnowitz R-squared[62]. This method is simply to obtain the R-squared from the regression of realized values of our variable on its forecasted ones. Before making the actual comparison, we explain how we proceed to obtain these forecasts based on the four models, namely the realized volatility, the simple GARCH, the ACD GARCH and the extended ACD GARCH.

The formula given by equation (110) is based on a long order autoregressive model. Assuming that we want to make forecasts based on data observed at fixed intervals, for example, every 5 minutes, then a forecasted value for the end of the next day is given by

$$\hat{v}_{t+288|t} = \hat{a}_1\hat{v}_{t+287|t} + \hat{a}_2\hat{v}_{t+286|t} + \hat{a}_3\hat{v}_{t+285|t} + \ldots + \hat{a}_{288}v_t + \hat{a}_{289}v_{t-1} + \hat{a}_{290}v_{t-2} + \ldots \qquad (115)$$

where the subscript index $t+288$, which means that there is 288 intervals of five minutes in one day. As we can see, (115) is simply a high order autoregressive process. Thus, our forecasts can be based on the estimation of that process.

The computation of forecasts from a simple GARCH (1, 1) can be done by using the formula given by equation (104).

The forecasts based on the ACD or extended ACD GARCH models are complicated due to the fact that we need expected values of durations. This problem might be bypassed by assuming that this expression has the same types of representation as the conditional durations. It can be represented by a simple ARMA process[63], which is the approach suggested by Engle and Russel (1998). First, we forecast the values of the durations based on the ARMA process and then, we include these values in the ACD GARCH, or in the extended ACD GARCH model. However, this manipulation increases

[62] *Other popular measures might be used, such as the mean absolute error (MAE), the root mean squared error (RMSE), the heteroskedasticy adjusted mean absolute error (HMAE), or the heteroskedasticy adjusted root mean squared error (HRMSE). HMAE is defined as* $\frac{1}{T}\sum_{t=1}^{T}\left|\left(1-\frac{Realized_t}{Forecast_t}\right.\right.$ *and* $HRMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\left(1-\frac{Realized_t}{Forecast_t}\right)^2\right.}$*, where the forecasted errors are adjusted for heteroskedasticity. For an application of the last two measures, see Andersen et al. (1999), Martens (2002).*

[63] *This is because we know that x_i is related to θ_i.*

AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2012. 5: 162-221
Notes on Nonlinear Dynamics. Racicot, F.E.

the number of computations that we have to perform to obtain a forecast. Another possibility would be to assume that the expected durations become constant after the last realized value, but this assumption appears to be quite unrealistic. The third approach suggested involves assuming that the durations can be represented by a log linear regression models that would look like this $x_i = e^{z_i \beta + \varepsilon_i}$. This approach leads to other similar types of modeling like the Cox proportional hazard model or the Weibull model. Here, we decide to follow the first approach. To be more specific, the conditional duration might be expressed in the form of an ARMA process. If we use an ARMA (1, 1) then it is defined by

$$x_i = v + \alpha x_{i-1} + \beta \varepsilon_{i-1} + \varepsilon_i \tag{116}$$

where $\varepsilon_i \equiv x_i - \theta_i$, which is a Martingale difference by definition (i.e. $\varepsilon_i = x_i - E_{i-1}(x_i)$). Forecasted values of x_i can be obtained from (116) and included in equation (107). Table 8 shows forecasts evaluation based on the GARCH models in comparisons of the realized volatility.

Table 8. **Forecasts evaluation of GARCH models and realized volatility**

Number of observations	Simple GARCH	ACD GARCH	Extended ACD GARCH	Realized Volatility
700	RMSE : 13.12	RMSE : 13.11	RMSE : 10.94	RMSE : 2.26
	MAE : 3.84	MAE : 3.84	MAE : 3.91	MAE : 2.03
	R2 : 0.0002	R2 : 0.0001	R2 : 0.0006	R2 : 0.0044
1400	RMSE : 14.83	RMSE : 14.83	RMSE : 12.47	RMSE : 2.23
	MAE : 4.35	MAE : 4.35	MAE : 4.34	MAE : 1.99
	R2 : 0.0001	R2 : 0.0001	R2 : 0.0003	R2 : 0.0024
2100	RMSE : 15.55	RMSE : 15.54	RMSE : 13.02	RMSE : 2.17
	MAE : 4.31	MAE : 4.31	MAE : 4.32	MAE : 1.97
	R2 : 0.00008	R2 : 0.00008	R2 : 0.0002	R2 : 0.0015

