CRDE

CAHIER 29-2001

AN EIGENFUNCTION APPROACH FOR VOLATILITY MODELING

Nour MEDDAHI

Centre de recherche et développement en économique

C.P. 6128, succursale Centre-ville
Montréal QC H3C 3J7

Téléphone : (514) 343-6557
Télécopieur : (514) 343-5831
crde@crde.umontreal.ca
http://www.crde.umontreal.ca/

Université de Montréal

CAHIER 29-2001

AN EIGENFUNCTION APPROACH FOR VOLATILITY MODELING

Nour MEDDAHI[1]

[1] Centre de recherche et développement en économique (C.R.D.E.), CIRANO and Département de sciences économiques (Université de Montréal), and CEPR

October 2001

The author would like to thank Torben Andersen, Federico Bandi, Tim Bollerslev, Marine Carrasco, Xiahong Chen, Peter Christoffersen, Frank Diebold, Jean-Marie Dufour, Rob Engle, Jean-Pierre Florens, Ron Gallant, Eric Ghysels, Silvia Gonçalves, Lars Hansen, Éric Jacquier, Benoit Perron, Éric Renault, Enrique Sentana, Neil Shephard, George Tauchen, and Nizar Touzi for helpful discussions when he was writing this paper and/or for comments on an earlier draft. The author is particularly indebted to Bryan Campbell, René Garcia and especially to Christian Gouriéroux. He thanks seminar participants at CIRANO, Triangle (Duke, NCS, and UNC), University of Chicago, MITACS workshop, C.R.D.E. conference on “Modeling, Estimating and Forecasting Volatility”, April 2001, CIRANO conference on “Financial Mathematics and Econometrics”, June 2001, Econometric Society European Meeting, Lausanne, August 2001, and the “International Conference on Modeling and Forecasting Financial Volatility”, Perth, September 2001, for comments. Many thanks also go to Neil Shephard for providing the data and Brahim Boudarbat for excellent research assistance. The author acknowledges financial support from FCAR, IFM2, and MITACS, and is solely responsible for any remaining errors.

RÉSUMÉ

Dans cet article, nous proposons une nouvelle approche pour la modélisation de la volatilité en temps discret et continu. Nous adoptons la même approche que la littérature de la volatilité stochastique en supposant que la volatilité est une fonction d'une variable d'état. Néanmoins, au lieu de supposer que la fonction de lien est donnée de manière ad hoc (par exemple, exponentielle ou affine), nous supposons que c'est une combinaison linéaire des fonctions propres de l'opérateur espérance conditionnelle (ou générateur infinitésimal) associé à la variable d'état en temps discret (ou continu). Les modèles populaires exponentiels et racine carrée sont des exemples où les fonctions propres sont respectivement les polynômes de Hermite et de Laguerre. L'approche par fonctions propres a au moins six avantages : i) elle est générale puisque toute fonction de carré intégrale peut être écrite comme combinaison linéaire des fonctions propres; ii) l'orthogonalité des fonctions propres permet d'utiliser les interprétations usuelles de l'analyse en composantes principales linéaires; iii) les dynamiques induites de la variance et du carré de l'innovation sont des ARMA et donc sont simples pour la prévision et l'inférence statistique; iv) plus important, cette approche génère des queues épaisses pour les processus de volatilité et de rendements; v) à l'opposé des modèles usuels, la variance de la variance est une fonction flexible de la variance; vi) ces modèles sont robustes vis-à-vis de l'agrégation temporelle.

Mots clés : volatilité, volatilité stochastique, générateur infinitésimal, espérance conditionnelle, fonctions propres, ARMA, queues épaisses, GMM

ABSTRACT

In this paper, we introduce a new approach for volatility modeling in discrete and continuous time. We follow the stochastic volatility literature by assuming that the variance is a function of a state variable. However, instead of assuming that the loading function is ad hoc (e.g., exponential or affine), we assume that it is a linear combination of the eigenfunctions of the conditional expectation (resp. infinitesimal generator) operator associated to the state variable in discrete (resp. continuous) time. Special examples are the popular log-normal and square-root models where the eigenfunctions are the Hermite and Laguerre polynomials respectively. The eigenfunction approach has at least six advantages: i) it is general since any square integrable function may be written as a linear combination of the eigenfunctions; ii) the orthogonality of the eigenfunctions leads to the traditional interpretations of the linear principal components analysis; iii) the implied dynamics of the variance and squared return processes are ARMA and, hence, simple for forecasting and inference purposes; (iv) more importantly, this generates fat tails for the variance and returns processes; v) in contrast to popular models, the variance of the variance is a flexible function of the variance; vi) these models are closed under temporal aggregation.

Key words : volatility, stochastic volatility, infinitesimal generator, conditional expectation, eigenfunctions, ARMA, fat tails, GMM

1 Introduction

In 1963, Benoit Mandelbrot reported two important facts that characterize financial returns: fat tails and the clustering effect of the shocks. This lead researchers to introduce new models capturing these two empirical facts. The most important contribution is Engle (1982) who introduces the ARCH models and show that time varying volatility implies fat tails and clustering effect. The initial ARCH model was generalized and improved in several directions to better describe the data, in particular by Bollerslev (1986, GARCH), Nelson (1991, EGARCH) and Baillie, Bollerslev and Mikkelsen (1996, FIGARCH). Following the main idea of ARCH models, i.e. time-varying volatility, the Stochastic Volatility (SV) class of models was introduced by specifying the volatility as an unobservable state variable. Examples of SV models are the log-normal ones of Taylor (1986), Nelson (1988) and Harvey, Ruiz and Shephard (1994), and the stochastic autoregressive volatility models (SARV) of Andersen (1994) in discrete time, Wiggins (1987) and Melino and Turnbull (1990) in continuous time as well as square-root and affine models of Heston (1993) and Duffie, Pan and Singleton (2000) respectively.[1]

Nowadays, it is well accepted that ARCH and SV models capture well the clustering effect. However, this is not the case for the fat tails. In particular, recent works on exchange rates as Kim, Shephard and Chib (1998) and especially on asset returns, as Bakschi, Cao and Chen (1997), Andersen, Benzoni and Lund (2001), suggest that the popular models do not capture well the data tails like the kurtosis. This lead to some generalization of the popular models by including jumps in the returns or the volatility (e.g., Andersen, Benzoni, and Lund, 2001; Bates, 1996; Chernov et al., 1999; Duffie, Pan and Singleton, 1999; Eraker, Johannes and Polson, 1999) or by taking a multifactor model for the volatility (e.g., Gallant, Hsu and Tauchen, 1999; Meddahi and Renault, 1996). This lead also to consider alternative specification of the variance process (e.g., Jones, 2000; Chernov et al., 2001; Barndorff-Nielsen and Shephard, 2001).

In this paper, we will first explain why the popular models fail to capture the tails of the data. Then, we show how to extend them in order to generate processes with fat tails and describe well the dynamics of the data. As in the SV literature, we assume that the variance is a function of a state variable. However, instead of assuming that the loading function is ad hoc (e.g., exponential or affine), we assume that it is a linear combination of the eigenfunctions of the conditional expectation (resp infinitesimal generator) operator associated to the state variable in discrete (resp continuous) time.

To understand why the popular models fail to capture fat tails, let us make an important difference between the two facts. The fat tails are related to the unconditional distribution of the returns while the clustering effect is related to the conditional distribution of the returns. But these two distributions describe two very different characteristics. For instance, two different processes may have the same marginal Gaussian distribution while one is i.i.d. and the other one is a Gaussian AR(1) or a long-memory process. As a consequence, a volatility model will fit more and more the data if the unconditional and the conditional distributions are less and less linked.

[1] See Bollerslev, Engle and Nelson (1994) for a review of the ARCH literature, Ghysels, Harvey and Renault (1996) and Shephard (1996) for the SV one. See Andersen (1992) and Meddahi and Renault (2000) for a theoretical comparison between ARCH and SV and Kim, Shephard and Chib (1998) for an empirical one.

This is not the case of GARCH models where the strong persistence of the volatility suggests that in many empirical cases, returns fourth moment is infinite.[2] The EGARCH and especially SV models were an important improvement since they reduce the link between the unconditional and conditional distributions. In particular, all the moments are finite in these models. Two factors in volatility models also reduces this link since one factor captures the unconditional distribution of the returns (fat tails) while the second one captures the conditional distribution (volatility persistence). Empirical studies that consider two factor models in the volatility, e.g. Barndorff-Nielsen and Shephard (2001) and Chernov et al. (2001), find that one factor is strongly mean reverting while the other one is slowly mean reverting. Typically, the first one captures the fat tails while the second one captures the persistence of the volatility. Another approach that have this flexibility is considered in Chen, Hansen and Scheinkman (2000). For multivariate diffusion modeling purposes, they specify the marginal distribution of the process and the diffusion term. Then, the drift and the conditional distribution are deduced.[3]

In this paper, we consider a flexible approach for volatility modeling. We follow the main idea of the SV literature, that is by specifying the volatility as a function of a state variable. Thus, all the dynamics of the volatility and, hence, those of the returns, will be governed by the dynamic of the state variable. This state variable may be governed for instance by a Gaussian AR(1) process (as for log-normal models) or a square-root model (as for square-root and affine models). However, rather than specifying that the variance is equal to a specific function of the state variable as exponential (log-normal case) or identity (square-root and affine cases), we will assume that the variance is a flexible form of the state variable. Many choices of this flexible form may be considered. In this paper, we assume that the variance is a linear combination of the eigenfunctions of the conditional expectation (resp infinitesimal generator) operator associated to the state variable in discrete (resp continuous) time. To be more specific, assume that the state variable f_t that governs the volatility is a standardized Gaussian AR(1) process, i.e., $f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2} v_t$, v_t i.i.d. $\mathcal{N}(0,1)$ where $\mid \gamma \mid < 1$. Then if one considers the Hermite polynomials, $H_i(.)$, we have $E[H_i(f_{t+1}) \mid f_\tau, \tau \leq t] = \gamma^i H_i(f_t)$, i.e. the Hermite polynomial H_i is an AR(1) process. This justifies the eigenfunction terminology.. Moreover, the Hermite polynomials are uncorrelated and any square-integrable function of f_t is a linear combination of the former. Hence, specifying the variance as a linear combination of the Hermite polynomials is a flexible approach. In particular, this is the case for the exponential function. However, as we will see later, the exponential function puts an important weight (more than 75%) on the first eigenfunction, $H_1(f_t) = f_t$, and less on higher order polynomials that capture fat tails. This is mainly why the log-normal model does not capture well the fat tails since the unconditional distribution of $H_1(f_t)$ is Gaussian. In this paper we will set these weights as free parameters and estimate them. This will generate fat tails. It is important to understand that we are not considering approximation of the volatility in terms of linear combination of the Hermite polynomials. We assume that the true volatility is a linear combination of the Hermite

[2] Indeed, if ε_t is a GARCH process, then $\exists r^*$ such that $\forall r > r^*$ $E[\mid \varepsilon_t \mid^r] = +\infty$, which is clearly a limitation.
[3] Barkovec and Kluppelberg (2000) suggest also the same approach in the univariate case.

polynomials. This may be viewed as a reduced form of the volatility.[4] Then we make the usual statistical inference, as specification tests.

