Available online at www.sciencedirect.com
SCIENCE DIRECT

ELSEVIER Journal of Economic Dynamics & Control 27 (2003) 1699–1737

JOURNAL OF Economic Dynamics & Control

www.elsevier.com/locate/econbase

Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements

Eric Jondeau[a], Michael Rockinger[a,b,c,*]

[a] *Banque de France, 41-1391 DGE, 31, rue Croix des Petits Champs, 75049 Paris Cedex 01, France*
[b] *HEC-School of Management, Department of Finance, 78351 Jouy-en-Josas, France*
[c] *CEPR, France*

Abstract

Recent portfolio-choice, asset-pricing, value-at-risk, and option-valuation models highlight the importance of modeling the asymmetry and tail-fatness of returns. These characteristics are captured by the skewness and the kurtosis. We characterize the maximal range of skewness and kurtosis for which a density exists and show that the generalized Student-*t* distribution spans a large domain in the maximal set. We use this distribution to model innovations of a GARCH type model, where parameters are conditional. After demonstrating that an autoregressive specification of the parameters may yield spurious results, we estimate and test restrictions of the model, for a set of daily stock-index and foreign-exchange returns. The estimation is implemented as a constrained optimization via a sequential quadratic programming algorithm. Adequacy tests demonstrate the importance of a time-varying distribution for the innovations. In almost all series, we find time dependency of the asymmetry parameter, whereas the degree-of-freedom parameter is generally found to be constant over time. We also provide evidence that skewness is strongly persistent, but kurtosis is much less so. A simulation validates our estimations and we conjecture that normality holds for the estimates. In a cross-section setting, we also document covariability of moments beyond volatility, suggesting that extreme realizations tend to occur simultaneously on different markets. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C22; C51; G12

Keywords: Volatility; Skewness; Kurtosis; Generalized Student-*t* distribution; GARCH; Stock indices; Exchange rates; SNOPT

* Corresponding author. Tel.: +33-1-39677259; fax: +33-1-39677085.
E-mail address: rockinger@hec.fr (M. Rockinger).

PII: S0165-1889(02)00079-9

1. Introduction

This paper investigates the existence and persistence of conditional skewness and kurtosis of various financial series taken at the daily frequency. To do so, we build on Hansen (1994), who proposes a GARCH model, where residuals are modeled as a generalized Student-t distribution.[1] The generalized t distribution is asymmetric and allows fat-tailedness. We express skewness and kurtosis of Hansen's GARCH model as a function of the parameters of the generalized t distribution. These functions characterize the conditional evolution of skewness and kurtosis for a given dynamic specification of the underlying parameters.

A further theoretical contribution is the characterization, conditional on kurtosis being finite, of the largest possible domain of skewness and kurtosis for which a density exists with a zero mean and a unit variance. We achieve this characterization using results from the so-called Hamburger (1920) problem. These results go back to Stieltjes (1894) and are also related to the little Hausdorff (1921a, b) problem. Conditional on the assumption that kurtosis and, hence, skewness exist, we show that the moments of the generalized t distribution are within the maximal range of skewness and kurtosis. An advantage of using the generalized t distribution is that moments may become infinite. As a consequence, it is possible to determine those periods when higher moments cease to exist.

In financial applications it may be necessary to simulate from a generalized Student t. We show how such a simulation can be implemented. We use this random number generator to establish normality of the estimates of a well specified model.

The model is inspired by Engle (1982) and Bollerslev (1986) in the way volatility is described. For the generalized t distribution, the asymmetry and the fat-tailness parameters are modeled as a function of lagged innovations. Given that the generalized t distribution is defined over a certain range of the parameters only, it is necessary to impose some constraints on the dynamic specification. We discuss in detail possible specifications of the parameters. Eventually, we settle for an autoregressive structure where parameters are constrained via a logistic map. The estimations are made using a recent sophisticated sequential quadratic optimization algorithm implemented in SNOPT, developed by Gill et al. (1997, 1999).

Given the way the parameters appear in this type of model, we show that an autoregressive specification of the parameters may yield spurious persistence. Therefore, it may seem that series have persistence in skewness and kurtosis where there is none. We present an estimation strategy that takes care of this difficulty. In a numerical simulation, we show that the potential of misinterpretation is huge.

We estimate the model for five stock-index returns and six foreign-exchange returns sampled at a daily frequency. The foreign-exchange series start on July 26, 1991, the stock indices on August 23, 1971. All series end on May 14, 1999. We investigate the existence and persistence of moments in a time-series context and the comovements of moments in a cross-section context.

[1] More recent contributions are by Premaratne and Bera (1999) who build on the Pearson type IV family, and by Rockinger and Jondeau (2002) who use entropy densities.

Our findings are of importance for the various strands of finance literature which emphasize the importance of moments beyond the second one. Within a Capital Asset Pricing Model, interest in skewness and kurtosis goes back to the theoretical work by Rubinstein (1973) and its first empirical implementation by Kraus and Litzenberger (1976). Further work in that area is by Friend and Westerfield (1980); Barone-Adesi (1985); Sears and Wei (1985, 1988), and more recently by Tan (1991); Fang and Lai (1997), Korkie et al. (1997); Kan and Zhou (1999), Harvey and Siddique (1999, 2000). For emerging countries, Hwang and Satchell (1999) model risk premia using higher moments. With our model and estimation technique, it is possible to extend these models to conditional versions.

Another strand of literature, initiated by Mandelbrot (1963) and Fama (1963), suggests that the density of asset returns has tails such that moments beyond the first one do not exist. Even though most empirical work, e.g., of Loretan and Phillips (1994), shows that moments up to the third exist in unconditional returns, there remains the issue of the existence of conditional moments. This issue may be easily addressed with our model. It involves a set of parameters which are related to skewness and kurtosis. We estimate our model under the constraints that a density exists. The data then decides if, furthermore, skewness and even kurtosis exist. We show that skewness appears to exist most of the time, but not necessarily kurtosis. Our work is, therefore, also related to extreme value theory. An advantage over extreme value theory is that we are able to make a statement concerning the entire distribution rather than only one for the tails.

We show that all foreign-exchange series obey a similar dynamics of the asymmetry parameter, whereas the fat-tailedness parameter is found to be constant. For stock-index returns, various types of dynamics emerge. Some series behave similarly as exchange rates. Other series allow for a complex, often difficult to interpret, dynamic of the fat-tailedness parameter. These results suggest that most of the tail-fatness of financial data is generated by large repeatedly occurring events of a given sign. The dynamics of skewness is straightforward to interpret, whereas the one of kurtosis is difficult to understand. This finding suggests that theoretical models should be able to incorporate conditional skewness and kurtosis. They should also be able to deal with infinite kurtosis.

We also perform tests of adequacy of our selected models and show that the retained model provides a better fit of returns than models where higher moments are held constant. The gain in fit may be significant.

The structure of this paper is as follows. In Section 2, we present the generalized t distribution and we compute the first four moments of this distribution. We relate the issue of existence of moments to the little Hausdorff problem and discuss the set of skewness-kurtosis pairs which are generated by the generalized t distribution. Next, we describe our general model and we indicate how such a model may be estimated. In Section 3, we present the data. In Section 4, we discuss the parameter estimates and address several other specification issues. In Section 5, we consider the existence and persistence of conditional skewness and kurtosis. Then, we consider in a multi-country setting the covariability between markets of moments beyond volatility. In Section 6, we conclude with directions for further research. Analytical results and a Monte-Carlo experiment are reported in the appendix.

2. A model for conditional skewness and kurtosis

2.1. The generalized t distribution

Our model builds on the GARCH model of Engle (1982) and Bollerslev (1986).[2] Within this class of models, it is well known that residuals are non-normal. This result has led to the introduction of fat-tailed distributions. Nelson (1991) considers the generalized error distribution. Bollerslev and Wooldridge (1992) consider the case of a Student-*t* distribution.[3] Engle and Gonzalez-Rivera (1991) model the residual distribution non-parametrically. Even though these contributions recognize the fact that errors have fat tails, they do not render the tails time varying, i.e., the parameters of the error distribution are assumed to be constant over time.

Hansen (1994) is the first author to propose a model which allows for conditional higher moments. He achieves this by introducing a generalization of the Student-*t* distribution where asymmetries may occur, while maintaining the assumption of a zero mean and unit variance. By assuming that parameters are dependent on past realizations, he shows that parameters, and thus higher moments, may be made time varying.[4] In the finance literature, Harvey and Siddique (1999) introduce a non-central Student-*t* distribution. This distribution allows the modelization of skewness, but does not allow for a separate variation of skewness and kurtosis. Premaratne and Bera (1999) use the Pearson family of distributions.

The density of Hansen's generalized *t* distribution (GT) is defined by

$$gt(z|\eta,\lambda) = \begin{cases} bc\left(1+\dfrac{1}{\eta-2}\left(\dfrac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } z < -a/b, \\ bc\left(1+\dfrac{1}{\eta-2}\left(\dfrac{bz+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } z \geqslant -a/b, \end{cases} \tag{1}$$

where

$$a \equiv 4\lambda c\frac{\eta-2}{\eta-1}, \qquad b^2 \equiv 1+3\lambda^2-a^2, \qquad c \equiv \frac{\Gamma((\eta+1)/2)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)}.$$

If a random variable Z has the density $gt(z|\eta,\lambda)$, we will write $Z \sim GT(z|\eta,\lambda)$. Inspection of the various formulae reveals that this density is defined for $2<\eta<\infty$ and $-1<\lambda<1$. Furthermore, this density encompasses a large set of conventional densities. For instance, if $\lambda=0$, Hansen's distribution reduces to the traditional Student-*t*

[2] The literature concerning GARCH models is huge. Several reviews of the literature are available, i.e., Bollerslev et al. (1992), Bera and Higgins (1993), and Bollerslev et al. (1994).

[3] For a definition of the traditional Student-*t* distribution, see, for instance, Mood et al. (1982).

[4] Hansen does not discuss the link between parameters and higher moments.

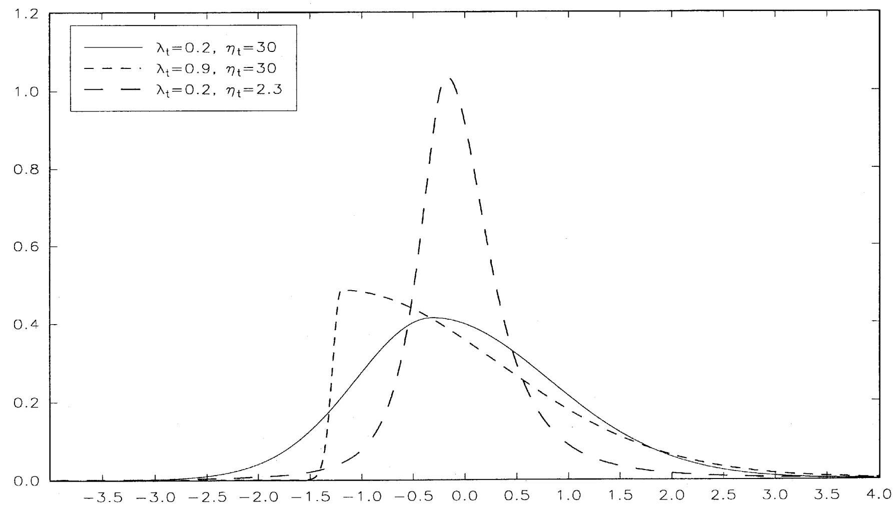


Fig. 1. Shape of the generalized t distribution.

distribution. We recall that the traditional Student-t distribution is not skewed. If in addition $\eta=\infty$, the Student-t distribution collapses to a normal density. Fig. 1 displays various densities obtained for different values of λ and η. We notice that λ controls skewness: If λ is positive, the probability mass concentrates in the right tail.

It is well known that the traditional Student-t distribution with η degrees of freedom allows the existence of all moments up to the ηth. Therefore, given the restriction $\eta > 2$, Hansen's distribution is well defined and its second moment exists. The higher moments are not given directly by the parameter η, but we were able to obtain formulae for these moments.

We show in Appendix A that, if $Z \sim GT(z|\eta,\lambda)$, then Z has zero mean and unit variance. Furthermore, in the appendix, we derive the formulae for skewness and kurtosis. Introducing the notations

$$m_2 = 1 + 3\lambda^2,$$

$$m_3 = 16c\lambda(1+\lambda^2)\,\frac{(\eta-2)^2}{(\eta-1)(\eta-3)} \quad \text{if } \eta > 3,$$

$$m_4 = 3\,\frac{\eta-2}{\eta-4}(1+10\lambda^2+5\lambda^4) \quad \text{if } \eta > 4,$$

we obtain that

$$\mathrm{E}[Z^3] = [m_3 - 3am_2 + 2a^3]/b^3, \tag{2}$$

$$\mathrm{E}[Z^4] = [m_4 - 4am_3 + 6a^2m_2 - 3a^4]/b^4. \tag{3}$$

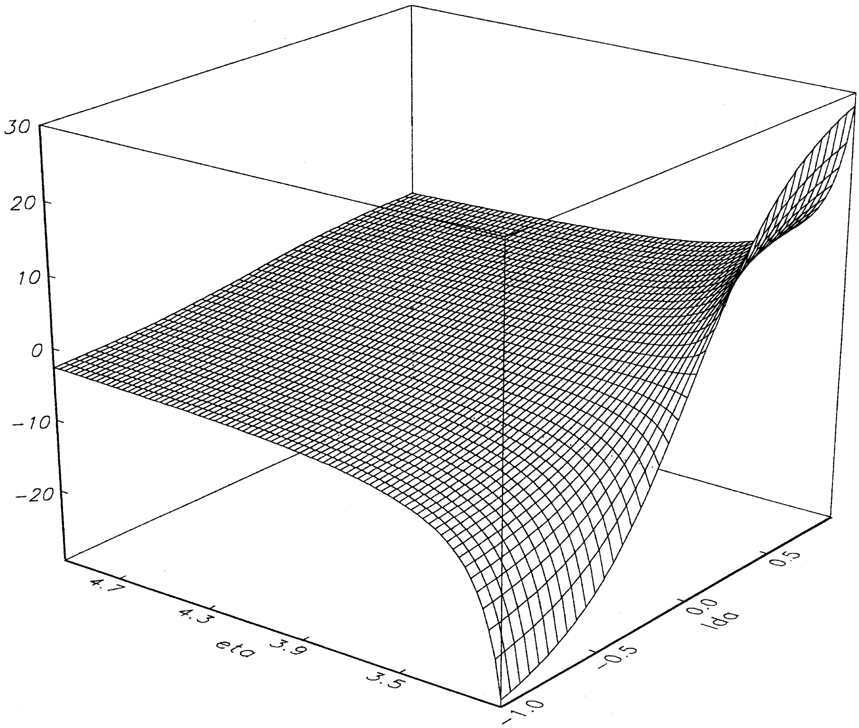


Fig. 2. Skewness surface for values of λ and η.

We recall that skewness and kurtosis are defined as

$$S[Z] = \mathrm{E}\left[\frac{(Z - \mathrm{E}[Z])^3}{(\mathrm{Var}[Z])^{3/2}}\right], \qquad K[Z] = \mathrm{E}\left[\frac{(Z - \mathrm{E}[Z])^4}{(\mathrm{Var}[Z])^2}\right].$$

Hence, since Z has zero mean and unit variance, we directly obtain for skewness $S[Z] = \mathrm{E}[Z^3]$, and for kurtosis $K[Z] = \mathrm{E}[Z^4]$.