Comparing the RMSE[64] to the MAE or to the R^2 of the Mincer-Zarnowitz[65] (1969) regression, the realized volatility method outperforms all the GARCH models. It should be also noted that none of the numbers presented in this table are significant. However, in the case of realized volatility, the t statistics of Mincer-Zarnowitz are near significance

[64] *Alexander (2001) disagrees with the common usage of criteria such as RMSE in the context of volatility forecast evaluation. The author observed that this criterion should be used only for mean forecast evaluation. In the context of volatility forecast evaluation, we should view this measure as a simple distance metric.*

[65] *The Mincer-Zarnowitz regressions are obtained by regressing the ex post realized values of the variable under scrutiny on the forecasted values of this variable plus a constant term. In our case, we forecasted the IBM prices for different sample sizes: 700, 1400 and 2100, and then we did the regressions of the related realized values on the forecasted ones, including a constant term (i.e. $y_t^* = c + \beta y_{ft}^* + \varepsilon_t$). The resulting R^2 are shown in Table 8.*

level[66]. The poor performance of all the suggested methods for forecasting volatility is not surprising. When using high frequency data, simple GARCH models are notoriously known to perform badly (Bellerslev and Wright, 2001). The performance of the simple realized volatility method is better than the ACD GARCH one. The combination of these two models could improve the forecasting power. This model combination is suggested, for instance, in Donaldson and Kamstra (1997). They add an Artificial Neural Network component to a standard GARCH and they apply this model to forecast the S&P500. Their model shows improvement in the forecasting power over standard volatility models. Martens (2002) makes a similar suggestion in his financial econometrics application on futures.

8.4.4. A possible application of UHF models in finance: The case of variance and volatility swaps

In section 3.1.3, we briefly described the process of the VIX calculation and discovered that it uses a realized volatility estimate. The pricing of volatility swaps on the VIX actually use this estimate to do so. We should start by describing the pricing of a variance swap. Let S be the underlying security, a variance swap with a notional amount N can be represented by (Neftci, 2008)

$$V(T_1,T_2)=[\sigma^2_{T_1,T_2}-F^2_{t_0}](T_2-T_1)\,N \qquad (117)$$

where $\sigma^2_{T_1,T_2}$ is a measure of realized variance rate of S_t, $t \in [T_1,T_2]$ and it can be viewed as a floating rate that will be observed only at T_2. F_{t_0} is the fixed volatility rate of S_t and is quoted at time t.

Here, we shall take a closer look at the floating and fixed legs of this swap.

The floating leg of this swap is given by

$$\sigma^2_{T_1,T_2}(T_2-T_1)=lim_{\delta\to 0}\,\Sigma^n_{i=1}\left(\frac{S_{t_i}-S_{t_{i-1}}}{S_{t_i}}-\mu\delta\right)^2=\int_{T_1}^{T_2}\left(\frac{1}{S_t}dS_t-\mu\delta\right)^2=\int_{T_1}^{T_2}\sigma_t^2 \qquad (118)$$

where $\delta = t_i - t_{i-1}$. Equation (118) is similar to the one that we presented for the realized variance (see equation 108). We could also suggest another estimator, which is given by equation (114). This estimator can be seen as an approximation of (108).

For the fixed leg, we determine the F_{t_0}, which gives the fair value of the variance swap. Note that the variance swap is designed so that its fair value is equal to 0 at time t_0. Thus, $F^2_{t_0}$ is the variance (value) that makes the fair value of the swap equal to zero.