The previous decomposition is not specific to Gaussian processes and Hermite polynomials. Indeed there is a general theory in terms of eigenfunctions of the conditional expectation operator of a state variable f_t in both discrete and continuous time. In the Gaussian case, the eigenfunctions are the Hermite polynomials while in the square-root model of Heston (1993), the eigenfunctions are the Laguerre polynomials. It turns out that Heston (1993) sets the variance equal to the first eigenfunction. Thus, the first eigenfunction explains 100% of the volatility while its marginal distribution is Gamma, i.e. with thin tails. As a consequence, one will have more tails by adding higher order polynomials in the variance decomposition. Finally, in continuous time, the eigenfunctions are also those of the infinitesimal generator of the diffusion; see Hansen and Scheinkman (1995) and Hansen, Scheinkman and Touzi (1998).

Indeed, our approach was motivated by these works and by the subsequent works that consider these eigenfunctions as nonlinear principal components (Darolles, Florens and Renault, 1998; Darolles, Florens and Gouriéroux, 1998; Chen, Hansen and Scheinkman, 2000). Instead of being nonparametric as in these papers, we adopt a parametric approach. Therefore, the estimation and implementation of our approach are simpler. Besides, we consider latent variables (SV models) and, hence, the observable processes that we consider are not Markovian. Thus, on one hand, by being parametric, we are less general than these papers in the Markovian case and, on the other hand, we are more general since we can consider non Markovian processes.

In summary, our approach of volatility modeling consists on considering a state variable with simple dynamics and specifying the variance as a linear combination of the eigenfunctions associated to the state variable. This approach is general and encompasses all the popular models. Moreover, while our first interest is modeling volatility, it is clear that our approach is a general one for nonlinear state space modeling.

The rest of the paper is organized as follows. In section 2, we explain why the log-normal and square-root models fail to capture fat tails. Then we introduce the Hermite SV model, HSV, (resp Laguerre, LSV) where the variance is a linear combination of Hermite (resp Laguerre) polynomials of a Gaussian AR(1) (resp square-root) process. In this section, we study the properties of these models. In particular: we characterize the moments of discrete time models while Meddahi (2001-a) does for continuous time ones; we consider an important robustness property in terms of the long run (Conley et al., 1997; Conley, Hansen and Liu, 1997); we use the Fiorentini and Sentana's (1998) persistence to derive the persistence of the volatility and the squared residuals; following Meddahi and Renault (2000), we give the semiparametric models that Meddahi (2001-b) showed that they are closed under temporal aggregation; we characterize the asymptotic behavior of the kurtosis of the aggregated process; finally, we show that the variance of the variance has a nice property that usual models do not have, which was criticized by Nelson and Foster (1994). In section 3, we recap the general theory of eigenfunctions of conditional expectations and infinitesimal generator operators and provide several examples. In section 4, we introduce the general class of the Eigenfunctions Stochastic Volatility (ESV) models

[4] I thank Lars Hansen for giving me this argument.

and study their properties. Moreover, we propose a new approach for multifactor modeling. The last section concludes and proposes several extensions while the Appendix provides the proofs.

2 The Hermite and Laguerre stochastic volatility models

In this section, we explain in the first subsection why the log-normal and square-root SV models fail to capture fat tails in both discrete and continuous time. Then we introduce the Hermite and Laguerre SV model in discrete and continuous time in two different subsections. Their properties are studied.

2.1 Understanding why the log-normal and square-root SV models fail to capture fat tails

2.1.1 The log-normal SV model

Consider the most popular SV model in discrete time, i.e. the log-normal model considered by Taylor (1986), Nelson (1988) and popularized by Harvey, Ruiz and Shephard (1994):

$$\varepsilon_t = \sigma_{t-1} u_t, \quad \text{with} \tag{2.1}$$

$$\log(\sigma_t^2) = \omega + \gamma \log(\sigma_{t-1}^2) + \sigma_v v_t, \quad \text{where} \tag{2.2}$$

$$(u_t, v_t) \;\; \text{i.i.d.} \;\; \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{2.3}$$

As we pointed out in the introduction, the usual SV model fails to capture the tails of the data. For instance, Kim, Shephard and Chib (1998) compare the log-normal SV with Gaussian and Student GARCH(1,1). They show by a likelihood comparison that the log-normal SV dominates the Gaussian GARCH(1,1) but that the Student GARCH(1,1) model is as good as (e.g., Pound-US$ exchange rate) or better (Yen-US$ exchange rate) than the log-normal one. To understand why the log-normal fails to capture very fat tails, let us compute the kurtosis of the returns. Simple calculus show that

$$\text{Kurtosis}[\varepsilon_t] = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} = E[u_t^4]\exp(\sigma^2) = \text{Kurtosis}[u_t]\frac{E[\sigma_t^4]}{E[\sigma_t^2]^2} \quad \text{where} \quad \sigma^2 \equiv \frac{\sigma_v^2}{1-\gamma^2}. \tag{2.4}$$

From this equation, it is clear that there are two ways to increase the kurtosis of the residuals ε_t. The first one is to assume that the process u_t has fat tail as for Student distribution. This approach is considered in discrete time in the ARCH literature by Bollerslev (1987) and in the SV literature by Jacquier, Polson and Rossi (1999) and Chib, Nardari and Shephard (2000). Observe that in continuous time, a Gaussian assumption of the standardized residuals is crucial since they are assumed to be Brownian motions and, hence, Gaussian. Moreover, while in the GARCH case, assuming that the standardized residuals are Student is not ambiguous, it is in the SV case. To be more specific, assume that

$$\varepsilon_t = \sigma_{t-1} u_t \quad \text{where} \quad u_t \sim T(\nu).$$

But a $\mathrm{T}(\nu)$ random variable u_t may be represented (see, e.g., Jacquier, Polson and Rossi, 1999) by $u_t = \tilde{u}_t\sqrt{z_t}$ where $\tilde{u}_t$ and z_t are two independent random variables that are respectively $\mathcal{N}(0,1)$ and inverse Gamma (i.e., ν/z_t is χ^2_ν). Then, the process ε_t is given by

$$\varepsilon_t = \tilde{\sigma}_{t-1}\tilde{u}_t \quad \text{where} \quad \tilde{\sigma}_{t-1} = \sigma_{t-1}\sqrt{z_t}.$$

Therefore, the conditional variance of ε_t given the sigma algebra generated by the past of ε_{t-1} and σ_{t-1}, i.e. $\sigma(\varepsilon_\tau, \sigma_\tau, \tau \leq t-1)$, is σ^2_{t-1} while the conditional variance of ε_t given $\sigma(\varepsilon_\tau, \tilde{\sigma}_\tau, \tau \leq t-1)$ is $\tilde{\sigma}^2_{t-1}$. The problem in the SV case is that these two informations are not available for the econometrician. Therefore, there is an identification problem and we need an identification assumption. For instance, by assuming that the log-variance is a Gaussian AR(1) process, Jacquier, Polson and Rossi (1999) and Chib, Nardari and Shephard (2000) identify the model. Observe that the model with Student distribution is a model with two factors in the volatility. Moreover, this is a multiplicative factors model. Finally, in continuous time, this looks like a subordinated process (Clark, 1973). See our multifactor approach in section 4.

Here, and in all our approach, we also allow the standardized residual to be Student or any other random variable. Thus, we are trying to explain why the volatility model, i.e. the process σ^2_t, fails to capture fat tails and how to improve it for a given process u_t.

Assume that the kurtosis of u_t is fixed. Hence, the kurtosis of ε_t is governed by σ^2 which is the variance of the log-variance. In practise, by using the empirical results of Kim, Shephard and Chib (1998), we can show that in the case of exchange rates, the value of $\exp(\sigma^2)$ is smaller than 2. In another analysis, Jacquier, Polson and Rossi (1994) show that for all the stock returns they consider, $\exp(\sigma^2)$ is smaller than 2.2. As a consequence, in the log-normal model, the kurtosis of the returns is less that 2.5 times the kurtosis of the standardized residual, u_t, which is assumed to be Gaussian. The equality between the first term and the last one in (2.4) is true in a general case since this needs only that u_t is independent from the past of σ_{t-1} and ε_{t-1}. Thus, to increase the kurtosis of ε_t, one has to increase the second moment of the variance process, i.e. $E[\sigma_t^4]$. This is what we consider now.

In the log-normal model, the variance σ^2_t is assumed to be a Gaussian AR(1) process. Let us define the state variable f_t by

$$f_t \equiv \frac{\log(\sigma_t^2) - \mu}{\sigma} \quad \text{where} \quad \mu \equiv E[\log(\sigma_t^2)]) = \frac{\omega}{1-\gamma}.$$

Then, it is easy to show that

$$f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2}\; v_t \quad \text{and} \quad f_t \;\sim \mathcal{N}(0,1). \tag{2.5}$$

The first part of (2.5) means that the process f_t is Gaussian AR(1) while the second part says that the unconditional distribution of f_t is $\mathcal{N}(0,1)$. It is important to notice that the unconditional distribution of f_t does not depend on the persistence parameter γ. The variance process σ^2_t is related to f_t by:

$$\sigma_t^2 = \exp(\mu)\exp(\sigma f_t).$$

To see how one can increase the second moment of σ^2_t from the previous formula, we will consider an expansion of $\exp(\sigma f_t)$. One can for instance consider the Taylor expansion on f_t^i, $i = 0, 1, ...$

We prefer to do the expansion in terms of the Hermite polynomials. The main reason is that in the Taylor expansion, the monomials f_t^i are correlated while Hermite polynomials are not. Moreover, the dynamics of f_t^i are more complicated than Hermite's ones (ARMA(i,i-1) and AR(1) respectively).

The Hermite polynomials are characterized by

$$H_0(x) = 1, \; H_1(x) = x \text{ and } \forall i > 1, H_i(x) = \frac{1}{\sqrt{i}}\{xH_{i-1}(x) - \sqrt{i-1}H_{i-2}(x)\}. \tag{2.6}$$

Then, it is well known that

$$E[H_i(f_{t+1}) \mid \varepsilon_\tau, f_\tau, \tau \leq t] = \gamma^i H_i(f_t) \quad \text{and} \quad E[H_i(f_t)H_j(f_t)] = \delta_{ij}. \tag{2.7}$$

The first part of (2.7) says that any Hermite polynomial is an AR(1). However, it is heteroskedastic. In the Appendix, we give the form of the heteroskedasticity. Since $H_0(f_t) = 1$, the second part of (2.7) implies that $E[H_i(f_t)] = 0$ for $i \geq 0$. It turns out that we have the relation (in mean square)

$$\exp(\sigma f_t - \frac{\sigma^2}{2}) = \sum_{i=0}^{\infty} \frac{\sigma^i}{\sqrt{i!}} H_i(f_t), \quad \text{m.s.} \tag{2.8}$$

As a consequence, the variance process is a linear combination of the Hermite polynomials with

$$\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t), \quad \text{where} \quad a_i = \exp(\mu + \frac{\sigma^2}{2})\frac{\sigma^i}{\sqrt{i!}}. \tag{2.9}$$

This decomposition is crucial to understand why the log-normal volatility model fails to generate fat tails. In this decomposition, the main coefficient is σ since this governs a_i, that is the relative importance of each Hermite polynomial, as well as the speed of convergence of a_i to zero.