We notice, at this stage, that the density and the various moments do not exist for all parameters. Given the way asymmetry is introduced, we must have $-1 < \lambda < 1$. The density g exists if $\eta > 2$. From Eq. (2), it follows that skewness exists if $\eta > 3$ and, from Eq. (3), we obtain that kurtosis is defined if $\eta > 4$.[5]

We define as $\mathscr{D}$ the domain $(\eta, \lambda) \in]2, +\infty[\times]-1, 1[$. Given these restrictions on the underlying parameters, it is clear that the range of skewness and kurtosis will also be restricted to a certain domain. Figs. 2 and 3 trace the skewness and kurtosis surface for given values of λ and η. Focusing on Fig. 2, we notice that, as η approaches 3, skewness becomes very large. On the other hand, when one slightly increases η, say

[5] In empirical applications, we will only impose that $\eta > 2$ and let the data decide for itself if for a given time period a given moment exists.

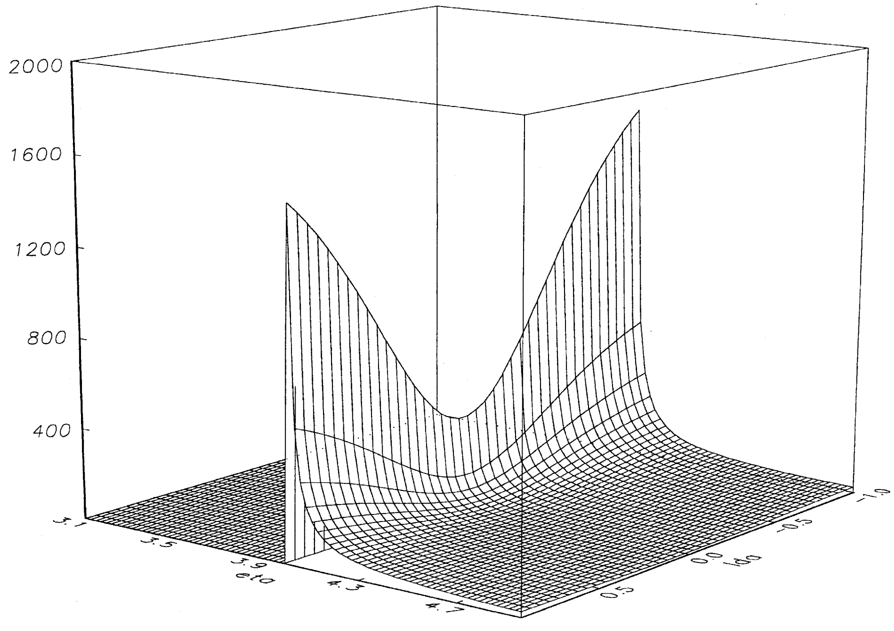


Fig. 3. Kurtosis surface for values of λ and η.

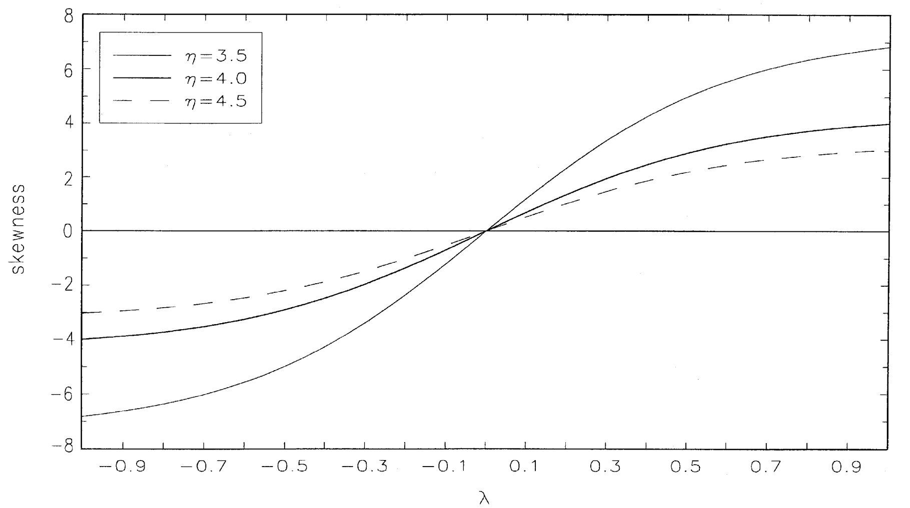


Fig. 4. Skewness for various values of η.

beyond 4, the surface strongly levels out. When we consider kurtosis, displayed in Fig. 3, we verify a degeneracy as η reaches its boundary value of 4. To get a better feel for the range of skewness as λ varies between -1 and 1, we trace in Fig. 4 various curves corresponding to selected values of η. For the case where η takes the value 4.5, i.e. kurtosis exists, we notice that skewness is rather restricted ranging between -3 and 3. If η decreases, kurtosis ceases to exist and skewness may take larger and larger values before it also ceases to exist. Symmetrically, if η increases, as the tails of the density become thinner, the range of skewness becomes smaller.

This last picture illustrates the fact that, for a given level of kurtosis, only a finite range of skewness exists. This feature raises the question of existence of a density for given moments. We address this question in the next section.

2.2. The moment problem

The question of the existence of a non-decreasing function α for a sequence of non-central moments μ_j, $j = 1, \ldots, M$, such that

$$\mu_j = \int_l^u z^j \, d\alpha(z) \tag{4}$$

has already been addressed in the functional analysis literature. There are essentially two approaches. The first approach is discussed in Widder (1946). The case $l = 0$ and $u = \infty$ has been investigated by Stieltjes (1894) and was motivated by a problem issued from physics.[6] The case $l=0$ and $u=1$ has been studied by Hausdorff (1921a, b), and is called the little Hausdorff problem. The situation of interest for us, $l=-\infty$ and $u=+\infty$, has been studied by Hamburger (1920). A second approach to the moment problem is discussed in Baker and Graves-Morris (1996) and involves Padé approximants.

Concerning the uniqueness of a solution to Eq. (4), Widder (1946) provides an example that demonstrates the multiplicity of a solution. Next, there is the question of conditions that must be satisfied by μ_j to ensure existence of a solution to Eq. (4). The answer is that the sequence μ_j must be positive definite (Widder, 1946, p. 134, Theorem 12.a). This means that the following number and sequence of determinants must satisfy

$$\mu_0 \geqslant 0 \qquad \begin{vmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix} \geqslant 0, \qquad \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} \geqslant 0, \ldots \, .$$

In particular, for the four-moment problem, with $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_2 = 1$, this implies the following relation between skewness μ_3 and kurtosis μ_4:

$$\mu_3^2 < \mu_4 - 1 \quad \text{with } \mu_4 > 0. \tag{5}$$

This relation confirms that, for a given level of kurtosis, only a finite range of skewness may be spanned.

Fig. 5 displays the skewness–kurtosis boundary ensuring the existence of a density. The curve ABC corresponds do the theoretical domain of maximal size (5). The curve DEF corresponds to the domain of skewness and kurtosis, which is attainable with a

[6] The first moment appears as a center of gravity and the second moment is interpreted as the inertia.

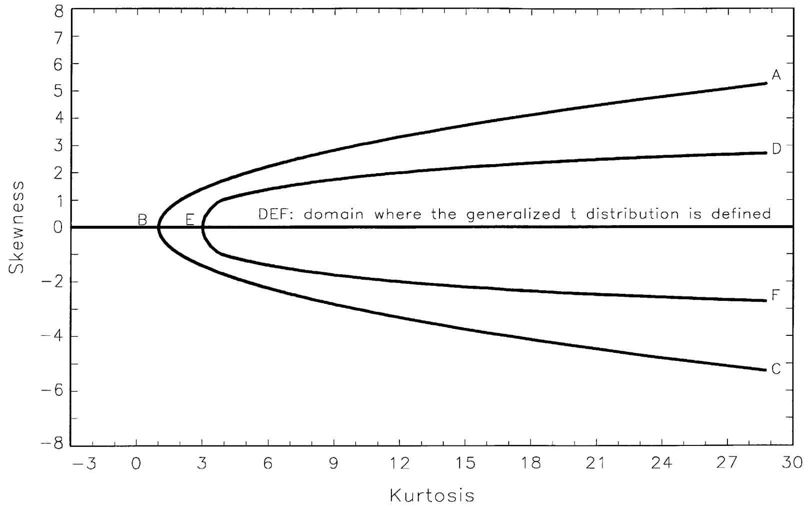


Fig. 5. Skewness–kurtosis boundary for the generalized t distribution.

generalized t distribution, assuming $\eta > 2$.[7] We notice that the kurtosis is bounded from below by 3, indicating that the generalized t distribution does not allow tails to be thinner than those of the normal distribution.

We will call $\mathscr{E}$ the domain corresponding to DEF in Fig. 5, i.e., that is spanned by skewness and kurtosis if both moments exist. We notice that the relation between $\mathscr{D}$ and $\mathscr{E}$ is not bijective. In particular, those points that are located in $\mathscr{D}$ but where $\eta < 4$ have no counterpart in $\mathscr{E}$. The logic is that there are points in $\mathscr{D}$ where skewness or kurtosis cease to exist, whereas in $\mathscr{E}$ skewness and kurtosis are finite by construction. It is only when one reduces the domain $\mathscr{D}$ to $]4, +\infty[\times] - 1, 1[$ that the relation is bijective.

Given these conditions on the parameters of the generalized t distribution, to ensure existence of moments, we now consider our general model.

2.3. A model for time-varying skewness and kurtosis

Let r_t, for $t = 1, \ldots, T$, be realizations of a variable of interest. For exchange-rate and stock-market data, this variable will be a log-return.[8] We assume that

$$r_t = \mu_t + y_t, \tag{6}$$

$$y_t = \sigma_t z_t, \tag{7}$$

[7] To construct the curve DEF, we notice that λ intervenes as a square in the expression of kurtosis. On the other hand, the skewness expression is linear in λ. Therefore, skewness will be symmetric in that changing λ to $-\lambda$ will change S into $-S$. As a consequence, we only construct the upper bound of DEF by taking a fine grid for η and by selecting a boundary value for λ, such as 0.999. The lower bound is obtained symmetrically.

[8] If S_t is the value of the index on date t, we define $r_t = 100 \ln(S_t/S_{t-1})$.

Table 1
Possible specifications of the model. g represents the logistic map

Model M1:	$\begin{cases} \eta_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^-, \\ \lambda_t = a_2 + b_2^+ y_{t-1}^+ + b_2^- y_{t-1}^-. \end{cases}$
Model M2:	$\begin{cases} \tilde{\eta}_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^-, \\ \tilde{\lambda}_t = a_2 + b_2^+ y_{t-1}^+ + b_2^- y_{t-1}^-, \\ \eta_t = g_{[2,30[}(\tilde{\eta}_t),\ \lambda_t = g_{]-1,1[}(\tilde{\lambda}_t). \end{cases}$
Model M3:	$\begin{cases} \eta_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^- + c_1 \eta_{t-1}, \\ \lambda_t = a_2 + b_2^+ y_{t-1}^+ + b_2^- y_{t-1}^- + c_2 \lambda_{t-1}. \end{cases}$
Model M4:	$\begin{cases} \tilde{\eta}_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^- + c_1 \tilde{\eta}_{t-1}, \\ \tilde{\lambda}_t = a_2 + b_2^+ y_{t-1}^+ + b_2^- y_{t-1}^- + c_2 \tilde{\lambda}_{t-1}, \\ \eta_t = g_{[2,30[}(\tilde{\eta}_t),\ \lambda_t = g_{]-1,1[}(\tilde{\lambda}_t). \end{cases}$
Model M5:	$\begin{cases} \tilde{\mu}_{3t} = a_1 + b_1 y_{t-1}^3, \\ \tilde{\mu}_{4t} = a_2 + b_2 y_{t-1}^4, \\ (\mu_{3t}, \mu_{4t}) = G(\tilde{\mu}_{3t}, \tilde{\mu}_{4t}), \\ (\eta_t, \lambda_t) = F^{-1}(\mu_{3t}, \mu_{4t}). \end{cases}$
Model M6:	$\begin{cases} \tilde{\mu}_{3t} = a_1 + b_1 y_{t-1}^3 + c_1 \tilde{\mu}_{3t-1}, \\ \tilde{\mu}_{4t} = a_2 + b_2 y_{t-1}^4 + c_2 \tilde{\mu}_{4t-1}, \\ (\mu_{3t}, \mu_{4t}) = G(\tilde{\mu}_{3t}, \tilde{\mu}_{4t}), \\ (\eta_t, \lambda_t) = F^{-1}(\mu_{3t}, \mu_{4t}). \end{cases}$

$$\sigma_t^2 = a_0 + b_0^+ (y_{t-1}^+)^2 + b_0^- (y_{t-1}^-)^2 + c_0 \sigma_{t-1}^2, \tag{8}$$

$$z_t \sim GT(z_t | \eta_t, \lambda_t). \tag{9}$$

Eq. (6) decomposes the return at time t into a conditional mean, μ_t, and an innovation, y_t. In Eq. (7), we define this innovation as the product between conditional volatility, σ_t, and a residual, z_t. Eq. (8) determines the dynamics of volatility. We use the notation $y_t^+ = \max(y_t, 0)$ and $y_t^- = \max(-y_t, 0)$. Such a specification has been suggested by Glosten et al. (1993), and by Zakoïan (1994). In Eq. (9), we specify that residuals follow a generalized t distribution with time-varying parameters (η_t, λ_t).

We now wish to discuss various possible specifications for the dynamics of η_t and λ_t. In Table 1, we display several specifications that could be used to describe η_t and λ_t. Many other specifications could be given, involving further lags or less linear relations. We emphasize these specifications, since they will highlight the difficulties that may be encountered. We assume that the coefficients are such that stability of the dynamics is guaranteed.

Model M1 specifies directly η_t and λ_t as functions of past positive and negative realizations. The advantage of this specification is that no further non-linear map is required to obtain a description of the parameters. A drawback of this specification is that its estimation is cumbersome since the constraints $2 < \eta_t$ and $-1 < \lambda_t < 1$ must be numerically imposed. Furthermore, nothing guarantees, out of sample, that η_t and λ_t will be well defined. Ad-hoc techniques, such as truncation at the boundaries could be devised for forecasting purposes.

Model M2 specifies a dynamic for unconstrained $\tilde{\eta}_t$ and $\tilde{\lambda}_t$. These unrestricted parameters get mapped into the authorized domain $\mathscr{D}$ via a logistic map. Many of the drawbacks of model M1 disappear. However, one consequence is that the impact of extreme realizations gets dampened since the logistic map tends to flatten the response of variables located in its tails.