[66] *The t statistics are 1.75 (0.08), 1.82 (0.06) and 1.80 (0.07) for sample sizes of 700, 1400 and 2100, respectively. Their corresponding p-values are given in parenthesis.*

From the fundamental theorem of asset pricing[67], we can find the proper measure $\tilde{P}$, which is the risk-neutral measure that gives

$$E_{t_0}^{\tilde{P}}\left[\sigma_{T_1,T_2}^2 - F_{t_0}^2\right](T_2 - T_1)N = 0 \tag{119}$$

Assuming that markets are complete and the continuously compounded risk-free spot rate r is constant. Then, the money market account (i.e. it give $1\$$ at T_2) can be used for normalization. Rearranging (119) obtains

$$F_{t_0}^2 = E_{t_0}^{\tilde{P}}\left[\sigma_{T_1,T_2}^2\right] \tag{120}$$

Substituting (118) in (120) obtains

$$F_{t_0}^2 = \frac{1}{(T_2 - T_1)} E_{t_0}^{\tilde{P}}\left[\int_{T_1}^{T_2}\left(\frac{1}{S_t} dS_t - \mu\delta\right)^2\right] \tag{121}$$

The integral inside the expectation of (121) can be evaluated using

$$F_{t_0}^2 = \frac{1}{(T_2 - T_1)} E_{t_0}^{\tilde{P}}\left[\Sigma_{i=1}^{n}\left(\frac{S_{t_i} - S_{t_{i-1}}}{S_{t_i}} - \mu\delta\right)^2\right] \tag{122}$$

The risk-neutral expectation can be evaluated via Monte Carlo simulation[68].

Volatility swaps can be valued similarly as (117). For instance, a volatility swap on the VIX, which trades on the Chicago Futures Exchange, can be valued as follows (McDonald, 2006)

$$v(T_1, T_2) = \left[VIX_{T_1,T_2} - F_{t_0}\right](T_2 - T_1)N \tag{123}$$

The difference is that we might not suggest a UHF methodology for valuating (123). This could be the case here, but as we have seen UHF models might be of some use for the valuation of volatility swaps. We leave that subject for further investigation.

9. Conclusion

We have reviewed several econometric and chaotic models that can be used as DGP's of financial time series. In the applied finance literature, econometric models have been

[67] *The fundamental theorem of asset pricing implies that if some state prices (Q^i) exist, then the prices that we are evaluating (S_{kt_0}) are arbitragefree. The theorem implies three important results: 1) The risk-neutral (risk-adjusted) probabilities are obtained from the state prices; 2) All properly normalized asset prices have a Martingale property under the selected synthetic probability $\tilde{p}^k$. If y_t is a stochastic process that has the property $y_t = E_t^{\tilde{p}k}[y_t]$ then, y_t is a Martingale; 3) Every synthetic probability leads to a particular expected return for the asset price under consideration (Neftci, 2008).*

[68] *For pricing variance swaps via Monte Carlo simulation, see Rostan et al. (2012).*

more popular than the chaotic ones. One reason that might explain this is the fact that models and tests that have emerged from chaos theory are generally built on multidimensional spaces, which complicates their computations. A successful related application can be found in the literature. It concerns the Brock, Dechert and Scheinkman (1996) or the BDS test. This test is quite powerful for detecting several types of nonlinearities, particularly the nonlinear deterministic ones.

The econometric models like the GARCH can be used not only to fit a particular series, but also for pricing derivatives. We have also discussed applications of these models to data observed at very high frequency. We observed that it is possible to use ARCH models to forecast volatility at very high frequency using a simple modification to account for the irregularity of information arrivals. The usual maximum likelihood method was used to estimate the parameters of the UHF-GARCH. We have also presented an application comparing UHF-GARCH models to the realized volatility concept. In terms of RMSE, MAE and Mincer-Zarnowitz (1969) criterions, the realized volatility has a better performance than any of the UHF-GARCH models proposed by Engle and Russel (1998) and by Engle (2000). The developments of the idea of using UHF data to do econometric inferences have not stopped there. New research has emerged and the concept of realized volatility has been extended to model bivariate phenomenon; this is referred to as the realized correlation (Russi, 2012). This idea uses an extension of the realized kernels developed by Barndorff-Nielsen *et al.* (2008a). The application of realized kernels to non-synchronous trading seems to yield relevant results (Barndorff-Nielsen *et al.*, 2008b), because it provides an efficient use of the information (Gatheral and Oomen, 2010).

Using nonparametric methods, as realized kernels, applied to compute volatility or correlation seems to be a good addition to our econometric tools. It provides an alternative or a validation process to our basic parametric GARCH or MGARCH. As we presented, we used the univariate EGARCH and MGARCH models to test for the leverage effect and the volatility feedback, respectively. Both of the parametric models seem to confirm previous research.