Consider the case of Pound-US$ and Yen-US$ of Kim, Shephard and Chib (1998). By using their estimates, the value of σ is .75 and .628 respectively. As a consequence, the sequence a_i is decreasing and converges very quickly to zero. In Table 1, we give a_i, $i = 0, 1, ..., 10$ for the two data sets. Since $H_0 = 1$ and the Hermite polynomials are orthogonal with variance equal to one, we can use the usual variance decomposition in the factor analysis to give the relative weight of each component H_i in (2.9) as well as the cumulative weights. More precisely, the relative weight w_i of the polynomial H_i in (2.9) and the cumulative weights cum_i of the $H_1, ..., H_i$, in this decomposition are given by

$$w_i = \frac{a_i^2}{\sum_{j=1}^{\infty} a_j^2} \quad \text{and} \quad cum_i = \sum_{j=1}^{i} w_j, \quad \text{for} \quad i \geq 1. \tag{2.10}$$

The third and seventh column of Table 1 means that the relative weight of the first Hermite polynomial, i.e. $H_1(f_t) = f_t$, is almost 75% and 82% respectively. This is very large if one wants to capture fat tails since the distribution of $H_1(f_t)$ is a $\mathcal{N}(0, 1)$. This is why log-normal volatility models fail to capture fat tails. Moreover, Table 1 tells us that the relative weight of the second polynomial, $H_2(f_t)$, is around 22% and 16% respectively. Table 1 means also that the first four polynomials explain almost all the variation of the volatility. Finally, observe that since the persistence of the Hermite polynomials is decreasing with i (see Table 1), for long horizon predictability of the variance, the relative weight of $H_1(f_t)$ is bigger than 75% and 82% respectively.

It turns out that the log-normal continuous time model (Wiggins, 1987; Melino and Turnbull, 1990) has the same drawback. To see this, consider the model

$$dy_t = \sigma_t \left[\sqrt{1-\rho^2} dW_t^{(1)} + \rho dW_t^{(2)}\right] \tag{2.11}$$

$$d\log(\sigma_t^2) = k[\theta - \log(\sigma_t^2)]dt + \sigma dW_t^{(2)}. \tag{2.12}$$

As for the discrete time case, let us define the state variable f_t by

$$f_t \equiv \frac{\sqrt{2k}}{\sigma}(\log \sigma_t^2 - \theta).$$

Then, we have $\log(\sigma_t^2) = \theta + \frac{\sigma}{\sqrt{2k}} f_t$,

$$df_t = -kf_t dt + \sqrt{2k}\; dW_t^{(2)} \tag{2.13}$$

with $f_t \sim \mathcal{N}(0,1)$. As a consequence, by using (2.8), we have

$$\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t), \quad \text{where} \quad a_i = \exp(\theta + \frac{\sigma^2}{4k}) \frac{(\sigma/\sqrt{2k})^i}{\sqrt{i!}}. \tag{2.14}$$

Let us now consider the empirical analysis of Andersen, Benzoni and Lund (2001). They use the Efficient Method of Moment (EMM, Gallant and Tauchen, 1996) to estimate various continuous time SV models on the S&P500 returns including the log-normal one with and without leverage effect. By using their estimates, we report in Table 2 the coefficients a_i in the decomposition (2.14) as well as the relative and cumulative weights of the Hermite polynomials for the models without and with leverage effect respectively. Again, the first eigenfunction $H_1(f_t)$ has a very important weight, especially in the model without leverage effect (almost 95%). Moreover, the first four eigenfunctions explain almost all the variance. Thus, the log-normal model does not capture fat tail. This is the main reason Andersen, Benzoni and Lund (2001) reject it.

It is well known that stock returns, and especially index returns, exhibit leverage effect (Black, 1976; Nelson, 1991). Hence, a model that captures this fact will better fit the data. Table 2 suggests that another advantage of incorporating leverage effect is that it reduces the link between the unconditional and conditional distribution of the volatility. In particular the mean (a_0) and the second moment ($\sum_{i=0}^{\infty} a_i^2$) of the variance are higher in the leverage model.

2.1.2 The square-root SV model

Another popular SV model in the continuous time literature is the Heston (1993) model where the variance process σ_t^2 is square-root, i.e.

$$dy_t = \sigma_t \left[\sqrt{1-\rho^2} dW_t^{(1)} + \rho dW_t^{(2)}\right] \qquad \text{where}$$

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta\sigma_t dW_t^{(2)}, \qquad k > 0.$$

Define the real α and the process f_t by

$$\alpha = \frac{2k\theta}{\eta^2} - 1, \quad f_t = \frac{2k}{\eta^2}\sigma_t^2. \tag{2.15}$$

Then, by Ito's Lemma, we have

$$df_t = k(\alpha + 1 - f_t) + \sqrt{2k}\sqrt{f_t}dW_t^{(2)}. \tag{2.16}$$

It turns out that the diffusion (2.16) admits as eigenfunctions the Laguerre polynomials $L_i^{(\alpha)}(f_t)$ associated to the eigenvalues $\delta_i = ki$. The Laguerre polynomials $L_i^{(\alpha)}$ are characterized by

$$\binom{i+\alpha}{i}^{1/2} i\, L_i^{(\alpha)}(x) = \binom{i-1+\alpha}{i-1}^{1/2} (-x+2i+\alpha-1)L_{i-1}^{(\alpha)}(x) - \binom{i-2+\alpha}{i-2}^{1/2} (i+\alpha-1)L_{i-2}^{(\alpha)}(x), \tag{2.17}$$

$$\text{where} \quad L_0^{(\alpha)}(x) = 1, \quad L_1^{(\alpha)}(x) = \frac{1+\alpha-x}{\sqrt{1+\alpha}}.$$

Thus, the variance process σ_t^2 is a linear combination of the constant and the first eigenfunction:

$$\sigma_t^2 = \theta - \frac{\theta}{\sqrt{\alpha+1}} L_1^{(\alpha)}(f_t), \quad \text{or}$$

$$\sigma_t^2 = a_0 L_0^{(\alpha)}(f_t) + a_1 L_1^{(\alpha)}(f_t) \quad \text{where} \quad a_0 = \theta \quad \text{and} \quad a_1 = -\frac{\sqrt{\theta}\eta}{\sqrt{2k}}. \tag{2.18}$$

In other words, in this model, the first eigenfunction explains 100% of the variance while its marginal distribution is $\gamma(\alpha+1, \theta/(\alpha+1))$[5] and, hence, with thin tails. This is the main reason why this model is rejected by Andersen, Benzoni and Lund (2001).

Observe that the affine model of Duffie, Pan and Singleton (2000) has also a variance process which is a linear function of the state variable. More precisely, assume that

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \sqrt{\eta_0 + \eta^2\sigma_t^2}dW_t^{(2)}, \qquad k > 0.$$

Assume that $\eta > 0$. Define $\tilde{\theta}$, $\tilde{\alpha}$ and $\tilde{f}_t$ by

$$\tilde{\theta} = \theta + \frac{\eta_0}{\eta^2}, \quad \tilde{\alpha} = \frac{2k\tilde{\theta}}{\eta^2} - 1 \quad \text{and} \quad \tilde{f}_t = \frac{2k}{\eta^2}(\frac{\eta_0}{\eta^2} + \sigma_t^2).$$

Then, by Ito's Lemma, it is easy to show that $\tilde{f}_t$ is characterized by

$$d\tilde{f}_t = k(\tilde{\alpha} + 1 - \tilde{f}_t) + \sqrt{2k}\sqrt{\tilde{f}_t}dW_t^{(2)}.$$

and that the variance process σ_t^2 is given by

$$\sigma_t^2 = \tilde{a}_0 L_0^{(\tilde{\alpha})}(\tilde{f}_t) + \tilde{a}_1 L_1^{(\tilde{\alpha})}(\tilde{f}_t) \quad \text{where} \quad \tilde{a}_0 = \theta \quad \text{and} \quad \tilde{a}_1 = -\frac{\tilde{\theta}}{\sqrt{\tilde{\alpha}+1}} = -\frac{\sqrt{\eta^2\theta + \eta_0}}{\sqrt{2k}}.$$

Note that since the variance process is not observable, there is an identification problem. In particular, one can identify only k, θ and $\tilde{a}_1$. Hence, one can identify $(k, \theta, \eta^2 + \frac{\eta_0}{\theta})$.[6]

[5] A positive random variable X is called $\gamma(a, b)$ if its density function is $p_X(x) = x^{a-1}\exp(-x/b)/(\Gamma(a)b^a)$ where $\Gamma(.)$ is the Gamma function defined by $\Gamma(a) = \int_0^{+\infty} \exp(-x)x^{a-1}dx$; see Johnson, Kotz and Balakrishnan (1994), page 337.

[6] Another potential statistical problem in this model is that the support of σ_t^2 depends on unknown parameters.

2.1.3 Our approach

The main idea of our approach is to relax the restrictions on the parameters a_i in (2.9) and (2.14) for the log-normal model in discrete and continuous time respectively and in (2.18) for the square-root model. Consider the log-normal case. Since the sample size is finite, one has to specify a parametric model. Two approaches may be considered. The first one is to parametrize the sequence a_i by a finite parameter θ and estimate θ.[7] The log-normal model is a special case. In order to have fat tails, one has to allow the relative weight of the first polynomial H_1 to be small. The second approach is to set the variance as a finite linear combination of Hermite polynomials and then estimate the coefficients a_i. In the empirical part, we will follow the second approach. However, our model allows the number of Hermite polynomials to be infinite. In the square-root case, in order to have fat tails, one has to put high order Laguerre polynomials in the variance decomposition (2.18).
To show the usefulness of our approach, consider the model defined by (2.1), (2.3), (2.5), where

$$\sigma_t^2 = a_0 + a_2 H_2(f_t).$$

Let us estimate this model with the Pound-US$ and Yen-US$ exchange rates data considered by Kim, Shephard and Chib (1998).[8] Assume that there is no leverage effect ($\rho = 0$) which is the case for exchange rates. Here, the coefficient a_0 and a_2 are free and $\theta = (a_0, a_2, \gamma)$ is the parameter of interest. We call this model the Hermite SV model of order 2, HSV(2). The reason why we consider this simple model is the following: We know that a GARCH(1,1) describes well the dynamics of the volatility. Hence, assuming that the squared-residual process is an ARMA(1,1) is a reasonable assumption. It turns out that this means that we need only one Hermite polynomial in the variance decomposition (see below). Moreover, for variance positivity, we need that the higher polynomial has an even degree. Thus we consider a HSV(2) where a_1 is zero.[9]

Many inference methods may be considered, in particular simulated methods (Duffie and Singleton 1993; Gallant and Tauchen, 1996; Gouriéroux, Monfort and Renault, 1993) or Bayesian methods (Jacquier, Polson and Rossi, 1994; Kim, Shephard and Chib, 1998). In this paper, we consider the Generalized Method of Moment (GMM) of Hansen (1982). It is well known that GMM estimation for SV models based on ad hoc marginal moments is less efficient than the simulated and Bayesian methods (see, e.g., Andersen and Sorensen, 1996; Andersen, Chung and Sorensen, 1999; Gallant and Tauchen, 1999; Jacquier, Polson and Rossi, 1994). However, we did GMM inference for its simple implementation. Moreover, in this paper, the comparison between alternative models is considered through the tail behavior, in particular the fourth moment. Therefore incorporating this moment in the estimation procedure makes the comparison easier.[10]

[7] Recentley, Giraitis, Kokoszka and Leipus (2000) introduced the ARCH(∞) model where the variance is a weighted infinite sum of the past squared-residual and study their probabilistic and statistical properties.