Model M3 specifies parameters η_t and λ_t as an auto-regressive structure. Such a specification suffers from a severe drawback. For data with sufficient variability, as the sample increases, the model is likely to degenerate to a solution where $b_2^+ = b_2^- = 0$. To understand why this is so, consider the simpler specification $\lambda_t = a_2 + b_2 y_{t-1} + c_2\lambda_{t-1}, \lambda_0$ given. For $|c_2| < 1$, we may write as an approximation, if t is sufficiently large, that

$$\lambda_t = a_2/(1 - c_2) + b_2 \sum_{s=0}^{\infty} c_2^s y_{t-1-s}.$$

From this expression, we see that λ_t has mean $a_2/(1 - c_2)$ and variance $b_2^2 \operatorname{Var}[\sum_{s=0}^{\infty} c_2^s y_{t-1-s}]$. This shows that the restriction $-1 < \lambda_t < 1$ will only be satisfied if $b_2 = 0$, otherwise, with some probability the constraint will be violated for some observation.[9]

Model M4 is similar to M3, yet, it uses a non-linear map to constrain the parameters to $\mathscr{D}$. The model is estimable, yet, some care must be taken in the interpretation of the estimates. For instance, if one estimates M4 and finds, for instance, that b_1^+ and b_1^- are not statistically different from 0, in this case the model reduces to

$$\tilde{\eta}_t = a_1 + c_1\tilde{\eta}_{t-1}.$$

At this stage, it may even be that c_1 is statistically significant. This may lead to the conclusion that there is persistence in the $\tilde{\eta}_t$. Such a conclusion would be erroneous however. Indeed, if actual observations y_{t-1} do not matter, then, starting from some initial $\tilde{\eta}_0$, the series of $\tilde{\eta}_t$ will quickly converge to its stationary level given by

$$\tilde{\eta}^* = a_1/(1 - c_1).$$

In other words, a model where one would have estimated $\tilde{\eta}_t = \tilde{\eta}^*$ (with $b_1^+ = b_1^- = c_1 = 0$) could not be distinguished from the one obtained earlier. This implies that there exists an entire class of parameters (a_1, c_1), all satisfying $(1 - c_1)\tilde{\eta}^* = a_1$, for which the model's characteristics are indistinguishable. The algorithm converges to one solution at random. In the appendix we present results from a Monte-Carlo simulation showing

[9] We verified this result by implementing model M2 with the constraints $2 < \eta_t$ and $-1 < \lambda_t < 1$ imposed numerically. The algorithm either converged to a solution with $b_2 = 0$ or aborted with a message hinting to a degeneracy of the problem.

that in this case the null hypothesis $c_1 = 0$ can be rejected as often as 50% of the estimations. To avoid this type of spurious finding, it is recommended to estimate M2 before M4 and to verify that past observations affect $\tilde{\eta}_t$ or $\tilde{\lambda}_t$. In no way should one trust in an estimation where c_1 or c_2 is statistically significant, yet, the parameters on the lagged innovations are not statistically significant.

We found that a further diagnostic to detect this behavior consists in changing the value of $\tilde{\eta}_1$ or of the initial value of the parameters in the numerical estimation. If the algorithm converges to significantly different values, then care should be taken.

In specification M5, the third and fourth non-central moments, μ_{3t} and μ_{4t}, get specified using actual observations. Leaving aside the observation made by Korkie et al. (1997) that a specification involving standardized innovations leads to less findings of spurious persistence in the third non-central moment, for the model to be well defined; it must be that μ_{3t} and μ_{4t} belong to the domain $\mathscr{E}$. This implies a potentially highly non-linear map, G that maps some unrestricted $\tilde{\mu}_{3t}$ and $\tilde{\mu}_{4t}$ into $\mathscr{E}$. Furthermore, in order to obtain λ_t and η_t from μ_{3t} and μ_{4t}, it is necessary to invert a highly non-linear map that we call F in the table. Even though such an inversion could be done in theory, it will lead to a slow algorithm and also to a rather unstable estimation since the analytic computation of gradients may not be feasible. Last, with such a specification, it is implicitly assumed that skewness and kurtosis are finite at each point of time. This observation is at odds with results from extreme value theory. These observations suggest that modeling directly skewness and kurtosis may be eventually the right thing to do. However, given the many difficulties this estimation involves, we will settle for a less complex parameterization.

Model 6 presents the same difficulties as M5 with the added complication discussed already for model M3, in that one may find spurious dependence of skewness due to a lack of significant b_1 or b_2 estimates.

Given these theoretical considerations, we decided to build an estimation strategy where we build up from an unconditional model to model M4, estimating model M2 as an intermediated step.

2.4. Monte-Carlo experiments

It may be necessary to simulate returns that are distributed following the above presented model. Such simulations may be useful for the purpose of validating the estimation of the model.[10] In another application, trajectories of prices could be simulated. These trajectories could be further used to stress testing financial models. In this section, we show how to simulate data distributed as a generalized Student-t.

First, we recall that the conventional Student-t distribution is defined by

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{1}{\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2},$$

[10] In Appendix C, we show that if the estimated model is well specified, the conventional asymptotic normality of the maximum-likelihood estimator holds.

where n is the degree-of-freedom parameter. Numerical evaluation of the cumulative distribution function (cdf) of the conventional Student-t is well known. Procedures are provided in most software packages, and in particular in Fortran, which is the language used in this study. We write the cdf of a Student-t with n degrees of freedom as

$$A_n(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x.$$

The following proposition presents the cdf of the generalized t distribution.

Proposition 1. *Defining* $D(t) = \Pr[Z < t]$, *where* Z *follows the density* (1), *yields*

$$D(t) = \begin{cases} (1-\lambda)A_\eta\left(\dfrac{bt+a}{1-\lambda}\sqrt{\dfrac{\eta}{\eta-2}}\right) & \text{if } t < -a/b, \\[2ex] (1+\lambda)A_\eta\left(\dfrac{bt+a}{1+\lambda}\sqrt{\dfrac{\eta}{\eta-2}}\right) - \lambda & \text{if } t \geqslant -a/b. \end{cases} \tag{10}$$

Proof. Suppose that $t < -a/b$. Given the definition of $D(t)$, we have

$$\begin{aligned} D(t) &= \int_{-\infty}^{t} bc\left(1 + \frac{1}{\eta-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} \mathrm{d}z \\ &= (1-\lambda)\int_{-\infty}^{(bt+a)/(1-\lambda)} \frac{\Gamma((\eta+1)/2)}{\Gamma(\eta/2)} \frac{1}{\sqrt{\pi(\eta-2)}}\left(1+\frac{u^2}{\eta-2}\right)^{-(\eta+1)/2} \mathrm{d}u \\ &= (1-\lambda)A_\eta\left(\frac{bt+a}{1-\lambda}\sqrt{\frac{\eta}{\eta-2}}\right). \end{aligned}$$

The second equation follows from a change of variable involving $u = (bz+a)/(1-\lambda)$. The last equation follows from a trivial change of variable. In the case where $t = -a/b$, we obtain that

$$D(t) = \frac{(1-\lambda)}{2}.$$

For $t > -a/b$, we have

$$D(t) = D(-a/b) + \int_{-a/b}^{t} bc\left(1 + \frac{1}{\eta-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} \mathrm{d}z.$$

The result now follows from a computation analogous to the case $t < -a/b$. □

It is easy to verify that $D(-\infty) = 0$, and $D(\infty) = 1$.

This proposition shows how the cdf of the generalized Student-t can be evaluated using readily available procedures. In order to simulate data from the generalized Student-t, we use the inverse cdf technique. This requires inversion of Eq. (10). Setting $u = D(t)$, that will be distributed, in a simulation, as an uniform variate with support

[0, 1], we obtain that

$$t = \begin{cases} \dfrac{1}{b}\left[(1-\lambda)\sqrt{\dfrac{\eta-2}{\eta}}\,A_\eta^{-1}\left(\dfrac{u}{1-\lambda}\right) - a\right] & \text{if } u < \dfrac{1-\lambda}{2}, \\ \dfrac{1}{b}\left[(1+\lambda)\sqrt{\dfrac{\eta-2}{\eta}}\,A_\eta^{-1}\left(\dfrac{u+\lambda}{1+\lambda}\right) - a\right] & \text{if } u \geqslant \dfrac{1-\lambda}{2}. \end{cases}$$

To obtain a generalized Student-t distributed realization, it suffices to simulate u uniform and to use these formulae to compute t. At this stage, we have gathered all the necessary theoretical tools. In the next section, we present the results of the empirical work.

3. Data

3.1. The data set

In this study, we investigate the time-series behavior of five stock indices and of six foreign exchange rates. We use the following symbols for stock indices: S&P, NIK, DAX, CAC, and FTSE for the S&P 500, the Nikkei, the Deutsche Aktien Index, the CAC40, and the FTSE 100, respectively.

For exchange rates, we use DM–US, YEN–US, UK–US, FF–US for the amount of Deutsche Mark, Japanese Yen, British Pound, and French Franc necessary to purchase one US dollar. Furthermore, we use SFR–DM for the Swiss Franc-Deutsche Mark and CAN–US for the Canadian dollar to US dollar. The stock-market indices have been obtained from Datastream. They cover the period from August 23, 1971 to May 14, 1999 for a total of 7169 observations. All these indices have been used in other studies, such as Jorion (1995). The exchange rates have been provided by a large bank. They cover the period from July 26, 1991 to May 14, 1999, representing 1969 observations. [11]

3.2. Descriptive statistics

Table 2 displays several sample statistics for stock indices and for exchange rates. We notice that the standard deviation of exchange rates tends to be smaller than for indices. For the stock indices, we find a negative skewness for the S&P, DAX, and CAC, indicating the presence of sharp drops in stock prices. Exchange rates display a wide range of possible skewness, which ranges from -1.3889 for the YEN–US to 0.3868 for the UK–US. This translates the fact that, over the sample considered, on certain occasions, the Yen appreciated sharply whereas the pound depreciated. Skewness

[11] We did not do any data-snooping in this study. That is we did not try to select a particular sample length to obtain *nicer* empirical results. Also, we did not drop any series for which our model may not have worked. The inclusion of little investigated data sets such as the Swiss Franc–Deutsche Mark or the Canadian dollar–US dollar exchange rate was motivated by the question whether less liquid markets are subject to different dynamics.

Table 2
Descriptive statistics for stock-index and foreign-exchange returns. This table presents descriptive statistics. Day 1 indicates when a given series starts. All series end with May 14, 1999. N. Obs is the number of observations for a given series. Mean, std. dev., min, and max stand for the mean, $\bar{r}$, the standard deviation, σ_r, the minimum, and the maximum. Sk, stsk, ku, stku stand for skewness and its studentized version as well as for kurtosis and its studentized version. Skewness and kurtosis are computed using the formulae

$$sk = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_t-\bar{r}}{\sigma_r}\right)^3 \quad \text{and} \quad ku = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_t-\bar{r}}{\sigma_r}\right)^4.$$

Under normality of the data, these statistics are asymptotically normally distributed. J-B is the Jarque–Bera statistics. Engle(1) and Engle(5) are the Lagrange multiplier statistics to test for heteroskedasticity in the data. The statistics ac(1), ac(2), and ac(3) are the first three autocorrelations. QW(5) and QW(10) are the Box–Ljung statistics for autocorrelation, robustified following White (1980). These statistics follow a χ^2 with 5, respectively 10, degrees of freedom. The critical values at a significant level of 5% of a χ^2 with 5 and 10 degrees of freedom are 11.1 and 18.3.

	S&P	FTSE	DAX	CAC	NIK	DM–US	YEN–US	UK–US	FF–US	SFR–DM	CAN–US
Day1	23/08/71	23/08/71	23/08/71	23/08/71	23/08/71	26/07/91	26/07/91	26/07/91	26/07/91	26/07/91	26/07/91
N. Obs	7169	7169	7169	7169	7169	1969	1969	1969	1969	1969	1969
Mean	0.0353	0.0381	0.0315	0.0355	0.0240	0.0025	−0.0061	0.0021	0.0018	−0.0032	0.0119
Std. dev.	0.9487	1.0474	1.0735	1.0708	1.0825	0.6906	0.8167	0.6081	0.6559	0.2948	0.4400
Min	−22.8330	−13.0286	−13.7099	−13.9100	−16.1354	−3.9346	−9.9540	−2.8076	−3.8343	−2.6011	−2.5231
Max	8.7089	8.9434	7.2875	8.2254	12.4303	3.8580	4.6928	4.0173	3.4857	2.2939	3.5879
sk	−2.1074	−0.3377	−0.7330	−0.6782	−0.1925	−0.0015	−1.3889	0.3868	0.0139	−0.3233	0.2388
stsk	−72.8449	−11.6746	−25.3384	−23.4418	−6.6546	−0.0280	−25.1608	7.0068	0.2518	−5.8572	4.3266
ku	51.9623	10.3230	10.5955	10.5855	15.1877	2.8554	15.9310	3.5188	2.7497	8.6757	6.5804
stku	898.07	178.41	183.12	182.95	262.49	25.86	144.30	31.87	24.91	78.58	59.60
J-B	811841.53	31968.18	34176.64	34020.81	68946.56	668.89	21455.02	1064.94	620.36	6209.44	3571.26
Engle(1)	92.01	1818.82	304.30	65.35	370.74	12.58	358.41	78.36	18.46	47.25	120.68
Engle(5)	348.30	1911.81	649.71	544.12	525.44	67.32	370.18	187.36	90.35	52.10	224.21
ac(1)	0.0725	0.1504	0.0416	0.1177	0.0148	−0.0596	−0.0199	−0.0187	−0.0372	−0.0309	−0.2138
ac(2)	−0.0189	−0.0041	−0.0563	−0.0008	−0.0555	0.0228	−0.0173	0.0621	0.0255	−0.0301	−0.0657
ac(3)	−0.0199	0.0198	−0.0084	−0.0097	0.0129	−0.0056	−0.0537	−0.0188	−0.0059	−0.0367	0.0011
QW(5)	7.73	26.58	13.16	51.08	9.40	12.54	3.48	9.75	9.45	4.66	39.45
QW(10)	16.01	38.19	23.35	57.78	17.91	18.92	6.71	17.10	16.41	6.26	45.47

is significant for all series except for the DM–US and the FF–US. Given the effort to align both currencies, we can expect the two series to have a similar behavior. Turning to excess kurtosis, we find that all countries have a strongly significant statistic. This result translates the fact that exchange rates and stock returns have fatter tails than the ones of the normal distribution. Considering the Jarque–Bera statistic, which is distributed as a χ^2 with two degrees of freedom, we reject normality for all series.

The Engle-test statistics with lags 1 and 5, obtained by regressing squared returns on one lagged, respectively, five lagged, squared returns is distributed as a χ^2 with the degree of freedom equal to the number of lags. The strong significance of the statistics reveals the presence of heteroskedasticity in the data.

The Box–Ljung statistic, corrected for heteroskedasticity, tests for the existence of serial correlation of order 5 or 10. Even though we are able to detect serial correlation, the coefficient of correlation is always small. [12]

4. Estimation results

4.1. Searching for an optimal model

In Tables 3 and 4, we report the results of the various estimations. Table 3 focuses on stock markets, whereas Table 4 considers exchange rates. We started by estimating the GARCH model with asymmetry in the volatility specification and with an unconditional generalized t distribution.