Further research should be done to investigate if the use of aggregated data, instead of non-synchronous ones, creates a bias that could results in bad inferences. In our research, we followed Engle (2000) approach and used irregularly spaced UHF data to compute both realized volatility and UHF-GARCH. However, most studies (e.g. Bollerslev *et al.*, 2009) use some sort of aggregated data that is regularly spaced data observed over a five minutes range. When using irregularly spaced data some sort of correction must be done to the parametric model used in a similar fashion, as in Engle (2000). All things considered, if a researcher decides to use irregularly spaced UHF data to compute some sort of UHF-Multivariate GARCH correlation; then, which corrections should he performed to the model? Further research should be done to investigate this issue.

Acknowledgements

I would like to thank R. J. Wood and F. H. Sarlan for several useful discussions on the subject of nonlinear dynamics and financial markets. I acknowledge the research assistance of Ana Fernández Rodríguez.

References

- Alexander, C.O. and Chibumba, A.M. (1996). Multivariate orthogonal factor GARCH, *University of Sussex Discussion Papers in Mathematics*.
- Alexander, C.O. (2001). *Market Models*, Wiley, New York.
- Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.
- Amemiya, T. (1994). *Introduction to Statistics and Econometrics*, Harvard University Press, Cambridge, Massachusetts.
- Andersen, T.G. and Bollerslev, T. (1998). Ansering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**, pp. 885-905.
- Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica,* **71**, pp. 529-626.
- Andersen, T.G., Bollerslev, T. and Lange, S. (1999). Forecasting financial market volatility: Sampling frequency vis-à-vis forecast horizon, *Journal of Empirical Finance*, **6**, pp. 457-477.
- Anselin, L. (2001). Spacial Econometrics in: *A companion to Theoretical Econometrics*, Blackwell, Massachusetts.
- Ashley, R., Patterson, D. and Hinich, M.J. (1986). A diagnostic test for nonlinear serial dependence in time series fitting errors, *Journal of Time Series Analysis*, **7**, pp. 165-178.
- Barndorff-Nielsen, O.E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck models and their uses in financial economics, *JASA*, forthcoming.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2008a). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise, *Econometrica*, **76**, pp. 1481-1536.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2008b). Multivariate realized kernels: Consistent semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Oxford-Man Institute Working Paper*, pp. 1-42.
- Berndt, E.K., Hall, B.H., Hall, R.E. and Hausman, J.A. (1974). Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement,* **4**, pp. 653-665.

- Bhansali, W. (1998). *Pricing and Managing Exotic and Hybrid Options*, McGraw-Hill, New York.
- Bhargava, A. (1986). On the theory of testing for unit roots in observed time series, *Review of Economic Studies*, **53**, pp. 369-384.
- Black, F. (1976). Studies in stock price volatility changes, Proceedings of the 1976 Business Meeting of the Business and Economic statistics section, *American Statistical Association*, pp. 177-181.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, pp. 637-654.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, pp. 307-327.
- Bollerslev, T., Chou, R. Y. and Kroner, K.F. (1992). ARCH modelling in finance, *Journal of Econometrics,* **52**, pp. 5-59.
- Bollerslev, T., Engle, R. F. and Wooldridge, J.M. (1988). A capital asset pricing model with time varying covariances, *Journal of Political Economy*, **96**, pp. 116-131.
- Bollerslev,T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *The Review of Economics and Statistics*, **72**, pp. 498-505.
- Bollerslev, T., Engle, R.F. and Nelson, D.B. (1993). ARCH models, *UCSD Working paper*, pp. 93-49.
- Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews*, **11**, pp. 143-172.
- Bollerslev, T. and Wright, J.H. (2001). High-frequency data, frequency domain inference, and volatility forecasting, *Review of Economics and Statistics*, **83**, pp. 596-602.
- Bollerslev, T., Litvinova, J. and Tauchen, G. (2006). Leverage and volatility feedback effects in high frequency data, *Journal of Financial Econometrics*, **2**, pp. 1-37.
- Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia, *Review of Financial Studies*, **22**, pp. 4463-4492.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Holden- Day, California, Oakland.
- Breitung, J. (2002). Nonparametric tests for unit roots and cointegration, *Journal of Econometrics*, **108**, pp. 343-363.
- Breidt, F.J. and Carriquiry, A.L. (1996). Improved quasi-maximum likelihood estimators for stochastic volatility models, In: A. Zellner and J.S. Lee (Eds.), *Modeling and Prediction: Honoring Seymour Geisel*, Springer-Verlag, New York.
- Brock, W. A., Dechert, W. and Scheinkman, J. (1987). A test for independence based on the correlation dimension, *Working Paper*, University of Wisconsin, Madison, University of Houston, and University of Chicago.