[8] Kim, Shephard and Chib (1998) use the same data as Harvey, Ruiz, Shephard (1994), i.e. daily observations of weekday close exchange rates from 1/10/81 to 28/6/85. The exchange rates are the U.K. Pound and Japanese Yen, both against the U.S. Dollar.

[9] This assumption was not rejected by the data.

[10] We give later a more important reason (see the long run subsection).

We consider the following moments: $E[\varepsilon_t^2]$, $E[\varepsilon_t^4]$, $Cov[\varepsilon_t^2, \varepsilon_{t-j}^2]$ for $j = 15$ and 20. Long lags are used in order to account for highly persistent volatility. We will see later that, under Gaussianity of u_t, their theoretical formulae are given by

$$E[\varepsilon_t^2] = a_0, \quad E[\varepsilon_t^4] = 3(a_0^2 + a_2^2), \quad Cov[\varepsilon_t^2, \varepsilon_{t-j}^2] = a_2^2 \gamma^{2j}. \tag{2.19}$$

We use the GMM optimal weighting matrix based on the Newey-West (1987) procedure with 10 lags.

In table 3, we report the estimates and their standard errors of a_0, a_2 and γ for the two series. As usual, the persistence parameter is high. In Table 4, we report the empirical values of the variance, fourth moment and kurtosis of the data, as well as their implied values by the HSV(2) model and the log-normal by using Kim Shephard and Chib (1998) estimates. Since these authors used a Bayesian Maximum Likelihood, the did not necessarily match the fourth moment. Therefore, for a fair comparison, we also estimate the log-normal model by using the same GMM procedure and same monents as for the HSV model.[11] The implied marginal moments are also reported in Table 4.

The main message of Table 4 is that the simple Hermite HSV(2) model with one eigenfunction captures well the kurtosis of the data. We are not saying that this is not the case for log-normal models. In fact, since the GMM procedure that we use is not very efficient, one can not reject that the implied kurtosis of the log-normal model is different from the data kurtosis. However, it is of interest to note that the the more efficient (Bayesian) likelihood estimators of Kim, Shephard and Chib (1998) imply a much smaller kurtosis than the GMM estimators. This implicitly suggests that the log-normal model is misspecified. However, by the GMM estimation, even if the model is misspecified, the implied kurtosis is closer to the empirical kurtosis since we incorporate the fourth moment in the GMM estimation. Finally, the persistence parameter γ is not very different between both models. This is not surprising because most of the volatility models imply the same volatility persistence. To clearly establish the usefulness of our approach for generating fat tails, we need to compare the higher order moments implied by each model.

Before doing this, it is important to notice that if one considers another eigenfunction than the second one, $H_2(f_t)$, for instance $H_4(f_t)$, then one gets similar result since in this case we have

$$E[\varepsilon_t^2] = a_0, \quad E[\varepsilon_t^4] = 3(a_0^2 + a_4^2), \quad Cov[\varepsilon_t^2, \varepsilon_{t-j}^2] = a_4^2 \gamma^{4j}.$$

In other words, one has to replace a_2 by a_4 and γ by γ^2. Notice that the implied marginal moments of ε_t and covariances of ε_t^2 are the same for the two models. This is an additional reason for using a GMM inference procedure. In Table 6, we report the empirical standardized six moment as well as the implied ones by the different models.[12] For the HSV model, we report them for four models. They correspond to the cases $\sigma_t^2 = a_0 + a_i H_i(f_t)$, $i = 2, 4, 6$ and 8. From Table 6, it is clear that Hermite SV models generate fat tails.[13] This is very promising for

[11] We report the estimates in Table 5.

[12] We report $E[y_t^6/(E[u_t^6]E[y_t^2]^3)]$ by taking $E[u_t^6] = 15$ (Gaussian case). Lemma 1 in the Appendix provides the theoretical formulae.

[13] Again, in contrast to the results based on the Bayesian method, the log-normal model based on GMM estimates implied a sixth moment which close to the empirical one.

this approach, in particular for stock returns that exhibit much more tails than exchange rate returns. We are working on this empirical application by using the results of Meddahi (2001-a).

2.2 The Hermite and Laguerre SV models in discrete time

We will now introduce the Hermite and Laguerre SV models in discrete and continuous times.

2.2.1 The model

Definition 2.1. Discrete time Hermite SV model: *A process $\{\varepsilon_t\}$ is called a Hermite SV model of order p, HSV(p), with an underlying state variable process $\{f_t\}$ if we have:*

$$\varepsilon_t = \sigma_{t-1} u_t, \quad \text{with} \quad \sigma_{t-1}^2 = \sum_{i=0}^{p} a_i H_i(f_{t-1}) \quad \text{where} \tag{2.20}$$

$$f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2}\; v_t, \quad f_0 \sim \mathcal{N}(0,1), \quad \mid \gamma \mid < 1, \tag{2.21}$$

$$(u_t, v_t) \ \text{ i.i.d. } \ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \ \text{and} \tag{2.22}$$

$$\sum_{i=0}^{p} a_i^2 < \infty, \tag{2.23}$$

where H_i are the Hermite polynomials characterized by (2.6).

In other words, we say that a process ε_t is a Hermite SV model of order p if its conditional variance is a linear combination of the $H_0(f_{t-1})$, $H_1(f_{t-1})$,...,$H_p(f_{t-1})$. We will say that this is a linear combination of the first p Hermite polynomials, $H_1(f_{t-1})$,...,$H_p(f_{t-1})$, by implicitly assuming that the constant, $H_0(f_{t-1}) = 1$, is always included. The condition (2.23) ensures that the variance process is well defined when $p = \infty$. In this case, σ_{t-1}^2 is the limit in mean-square of $\sum_{i=0}^{p} a_i H_i(f_{t-1})$ when $p \to +\infty$. The coefficient ρ is the leverage effect coefficient. As for the log-normal case, the process u_t may be not Gaussian. In particular it may have more tails than the Gaussian distribution, a Student for example, and may be skewed. In any case, we need the second moment to be finite. We will assume that its fourth moment is finite when we compute the returns' kurtosis and the covariance structure of the squared returns.

The main assumption in the previous definition is (2.21), that is the state variable that governs the volatility, f_t, is a Gaussian AR(1) process. This is why we specify the variance process as a linear combination of Hermite polynomials. Under (2.21), the Hermite polynomials are uncorrelated and each polynomial is an AR(1) process. We will see later in sections 3 and 4 that when the dynamics of f_t are not given by (2.21), then one has to consider alternative functions than the Hermite polynomials. Finally, we assume that f_t is standardized, i.e. $E[f_t] = 0$ and $Var[f_t] = 1$, for identification purposes.

Since the Hermite polynomials are uncorrelated, the HSV model may be viewed as a multifactor model. However, here, the Hermite polynomials are uncorrelated but not independent. In section 4, we show how one can consider the independent multifactor models.

We now introduce the Laguerre SV model.

Definition 2.2. Discrete time Laguerre SV model: *A process $\{\varepsilon_t\}$ is called a Laguerre SV model of order p, LSV(p), with an underlying state variable process $\{f_t\}$ if we have:*

$$\varepsilon_t = \sigma_{t-1} u_t, \quad \text{with} \quad \sigma_{t-1}^2 = \sum_{i=0}^{p} a_i L_i^{(\alpha)}(f_{t-1}) \quad \text{where} \tag{2.24}$$

$$\frac{2}{1-\gamma} f_t \mid J_{t-1} \sim \chi^2(2\alpha+2, \frac{2}{1-\gamma}\gamma f_{t-1}), \quad f_0 \sim \gamma(\alpha+1,1), \quad -1 < \alpha \text{ and } 0 \le \gamma < 1, \tag{2.25}$$

where $\sum_{i=0}^{p} a_i^2 < \infty$ and $L_i^{(\alpha)}$ are the Laguerre polynomials characterized by (2.17).

In other words, the main difference between the HSV and LSV models is that the state variable f_t is a Gaussian AR(1) process while it is (the exact discretization of the) square-root.[14] As a consequence, the variance decomposition involves Hermite polynomials in the HSV model while it involves Laguerre polynomials in the LSV model.

2.2.2 Positivity

One of the main reasons of the popularity of log-normal SV (and EGARCH) models is that the specification of the variance as an exponential ensures its positivity. This is not the case for HSV models. However, since a linear combination of Hermite polynomials is a polynomial, say $Q(.)$, it is very easy to ensure the variance positivity. Indeed a necessary and sufficient condition is that all the real roots of $Q(.)$ have an even order. When the order of the polynomial Q is two, i.e. in the HSV(2) model, this means that

$$a_1^2 - 4\frac{a_2}{\sqrt{2}}(a_0 - \frac{a_2}{\sqrt{2}}) \le 0 \text{ and } a_2 > 0.$$

Of course, a sufficient condition is that $Q(.)$ is the square of a polynomial. For instance, Robinson and Zaffaroni (1998) consider a discrete time SV model with $\varepsilon_t = g_{t-1}u_t$, where g_{t-1} is a Gaussian AR(1) process. In this case, the variance process is g_{t-1}^2 and, hence, positive. Indeed, Robinson and Zaffaroni's model is a (constrained) HSV(2) model. Stein and Stein (1991) and Ho, Perraudin and Sorensen (1996) follow the same approach in continuous time. Another necessary and sufficient condition is that $Q(.)$ is the sum of two squared polynomials, i.e., $Q(x) = A(x)^2 + B(x)^2$.[15] In our model, we assume that $\sum_{i=0}^{p} a_i H_i(.)$ is positive with probability one. This means that the parameters are constrained.

An alternative parameterization of the volatility is to assume that the log-variance is not an affine function of f_t but a quadratic one, i.e.

$$\sigma_t^2 = \exp(\lambda_0 + \lambda_1 f_t + \lambda_2 f_t^2) \quad \text{with} \quad \lambda_2 < \frac{1}{2} \tag{2.26}$$

where f_t is a Ornstein-Uhlenbeck process. The log-normal model is a special example with $\lambda_2 = 0$. The restriction on λ_2 ensures that the second moment of σ_t^2 is finite.[16] Thus, this model is also a HSV model. Indeed, it is easy to show that this is a HSV(∞) with

[14] The process f_t was independently introduced by Gouriéroux and Jasiak (2001).

[15] For a nonnegative polynomial $Q(x)$, one can always write it as the product of two conjugate polynomials, i.e. $(A(x) + iB(x))(A(x) - iB(x))$. Thus, one has $Q(x) = A(x)^2 + B(x)^2$.