As a first step, we examine for all series and models the specification of volatility. A likelihood-ratio test of the restriction $b_0^+ = b_0^-$ could be rejected for all stock markets. For these markets we uniformly find that negative past innovations have a larger impact on subsequent volatility than positive innovations. This observation has been well documented, e.g., Nelson (1991), Campbell and Hentschel (1992), Glosten et al. (1993), Zakoïan (1994) Sentana (1995), and may be explained by Black's (1976) leverage hypothesis. Considering exchange rates, we notice that, except for the SFR–DM, no asymmetric impact of news exists. For the SFR–DM, a rather thinly traded currency, we find that a decrease of the exchange rate (an appreciation of the SFR with respect to the DM) is followed by an increase of volatility. [13]

Next, following the logic described in Section 2.3, we estimate a model using Eq. (8) for volatility and

$$\tilde{\eta}_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^-, \tag{11}$$

$$\tilde{\lambda}_t = a_2 + b_2^+ y_{t-1}^+ + b_2^- y_{t-1}^-, \tag{12}$$

$$\eta_t = g_{]2,+30]}(\tilde{\eta}_t), \quad \lambda_t = g_{]-1,1[}(\tilde{\lambda}_t). \tag{13}$$

g represents the logistic map.

[12] We filtered the data with an AR(5) auto-regression and estimated various specifications with and without the filtering. Since the estimations, involving filtered or non-filtered data, yielded similar results, we decided to report the results obtained for non-filtered data only.

[13] An explanation of this finding without further considering cross-country interest-rate differentials would be hazardous and will not be attempted here.

Table 3

Search for a best specification of stock-index returns. Our general model describes innovations y_t in the spirit of GARCH models, that is $y_t = \sigma_t z_t$. The disturbance z_t follows a generalized Student-t distribution with asymmetry parameter λ_t and tail-fatness parameter η_t. We have the following specification:

$$\sigma_t^2 = a_0 + b_0^+(y_{t-1}^+)^2 + b_0^-(y_{t-1}^-)^2 + c_0\sigma_{t-1}^2,\ \tilde{\eta}_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^- + c_1\tilde{\eta}_{t-1},\ \tilde{\lambda}_t = a_2 + b_2 y_{t-1}^2 + c_2\tilde{\lambda}_{t-1},\ \eta_t = g_{]2,30]}(\tilde{\eta}_t),\ \lambda_t = g_{]-1,1[}(\tilde{\lambda}_t).$$

	S&P				FTSE			DAX				CAC		NIK		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Time	25.48	27.35	37.84	27.96	30.76	26.2	23.18	26.63	21.14	19.5	33.18	31.81	27.08	23.84	23.18	372.95
N.obs.	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168	7168
log-lik	−8631.83	−8625.09	−8621.22	−8616.36	−9369.99	−9364.37	−9360.08	−9483.16	−9480.45	−9476.24	−9482.18	−9542.74	−9540.79	−8590.84	−8583.29	−8585.92
a_0	0.0088	0.0084	0.0074	0.0074	0.0192	0.0193	0.0191	0.0160	0.0160	0.0167	0.0162	0.0235	0.0244	0.0134	0.0133	0.0137
	0.0024	0.0024	0.0022	0.0022	0.0041	0.0039	0.0041	0.0039	0.0040	0.0040	0.0040	0.0053	0.0055	0.0030	0.0030	0.0031
b_0^+	0.0298	0.0331	0.0385	0.0384	0.0642	0.0625	0.0615	0.0605	0.0644	0.0623	0.0603	0.0691	0.0641	0.0573	0.0630	0.0585
	0.0047	0.0053	0.0061	0.0062	0.0088	0.0083	0.0081	0.0084	0.0094	0.0084	0.0085	0.0104	0.0104	0.0105	0.0119	0.0105
b_0^-	0.0709	0.0737	0.0760	0.0759	0.0915	0.0896	0.0922	0.0995	0.1035	0.1053	0.0997	0.1180	0.1121	0.1856	0.1870	0.1912
	0.0115	0.0116	0.0118	0.0113	0.0111	0.0102	0.0112	0.0149	0.0156	0.0155	0.0151	0.0177	0.0181	0.0249	0.0260	0.0254
c_0	0.9390	0.9373	0.9367	0.9366	0.9014	0.9023	0.9017	0.9052	0.9020	0.9008	0.9046	0.8900	0.8940	0.8777	0.8750	0.8740
	0.0086	0.0086	0.0083	0.0081	0.0119	0.0113	0.0118	0.0129	0.0134	0.0132	0.0131	0.0145	0.0146	0.0150	0.0160	0.0153
a_1	−1.6902	−1.5136	−0.562	−0.5191	−0.4841	−0.842	−0.3970	−1.1081	−0.9819	−1.0778	−1.6772	−2.0500	−2.3124	−2.3497	−2.3172	−2.3316
	0.1498	0.1668	0.1975	0.2268	0.3618	0.4444	0.3869	0.2237	0.2585	0.2278	0.4968	0.1355	0.2744	0.1259	0.1391	0.1256
b_1^+	—	−0.5457	−0.6079	−0.5615	—	0.0377	—		−0.3242	—	0.4369		0.4855		−0.2296	—
		0.2530	0.1729	0.1746		0.2559			0.1817		1.8455		0.4920		0.1184	
b_1^-	—	−0.0882	−0.1241	−0.0653	—	2.9316	—	—	−0.0662	—	−0.1027	—	0.3142	—	0.1044	—
		0.0525	0.1385	0.0214		2.4718			0.1134		0.1976		0.4700		0.1189	
c_1	—		0.5568	0.5999	—	—	—	—	—	—	−0.5569	—	—	—	—	—
			0.1385	0.1440							0.2554					
a_2	−0.0257	−0.0270	−0.0228	−0.0062	−0.1274	−0.1354	−0.0259	−0.0749	−0.0695	−0.0188	−0.0712	−0.0246	−0.0254	−0.1049	−0.1088	−0.0639
	0.0271	0.0271	0.0272	0.0088	0.0364	0.0365	0.0092	0.0345	0.0352	0.0113	0.0351	0.0252	0.0253	0.0264	0.0264	0.0295
b_2	—	0.0453	—	0.0626	—	0.0670	0.0665	—	0.0619	0.0777	—	—	−0.0356	—	0.0736	0.0731
		0.0313		0.0212		0.0301	0.0172		0.0306	0.0234			0.0275		0.0333	0.0266
c_2	—	—	—	0.6961	—	—	0.8065	—	—	0.7087	—	—	—	—	—	0.3920
				0.1009			0.0514			0.1017						0.2396

Table 4
Search for a best specification of foreign-exchange returns. The general model is the same as in Table 2.

	DM–US			YEN–US			UK–US			FF–US			SFR–DM		CAN–US		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Time	7.42	5.49	5	6.82	7.2	4.72	10.17	6.92	6.59	8.35	6.1	5.11	6.54	9.17	10.32	6.86	212.57
N.obs.	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969	1969
Lik	−1922.97	−1917.46	−1915.89	−2045.38	−2033.90	−2034.15	−1550.50	−1548.41	−1542.04	−1807.83	−1804.92	−1802.77	−193.75	−192.50	−823.37	−817.62	−818.06
a_0	0.0138	0.0148	0.0141	0.0101	0.0134	0.0162	0.0057	0.0057	0.0056	0.0118	0.0123	0.0116	0.0069	0.0075	0.0057	0.0075	0.0075
	0.0056	0.0061	0.0057	0.0056	0.0059	0.0059	0.0031	0.0031	0.0030	0.0054	0.0057	0.0053	0.0027	0.0031	0.0036	0.0050	0.0047
b_0^+	0.0471	0.0559	0.0531	0.0609	0.0799	0.0816	0.0920	0.0925	0.0963	0.0658	0.0701	0.0680	0.0166	0.0102	0.1525	0.2027	0.2082
	0.0148	0.0173	0.0167	0.0187	0.0186	0.0177	0.0297	0.0309	0.0271	0.0189	0.0195	0.0192	0.0178	0.0192	0.0605	0.0865	0.0749
b_0^-	0.0786	0.0930	0.0862	0.0745	0.1082	0.1325	0.0788	0.0814	0.0847	0.0876	0.0995	0.0930	0.1242	0.1305	0.1045	0.1404	0.1383
	0.0158	0.0203	0.0173	0.0280	0.0205	0.0121	0.0206	0.0215	0.0214	0.0192	0.0236	0.0200	0.0342	0.0391	0.0413	0.0644	0.0544
c_0	0.9099	0.8988	0.9043	0.9182	0.8930	0.8761	0.9077	0.9072	0.9047	0.8996	0.8918	0.8972	0.8423	0.8314	0.8460	0.8086	0.8078
	0.0204	0.0234	0.0214	0.0275	0.0227	0.0173	0.0259	0.0265	0.0242	0.0246	0.0264	0.0243	0.0412	0.0481	0.0624	0.0846	0.0740
a_1	−1.7935	−1.5352	−1.8030	−2.2135	−2.2333	−2.1875	−2.3562	−2.4309	−2.3674	−1.8582	−1.6569	−1.8713	−1.8301	−1.7525	−1.9716	−2.1087	−2.1085
	0.2247	0.3210	0.2230	0.2042	0.2243	0.2050	0.2027	0.2762	0.2057	0.2183	0.3450	0.2158	0.2572	0.2825	0.2059	0.2974	0.2105
b_1^+	—	−0.1953	—	—	−0.2661	—	—	0.1344	—	—	−0.0344	—	—	1.9295	—	0.3044	—
		0.3524			0.3923			0.4667			0.4357			3.3566		0.5582	
b_1^-	—	−0.7075	—	—	0.0971	—	—	0.0797	—	—	−0.6404	—	—	−0.5826	—	−0.2822	—
		0.4881			0.0829			0.4448			0.5837			0.5528		0.4728	
a_2	−0.1304	−0.1313	−0.0513	−0.1539	−0.1494	−0.0861	0.0726	0.0661	0.0128	−0.1101	−0.1118	−0.0358	−0.0628	−0.0560	0.0072	0.0145	0.0147
	0.0698	0.0671	0.0344	0.0579	0.0567	0.0462	0.0543	0.0518	0.0081	0.0659	0.0647	0.0251	0.0642	0.0644	0.0747	0.0739	0.0689
b_2	—	0.2422	0.2731	—	0.3498	0.3802	—	0.1895	0.1988	—	0.1742	0.2164	—	0.0940	—	0.4009	0.4164
		0.0992	0.0897		0.0727	0.0655		0.0891	0.0441		0.0823	0.0819		0.2013		0.1495	0.1128
c_2	—	—	0.6386	—	—	0.3191	—	—	0.8452	—	—	0.6991	—	—	—	—	0.0462
			0.1462			0.1332			0.0326			0.1018					0.1795

Inspection of the coefficients b_2^+ and b_2^-, followed by a formal test, revealed that the hypothesis $b_2^+ = -b_2^-$ could not be rejected for any of the series. This implies that, subsequent to positive innovations, the asymmetry parameter tends to be larger and that subsequent to negative returns, the λ_t parameter becomes smaller. As a consequence, we re-estimated for all series the model with equation (12) replaced by

$$\tilde{\lambda}_t = a_2 + b_2 y_{t-1}.$$

4.2. Discussion of the estimations for stock-index returns

At this stage, rather than estimating all sorts of models and test their restrictions, we adopted a strategy taking care of spurious persistence. This leads us to a discussion where we have to consider each series one by one since no single pattern emerges.

For the S&P series, we notice in column 2, that the parameter b_2 is not significant. This means that an autoregressive specification of $\tilde{\lambda}_t$ could be misleading. However, since b_1^+ is statistically significant, taking the value -0.5457, we estimate an autoregressive specification for $\tilde{\eta}_t$.

As column 3 shows, the parameter η_t displays some persistence since c_1 takes the value 0.5568. Given that b_1^+ and b_1^- remain significant, we conclude that this persistence will not be spurious. We further estimated a model where we used an autoregressive specification of η_t and a dynamic specification $\tilde{\lambda}_t = a_2 + b_2 y_{t-1}$. We found that b_2 became significant. This led us to estimate a model with autoregressive specification of $\tilde{\lambda}_t$. The result of this estimation is presented in column 4. We notice that both the degree of freedom, $\tilde{\eta}_t$, and the asymmetry parameter, $\tilde{\lambda}_t$, are autoregressive.

Next, we turn to the FTSE. Inspection of column 6 of Table 3 shows that neither b_1^+ nor b_1^- are statistically significant. However, since b_2 is statistically significant, we estimated an autoregressive specification for $\tilde{\lambda}_t$. The result is reported in column 7. We further verified that b_1^+ and b_1^- remained non significant, even when we introduced the autoregressive specification of $\tilde{\lambda}_t$. Thus, for the FTSE a model with constant $\tilde{\eta}_t$ and autoregressive $\tilde{\lambda}_t$ appears optimal such as presented in column 7.

For the DAX, the model presented in column 9 hints at a possible autoregressive specification for $\tilde{\lambda}_t$. In column 10, we present the estimation where we allow for an autoregressive specification for $\tilde{\lambda}_t$, keeping $\tilde{\eta}_t$ constant. We extended that model allowing for a dynamic and even autoregressive specification of $\tilde{\eta}_t$. The parameters were close to the ones reported in column 11 and non-significant. As a consequence, we will use for the DAX as best specification the one presented in column 10.

To illustrate what we mean with spurious dependence, let us again consider the estimation of the DAX. As an exercise we estimated the dynamics of $\tilde{\eta}_t$ and $\tilde{\lambda}_t$ as an autoregressive process. We obtained

$$\tilde{\eta}_t = \underset{(0.465)}{-1.775} - \underset{(0.218)}{0.062}\, y_{t-1}^+ - \underset{(0.121)}{0.153}\, y_{t-1}^- - \underset{(0.083)}{0.885}\, \tilde{\eta}_{t-1}^+,$$

$$\tilde{\lambda}_t = \underset{(0.011)}{-0.0162} + \underset{(0.023)}{0.069}\, y_{t-1} + \underset{(0.115)}{0.726}\, \tilde{\lambda}_{t-1}.$$

Inspection of the parameters reported in column 10 of Table 3 shows that the dynamics of $\tilde{\lambda}_t$ is similar to the one reported previously. When we compare the new parameters of the $\tilde{\eta}_t$ dynamics with the ones displayed in column 11, we find that the estimates behave rather erratically. Interestingly, now we even find that c_1, with a value of -0.885, is statistically significant. In such a situation, one might have concluded that $\tilde{\eta}_t$ reverts strongly. Since b_1^+ and b_1^- are not statistically significant, however, this finding turns out to be spurious.

The discussion concerning the CAC is short. Estimation of the model of column 13 and comparison of the log-likelihoods of the models presented in columns 12 and 13 show that a dynamic specification of $\tilde{\lambda}_t$ and $\tilde{\eta}_t$ does not improve the fit. In other words, the CAC displays no time-varying skewness and kurtosis of a nature that can be captured by this model.

For the Nikkei, we report the estimates for the basic models in columns 14 and 15 of Table 3. There is potentially no dynamics in $\tilde{\eta}_t$ and $\tilde{\lambda}_t$. As a consequence, we re-estimate the model with a possible autoregressive specification for $\tilde{\lambda}_t$. As column 16 displays, the parameter associated with $\tilde{\lambda}_{t-1}$ turns out to be non significant. We did a similar estimation with an autoregressive structure for $\tilde{\eta}_t$, yet, we retain as preferred specification the one presented in column 15.