- Brock, W.A., Hsieh, D.A. and LeBaron, B. (1991). *Nonlinear Dynamics, Chaos, and Instability*, MIT Press, Massachusetts.
- Brock, W.A., Scheinkman, J.A., Dechert, W.D. and LeBaron, B. (1996). A test for independence based on the correlation dimension, *Econometric Reviews*, **15**, pp. 197-235.
- Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2ed., Springer-Verlag, New York.
- Brooks, C. (2008), *Introductory econometrics for finance*, 2 ed., Cambridge University Press, Cambridge.
- Byström, H. (2000). Orthogonal GARCH and covariance matrix forecasting and a Stress Scenario: The nordic stock markets during the Asian financial crisis 1997-98, *Discussion paper*, Department of Economics, Lund University.
- Campbell, J.Y., Lo, A.W. and Mackinlay, A.C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.
- Christie, A. (1982). The stochastic behavior of common stock variances : Value, leverage and interest rate effects, *Journal of Financial Economics*, **10**, pp. 407-432.
- Dacorogna, M.M., Gençay, R., Müller, U., Olsen, R.B. and Pictet, O.V. (2001). *An Introduction to High-Frequency Finance*, Academic Press, London.
- Davidson, R. and Mackinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York.
- Dendney, A.K. (1991). Leaping into Lyapunov space, *Scientific American,* **265**, pp. 178-180.
- Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, pp. 427-431.
- Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, **49**, pp. 1057-1072.
- Donaldson, R.G. and Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance*, **4**, pp. 17-46.
- Duan, J.-C. (1995). The GARCH option pricing model, *Mathematical Finance*, **5**, pp. 13–32.
- Duan, J.-C.(1996). Craking the smile, *RISK*, **9**, pp. 55-59.
- Duan, J.-C., Gauthier, G. and Simonato, J.G. (1999). An analytical approximation for the GARCH option pricing model, *Journal of Computational Finance*, **2**, pp. 75-116.
- Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment, *Journal of Finance*, **47**, pp. 905-927.
- Enders, W. (2004). *Applied EconometricTime Series*, sec. ed., Wiley, New York.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, **50**, pp. 987-1008.

- Engle, R.F. (1995). *ARCH: Selected Readings*, Advanced Texts in Econometrics, Oxford University Press, Oxford.
- Engle, R.F., and Russel, J.R. (1998). Autoregressive conditional duration : A new model for irregularly spaced transaction data, *Econometrica*, **66**, pp. 1127-1162.
- Engle, R.F.(2000). The econometrics of ultra-high-frequency data, *Econometrica*, **68**, pp. 1-22.
- Engle, R.F.(2002). Dynamic conditional correlation: A simple class of multivariate GARCH models, *Journal of Business and Economic Statistics*, **20**, pp. 339-350.
- Engle, R.F. and Kroner, K. (1995). Multiplicative simultaneous generalized ARCH, *Econometric Theory*, **11**, pp. 122-150.
- Figlewski, S. and Wang, X. (2000). Is the leverage effect a leverage effect?, *Technical Report*, NYU Stern School of Business and City University of Hong Kong.
- Foerster, S.R. and Sapp, S.G. (2006). The changing role of dividends: a firm-level study from the nineteenth to the twenty-first century, *Canadian Journal of Economics*, **39**, pp. 1316-1344.
- Franses, P.H. and Van Dijk, D. (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.
- Fuller, W.A. (1996). *Introduction to Statistical Time Series*, Wiley, New York.
- Haug, E.G. (2007). *Derivatives: Models on Models,* Wiley, New York.
- Huptas, R. (2009). Intraday seasonality in analysis of UHF financial data: Models and their empirical verification, *Dynamic Econometrics Models*, **9**, pp. 129-140.
- Getheral, J. and Oomen, R.C.A. (2010). Zero-intelligence realized variance estimation, *Finance and Stochastics*, **14**, pp. 249-283.
- Glosten, L.R., R. Jaganathan and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, **48**, pp. 1779-1802.
- Gouriéroux, C. and Monfort, A. (1990). *Série Temporelles et Modèles Dynamiques*, Economica, Paris.
- Gouriéroux, C. (1992). *Modèles ARCH et Applications Financières*, Economica, Paris.
- Gouriéroux, C., Jasiak, J. and Le Fol, G. (1999). Intra-day market activity. *Journal of Financial Markets*, **2**, pp. 193-206.
- Gouriéroux, C. and Jasiak, J. (2001). Durations, in: *A Companion to Theoretical Econometrics*, chap. 21, Blackwell, Massachusetts.
- Gouriéroux, C. and Jasiak, J. (2001), *Financial Econometrics*, Princeton Series in Finance, Princeton University Press, Princeton.
- Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors, *Physica*, **9**D, pp. 189-208.