[16] Alternative processes like square-root may be considered. In this case, the assumptions for the square-integrability of the variance are different.

$$a_0 = \exp(\lambda_0 + \frac{\lambda_1^2}{2(1-2\lambda_2)})\frac{1}{\sqrt{1-2\lambda_2}};\ a_1 = \frac{\lambda_1}{1-2\lambda_2}a_0 \text{ and}$$

$$a_i = \frac{1}{\sqrt{i}(1-2\lambda_2)}[\lambda_1 a_{i-1} + (\sqrt{i-2} - (1-2\lambda_2)\sqrt{i-1})a_{i-1}] \text{ for } \ i \geq 2. \tag{2.27}$$

Observe that in this case, it is easy to compute $E[\sigma_t]$[17] and also the coefficient in the decomposition of σ_t in terms of Hermite polynomials. These coefficients have the same form as (2.27) by replacing $\lambda_0, \lambda_1, \lambda_2$ by $\lambda_0/2, \lambda_1/2, \lambda_2/2$ respectively. One can also incorporate higher order monomials in the exponential function. However, since we need (square) integrability of the variance process, all the even monomials higher than 3 are precluded while odd monomials may be considered if one assumes that the corresponding coefficients are nonpositive.

2.2.3 The moments

We now compute the moments of the HSV and LSV models. We compute the first four moments, the covariance structures of the squared residual and conditional moment restrictions. Observe that all the results in this section are special examples of those in section 4. Therefore, for the proof of the results, one has to read the proofs of the corresponding results in the fourth section.

We start by computing the marginal moments. We recall that, for a given random variable z, with finite fourth moment, the later are defined by

$$Var[z] = E[(z - E[z])^2];\ Skew[z] = \frac{E[(z-E[z])^3]}{(Var[z])^{3/2}};\ Kurt[z] = \frac{E[(z-E[z])^4]}{(Var[z])^2}.$$

Then, if one considers a HSV(p) or an LSV(p) model where the fourth moment of u_t is finite, one gets

$$E[\varepsilon_t] = 0;\ Var[\varepsilon_t] = a_0;\ Kurt[\varepsilon_t] = Kurt[u_t]\left(1 + \frac{\sum_{i=1}^{p} a_i^2}{a_0^2}\right). \tag{2.28}$$

As usual in volatility models, time-varying volatility, i.e. $a_i \neq 0$ for some $i \geq 1$, implies that the returns have fatter tails than the standardized residual u_t. Observe that we did not give the form of the skewness. As in the GARCH case, the main reason is that it is not always easy to compute $E[\sigma_t^{3/2}]$. However, if one specifies the volatility model is terms the standard deviation, that is assume that the standard deviation is a linear combination of the eigenfunction, then we can compute $E[\sigma_t^{3/2}]$ of this ESV model. In any case, if the skewness of u_t is zero, then this is also the case for ε_t.

Time-varying volatility means that the squared residuals are correlated. Indeed, if we assume that there is no leverage effect, i.e. u_t and f_t are independent, then we have in both HSV and LSV models:

$$Cov[\varepsilon_t^2, \varepsilon_{t-j}^2] = \sum_{i=1}^{p} a_i \gamma^{ij}. \tag{2.29}$$

We have:

$$E[\varepsilon_t \mid \varepsilon_\tau, \tau \leq t-1] = 0. \tag{2.30}$$

[17]Another model where this is the case is obtained by specifying the conditional standard deviation as a linear combination of the eigenfunctions. Conditional standard deviation models are considered by Taylor (1986) and Schwert (1989). An advantage of these models is their robustness against outliers (Davidian and Caroll, 1987).

Let L be the Lag operator and assume that p is finite. Then we have:

$$E[\prod_{i=1}^{p}(1-\gamma^{i}L)[\varepsilon_t^2 - a_0] \mid \varepsilon_\tau, \tau \leq t-p-1] = 0. \tag{2.31}$$

The restriction (2.30) is the usual martingale difference sequence assumption of the innovation ε_t. The equation (2.31) is a multiperiod conditional moment restriction. Such restrictions introduced by Hansen (1985) are useful for inference purposes. It is important to notice that they are based on observable. Hence, one can estimate the parameter of interest by the usual GMM rather than, e.g., simulated methods. This approach is introduced in Meddahi and Renault (1996, 2000). Indeed, following Andersen (1994), these authors called the models that follow (2.31) Square-Root Stochastic Autoregressive Variance models of order p, SR-SARV(p). Thus HSV(p) are included in the class of SR-SARV(p) which is closed under temporal aggregation (see below). Another advantage of (2.31) is that one can estimate the coefficient of interest, in particular the persistence parameter γ, more efficiently than by (2.29). This is important because the squared residual process ε_t^2 is heteroskedastic. Thus, one can choose appropriate instruments in the GMM estimation based on (2.31) to estimate more efficiently the parameters. For efficient estimation in multiperiod conditional moment restrictions, see Hansen (1985), Hansen, Heaton and Ogaki (1988) and Hansen and Singleton (1996). Finally, observe that the restriction (2.31) is true even if the fourth moments of u_t and ε_t are not finite.

When the fourth moment of ε_t is finite, then (2.31) implies that the process w_t defined by

$$w_t \equiv \prod_{i=1}^{p}(1-\gamma^{i}L)[\varepsilon_t^2 - a_0] \tag{2.32}$$

is a weak moving average of order p, MA(p), and, hence, ε_t^2 is a weak ARMA(p,p) model with autoregressive coefficient γ^1, γ^2,...,γ^p. As a consequence, ε_t is a weak GARCH(p,p) of Drost and Nijman (1993); for a detailed comparison between Bollerslev's GARCH, weak GARCH and SR-SARV models, see Meddahi and Renault (2000). Observe that for efficiency purposes, one can compute the moving average parameters which depend on $\gamma, a_1, ..., a_p$, i.e. the ARMA model is constrained.

In some cases, one knows that a Hermite polynomial H_i is not involved in the variance decomposition. In this case, the term $1-\gamma^{i}L$ has to be removed from (2.31) and the conditional information is $\sigma(\varepsilon_\tau, \tau \leq t-p)$. Moreover, ε_t^2 is an ARMA(p-1,p-1). For instance, if one considers a HSV(2) with $a_1 = 0$ as we did in the previous empirical analysis, then we have the following restriction

$$E[(1-\gamma^2 L)[\varepsilon_t^2 - a_0] \mid \varepsilon_\tau, \tau \leq t-2] = 0.$$

2.2.4 The long run

The equation (2.28) implies that the first four marginal moments[18] depend only on the reals a_i and not on the dynamics of the eigenfunctions. This property is interesting for two reasons. First, this reduces the link between the marginal and conditional densities of the returns. This

[18]It is clear that this is also the case for higher moments.

is in contrast to GARCH models where the high persistence implies in many empirical cases that the fourth moment is not finite (if one assumes that the standardized residuals are i.i.d.). Note that Barndorff-Nielsen and Shephard's (2001) SV model has also this property and that this is in line with the diffusion modeling approach of Chen, Hansen and Scheinkman (2000). The second reason is that economists are more concerned with the long run, i.e. the marginal distribution. Thus, even if we specify a wrong volatility model, i.e. we consider a wrong state variable f_t and, hence, wrong eigenfunctions, we still have the correct marginal moments and distribution. This is the case for instance when the observations are a subordinated process (Conley et al., 1997; Conley, Hansen and Liu, 1997).

Another robust property of our model is due to (2.29). This proposition implies that the covariance structure of the squared residuals is the same for both Hermite and Laguerre models (with the same coefficients a_i, $i = 0, 1, ..., p$, and γ).[19] In other words, these two models imply the same first four marginal moments and the same covariance structure of the squared residuals. Observe that most of the empirical volatility works are concerned only with characteristics, i.e. try to match these moments. This is an additional reason of using the GMM. Of course, one may discriminate between these models if one considers marginal moments of order higher than 4 or other covariance function (e.g. $Cov(\varepsilon_t^4, \varepsilon_{t-h}^2)$). This means that in applied work, one can start by matching these moments. This will indicate how many eigenfunctions one has to include. Then, one has to consider higher order moments or covariances to find the best model between Hermite and Laguerre ones.

2.2.5 Persistence

In this section, we consider the persistence of the variance process and the squared residual. Of course, there are many definitions of persistence, in particular in volatility models (e.g., Nelson, 1990-a; Bollerslev and Engle, 1993; Gallant, Rossi and Tauchen, 1993). In this paper, we consider the definition of Fiorentini and Sentana (1998). The main reason is that this definition combined with the eigenfunctions allows us to make an interesting decomposition of the variance process in terms of the eigenfunctions ones.

Fiorentini and Sentana (1998) consider the following definition of persistence. Let a process x_t with a conditional mean m_{t-1}, then the persistence of x_t is given by

$$P(x_t) \equiv \frac{Var(x_t)}{Var(x_t) - Var(m_{t-1})}. \tag{2.33}$$

Thus, if x_t is an AR(1) process which autoregressive coefficient is ρ, then its persistence is $1/(1-\rho^2)$. Thus, if $\rho = 0$ (white noise), then the persistence is one while it goes to infinity when ρ goes to one. Finally, the persistence of an ARMA(1,1) is $1 + (\rho - \beta)^2/(1-\rho^2)$ where ρ and β are respectively the autoregressive and moving average parameters.

By using this definition, we can compute the persistence of the variance process and squared

[19] This is also the case for the Jacobi model that we consider in the fourth section.

residual process of a HSV or LSV model. Indeed we have

$$P(\sigma_t^2) = \frac{\sum_{i=1}^p a_i^2}{\sum_{i=1}^p (1-\gamma^{i2}) a_i^2} = \sum_{i=1}^p \omega_i P(E_i(f_t)) \quad \text{where} \quad \omega_i = \frac{a_i^2}{\sum_{i=1}^p a_i^2}. \tag{2.34}$$

Thus, the persistence of the variance process is a weighted average of the polynomials persistence with the same weights than the variance decomposition.

Now, assume that $E[u_t^4] < \infty$, then the persistence of the squared residual process, ε_t^2, is

$$P(\varepsilon_t^2) = \frac{E[u_t^4] - (a_0^2 / \sum_{i=0}^p a_i^2)}{E[u_t^4] - 1} = \frac{E[u_t^4]}{E[u_t^4] - 1}\left(1 - \frac{E[\varepsilon_t^2]^2}{E[\varepsilon_t^4]}\right). \tag{2.35}$$

We already show that the variance process is ARMA(p,p-1) process while the squared residual process is an ARMA(p,p) process. On the other hand, the GARCH(1,1) model which is the benchmark volatility model and which successfully describes the dynamics of the volatility, implies that the variance process is an AR(1) while the squared residual process is an ARMA(1,1). Thus, it is of interest to derive the implied autoregressive coefficient of an AR(1), $\bar{\gamma}$, which persistence is the same then the conditional variance process. We show that:

$$\bar{\gamma}^2 = \sum_{i=1}^p \omega_i \gamma^{i2}, \tag{2.36}$$

i.e., the variance process has the same persistence than an AR(1) which squared autoregressive coefficient is a weighted average of the squared autoregressive coefficients of the eigenfunctions.