Broadly speaking, we obtain that the degree of freedom, $\tilde{\eta}_t$, can be considered as constant over time for the FTSE, the DAX, and the CAC. It is found to be significantly persistent for the S&P only. As far as the asymmetry parameter, $\tilde{\lambda}_t$, is concerned, it is significantly persistent (with an autoregressive parameter as high as 0.7) for the S&P, the FTSE, and the DAX. Parameter b_2, associated with y_{t-1}, is positive in most countries and close to 0.07, indicating that a positive (negative) shock yield an increase (decrease) in the degree of freedom.

Note that we investigated the estimation of the model over subsamples. Our finding is that, since extreme events that drive skewness and kurtosis become rarer, standard errors increase and make it more difficult to identify the dynamics of the parameters. We found, however, that the estimates remained similar.

4.3. Discussion of estimations for foreign-exchange returns

Following the strategy described previously, in order to obtain reasonable estimates, it is necessary to start from a specification where it is known that past innovations matter. For the DM–US$ pair, inspection of column 2 of Table 4, shows that an autoregressive specification for $\tilde{\eta}_t$ does not seem appropriate but possibly for $\tilde{\lambda}_t$.

Column 3 displays the coefficients obtained for an autoregressive specification of $\tilde{\lambda}_t$. We notice that $\tilde{\lambda}_t$ may indeed be modeled as an autoregressive process since the parameter c_2 takes the value of 0.638 with a standard error of 0.1462.

The result for the pairs YEN–US, UK–US and FF–US are very similar to the ones just reported. We found no dynamics at the level of $\tilde{\eta}_t$, but the autoregressive component for $\tilde{\lambda}_t$ turns out to be significant. For theses series, we always obtain a positive c_2 coefficient with a value between 0.31 and 0.85.

The results for the SFR–DM, reported in column 14, are disappointing. We cannot reject the model with constant skewness and kurtosis.

The results for CAN–US, reported in column 14, reveal that $\tilde{\eta}_t$ is best taken as a constant. In column 17 we show that the parameter c_2 is not significant. Hence, for this series an autoregressive specification of $\tilde{\lambda}_t$ does not seem correct. Eventually, we settled for a specification (not reported) where $b_1^+ = b_1^- = 0$, and a_2 and b_2 were estimated.

Broadly speaking, the degree of freedom, $\tilde{\eta}_t$, is constant over time for all foreign-exchange returns. By contrast, the asymmetry parameter, $\tilde{\lambda}_t$, is significantly persistent (with an autoregressive parameter ranging between 0.3 and 0.85), except for CAN–US. Parameter b_2 is positive in most countries, between 0.2 and 0.4, with the exception of SFR–DM.

To conclude this section, we wish to comment on the dynamics of volatility, once skewness and kurtosis have been modeled. Inspection of the parameters b_0^+ and b_0^- reveals for all stock-index and exchange-rate series great stability for the estimates and their standard errors. These inspections confirm the observation of Harvey and Siddique (1999) that the asymmetries at the level of volatility are not altered as higher moments get specified.

4.4. Testing the adequacy of the model

To assess the quality of the estimates, we present in Table 5 an analysis of orthogonality conditions that should be satisfied by the residuals. The use of orthogonality conditions to test the specification of a model has been advocated by Newey (1985) and can be found in many subsequent contributions, e.g., Harvey and Siddique (1999). We group these orthogonality conditions in order to focus on the dependency of higher moments.

$$\mathrm{E}[z_t] = 0, \quad \mathrm{E}[z_t^2 - 1] = 0, \quad \mathrm{E}[z_t^3] = 0, \quad \mathrm{E}[z_t^4] - 3 = 0, \tag{14}$$

$$\mathrm{E}[(z_t^2 - 1)(z_{t-j}^2 - 1)] = 0, \quad j = 1, \ldots, 4, \tag{15}$$

$$\mathrm{E}[(z_t^3)(z_{t-j}^3)] = 0, \quad j = 1, \ldots, 4, \tag{16}$$

$$\mathrm{E}[(z_t^4 - 3)(z_{t-j}^4 - 3)] = 0, \quad j = 1, \ldots, 4. \tag{17}$$

Hansen (1982) shows how theses conditions can get tested. It is possible to test each condition individually or it is possible to test the various groups with a Wald test. For brevity, we present the means associated with conditions (14) and associated standard errors. We also present Wald tests for joint significance of higher moments given by conditions (15)–(17). The results of these estimations are presented in Table 5.

In the first two rows of Table 5, we present measures of skewness and excess kurtosis. We notice for excess kurtosis (defined as kurtosis-3) a strong reduction when compared with the values presented for the raw data in Table 2. Inspection of the test for the first two unconditional moments shows that we cannot reject the null hypothesis

Table 5
Skewness, kurtosis, as well as individual and group-wise tests of orthogonality conditions. We first present the skewness and kurtosis of the residuals. We used the same formula for skewness and kurtosis as in Table 2. The statistics Q_3 corresponds to a Wald test for joint significance of the orthogonality conditions $\mathrm{E}[(z_t^2-1)(z_{t-j}^2-1)]=0,\ j=1,\dots,4$. Similarly, Q_4 tests for joint significance of $\mathrm{E}[(z_t^3)(z_{t-j}^3)]=0,\ j=1,\dots,4$. Last, Q_5 tests $\mathrm{E}[(z_t^4-3)(z_{t-j}^4-3)]=0,\ j=1,\dots,4$.

	S&P	FTSE	DAX	CAC	NIK	DM–US	YEN–US	UK–US	FF–US	SFR–DM	CAN–US
Skewness	−0.319	−0.146	−0.852	−0.578	−0.126	−0.847	0.068	−0.202	−0.509	0.075	−0.186
Kurtosis	6.056	3.955	14.359	11.055	8.334	11.173	5.987	5.647	5.704	5.799	5.290
$\mathrm{E}[z]=0$	−0.005	−0.003	−0.004	0.000	0.006	−0.007	0.005	−0.003	0.002	−0.016	−0.005
se	0.012	0.012	0.012	0.012	0.012	0.023	0.022	0.023	0.022	0.022	0.023
$\mathrm{E}[z^2]-1=0$	−0.022	0.006	0.041	0.011	−0.023	0.050	−0.035	−0.007	−0.033	−0.044	−0.006
se	0.026	0.021	0.045	0.038	0.031	0.076	0.049	0.048	0.047	0.047	0.046
$\mathrm{E}[z^3]=0$	−0.323	−0.158	−0.918	−0.585	−0.104	−0.941	0.077	−0.213	−0.479	0.011	−0.201
se	0.172	0.084	0.595	0.371	0.212	0.637	0.204	0.219	0.188	0.196	0.191
$\mathrm{E}[z^4]-3=0$	2.805	1.012	12.621	8.341	4.944	9.438	2.575	2.538	2.316	2.203	2.230
se	1.575	0.458	8.721	4.366	1.818	6.005	0.931	1.180	0.807	0.968	0.891
Q_3	7.415	3.163	5.881	1.813	7.328	4.297	14.899	7.101	0.142	7.554	6.168
p-value	0.12	0.53	0.21	0.77	0.12	0.37	0.00	0.13	1.00	0.11	0.19
Q_4	4.478	22.792	1.432	10.688	2.419	5.475	4.699	0.703	3.072	5.087	0.841
p-value	0.35	0.00	0.84	0.03	0.66	0.24	0.32	0.95	0.55	0.28	0.93
Q_5	4.107	5.465	6.755	4.918	8.903	6.238	25.294	13.171	8.463	14.181	16.816
p-value	0.39	0.24	0.15	0.30	0.06	0.18	0.00	0.01	0.08	0.01	0.00

Table 6
Existence of skewness and kurtosis. We count for each series how often skewness and kurtosis cease to exist, i.e. $\eta_t \leqslant 3$, respectively, $\eta_t \leqslant 4$.

	S&P500	FTSE	DAX	CAC	NIK	
No skewness	53	0	0	0	19	
Percentage of sample	0.74%	0.00%	0.00%	0.00%	0.27%	
No kurtosis	312	0	0	0	501	
Percentage of sample	4.36%	0.00%	0.00%	0.00%	6.98%	
	DM–US	YEN–US	UK–US	FF–US	SFR–DM	CAN–US
No skewness	5	5	9	5	0	1
Percentage of sample	0.25%	0.25%	0.46%	0.25%	0.00%	0.05%
No kurtosis	27	50	31	25	1	8
Percentage of sample	1.37%	2.54%	1.57%	1.27%	0.05%	0.41%

that residuals have a zero mean and a unit variance for any series. A test of joint nullity of skewness and excess kurtosis gets rejected in many cases. One could expect these rejections to occur, since it is precisely higher moments that we wish to capture in this study. Noticeable exceptions are the S&P where only the third moment, with a value of -0.323 is marginally significant. For the CAC, the third and fourth moments are such that the hypothesis of normally distributed residuals cannot be rejected.

We now turn to the Wald tests of joint significance in the orthogonality conditions presented in Eqs. (15)–(17). The statistic Q_3 corresponds to a joint test for heteroskedasticity. The p-values show that we cannot reject homoskedasticity for all series with the exception of the Yen–US exchange rate. The statistics Q_4 and Q_5 test for dependency of the third, respectively, the fourth moment. Only for the CAC and the Nikkei can we reject independence of the third moment. Rejection of independence of kurtosis occurs for nearly all exchange rates. We marginally reject for the FTSE.

5. Analysis of the dynamics of skewness and kurtosis

In the previous section, we estimated the dynamics of parameters η_t and λ_t. Even though these parameters are related to skewness and kurtosis, the relation is a highly non-linear one. For this reason, in order to proceed one step further, we now consider the evolution of skewness and kurtosis through time, using Eqs. (2) and (3). Next, we analyze cross-sectional movements between various markets in terms of skewness as well as kurtosis. Our model, therefore, extends in a certain sense the one by Kroner and Ng (1998).

5.1. Existence of moments

We consider now the existence of conditional skewness and kurtosis in actual data. Inspection of formulae (2) and (3) suggests that third and fourth moments of innovations will only exist for $\eta_t > 3$ and $\eta_t > 4$, respectively. Table 6 reports, for each

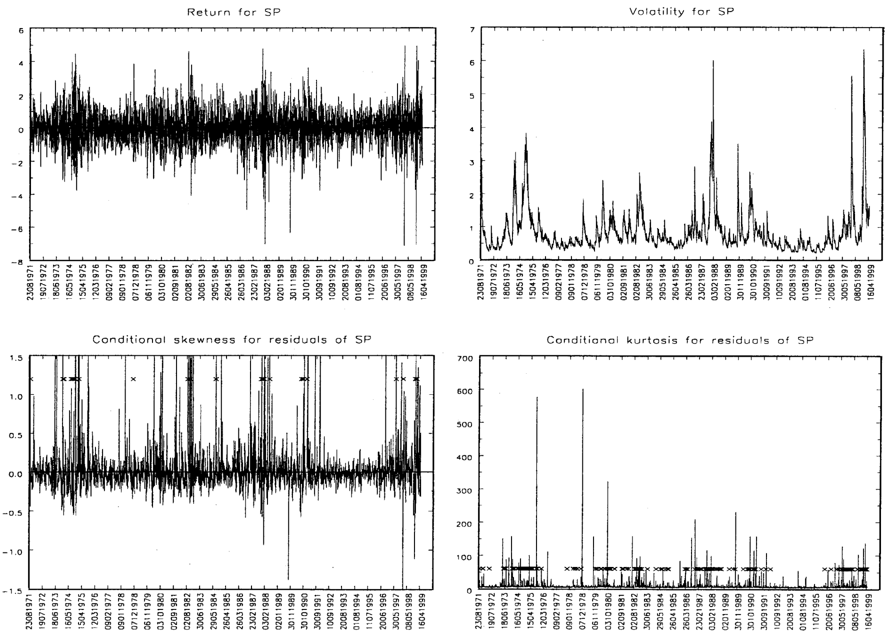


Fig. 6.

series, the number of times that $\eta_t \leqslant 3$ or $3 < \eta_t \leqslant 4$. In the former case, skewness and kurtosis do not exist, whereas in the latter case, only kurtosis fails to exist. Inspection of this table shows that for stock markets, the percentage of days when skewness did not exist ranges between 0.74% for the S&P to 0.27% for the Nikkei. The corresponding percentages for kurtosis range between 4.36% for the S&P and 6.98% for the Nikkei. This finding suggests that the conclusion reached by Loretan and Phillips (1994) that the fourth unconditional moment of stock-index returns fails to exist is due to some episodes where higher moments cease to exist.

For exchange rates, we notice that the number of days when skewness or kurtosis fails to exist is much smaller.

5.2. Persistence of skewness and kurtosis

As an illustration, we display in Figs. 6 and 7 the raw returns, the volatility, σ_t, the conditional skewness, s_t, and the conditional kurtosis, k_t, of the residuals of our model for the S&P and the FTSE, respectively. Notice that s and k are computed with formulae (2) and (3). Since skewness only exists if $\eta_t > 3$, we mark days where this condition fails with a star. Similarly, kurtosis exists only if $\eta_t > 4$. Again, dates where this fails are marked with a star. We would like to mention that if η_t approaches 4, the kurtosis exists, yet may take very large values. Inspection of these conditional evolutions with the raw returns series shows that extreme realizations are rather well captured by our model.

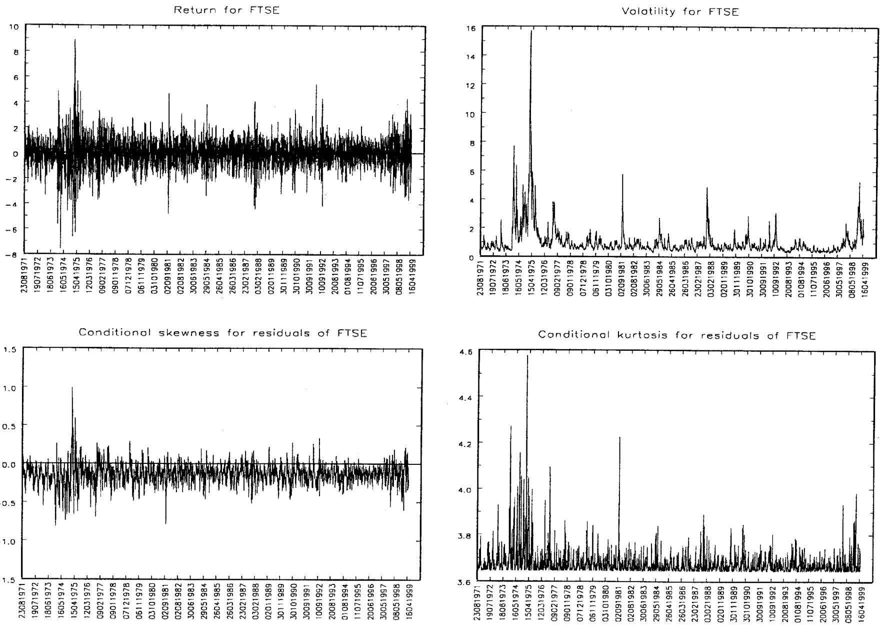


Fig. 7.

The average kurtosis, for those days where kurtosis exists is 7.66. This value should be compared with the one of 6.056 that we presented in Table 5, corresponding to the kurtosis of standardized residuals. Even though kurtosis seems strongly varying, its average value seems very reasonable. Careful inspection of the figures of skewness and kurtosis shows evidence for persistence of skewness and kurtosis.