- Greene, W.H. (2008). *Econometric Analysis*, 6th ed., Prentice-Hall, New Jersey.
- Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press, New Jersey.
- Hentschel, L. (1995). All the family : Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics*, **39**, pp. 71-104.
- Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with application to bond and currency options, *Review of Financial Studies*, **6**, pp. 327-343.
- Heston, S.L. and Nandi, S. (2000). A closed-form GARCH option valuation model, *Review of Financial Studies*, **31**, pp. 585-625.
- Hinich, M.J. (1982). Testing estimation and linearity of a stationary time series, *Journal of Time Series Analysis*, **3**, pp. 169-176.
- Hull, J.C. (2012). *Options, Futures, and Other Derivatives*, 7th Edition, Prentice-Hall, New Jersey.
- Hull, J.C. and White, A. (1987) The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **42**, pp. 381-400.
- Jasiak, J. (1999). Persistence in intertrade durations, *Finance*, **19**, pp. 166-195.
- Jäckel, P. (2002). *Monte Carlo Methods in Finance*, Wiley, Chichester.
- James, J. and Webber, N. (2000). *Interest Rate Modelling*, Wiley, New York.
- Khemiri, R. (2012). Volume and volatility in foreign exchange market microstructure: a Markov switching approach, *Applied Financial Economics*, **22**, pp. 1121–1133.
- Kraft, D.F. and Engle, R.F. (1983). Autoregressive conditional heteroskedasticity in multiple time series, *Unpublished manuscript*, Department of Economics, University of California, San Diego.
- Koutmos, G. (1996), Modelling the dynamic interdependence of major european stock markets, *Journal of Business Finance & Accounting*, **23**, pp. 975-988.
- Kwaitkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics,* 54, pp. 159-178.
- Lawrence, A.J. and Lewis, P.A.W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with Discussion), *Journal of the Royal Statistical Society*, Series B, **47**, pp. 165-202.
- Lee, T.H., White, H. and Granger, C.W.J. (1993). Testing for neglected nonlinearity in time series models : A comparison of neural network methods and alternative tests, *Journal of Econometrics* , **56**, pp. 269-290.
- MacDonald, R. (2006). *Derivatives Markets*, Sec. ed., Pearson.
- Mackey, M. and Glass, L. (1977). Oscillation and chaos in physiological control systems, *Science*, **50**, pp. 287-289.