Consider now the persistence of the squared residual process. ε_t^2 is equal to the conditional variance plus a noise. Therefore, if one made the assumption that the conditional variance is an AR(1), the squared residual process is an ARMA(1,1) which autoregressive coefficient is the same that the AR(1) process. Thus it is interesting to compute the moving average coefficient implied by the empirical persistence of the squared residual in the HSV or LSV model. This is easy to derive since the Fiorentini and Sentana's persistence for an ARMA(1,1) is $1 + (\bar{\gamma} - \beta)^2/(1 - \bar{\gamma}^2)$.

As an example, consider the log-normal model. It is easy to show that in this case, (2.34) and (2.35) imply that

$$P(\sigma_t^2) = \frac{1 - \exp[-\sigma^2/(1-\gamma^2)]}{1 - \exp[-\sigma^2]}, \qquad P(\varepsilon_t^2) = \frac{3 - \exp[-\sigma^2/(1-\gamma^2)]}{2}.$$

By using the empirical results of Kim et al. (1998), we get that the persistence of the variance and the squared residuals are $P(\sigma_t^2) = 17.560$ $P(\varepsilon_t^2) = 1.218$. Therefore the autoregressive coefficient $\bar{\gamma}$ and the moving average parameter β are $\bar{\gamma} = 0.972$ and $\beta = .860$. Interestingly, the values $\beta = .860$ and $\alpha \equiv \gamma - \beta = .112$ are very close from ones obtained for a GARCH(1,1) estimated by a Gaussian QMLE. This confirms our claim made in the introduction that all the volatility models describe well the volatility (and squared residuals) dynamics. Indeed, by using Fiorentini and Sentana's persistence, we obtain that these dynamics are very close.

2.2.6 Temporal aggregation

As we already mentioned, HSV and LSV models are SR-SARV. Meddahi and Renault (1996, 2000) show that this class of models is closed under temporal aggregation, cross-sectional aggregation and information reduction. Since HSV(p) are SR-SARV(p), this implies that if ε_t is a HSV(p) or LSV(p) at daily frequency for example, then the aggregated process $\sum_{i=1}^{5} \varepsilon_{t-i}$ is a SR-SARV(p) but not a HSV(p) or LSV(p) model. It is of interest to characterize the assumptions of the HSV and LSV models that are closed under temporal aggregation. Meddahi and Renault (2000) show that to have a model which is closed under temporal aggregation, one has to make two assumptions: i) the variance process is a marginalization of a VAR(1); ii) one has to make semiparametric assumptions on the standardized residual u_t, i.e. the conditional mean and variance are 0 and 1 respectively. The first assumption holds for the HSV and LSV models but not the second one since we assume that u_t is i.i.d.; hence, HSV and LSV models are not closed under temporal aggregation. However, if we relax the i.i.d. assumption, then we will have the temporal aggregation result. Meddahi (2001-b) calls these models the semiparametric HSV or LSV models. He shows that semiparametric HSV (and LSV) models are closed under temporal aggregation (see Section 4).

Diebold (1988) suggested that the aggregated process converges to a Guassian process. Therefore, we also characterize the asymptotic behavior of the kurtosis of the aggregated process. Indeed, let ε_t be a HSV or LSV model without leverage effect and consider the process $\{\varepsilon_{\tau m}^{(m)}\}$ defined by $\varepsilon_{\tau m}^{(m)} \equiv \sum_{i=0}^{m-1} \varepsilon_{tm-i}$ where m is an integer. Then, we show in Section 4 that

$$Kurt(\varepsilon_{\tau m}^{(m)}) - 3 = \frac{1}{m}(kurt(\varepsilon_t) - 3) + \frac{6}{a_0^2} \sum_{i=1}^{p} a_i^2 \frac{\gamma^i}{(1-\gamma^i)^2} \frac{\gamma^{im} - m\gamma^i + m - 1}{m^2}. \tag{2.37}$$

Moreover, when m $\longrightarrow +\infty$, we have

$$Kurt(\varepsilon_{\tau m}^{(m)}) - 3 = \frac{1}{m}\left((kurt(\varepsilon_t) - 3) + \frac{6}{a_0^2} \sum_{i=1}^{p} a_i^2 \frac{\gamma^i}{(1-\gamma^i)^2}\right) + \mathrm{o}(\frac{1}{m}). \tag{2.38}$$

The equation (2.37) means that the excess kurtosis of the aggregated process depends on two terms. The first one is the excess kurtosis of the non aggregated process while the second one is due to the time-varying volatility. This means that the volatility reduces the speed of convergence of the aggregated process to Gaussianity. Since the volatility persistence decreases when one aggregates the process, one may also expect the same result for the kurtosis, i.e. (2.37) is a nonincreasing function of m. It turns out that this is not the case. It is easy to show that for some parameters $\{a_i, i \in \mathbf{N}\}$ and γ, (2.37) starts by increasing and then decreases. In fact, Nelson (1996) already shows in a different context that for some examples, reducing the information may increase the kurtosis. It turns out that temporal aggregation reduces the information set. Consider now the kurtosis asymptotic behavior of the aggregated process. (2.38) says that the convergence arises at the speed $1/m$. Moreover, the term governing the leading coefficient is also a combination of two terms due to the excess kurtosis and the time-varying volatility effect of the high frequency process. Note that the second effect is due to both a_i and λ_i. In particular, if some eigenfunction of high order (thus γ^i is small) has an important weight (a_i is big), then this will reduce the speed of convergence to Gaussianity.

2.3 The Hermite and Laguerre SV models in continuous time

2.3.1 The model and exact discretization

Definition 2.3. Continuous time Hermite and Laguerre SV models: *A continuous time process $\{y_t\}$ is called a Hermite (resp Laguerre) SV model of order p, HSV(p) (resp LSV(p)), with an underlying state variable process $\{f_t\}$ if:*

$$dy_t = \sigma_t \left[\sqrt{1-\rho^2}dW_t^{(1)} + \rho dW_t^{(2)}\right] \quad \text{with} \quad \sigma_t^2 = \sum_{i=0}^{p} a_i E_i(f_t) \quad \text{where} \tag{2.39}$$

f_t is the Ornstein-Uhlenbeck (resp square-root) process defined by (2.13) (resp (2.16)), E_i the Hermite (resp Laguerre) polynomials defined by (2.6) (resp (2.17)), and $\sum_{i=0}^{p} a_i^2 < \infty$.

In other words, HSV and LSV continuous time models are defined as in discrete time, that is the variance process is a linear combination of the Hermite and Laguerre polynomials respectively.[20] One needs to impose the same positivity assumption as in the discrete time case. Unconditional and conditional moments of these models are derived in Meddahi (2001-a). Since the semiparametric discrete time model are closed under temporal aggregation, this suggests that exact discretization of continuous time HSV(p) (resp LSV(p)) is a semiparametric HSV(p) (resp LSV(p)). It is established in the following proposition by Meddahi (2001-b):

Proposition 2.1 Exact discretization. *Consider $\{y_t\}$ a continuous time HSV(p) (resp LSV(p)) model where $\{f_t\}$ is the underlying state variable. Then, for any sampling interval h, the associated discrete time process $\varepsilon_{th}^{(h)} \equiv y_{th} - y_{(t-1)h}$ is a semiparametric HSV(p) (resp LSV(p)) model w.r.t. $J_{th}^{(h)} = \sigma(\varepsilon_{\tau h}^{(h)}, f_{\tau h}, \tau \leq t)$ with the same underlying state variable process, i.e. $f_{\tau h}^{(h)} = f_{\tau h}$. More precisely, we have:*

$$\sigma_{(t-1)}^{(h)2} \equiv Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}] = a_0 h E_0(f_{(t-1)h}^{(h)}) + \sum_{i=1}^{p} \frac{a_i(1-\exp(-kih))}{ki} E_i(f_{(t-1)h}^{(h)}). \tag{2.40}$$

2.3.2 Variance of the variance

In their filtering work, Nelson and Foster (1994) highlighted the importance of the variance of the variance. In particular, they show that if one considers a filter of the volatility process, which is always what we do in practice since we do not know the true data generating process, then the variance of the filter variance process has to be the same as the variance of the true variance to achieve optimality. Moreover, if this is not the case, the filter may be very worse. It turns out that for the popular SV models, namely lognormal, GARCH and square-root, the variance of the variance is fixed, quadratic in the first two cases and linear in the last one.[21] Therefore, Nelson and Foster (1994) criticize these models and recommend the development of volatility models where the variance of the variance is flexible such that the data will tell us what is the variance of the variance.

[20] Observe that there is not a problem of the existence of a solution of y_t since: i) the existence of f_t and, hence, σ_t hold; ii) $y_t = y_0 + \int_0^t \sigma_u \left[\sqrt{1-\rho^2}dW_u^{(1)} + \rho dW_u^{(2)}\right]$.

[21] Observe that when the variance of the variance is quadratic, the log-variance is homoskedastic.

The CEV model considered by Jones (2000) is flexible since this process is given by[22]

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \sigma\sigma_t^{\psi}dW_t, \quad \frac{1}{2} < \psi.$$

When $\psi = 1/2$ (resp $\psi = 1$), the CEV process coincides with the square-root (resp GARCH) one. However, it is easy to show that for $\psi < 1$, the tails of the unconditional distribution of the σ_t^2 are a decreasing function of ψ. Therefore, one will prefer empirically $\psi = .5$ or $\psi > 1$.[23] In Gallant, Hsu and Tauchen (1999) and Chernov et al. (2001), the variance process is given by

$$\sigma_t^2 = \exp(a + f_t) \quad \text{where} \quad df_t = k(\theta - f_t)dt + \sigma f_t dW_t.$$

Therefore, by using Ito's Lemma, one gets that the variance of the variance is given by $\sigma^2(\sigma_t^2)^2[\log(\sigma_t^2)]^2$. Thus, this is also a fixed function.[24] Therefore, it is of interest to find volatility models where the variance of the variance is a power function, $\sigma_t^{2\psi}$, with $\psi < 1$.

Let us now consider the Hermite and the Laguerre SV models. Observe that while our results here are for the continuous time models, the same ones hold in discrete time. For the simplicity of the exposition, consider the stochastic differential equation of the monomials f_t^i with $i > 1$. Then, by using again Ito's Lemma, one gets that the variance of f_t^i is given respectively by

$$\text{Var of } (\mathrm{f_t^i}) = 2ki^2(f_t^i)^{2\lambda_i} \quad \text{where} \quad \lambda_i = 1 - \frac{1}{i}, \text{ and}$$

$$\text{Var of } (\mathrm{f_t^i})_t = 2ki^2(f_t^i)^{2\lambda_i} \quad \text{where} \quad \lambda_i = 1 - \frac{1}{2i}.$$

Observe that in both cases, λ_i is an increasing function and that the limit of Var of $(\mathrm{f_t^i})$ when $i \to +\infty$ is 2. Thus, since the Hermite and Laguerre polynomials of order n are linear combination of the monomials f_t^i, $i \leq n$, the variance of the variance of these polynomials is a smooth function of the variance. Moreover, by increasing n, one increases the variance of the variance. At the limit, i.e. when $n \to +\infty$, the variance of the variance is quadratic as in GARCH and log-Normal SV cases. In conclusion, the variance of the variance of a Hermite or Laguerre SV model is (almost) a power function, $\sigma_t^{2\psi}$, with ψ smaller than one and this power increases with p and goes to one when $p \to +\infty$.