To gain further insight on persistence in third and fourth moments, Table 7 reports transition probability matrices of skewness and kurtosis for selected series. The foreign-exchange series tend to have very similar dynamics, for this reason, we only present the results for the UK–US exchange rate. To construct the transition probability matrices, we rank the value of a higher moment into one of four possible categories. First, entries with finite values were classified into four intervals corresponding to the quartiles (I_j, $j = 1, \ldots, 4$). Second, entries with infinite values were counted in interval 4. The element in row a and column b of a transition probability matrix, (a, b), measures the percentage of times that a series moves from a skewness (or kurtosis) in quartile a at time $t - 1$ to quartile b at time t. Consideration of transition probabilities circumvents some drawbacks of serial-correlation coefficients in the presence of outliers.[14] In the absence of autocorrelation, each element of the transition matrix would have the same value ($1/16 = 6.25\%$). In case of a positive correlation, elements along the principal diagonal are larger than off-diagonal elements.

[14] This is a well-known problem. The correlation coefficient can be biased, if two extreme values occur at two consecutive dates. With transition matrices, outliers cannot bias correlation measures, since they only appear as elements of an interval.

Table 7
Transition probabilites for skewness and kurtosis. $I_t(j)$ designs the interval to which skewness, respectively, kurtosis belongs at time t. The intervals $I(j)$ for $j = 1, \ldots, 4$, correspond to the four quartiles starting with the smallest. The interval $j = 4$ also contains those situations where skewness or kurtosis does not exist ($\eta_t \leqslant 3$, respectively, $\eta_t \leqslant 4$). An element in row a and column b of each matrix measures the percentage of times one moves from a skewness or kurtosis in the quartile a to a quartile b.

	Skewness					Kurtosis			
	$I(1)$	$I(2)$	$I(3)$	$I(4)$		$I(1)$	$I(2)$	$I(3)$	$I(4)$
		S&P					S&P		
$I_{-1}(1)$	16.63	5.66	1.47	1.03	$I_{-1}(1)$	15.21	3.55	2.92	2.25
$I_{-1}(2)$	5.34	10.84	5.89	2.75	$I_{-1}(2)$	8.69	8.18	3.88	3.18
$I_{-1}(3)$	2.08	6.44	10.41	5.89	$I_{-1}(3)$	0.01	12.01	6.69	5.21
$I_{-1}(4)$	0.75	1.89	7.04	15.88	$I_{-1}(4)$	0.00	0.20	10.45	17.56
		FTSE					FTSE		
$I_{-1}(1)$	18.23	5.66	0.88	0.22	$I_{-1}(1)$	16.27	6.91	1.62	0.20
$I_{-1}(2)$	5.50	12.26	6.20	1.03	$I_{-1}(2)$	6.83	10.87	6.02	1.28
$I_{-1}(3)$	1.12	6.09	11.90	5.89	$I_{-1}(3)$	1.54	6.30	11.62	5.54
$I_{-1}(4)$	0.15	0.99	6.02	17.84	$I_{-1}(4)$	0.35	0.92	5.74	17.97
		DAX					DAX		
$I_{-1}(1)$	16.30	5.96	2.18	0.54	$I_{-1}(1)$	11.03	8.20	4.40	1.37
$I_{-1}(2)$	6.41	10.60	5.89	2.10	$I_{-1}(2)$	8.13	8.37	6.27	2.23
$I_{-1}(3)$	1.80	6.30	10.34	6.56	$I_{-1}(3)$	4.15	6.09	8.63	6.13
$I_{-1}(4)$	0.49	2.14	6.58	15.80	$I_{-1}(4)$	1.68	2.35	5.70	15.27
		NIK					NIK		
$I_{-1}(1)$	7.75	4.90	5.03	7.23	$I_{-1}(1)$	7.03	4.75	3.62	7.88
$I_{-1}(2)$	6.03	7.49	6.48	5.17	$I_{-1}(2)$	5.87	8.53	5.42	7.00
$I_{-1}(3)$	5.14	6.76	7.36	5.45	$I_{-1}(3)$	3.90	5.75	4.87	5.24
$I_{-1}(4)$	6.01	6.02	5.82	7.35	$I_{-1}(4)$	6.49	7.78	5.84	10.03
		UK–US					UK–US		
$I_{-1}(1)$	18.02	5.03	1.73	0.20	$I_{-1}(1)$	12.13	7.72	4.11	1.02
$I_{-1}(2)$	4.87	12.18	6.50	1.47	$I_{-1}(2)$	8.32	9.29	5.58	1.78
$I_{-1}(3)$	1.83	6.45	11.88	4.77	$I_{-1}(3)$	3.60	6.40	9.14	5.84
$I_{-1}(4)$	0.25	1.32	4.87	18.58	$I_{-1}(4)$	0.86	1.62	6.14	16.40
		DM–US					DM–US		
$I_{-1}(1)$	13.30	6.95	3.50	1.22	$I_{-1}(1)$	8.68	8.53	5.38	2.34
$I_{-1}(2)$	7.16	7.82	6.90	3.15	$I_{-1}(2)$	8.43	7.41	5.84	3.35
$I_{-1}(3)$	3.50	7.06	8.43	5.99	$I_{-1}(3)$	5.53	5.63	6.85	6.95
$I_{-1}(4)$	1.02	3.20	6.09	14.67	$I_{-1}(4)$	2.28	3.45	6.90	12.39
		YEN–US					YEN–US		
$I_{-1}(1)$	9.80	6.04	4.82	4.31	$I_{-1}(1)$	6.35	6.85	6.75	4.97
$I_{-1}(2)$	6.70	6.14	7.01	5.18	$I_{-1}(2)$	6.65	6.85	6.45	5.08
$I_{-1}(3)$	4.87	6.85	6.60	6.65	$I_{-1}(3)$	6.85	6.19	6.19	5.74
$I_{-1}(4)$	3.60	5.99	6.50	8.88	$I_{-1}(4)$	5.13	5.08	5.58	9.24

We first consider transition matrices for skewness (left part of Table 7). In most cases, we obtain a positive relation between s_{t-1} and s_t. For instance, if we consider the S&P, we notice that with 16.63% probability, skewness stays in the first quartile. The probability of moving from the first quartile to the second, third and fourth quartile is 5.66%, 1.47%, and 1.03% respectively. It is possible to summarize the persistence of a given moment by summing the diagonal elements of the matrix. If there was no persistence of a given moment, each diagonal should sum up to 25%. For the various series displayed in Table 7 we obtain 53.76% (S&P), 60.22% (FTSE), 53.03% (DAX), 29.94% (NIK), and eventually 60.66% for the UK–US. This shows that a given level of skewness tends to be followed by a similar level of skewness. Note that, for the CAC, since there is no dynamic of λ_t and η_t, skewness and kurtosis are constant over time.

Now, we turn to the persistence of kurtosis presented in the right part of Table 8.

Persistence tends to be somewhat weaker for kurtosis than for skewness. The sum of diagonal elements is 47.64% (S&P), 56.73% (FTSE), 43.30% (DAX) and 46.95% (UK–US). For the Nikkei, the transition probability matrix is dominated by the (4, 4) element corresponding to all those days when kurtosis failed to exist for that market.

This investigation shows that there is evidence of persistence in skewness but much less so for kurtosis.

5.3. Cross-index variability of higher moments

An abundant literature has documented volatility comovements (see, e.g., Hamao, et al., 1990; or Susmel and Engle, 1994, for stock markets). More recently, some authors stated that correlation between markets may increase during periods of high volatility (Longin and Solnik, 1995; Ramchand and Susmel, 1998). Now, we wish to address the issue of comovements between markets in terms of skewness and kurtosis. If skewness varies jointly between two markets this suggests an increase in the probability of occurrence of a large event with the same sign on both markets. If kurtosis varies jointly, then there will be an increase in the probability of occurrence of a large event on both markets, whatever the direction of the shock.

To measure the correlation between joint realizations of skewness or kurtosis for pairs of series, we construct frequency matrices. These matrices are constructed in the following manner. We consider four intervals. First, we classify all realizations where $\eta_t < 3$ ($\eta_t < 4$), i.e., where skewness (respectively kurtosis) does not exist, as belonging to the fourth interval, I_4. All other realizations get then classified according to the quartile to which they belong. The fourth interval, therefore, contains those observations where either skewness or kurtosis did not exist or belonged to the upper quartile. The left part of Table 8 contains the frequency matrices of joint realizations of skewness for a series a and a series b. The right part contains the kurtosis matrices. To illustrate, the element of the first matrix, 6.33 indicates that 6.33% of all realizations were such that the skewness of the S&P belonged to the second quartile whereas, simultaneously, the FTSE had a skewness in the first quartile.

Table 8
Joint skewness and joint kurtosis classification. $I(j)$ designs the interval to which skewness or kurtosis of a given series belongs. The intervals $I(j)$ for $j = 1, \ldots, 4$, correspond to the four quartiles starting with the smallest. Interval $I(4)$ also contains those cases where skewness, respectively kurtosis, does not exist. An element in row a and column b of each matrix measures the percentage of times one observes a skewness (kurtosis) in quartile a for the first mentioned country and a skewness (kurtosis) in quartile b for the second mentioned country.

	Joint skewness					Joint kurtosis			
	$I_b(1)$	$I_b(2)$	$I_b(3)$	$I_b(4)$		$I_b(1)$	$I_b(2)$	$I_b(3)$	$I_b(4)$
		SP and FTSE					SP and FTSE		
$I_a(1)$	11.30	6.09	4.60	2.82	$I_a(1)$	4.22	5.81	7.12	6.77
$I_a(2)$	6.33	7.60	6.51	4.39	$I_a(2)$	5.13	5.85	6.13	6.83
$I_a(3)$	4.18	6.34	7.37	6.93	$I_a(3)$	6.12	6.49	5.75	5.57
$I_a(4)$	3.20	4.97	6.52	10.87	$I_a(4)$	9.54	6.84	5.99	5.82
		SP and DAX					SP and DAX		
$I_a(1)$	10.36	6.16	4.75	3.53	$I_a(1)$	5.17	6.19	6.51	6.06
$I_a(2)$	6.24	7.44	6.03	5.10	$I_a(2)$	6.06	5.52	5.98	6.38
$I_a(3)$	4.51	5.82	7.54	6.94	$I_a(3)$	6.40	6.23	5.70	5.61
$I_a(4)$	3.88	5.57	6.68	9.43	$I_a(4)$	7.37	7.07	6.82	6.94
		SP and NIK					SP and NIK		
$I_a(1)$	8.39	6.62	5.28	4.51	$I_a(1)$	6.23	7.11	4.57	6.02
$I_a(2)$	6.47	6.26	6.06	6.03	$I_a(2)$	5.91	6.55	5.06	6.42
$I_a(3)$	5.21	6.33	6.61	6.68	$I_a(3)$	5.60	6.44	4.83	7.07
$I_a(4)$	4.86	5.96	6.76	7.97	$I_a(4)$	5.54	6.72	5.31	10.63
		DM–US and YEN–US					DM–US and YEN–US		
$I_a(1)$	12.34	7.06	3.81	1.78	$I_a(1)$	7.46	6.80	6.29	4.42
$I_a(2)$	6.14	7.51	6.65	4.72	$I_a(2)$	7.31	7.01	5.99	4.72
$I_a(3)$	3.81	6.14	8.22	6.80	$I_a(3)$	6.60	6.70	6.35	5.33
$I_a(4)$	2.69	4.31	6.29	11.73	$I_a(4)$	3.60	4.52	6.35	10.56
		DM–US and UK–US					DM–US and UK–US		
$I_a(1)$	13.76	6.80	2.89	1.52	$I_a(1)$	4.97	6.95	7.31	5.74
$I_a(2)$	6.85	8.27	6.19	3.71	$I_a(2)$	6.70	6.29	6.50	5.53
$I_a(3)$	3.71	6.90	8.73	5.63	$I_a(3)$	7.06	7.06	5.13	5.74
$I_a(4)$	0.66	3.05	7.16	14.16	$I_a(4)$	6.24	4.72	6.04	8.02
		DM–US and FF–US					DM–US and FF-US		
$I_a(1)$	22.34	2.44	0.20	0.00	$I_a(1)$	19.59	4.72	0.56	0.10
$I_a(2)$	2.54	18.53	3.65	0.30	$I_a(2)$	4.72	16.09	4.01	0.20
$I_a(3)$	0.05	3.96	18.38	2.59	$I_a(3)$	0.56	4.01	17.77	2.64
$I_a(4)$	0.05	0.10	2.74	22.13	$I_a(4)$	0.10	0.20	2.64	22.08

Again, inspection of the sum of the diagonals is indicative of the correlation of the two series of skewness or kurtosis. For skewness, we obtain the following sums: 37.14% (S&P/FTSE), 34.78% (S&P/DAX), 29.23% (S&P/NIK), 44.92% (DM–US/UK–US), 81.37% (DM–US/FF–US), and 36.80% (DM–US/Yen–US). Similarly, we obtain for kurtosis the sums: 21.65% (S&P/FTSE), 23.32% (S&P/DAX), 28.24% (S&P/NIK), 24.42% (DM–US/UK–US), 75.53% (DM–US/FF–US), and 31.37% (DM–US/Yen–US). These numbers show that, for most series considered, the figures are larger than 25%, the value which would hold if there was no correlation. The fact that the elements $(1,1)$ and $(4,4)$ of the matrices take particularly large values suggests that large returns with the same sign tend to occur jointly. The link between joint large realizations is particularly strong for the S&P/FTSE pair, both at the level of skewness and kurtosis.

Therefore, once again, our results indicate that comovements of moments beyond volatility are more intensive during agitated periods. Comovements of volatility have been pointed out by Longin and Solnik (1995) and Ramchand and Susmel (1998).

To illustrate this section, we display in Fig. 8 scatterplots of skewness and kurtosis for the DM–US and FF–US pair, for which correlation is particularly strong between higher moments. Inspection of the figure shows that the relation is clearly positive. In the absence of outliers, the estimation of the correlation coefficient is not biased. Therefore, we regress the DM–US moment on the corresponding FF–US moment. We find a parameter estimate of 1.118 (with standard error of 0.010) for skewness and 0.921 (with standard error of 0.011) for kurtosis. Adjusted R^2 are as high as 0.87 and 0.77, respectively.

In this section, we have shown that there is evidence that large events generating skewness tend to occur simultaneously for stock markets.