- Manda, K. (2010). Stock market volatility during the 2008 financial crisis, *Working Paper*, Glucksman Institute for Research in Securities Markets, pp. 1-32.
- Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business*, **36**, pp. 392-417.
- Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency Data, *Journal of Futures Markets*, **22**, pp. 497-518.
- McLeod, A.I. and W.K. Li (1983), 'Diagnostic checking ARMA time series models using a squared residual autocorrelations', *Journal of Time Series Analysis*, **4**, pp. 169-176.
- Milhøj, A. (1987b), A multiplicative parameterization of ARCH models, *Unpublished manuscript*, Departement of Statistics, University of Copenhagen.
- Mills, T.C. (1990). *Time Series Techniques for Economist*, Cambridge University Press, Cambridge.
- Mills, T.C. (1999). *The Econometric Modelling of Financial Time Series*, Sec. ed., Cambridge University Press, Cambridge.
- Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts, in: J. Mincer (Ed.), *Economic Forecasts and Expectations*, NBER, Cambridge, pp. 3-46.
- Mun, J. (2006). *Real Option Analysis*, Wiley, New York.
- Neftci, S. (2008). *Principles of Financial Engineering*, Sec. ed., Academic Press, New York.
- Nelson, D.B. (1990). Conditional heteroskedasticity in asset returns : A new approach, *Econometrica*, **59**, pp. 347-370.
- Pagan, A. (1996). The econometrics of financial markets, *Journal of Empirical Finance,* **3**, pp. 15-102.
- Peters, E. (1991). *Chaos and Order in the Capital Markets*, Wiley, New York.
- Peters, E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, Wiley, New York.
- Pilopovic, D. (1998). *Energy Risk*, McGraw-Hill, New York.
- Priestley, M. (1980). State-dependent models : A general approach to non-linear time series analysis, *Journal of Time Series Analysis*, **1**, pp. 47-71.
- Priestly, M. (1981). *Spectral Analysis and Time Series*, Academic Press, London.
- Racicot, F.E. (2000). Notes on nonlinear dynamics, *Working Paper*, No. 16-2000, CRG-ESG, UQAM.
- Racicot, F.E, Théoret, R. and Coën, A. (2008). Forecasting irregularly spaced UHF financial data: Realized volatility vs UHF-GARCH models, *Int Adv Econ Res*, **14**, pp. 112-124.
- Racicot, F.E, Théoret, R. and A. Coën (2007). Forecasting UHF financial data: Realized volatility versus UHF-GARCH models, *Int Adv Econ Res*, **13**, pp. 243-244.
- Racicot, F.E. (2009). Volatility, In: *Encyclopedia of Alternative Investment*, CRC Press, London, p. 516.

- Racicot, F.E. and Théoret, R. (2009). Integrating volatility factors in the analysis of hedge fund alpha puzzle, *Journal of Asset Management*, **10**, pp. 37-62.

- Racicot, F.E. (2011). Low-Frequency Components and the Weekend Effect Revisited: Evidence from Spectral Analysis, *AESTIMATIO,The IEB International Journal of Finance*, **3**, pp. 8-27.

- Rachev, S.T., Mittnik, S., Fabozzi, F.J., Focardi, S.M. and Jasic, T. (2007). *Financial Econometrics : From Basics to Advanced Modeling Techniques*, Wiley, New York.

- Rostan, P., Rostan, A., Ait El Trach, A. and Mercier, S. (2012). Pricing variance and volatility swaps: a Monte Carlo simulation technique benchmarked to two closed-form solutions, *AESTIMATIO, The IEB International Journal of Finance*, **5**, pp. 32-57.

- Rouah, F.D. and Vainberg, G. (2007), *Option princing models & volatility models using Excel-VBA*, Wiley, New York.

- Russi, G. (2012). Estimating the Leverage effect using High frequency data, *Review of Economics & Finance*, **1**, pp. 1-24.

- Salazar, J. and Lambert, A. (2010). Fama and McBeth revisited : A critique, *AESTIMATIO, The IEB International Journal of Finance*, **1**, pp. 48-71.

- Sims, C. (1984). Martingale-like behavior of prices and interest rates, *Working Paper*, University of Minnesota.

- Stanley, H.E., Gopikrishnan, P., Plerou, V. and Amaral, L.A.N. (2000). Quantifying Fluctuations in Economic Systems by Adapting methods of Statistical Physics, *Physica A*, **287**, pp. 339-361.

- Stigum, M. (1990). *The Money Market*, Third ed., Dow Jones-Irwin, Homewood.

- Sun, Y. and Wu, X. (2010). A nonparametric study of dependence between S&P500 index and the market volatility index (VIX), *Beijing: EFMA symposium on Asian finance*, April, pp. 1-21.

- Taylor, S.J. (1986). *Modeling Financial Time Series*, Wiley, Chichester.

- Tsay, R. (1986). Nonlinearity tests for time series, *Biometrica*, **73**, pp. 464-466.

- Tsay, R. (2005). *Analysis of financial time series*, Wiley, New York.

- Wilmott, P. (2000). *Paul Wilmott on Quantitative finance*, vol. 2, Wiley, Chichester.

- White, H. (2001). *Asymptotic Theory for Econometricians*, Sec. ed., Academic Press, Orlando.