2.3.3 Option pricing

An important advantage of affine models (Heston, 1993; Duffie, Pan and Singleton, 2000) is that closed form option pricing formulae are available. This is not the case for the other models.[25] In this paper, we do not address this issue. However, we think that this is possible in the eigenfunctions framework. The main reason is that prices of financial derivatives are usually characterized by differential equation (see, e.g., Duffie, 1992) while the eigenfunctions were developed for solving non-linear differential equations. In general, the solution are characterized

[22]Meddahi and Renault (1996, 2000) consider the CEV model with $\psi \leq 1$.

[23]By using a Bayesian method, Jones (2000) estimates that ψ is around 1.5. Note that in CEV processes with $\psi > 1$, $E[(\sigma_t^2)^r]$ is not finite for $r > 2\psi - 1$. Thus, this is also the case for $E[\epsilon_t^{2r}]$ where ε_t is, say, a daily return.

[24]Observe that while the variance of the variance in CEV process with $\psi > 1$ is greather than the one in Gallant, Hsu and Tauchen (1999), the tails of the unconditional distribution of the latter are fatter than the former.

[25]However, this is the case for the Levy SV model of Barndorff-Nielsen and Shephard (2001).

as a linear combination, in general infinite, of the eigenfunctions. Indeed, Davydov and Linetsky (2000) consider this approach for pricing options on scalar diffusions. Moreover, an application of the eigenfunctions considered in Wong (1964) was computing the distributions (or moments) of functional forms of Markov process. In particular, Wong (1964) considers functionals of the form

$$Y(t) = \int_0^t f[X(\tau)]d\tau$$

where $X(\tau)$ is a scalar stationary Markov process, with special attention to the square-root process.[26]

Another potential interesting property of the eigenfunctions for option pricing is the following. We consider the affine case since this is the only example were there are closed form option pricing formulae. Let the variance process under the Objective and Risk neutral probabilities given respectively by

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta\sigma_t dW_t^{(2)} \quad \text{and}$$

$$d\sigma_t^2 = k^*(\theta^* - \sigma_t^2)dt + \eta\sigma_t dW_t^{(2)*}.$$

It turns out that the risk premia form is specified in Heston (1993) such that

$$k^* = k + \lambda \quad \text{and} \quad \theta^* = \frac{k\theta}{k+\lambda}$$

where λ is the main parameter in the risk premia specification. In this case, the α and α^* parameters that characterize the Laguerre polynomials (2.15) coincide since

$$\alpha^* = \frac{2k^*\theta^*}{\eta^2} - 1 = \frac{2k\theta}{\eta^2} - 1 = \alpha.$$

As a consequence, the eigenfunctions are the same in both Objective and Risk neutral probabilities. Moreover, the marginal distributions of the state variables f_t and f_t^* defined in (2.15) are the same.[27] Therefore, making similar assumptions about the eigenfunctions and the state variable f_t in our Eigenfunction SV model may be useful for option pricing purposes.

3 Eigenfunctions of the conditional expectation and infinitesimal generator operators

In this section, we recall some definitions and properties of the eigenfunctions of a conditional expectation operator in the discrete time case as well as the eigenfunctions of the infinitesimal generator of diffusions in the continuous time case. We give examples in both cases.

[26] Finally, let us observe that Lars Hansen presented in his lectures at CIRANO conference on "Financial Mathematics and Econometrics" in June 2001 a work in progress with José Scheinkman, where they define the pricing operator and study its spectral properties, that is the eigenfunctions of this operator.

[27] The marginal distribution of σ_t^2 is $\gamma(\alpha+1, \theta/(\alpha+1))$ under the Objective probability and $\gamma(\alpha+1, \theta^*/(\alpha+1))$ under the Risk neutral one. Thus, the skewness and kurtosis of σ_t^2 is the same in both probabilities. This is probably a restrictive assumption. Note however that this is not the case for the aggregated returns because the aggregation incorporates the mean reverting parameter (k or k^*); see Meddahi (2001-a).

3.1 The general theory

Let $\{f_t\}$ be a stationary Markovian discrete time process. Consider the following problem: is there a function $\phi(f_t)$, square-integrable, i.e. $E[\phi(f_t)^2] < \infty$, such that

$$E[\phi(f_{t+1}) \mid f_t] = \lambda \phi(f_t) \tag{3.1}$$

where λ is a real number? The solutions ϕ of (3.1) are called the eigenfunctions of the conditional expectation operator associated to the process $\{f_t\}$ and the reals λ are called the eigenvalues. The set of eigenvalues is called the spectrum of the operator.[28]

Indeed, as in the linear algebra case, the existence of real eigenfunctions and real eigenvalues solutions of (3.1) is not always ensured. However, complex ones are. In the following we will assume that the solutions are reals. It is well known in the linear algebra case, that this is ensured when the considered matrix is symmetric. The counterpart assumption in the conditional expectation operator case is that the state variable $\{f_t\}$ is time-reversible, i.e. the conditional distributions of f_t given f_{t-1} and of f_t given f_{t+1} are the same.[29] Therefore we make the following assumption which ensures that the eigenfunctions are real functions and the eigenvalues are real numbers:

Assumption A1. The stationary Markovian process $\{f_t\}$ is time-reversible.

Now, we will consider some interesting properties of the eigenfunctions. It is clear that if a function ϕ is an eigenfunction associated to an eigenvalue λ, then any non-zero function proportional to ϕ is also an eigenfunction associated to the same eigenvalue λ. Therefore we make a normalization assumption by considering the eigenfunction with $E[\phi(f_t)^2] = 1$. Observe that an obvious solution of (3.1) is any non-zero constant function associated to the eigenvalue one. This eigenvalue will be denoted λ_0 and the corresponding eigenfunction $E_0(f_t) \equiv 1$.

Since the eigenvalues are reals and that any eigenfunction is a square-integrable function of the stationary process $\{f_t\}$, (3.1) means that any eigenfunction $\phi(f_t)$ (not associated to λ_0) is an AR(1) process. However, in general, this process is an heteroskedastic one. The autoregression dynamics and the stationarity assumption mean that any eigenvalue $\lambda \neq \lambda_0$ is smaller than one in absolute value. We will now make an assumption on the set of the eigenvalues (spectrum):

Assumption A2. The spectrum of the conditional expectation operator is discrete and denoted $\{\lambda_i, i \in \mathbf{N}\}$ with $1 = \lambda_0 >\mid \lambda_1 \mid>\mid \lambda_2 \mid> ... >\mid \lambda_i \mid>\mid \lambda_{i+1} \mid ...$; the corresponding eigenfunctions are denoted $E_i(f_t)$, $i \in \mathbf{N}$.

A sufficient condition to ensure **A2** is that the conditional expectation operator is compact.

Consider two different eigenvalues λ_i and λ_j with corresponding eigenfunctions $E_i(f_t)$ and $E_j(f_t)$. Then, these two functions are orthogonal, i.e.

$$E[E_i(f_t)E_j(f_t)] = 0. \tag{3.2}$$

[28] Observe that this terminology is close to the one used in the linear algebra case, when one considers the eigenvector-eigenvalue problem.

[29] Recall that f_t is assumed to be Markovian. Therefore, the conditional distribution of f_t given f_{t-1} (resp f_{t+1}) is also the conditional distribution of f_t given $\{f_\tau, \tau \leq t-1\}$ (resp $\{f_\tau, \tau \geq t+1\}$).

As a consequence, any eigenfunction associated to an eigenvalue different from λ_0 is centered:

$$E[E_i(f_t)] = 0. \tag{3.3}$$

Consider now a square-integrable function g, i.e. $E[g(f_t)^2] < \infty$. Then we have the important following result:

$$g(f_t) = \sum_{i=0}^{\infty} a_i E_i(f_t) \ \text{ in mean} - \text{square} \ \text{ where } \ a_i = E[g(f_t)E_i(f_t)]. \tag{3.4}$$

In other words, any square-integrable function may be written as a linear combination of the eigenfunctions. Observe that

$$\sum_{i=0}^{\infty} a_i^2 = E[g(f_t)^2] < \infty. \tag{3.5}$$

Therefore, $g(f_t)$ is the limit in mean-square of $\sum_{i=0}^{p} a_i E_i(f_t)$ when $p \to +\infty$.

Let us now consider the continuous time case. Let $\{f_t\}$ be the stationary scalar diffusion characterized by

$$df_t = \mu(f_t) + \sigma(f_t)dW_t, \tag{3.6}$$

where W_t is a standard Brownian process. Let $\mathcal{A}$ be the infinitesimal generator operator, i.e.,

$$\mathcal{A}\phi(f_t) \equiv \mu(f_t)\phi'(f_t) + \frac{\sigma^2(f_t)}{2}\phi''(f_t) \tag{3.7}$$

where $\phi(f_t)$ is square-integrable function and twice differentiable. A function ϕ is called an eigenfunction of the infinitesimal generator $\mathcal{A}$ with a corresponding eigenvalue $-\delta$ if

$$\mathcal{A}\phi(f_t) = -\delta\phi(f_t). \tag{3.8}$$

It turns out that the properties of the eigenfunctions and spectrum of the infinitesimal generator operator are similar to those of a conditional expectation operator. In particular:

i) When $\{f_t\}$ is time-reversible, the eigenfunctions and eigenvalues are reals. Hansen, Scheinkman and Touzi (1998) show that under appropriate boundary protocol, stationary scalar diffusions are time-reversible.[30] So we make the time-reversibility assumption:

Assumption A1'. The process stationary process $\{f_t\}$ is time reversible.

ii) The non-zero constant functions are eigenfunctions associated to the eigenvalue zero (denoted δ_0).

iii) The spectrum of the $\mathcal{A}$ is not necessarily discrete. However, when $\mathcal{A}$ is compact, it is discrete. Therefore, we assume in the sequel that the spectrum is discrete:

Assumption A2'. The spectrum of the infinitesimal generator operator, $\mathcal{A}$, of $\{f_t\}$ is discrete and denoted $\{-\delta_i, i \in \mathbf{N}\}$ with $\delta_0 = 0$ and $\delta_0 < \delta_1 < \delta_2 < ... < \delta_i < \delta_{i+1}...$; the corresponding eigenfunctions are denoted $E_i(f_t)$, $i \in \mathbf{N}$.

iv) Two eigenfunctions associated to two different eigenvalues are orthogonal, i.e. (3.2) holds. As a consequence, any non-constant eigenfunction is centered, i.e. (3.3). Finally, any square integrable function $g(f_t)$ can be decomposed in terms of eigenfunctions, i.e. (3.4) and (3.5).

[30] See Florens, Renault and Touzi (1998) for a discussion of the observable implications of reversibility.