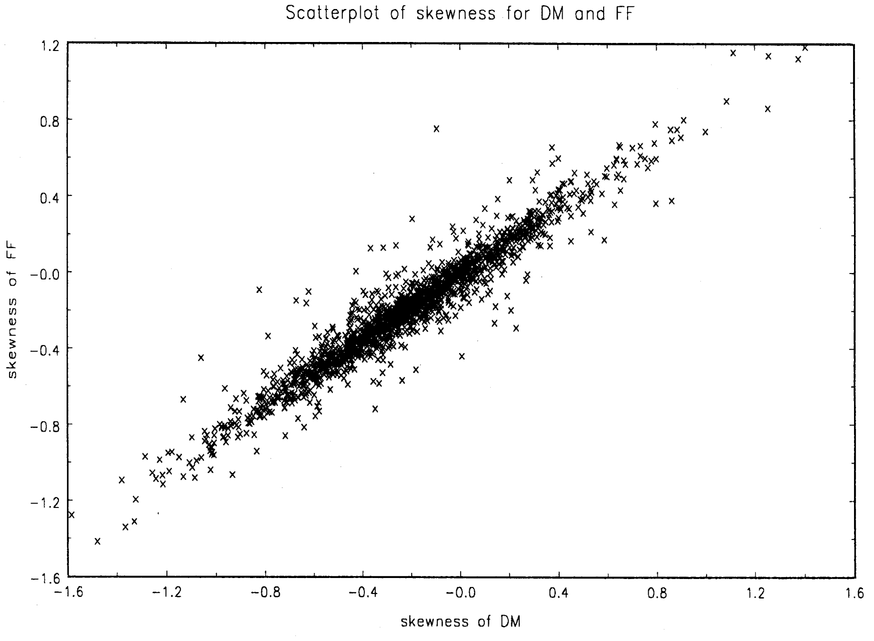
6. Conclusion

In this work, we implement a GARCH-type model where innovations are modeled according to Hansen's (1994) generalized Student-*t* distribution and where skewness and kurtosis are time dependent.

We discuss, from a theoretical perspective, various possible specifications. Eventually, we settle for a specification where the asymmetry and the fat-tailedness parameter have an autoregressive structure. The parameters get mapped into the domain where the density is well defined with a logistic map. Experiments with this specification revealed great numerical instability in the estimation. For certain series, the program would converge to different values for different initial values. Also the coefficient of the lagged parameters turned out to be often ill behaved. For one set of initial values, it would converge to a value such as 0.9 and appear statistically significant. In another run, it would take the value -0.8 and still be statistically significant. This difficulty led us to devise an estimation strategy that takes care of the finding of spurious persistence.

We also show, from a theoretical point of view, how our model can be simulated. In a Monte-Carlo experiment, we verify the validity of our interpretation of the finding of spurious persistence in higher moments. This Monte-Carlo simulation also allows us to establish normality of the parameters in well specified models.

The model is run over a large number of series. We find for many series that skewness and kurtosis are persistent. Furthermore, the modeling of skewness and kurtosis


Scatterplot of skewness for DM and FF
skewness of FF
skewness of DM
1.2
0.8
0.4
−0.0
−0.4
−0.8
−1.2
−1.6
−1.6
−1.2
−0.8
−0.4
−0.0
0.4
0.8
1.2
1.6

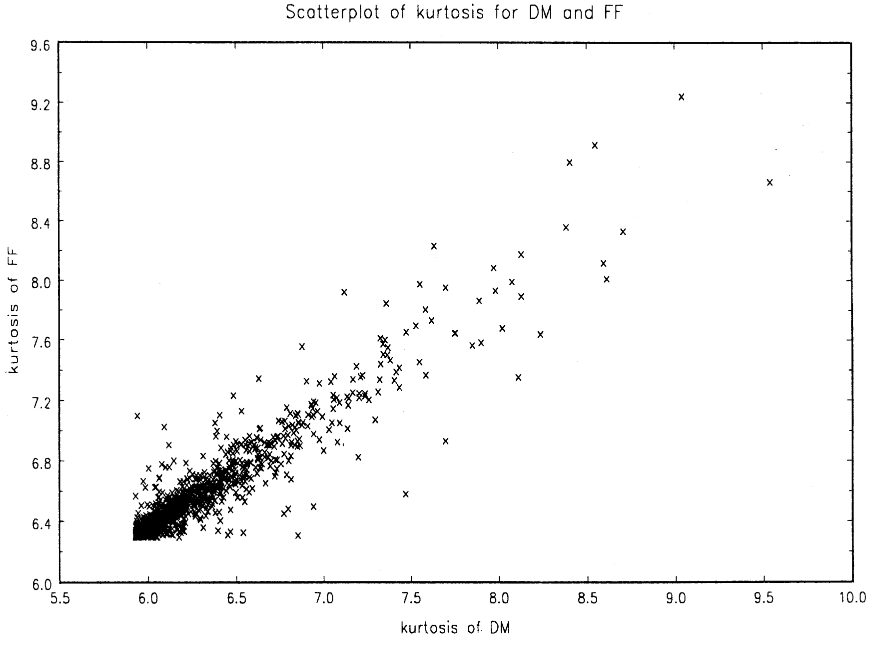

Scatterplot of kurtosis for DM and FF
kurtosis of FF
kurtosis of DM
9.6
9.2
8.8
8.4
8.0
7.6
7.2
6.8
6.4
6.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0

Fig. 8.

does not affect the dynamics of volatility, an observation which confirms Harvey and Siddique (1999).

Turning to cross-sectional skewness and kurtosis, we document that higher moments of stock-index and foreign-exchange returns are strongly related. We show that very large events of a given sign tend to occur jointly. In particular, this result indicates that crashes will tend to happen at the same time.

The model is presently univariate. As such it could be of value for option pricing, for stress testing, and for Value at Risk measurements. Also in cases, where Monte-Carlo simulations are required, our model can be used, since we show how to generate returns with time varying volatility, skewness and kurtosis.

In Rockinger and Jondeau (2001) we develop a multivariate framework. The way we obtain the multivariate density is by using copula functions that allow us to connect univariate models of the kind developed in this paper (see, for instance, Nelsen, 1999). This multivariate framework can be of value to test asset pricing models involving higher moments. For instance, in Rubinstein (1973), one needs to estimate comoments. The nth conditional comoment between some asset i, with return r_i, and the market portfolio with return r_0 is defined as

$$\mathrm{E}_t[(r_i - \mathrm{E}_t[r_i])(r_0 - \mathrm{E}_t[r_0])^{n-1}], \quad n \geqslant 2.$$

If $n=2$ ($n=3$) we obtain coskewness (respectively cokurtosis) with the market return. The authors' present work focuses on a model where the comoments are computed using a numerical integration.

$$\int (r_i - \mathrm{E}_t[r_i])(r_0 - \mathrm{E}_t[r_0])^{n-1} f_t(r_i, r_0)\, \mathrm{d}r_i\, \mathrm{d}r_0.$$

The density is modeled with a time-varying copula whereas the marginal distribution of returns is modeled using the framework developed in this study.

Acknowledgements

The second author acknowledges help from the HEC Foundation and the European Community TMR Grant: "Financial Market Efficiency and Economic Efficiency." We also wish to thank Sanvi Avouyi-Dovi, Anil Bera, Jean Cordier, Louis Ederington, Christian Gourieroux, Bob Korkie, Mark Rubinstein, Stephen Satchell, as well as Stephen Taylor for their comments. We are grateful to Sébastian Laurent for having pointed out an error. A very special thanks goes to Philip Gill for his explanations on optimization and ongoing help with SNOPT. This paper was written during a much appreciated stay of the second author at UCSD. We are grateful for suggestions from conference and seminar participants at the FMA Paris, the EFMA Lugano, and City University London. The usual disclaimer applies. The Banque de France does not necessarily endorse the views expressed in this paper.

Appendix A

In this appendix, we derive a certain number of theoretical results concerning Hansen's (1994) generalized t distribution. We will extensively use the following result of Gradshteyn and Ryzhik (1994, p. 341, 3.241.4):

$$\int_0^\infty x^{\mu-1}(p+qx^v)^{-(n+1)}\,\mathrm{d}x = \frac{1}{vp^{n+1}}\left(\frac{p}{q}\right)^{\mu/v}\frac{\Gamma(\mu/v)\Gamma(1+n-(\mu/v))}{\Gamma(1+n)}$$

defined for $0<\mu/v<n+1$, $p\neq 0$, and $q\neq 0$, where Γ is the gamma function for which $\Gamma(x)=(x-1)\Gamma(x-1)$ and $\Gamma(1/2)=\sqrt{\pi}$. We first use this lemma to verify that expression (1), given in the text, truly defines a density.

The starting point is the conventional Student-t distribution with η degrees of freedom defined by

$$f(x|\eta)=c\left(1+\frac{x^2}{\eta-2}\right)^{-(\eta+1)/2}, \quad x\in\mathscr{R}.$$

The constant c has to be set in such a manner that the probability mass integrates to 1. This requires that

$$\int_{x\in\mathscr{R}} f(x|\eta)\,\mathrm{d}x = c\int_{-\infty}^{0}\left(1+\frac{x^2}{\eta-2}\right)^{-(\eta+1)/2}\mathrm{d}x + c\int_0^\infty\left(1+\frac{x^2}{\eta-2}\right)^{-(\eta+1)/2}\mathrm{d}x = 1.$$

A change of variable in the left central integral, from x into $-x$, shows that the two integrals in the center are equal. A straightforward application of the lemma yields the well-known result that

$$c=\frac{\Gamma((\eta+1)/2)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)}.$$

We assume now that X follows a conventional Student-t distribution with η d.o.f. In order to introduce an asymmetry, Hansen considers the new random variable

$$Y=\begin{cases}(1-\lambda)X & \text{if } X\leqslant 0,\\ (1+\lambda)X & \text{if } X>0\end{cases}$$

with candidate density

$$gt(y|\eta,\lambda)=\begin{cases} c\left(1+\dfrac{1}{\eta-2}\left(\dfrac{y}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } y\leqslant 0,\\[2ex] c\left(1+\dfrac{1}{\eta-2}\left(\dfrac{y}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } y>0.\end{cases}$$

Straightforward computations show that the constant c is the same as for the conventional Student-t distribution. The first moment of Y is given by

$$\begin{aligned} a \equiv \mathrm{E}[Y] &= c\int_{-\infty}^{0} y\left(1+\frac{1}{\eta-2}\left(\frac{y}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} \mathrm{d}y \\ &\quad + c\int_{0}^{\infty} y\left(1+\frac{1}{\eta-2}\left(\frac{y}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} \mathrm{d}y \\ &= I_l^1 + I_r^1. \end{aligned}$$

We perform the change of variables $y=(1-\lambda)x$ and $(1+\lambda)x$ in the two integrals and apply the lemma to obtain

$$\begin{aligned} I_l^1 &= c(1-\lambda)^2\int_{-\infty}^{0} x\left(1+\frac{x^2}{\eta-2}\right)^{-(\eta+1)/2} \mathrm{d}x = -\frac{c}{2}(1-\lambda)^2(\eta-2)\frac{\Gamma((\eta-1)/2)}{\Gamma((\eta+1)/2)} \\ &= -\frac{c}{2}(1-\lambda)^2\,\frac{\eta-2}{\eta-1}. \end{aligned}$$

Similarly $I_r^1 = c/2(1+\lambda)^2(\eta-2)/(\eta-1) = -(1+\lambda)^2/(1-\lambda)^2 I_l^1$. Combination and simplification of the various terms yield the mean

$$a \equiv 4\lambda c\frac{\eta-2}{\eta-1}.$$

The second moment of Y follows using similar computations: $m_2 \equiv \mathrm{E}[Y^2] = I_l^2 + I_r^2$. The same change of variables as previously yields

$$I_l^2 = \frac{c}{2}(1-\lambda)^3(\eta-2)^{3/2}\,\frac{\Gamma\left(\frac{3}{2}\right)\Gamma((\eta-2)/2)}{\Gamma((\eta+1)/2)} = \frac{(1-\lambda)^3}{2}.$$

Also $I_r^2 = (1+\lambda)^3/(1-\lambda)^3 I_l^2$. After several simplifications, we get

$$m_2 = \mathrm{E}[Y^2] = 1 + 3\lambda^2.$$

Since $V[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2$, we obtain for the variance $b^2 \equiv V[Y] = 1 + 3\lambda^2 - a^2$.

Since the residuals of the GARCH model are assumed to have zero mean and unit variance, we introduce the random variable $Z=(Y-a)/b$ which will be centered, i.e., with mean 0, and reduced, i.e., with variance 1. The passage from Y to Z will not change the constant c, it is only necessary to multiply the density by the Jacobian of the transformation, that is b. The density of Z follows by replacing Y with $bZ+a$ and is displayed in formula (1) of the text.

These computations verify Hansen's. Our model involves, however, higher order moments that we compute now. The third moment of Y is given by $m_3 \equiv \mathrm{E}[Y^3] = I_l^3 + I_r^3$.

The same change of variables as previously yields

$$I_l^3 = -\frac{c}{2}(1-\lambda)^4(\eta-2)^2 \frac{\Gamma(2)\Gamma((\eta-3)/2)}{\Gamma((\eta+1)/2)} = -2c\frac{(1-\lambda)^4(\eta-2)^2}{(\eta-1)(\eta-3)},$$

where the first equality follows from a straightforward application of the lemma. The second equality follows from simple algebra. Also, $I_r^3 = -(1+\lambda)^4/(1-\lambda)^4 I_l^3$. Eventually, we obtain

$$m_3 = 16c\lambda(1+\lambda^2)\frac{(\eta-2)^2}{(\eta-1)(\eta-3)}$$

defined for $\eta > 3$. The third moment of Z may be obtained as a function of the various moments of Y. We obtain

$$\mathrm{E}[Z^3] = (\mathrm{E}[Y^3] - 3a\mathrm{E}[Y^2] + 2a^3)/b^3.$$

We now turn to the computation of the last moment of interest for this paper. A generalization for even higher moments can be easily obtained. The fourth moment may again be written as the sum of integrals: $m_4 = \mathrm{E}[Z^4] = I_l^4 + I_r^4$.

We have

$$I_l^4 = \frac{c}{2}(1-\lambda)^5(\eta-2)^{\frac{5}{2}} \frac{\Gamma\left(\frac{5}{2}\right)\Gamma((\eta-4)/2)}{\Gamma((\eta+1)/2)} = \frac{3}{2}(1-\lambda)^5\frac{\eta-2}{\eta-4}.$$

The various steps involved in the computation use the same techniques as previously. The first equality follows from the lemma and the second one from simple algebra. Also, it can be shown that $I_r^4 = (1+\lambda)^5/(1-\lambda)^5 I_l^4$. Regrouping terms, we obtain

$$m_4 = 3\frac{\eta-2}{\eta-4}(1+10\lambda^2+5\lambda^4)$$

defined for $\eta > 4$. We also obtain the associated moment of Z as

$$\mathrm{E}[Z^4] = (\mathrm{E}[Y^4] - 4a\mathrm{E}[Y^3] + 6a^2\mathrm{E}[Y^2] - 3a^4)/b^4.$$

We verified these formulae and their numerical implementation by computing the various moments via numerical integration of a generalized t distribution.

Appendix B

In the following, we present the computations of the gradient of the log-likelihood. To simplify notations, we focus on the gradient of a single observation. Summation of these gradients yields the sample gradients. We define $d = (by/\sigma + a)/(1-\lambda s)$, where s is a sign dummy taking the value of 1 if $by/\sigma + a < 0$ and -1 otherwise. We also define $v_1 = 1 + d^2/(\eta-2)$. We recall that the log-likelihood of an observation is

$$l = \ln(b) + \ln(c) - \ln(\sigma) - \frac{\eta+1}{2}\ln(v_1).$$

To obtain the gradients, we decompose the problem and make frequent use of the chain rule of differentiation. A necessary ingredient to obtain the gradients is

$$\frac{\partial l}{\partial \sigma} = -\frac{1}{\sigma} + \frac{\eta+1}{2}\frac{1}{v_1}\frac{2d}{\eta-2}\frac{by}{(1-\lambda s)\sigma^2}.$$

Next, to obtain $\partial l/\partial\lambda$, we note that

$$\frac{\partial a}{\partial\lambda} = 4c(\eta-2)/(\eta-1),$$

$$\frac{\partial b}{\partial\lambda} = (3\lambda - a\,\partial a/\partial\lambda)/b,$$

$$\frac{\partial d}{\partial\lambda} = \left(\frac{\partial b}{\partial\lambda}\frac{y}{\sigma} + \frac{\partial a}{\partial\lambda}\right)(1-\lambda s)^{-1} + s(by/\sigma + a)(1-\lambda s)^{-2},$$

$$\frac{\partial v_1}{\partial\lambda} = \frac{2d}{\eta-2}\frac{\partial d}{\partial\lambda},$$

so that

$$\frac{\partial l}{\partial\lambda} = \frac{1}{b}\frac{\partial b}{\partial\lambda} - \frac{\eta+1}{2}\frac{1}{v_1}\frac{\partial v_1}{\partial\lambda}.$$

To obtain $\partial l/\partial\eta$, we proceed similarly. We notice that $\partial c/\partial\eta = c\partial\ln(c)/\partial\eta$ and

$$\frac{\partial\ln(c)}{\partial\eta} = \frac{1}{2}\Psi\left(\frac{\eta+1}{2}\right) - \frac{1}{2}\frac{1}{\eta-2} - \frac{1}{2}\Psi\left(\frac{\eta}{2}\right),$$

$$\frac{\partial a}{\partial\eta} = 4\lambda(\eta-2)(\eta-1)^{-1}\frac{\partial c}{\partial\eta} + 4\lambda c[(\eta-1)^{-1} - (\eta-2)(\eta-1)^{-2}],$$

$$\frac{\partial b}{\partial\eta} = -\frac{a}{b}\frac{\partial a}{\partial\eta},$$

$$\frac{\partial d}{\partial\eta} = \left(\frac{\partial b}{\partial\eta}\frac{y}{\sigma} + \frac{\partial a}{\partial\eta}\right)(1-\lambda s),$$

$$\frac{\partial v_1}{\partial\eta} = -(\eta-2)^{-2}d^2 + 2(\eta-2)^{-1}d\,\frac{\partial d}{\partial\eta},$$

so that

$$\frac{\partial l_t}{\partial\eta} = \frac{1}{b}\frac{\partial b}{\partial\lambda} + \frac{\partial\ln(c)}{\partial\eta} - \frac{3}{2}\ln(v_1)(\eta+1)\frac{1}{v_1}\frac{\partial v_1}{\partial\eta},$$

where $\Psi(\cdot)$ is the derivative of the log of the gamma function. This derivative is known as the di-gamma function, which may be implemented with desired accuracy. The Fortran library IMSL also implements this function.

The passage from η_t to $\tilde{\eta}_t$ and from λ_t to $\tilde{\lambda}_t$ involves the logistic transform. The computation of $\partial\eta_t/\partial\tilde{\eta}_t$ and of $\partial\lambda_t/\partial\tilde{\lambda}_t$ is therefore trivial. For a GARCH specification of volatility with $\sigma_t^2 = a_0 + b_0 y_{t-1}^2 + c_0 \sigma_{t-1}^2$ we obtain

$$\frac{\partial l_t}{\partial a_0} = \frac{\partial l_t}{\partial \sigma}\left(1 + c_0\,\frac{\partial h_{t-1}}{\partial a_0}\right), \quad \frac{\partial l_t}{\partial b_0} = \frac{\partial l_t}{\partial \sigma}\frac{1}{2\sigma}\left(y_{t-1}^2 + c_0\,\frac{\partial h_{t-1}}{\partial b_0}\right),$$

$$\frac{\partial l_t}{\partial c_0} = \frac{\partial l_t}{\partial \sigma}\frac{1}{2\sigma}\left(h_{t-1} + c_0\,\frac{\partial h_{t-1}}{\partial c_0}\right),$$

$$\frac{\partial h_1}{\partial a_0} = 1 + c_0\,\frac{\partial h_0}{\partial a_0} = 1, \quad \frac{\partial h_2}{\partial a_0} = 1 + c_0.$$

For the specification $\tilde{\lambda}_t^2 = a_2 + b_2 y_{t-1}^2 + c_2 \tilde{\lambda}_{t-1}^2$, we obtain

$$\frac{\partial l_t}{\partial a_2} = \frac{\partial l_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \tilde{\lambda}_t}, \quad \frac{\partial l_t}{\partial b_2} = \frac{\partial l_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \tilde{\lambda}_t} y_{t-1}, \quad \frac{\partial l_t}{\partial c_2} = \frac{\partial l_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \tilde{\lambda}_t}\left(\tilde{\lambda}_t + c_2\,\frac{\partial \tilde{\lambda}_{t-1}}{\partial c_2}\right).$$

These equations may be computed recursively, starting from $\partial \tilde{\lambda}_{t-1}/\partial c_2 = 0$.

Appendix C

In this appendix, we report on a simulation experiment that we performed to assess the quality of the estimates. We first discuss the main model to be simulated and then present our conclusions.

For the reader's convenience, we present again the general model that we simulate and estimate:

$$r_t = y_t,$$

$$y_t = \sigma_t z_t,$$

$$\sigma_t^2 = a_0 + b_0^+ (y_{t-1}^+)^2 + b_0^- (y_{t-1}^-)^2 + c_0 \sigma_{t-1}^2,$$

$$z_t \sim GT(z_t | \eta_t, \lambda_t),$$

$$\tilde{\eta}_t = a_1 + b_1^+ y_{t-1}^+ + b_1^- y_{t-1}^- + c_1 \tilde{\eta}_{t-1},$$

$$\tilde{\lambda}_t = a_2 + b_2 y_{t-1}^2 + c_2 \tilde{\lambda}_{t-1},$$

$$\eta_t = g_{]2,+30]}(\tilde{\eta}_t), \ \lambda_t = g_{]-1,1[}(\tilde{\lambda}_t).$$

We have already presented in the main text how to generate numbers that are distributed as the generalized Student-t.

The first issue that we examine concerns the properties of the estimates obtained when the model is well specified. To address this issue we simulated 1,000 samples of size 5,000, where we set $b_1^+ = b_1^- = c_1 = 0$ and the other parameters were set as indicated in column 1 of Table 9. Out of these 1,000 simulations, we found that in 11.29% of the cases the algorithm could not converge with a relative gradient smaller than 10^{-6}. In another run with a precision of 10^{-4}, we always obtained convergence. On average, it took us 192 s before convergence was reached. The simulations and

Table 9
Results of Monte-carlo Experiments. Column 1 and 5 present the parameters used in a simulation corresponding to a model such as the one described in Table 3. We simulated 1,000 sample of size 5,000

	Experiment 1				Experiment 2			
	True parameter 1	Average estimate 2	Average standard error 3	Standard error estimate 4	True parameter 5	Average estimate 6	Average standard error 7	Standard error estimate 8
a_0	0.05	0.0932	0.0147	0.0160	0.05	0.0867	0.0165	0.0171
b_0^+	0.03	0.0372	0.0075	0.0076	0.03	0.0484	0.0096	0.0095
b_0^-	0.07	0.0949	0.0095	0.0081	0.07	0.1125	0.0127	0.0124
c_0	0.90	0.9060	0.0088	0.0085	0.90	0.8973	0.0107	0.0109
a_1	−1.00	−1.0035	0.2091	0.2448	−1.00	−0.9778	0.2203	0.2119
a_2	−0.02	−0.0296	0.0098	0.0125	−0.02	−0.0117	0.0286	0.0311
b_2	0.15	0.1395	0.0179	0.0196	0.00	−0.0006	0.0165	0.0227
c_2	0.80	0.7956	0.0294	0.0342	0.00	0.3977	0.2536	0.5214

associated estimations yield, for each parameter and its standard error, a sample of 1,000 observations, for which one can comment descriptive statistics.

Column 2 of Table 9 presents the average parameter estimates. We notice that a_1 and the dynamic of $\tilde{\lambda}_t$ is very well estimated. Inspection of the parameters b_0^+ and b_0^- in the volatility equation reveals that the estimate of b_0^- is slightly too high.

Next, we turn to the quality of the standard errors. Here we report standard errors obtained using the "sandwich method", i.e., using $I^{-1}JI^{-1}$ where I is the information matrix and J the matrix of outer products. We compare the average of the computed standard errors (see column 3) with the standard deviation of the estimated parameters (see column 4). We notice (with relief) that our standard errors seem to correspond to the actual measure. Inspection of the histogram of the parameter estimates displayed a normal distribution.

Even though we did not attempt a formal proof, these observations led us to conjecture that if the model is well specified, the likelihood estimates are distributed asymptotically normal.

The next experiment on which we wish to report concerns the degeneracy, mentioned in the main text, when $b_2 = 0$, yet where one estimates a model with autoregressive $\tilde{\lambda}_t$. To assess this situation, we performed another experiment where we set $a_2 = -0.02$, $b_2 = c_2 = 0$. We estimated a model with autoregressive dynamic for $\tilde{\lambda}_t$.

In column 6, we present the average of the estimated parameters. Concerning the parameters of interest, we find that b_2 is estimated well, whereas there is a bias for the parameter c_2. The average is 0.398. Even though the standard deviation of the parameter is huge (0.521), we find that, in 58.75% of the simulations, one would have rejected the null hypothesis that $c_2 = 0$ at the 5% significance level. This finding demonstrates that, if the parameter corresponding to lagged innovations in an autoregressive specification is not significant, the interpretation of a significant autoregressive parameter should be done with great care.

References

Baker Jr., G.A., Graves-Morris, P., 1996. Padé Approximants. Cambridge University Press, Cambridge.

Barone-Adesi, G., 1985. Arbitrage equilibrium with skewed asset returns. Journal of Financial and Quantitative Analysis 20 (3), 299–313.

Bera, A.K., Higgins, M.L., 1993. ARCH models: properties, estimation and testing. Journal of Economic Surveys 7 (4), 305–362.

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.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11 (2), 143–172.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52 (1–2), 5–59.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. IV. Elsevier Science, Amsterdam, pp. 2959–3038 (Chapter 49).

Campbell, J.Y., Hentschel, L., 1992. No news is goos news: an asymmetric model of changing volatility. Journal of Financial Economics 31 (3), 281–318.

Engle, R.F., 1982. Auto-regressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.

Engle, R.F., Gonzalez-Rivera, G., 1991. Semi-parametric ARCH models. Journal of Business and Economic Statistics 9 (4), 345–359.

Fama, E., 1963. Mandelbrot and the stable Paretian hypothesis. Journal of Business 36, 420–429.

Fang, H., Lai, T.-Y., 1997. Co-kurtosis and capital asset pricing. Financial Review 32 (2), 293–307.

Friend, I., Westerfield, R., 1980. Co-skewness and capital asset pricing. Journal of Finance 35 (4), 897–913.

Gill, Ph.E., Murray, W., Saunders, M.A., 1997. SNOPT: an SQP algorithm of large-scale constrained optimization. Report NA 97-2, Department of Mathematics, University of California, San Diego.

Gill, Ph.E., Murray, W., Saunders, M.A., 1999. User's guide for SNOPT 5.3: a Fortran package for large-scale nonlinear programming. Working paper, Department of Mathematics, UCSD, California.

Glosten, R.T., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48 (5), 1779–1801.

Gradshteyn, I.S., Ryzhik, I.M., 1994. Table of Integrals, Series, and Products, 5th Edition. Academic Press, New York.

Hamao, Y., Masulis, R.W., Ng, V., 1990. Correlations in price changes and volatility across international stock markets. Review of Financial Studies 3 (2), 281–307.

Hamburger, H., 1920. Über eine Erweiterung des Stieltjesschen momentproblems. Mathematische Zeitschrift 7, 235–319.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimator. Econometrica 50 (4), 1029–1054.

Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review 35 (3), 705–730.

Harvey, C.R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34 (4), 465–487.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55 (3), 1263–1295.

Hausdorff, F., 1921a. Summationsmethoden und momentfolgen, I. Mathematische Zeitschrift 9, 74–109.

Hausdorff, F., 1921b. Summationsmethoden und momentfolgen, II. Mathematische Zeitschrift 9, 280–299.

Hwang, S., Satchell, S.E., 1999. Modelling emerging market risk premia using higher moments. Journal of Finance and Economics 4 (4), 271–296.

Jorion, Ph., 1995. Predicting volatility in the foreign exchange market. Journal of Finance 50 (2), 507–528.

Kan, R., Zhou, G., 1999. A critique of the stochastic discount factor methodology. Journal of Finance 54 (4), 1221–1248.

Korkie, B., Sivakumar, R., Turtle, H.J., 1997. Skewness persistence: it matters, just not how we thought. Working paper, Faculty of Business, University of Alberta, Canada.

Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31 (4), 1085–1100.

Kroner, K.F., Ng, V.K., 1998. Modeling asymmetric comovements of asset returns. Review of Financial Studies 11 (4), 817–844.

Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960–1990? Journal of International Money and Finance 14 (1), 3–26.

Loretan, M., Phillips, P.C.B., 1994. Testing covariance stationarity under moment condition failure with an application to common stock returns. Journal of Empirical Finance 1, 211–248.

Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 35, 394–419.

Mood, A.M., Graybill, F.A., Boes, D.C., 1982. Introduction to the Theory of Statistics. McGraw-Hill, New York.

Nelsen, R.B., 1999. An Introduction to Copulas. Springer, Berlin, New York.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59 (2), 370–397.

Newey, W., 1985. Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Premaratne, G., Bera, A.K., 1999. Modeling asymmetry and excess kurtosis in stock return data. Working paper, Department of Economics, University of Illinois.

Ramchand, L., Susmel, R., 1998. Volatility and cross correlation across major stock markets. Journal of Empirical Finance 5 (4), 397–416.

Rockinger, M., Jondeau, E., 2001. Conditional dependency of financial series: an application of copulas. Working paper, Banque de France, NER#82.

Rockinger, M., Jondeau, E., 2002. Entropy densities with an application to autoregressive conditional skewness and kurtosis. Journal of Econometrics 106 (1), 119–142.

Rubinstein, M.E., 1973. The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis 8 (1), 61–69.

Sears, R.S., Wei, K.C.J., 1985. Asset pricing, higher moments, and the market risk premium: a note. Journal of Finance 40 (4), 1251–1253.

Sears, R.S., Wei, K.C.J., 1988. The structure of skewness preferences in asset pricing models with higher moments: an empirical test. Financial Review 23 (1), 25–38.

Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies 62 (4), 639–661.

Stieltjes, T.J., 1894. Recherches sur les fractions continues. Annales de la Faculté des Sciences de Toulouse 8 (1), 1–22.

Susmel, R., Engle, R.F., 1994. Hourly volatility spillovers between international equity markets. Journal of International Money and Finance 13 (1), 3–25.

Tan, K.-J., 1991. Risk return and the three-moment capital asset pricing model: another look. Journal of Banking and Finance 15 (2), 449–460.

White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48 (4), 817–838.

Widder, D.V., 1946. The Laplace Transform. Princeton University Press, Princeton, NJ.

Zakoïan, J.M., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and Control 18 (5), 931–955.