Finally, there is the following important result:

$$\forall h > 0, E[E_i(f_{t+h}) \mid f_t] = \exp(-\delta_i h) E_i(f_t). \tag{3.9}$$

As a consequence, each eigenfunction E_i of $\mathcal{A}$ is also an eigenfunction of the conditional expectation operator associated to a sample of the continuous time process $\{f_t, t \in I\!R^+\}$ observed at discrete time, say $f_{th}, t \in \mathbf{N}$. The corresponding eigenvalue $\lambda_i(h)$ is[31]

$$\lambda_i(h) = \exp(-\delta_i h). \tag{3.10}$$

Before giving examples, let us make three remarks. First of all, the spectrum of an operator may be continuous or mixed (i.e. discrete and continuous). The GARCH diffusion of Nelson (1990-b) is a particular case that we will consider. All the results developed before are still valid. The only difference is that in the expansion result (3.4), one has to consider an integral instead of a sum. Second, while we consider a univariate process f_t, the same theory holds for the multivariate case. The main difficulty is the time reversibility assumption. Time reversibility is a restrictive assumption in multivariate case. A sufficient condition is that the multivariate Markovian process is a function of independent univariate Markovian processes. This is the case of most models considered in the literature; see Hansen and Scheinkman (1995), Florens et al. (1998) and Chen et al. (2000). Finally, we implicitly assume that for a given eigenvalue, all the related eigenfunctions are proportional. If this assumption does not hold, we say that the eigenvalue has a multiple order. We exclude this in our setting since the assumption holds in all the considered examples.

3.2 Examples

Now we consider some examples in the discrete and continuous times. We already considered two eigenfunctions examples, that is Hermite and Laguerre polynomials in both discrete and continuous time. There are many other examples where the eigenfunctions are orthonormal polynomials.[32] The first reference that derives some of them in continuous time is probably Wong (1964). An excellent reference of the use of orthonormal polynomials for stochastic processes purposes is Schoutens (2000). We will follow these authors in the construction of eigenfunctions. Therefore, we first start in continuous time since the eigenfunctions of the infinitesimal generator are characterized by the differential equation (3.8).

3.2.1 Examples in continuous time

Let f_t the stationary solution of the following stochastic differential equation

$$df_t = (c + df_t)dt + \sqrt{af_t^2 + bf_t + c}\, dW_t. \tag{3.11}$$

Then we have five different cases of interest that correspond to the behavior of the roots of

$$af^2 + bf + c = 0. \tag{3.12}$$

[31] As a consequence, any eigenvalue of the conditional expectation operator is positive; see Florens et al. (1998).
[32] For references on orthogonal polynomials, see Chihara (1978) and Szego (1975).

i) If $a = b = 0$ (and $c \neq 0$), then (3.12) has no root and this example corresponds to the Ornstein-Uhlenbeck process. Therefore, the eigenfunctions are the Hermites polynomials, H_i, $i \in \mathbf{N}$, while the eigenvalues are βi, $i \in \mathbf{N}$. The support of f_t is $I\!R$, the marginal distribution is normal and the conditional distribution of f_t given f_{t-h} is also normal.

ii) If $a = 0$ and $b \neq 0$, then (3.12) has one real root, say f^*. Therefore, to ensure the nonnegativity of the diffusion term, we have to exclude $]-\infty, f^*[$ or $]f^*, \infty[$ depending on the sign of b. After an affine transformation, one may get the following diffusion

$$df_t = k(\theta - f_t)dt + \sigma\sqrt{f_t}\, dW_t.$$

In other words, this corresponds to the square-root process considered by Feller (1951) and popularized in Finance by Cox, Ingersoll and Ross (1984). Observe that the affine model of Duffie and Kan (1996) is also a special example of this case. Therefore, the eigenfunctions are the Laguerre polynomials, $\{L_i^{(\alpha)},\ i \in \mathbf{N}\}$ where $\alpha = 2k\theta/\sigma^2 - 1$, while the eigenvalues are $\{ki,\ i \in \mathbf{N}\}$. The support of f_t is $[0, \infty[$, the marginal distribution is $\gamma(\alpha+1, \theta/(\alpha+1))$ while the conditional distribution is chi-square.

Consider now the case where $a \neq 0$. There are three cases: the roots are real and different (iii), they are real and equal (iv) and they are complex and conjugate (v).

iii) If $a \neq 0$ and the roots of (3.12) are real and different (when $b^2 - 4ac > 0$). Then, an affine transformation of f_t gives

$$df_t = \frac{1}{2}[(\beta+1)(1-f_t) - (\alpha+1)(1+f_t)]dt + \sqrt{1-f_t^2}\, dW_t.$$

This process is known as the Jacobi diffusion (see Karlin and Taylor, page 335). The support of f_t is $[-1, 1]$ and the marginal distribution is Beta. The eigenfunctions are the Jacobi polynomials $P_i^{(\alpha,\beta)}$ (see, e.g., Szego, 1975) while the corresponding eigenvalues are $i(i+\alpha+\beta+1)$.

iv) If $a \neq 0$ and the roots of (3.12) are equal (when $b^2 - 4ac = 0$). Then, an affine transformation of f_t gives

$$df_t = k(\theta - f_t)dt + \sigma f_t\, dW_t.$$

In other words, this is the GARCH diffusion process considered by Nelson (1990-b) and introduced the first time by Wong (1964). The support of f_t is $[0, +\infty[$ while the marginal distribution is inverse Gamma (see Wong, 1964). The polynomial solutions of

$$k(\theta - f)\phi'(f) + \frac{\sigma^2 f^2}{2}\phi''(f) = -\delta\phi(f)$$

are known as the Bessel polynomials (see Chihara, 1978, page 182). However, there exists an integer i_0 such that for all $i > i_0$, the Bessel polynomials are not square integrable with respect to the stationary marginal density of f_t. As a consequence, the Bessel polynomials of order higher than i_0 are not in the domain of the infinitesimal generator and, hence, are not eigenfunctions. Fortunately, Wong (1964) derives the eigenfunctions of the diffusion infinitesimal generator by using results on Sturm-Liouville equations. In particular, he shows that the spectrum is mixed or continuous and that the eigenfunctions are hypergeometric functions.[33]

[33] Note that Hermite, Laguerre and Jacobi polynomials are particular hypergeometric functions.

v) If $a \neq 0$ and the roots of (3.12) are complex and conjugate (when $b^2 - 4ac < 0$). Then, the solutions of the differential equations are known as the Romanowski polynomials. However, as for the previous case, higher order moments are not defined. Therefore, the polynomials are not eigenfunctions.[34] Note that in a particular case, the marginal distribution of f_t is a Student.

Note that Wong (1964) shows that any diffusion which eigenfunctions are only polynomials is necessarily one of the cases i), ii) or iii), i.e. Ornstein-Uhlenbeck, square-root or Jacobi. In addition to all these examples, there are others where the eigenfunctions are not polynomials. Some of them are considered in Kessler and Sorensen (1999). In particular, they consider these two examples:

$$df_t = -\theta \tan(f_t)dt + dW_t, \quad \theta \geq \frac{1}{2} \text{ and}$$

$$df_t = \left\{ (\theta_1 - \theta_2)\cosh(\frac{f_t}{2} - (\theta_1 + \theta_2)\sinh(\frac{f_t}{2})) \right\} \cosh(\frac{f_t}{2})dt + 2\cosh(\frac{f_t}{2})dW_t, \ \theta_1 > 0, \ \theta_2 > 0.$$

In the first case, the spectrum is $\{i(\theta + i/2), i \in \mathbf{N}\}$ with corresponding eigenfunctions given by $\phi_i(f) = C_i^\theta(\sin(f))$ where C_i^θ is a Gegenbauer polynomial of order i. In this case, the support is $]-\pi/2, \pi/2[$ and the marginal density is proportional to $\cos(f)^{2\theta}$. In the second example, the spectrum is $\{(i/2)(i + 1 + \theta_1 + \theta_2), i \in \mathbf{N}\}$ with corresponding eigenfunctions given by $\phi_i(f) = P_i^{(\theta_1,\theta_2)}(\tanh(f/2))$ where $P_i^{(\theta_1,\theta_2)}$ is the Jacobi polynomial of order i. In this case, the marginal distribution is the generalized logistic distribution with density

$$B(\theta_1 + 1, \theta_2 + 1)e^{(\theta_1+1)x}(1 + e^x)^{\theta_1+\theta_2+2}$$

where $B(.,.)$ denotes the Beta function given by $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$.

3.2.2 Examples in discrete time

As we already mentioned, any discrete sample of a diffusion process is an additional example. In particular, the eigenfunctions of the conditional expectation operator coincide with ones of the infinitesimal generator while the eigenvalues are related by (3.10). In addition to these examples, we can mention some others where the eigenfunctions are polynomials. In particular (see Schoutens, 2000), the eigenfunctions may be the Meixner, Krawtchouk or the Charlier polynomials. In these case, the marginal distribution of f_t is Pascal, Binomial or Poisson respectively. Observe that the support of the random variable f is discrete in these cases.

Recently, Darolles, Gouriéroux and Jasiak (2001) introduced a large class of processes in discrete time called compound autoregressive (CAR). This class is characterized by an affine function of the conditional log-Laplace transform. This property is probably the main advantage of the square-root process. Therefore, CAR models share this property. When the process is time-reversible, these authors characterize the eigenfunctions of the conditional expectation operator and show that they are polynomials. Special examples are the square-root and the three discrete valued cases (Pascal, Binomial and Poisson).

[34] We have not found references that characterize the eigenfunctions and the spectrum.

4 The Eigenfunctions Stochastic Volatility models

In this section, we introduce the general approach of modeling volatility by eigenfunctions. We start by introducing the discrete time volatility model and give additional examples other than the Hermite and Laguerre ones. After that, we introduce the continuous time volatility model and give some additional examples. Then we consider the multifactor approach. In the last subsection, we relate our approach to the literature. More precisely, we give several examples of the literature that implicitly consider particular eigenfunctions models. Moreover, we consider a popular continuous time volatility model, namely the GARCH diffusion model of Nelson (1990-b), which is not included in the general theory. However, we show how our approach can be extended to incorporate this model.

4.1 The discrete time Eigenfunction Stochastic Volatility model

4.1.1 The model

Definition 4.1. Discrete time Eigenfunction SV model: *A process $\{\varepsilon_t\}$ is called an Eigenfunction SV model of order p, ESV(P), with an underlying Markovian state variable $\{f_t\}$ if:*

$$\varepsilon_t = \sigma_{t-1} u_t, \quad \text{with} \tag{4.1}$$

$$\sigma_{t-1}^2 = \sum_{i=0}^{p} a_i E_i(f_{t-1}) \quad \text{where} \tag{4.2}$$

$$\sum_{i=0}^{p} a_i^2 < \infty, \tag{4.3}$$

u_t is independent from $\{f_\tau, u_\tau, \tau \leq t-1\}$, identically distributed with zero mean and unit variance and $E_i(f_t)$, $i \in \mathbf{N}$, are the eigenfunction (with corresponding eigenvalues λ_i) of the conditional expectation operator associated to the state variable f_t given $\{u_\tau, f_\tau, \tau \leq t-1\}$, i.e.

$$E[E_i(f_t) \mid u_\tau, f_\tau, \tau \leq t-1] = \lambda_i E_i(f_{t-1}). \tag{4.4}$$

In other words, we say that a process $\{\varepsilon_t\}$ is an Eigenfunction SV model of order p, ESV(p), if its conditional variance process given $\{f_\tau, u_\tau, \tau \leq t-1\}$ is a linear combination of the eigenfunctions of the conditional expectation operator associated to the state variable f_t. This linear combination is in terms of the constant and the p first eigenfunctions. Observe that in the previous definition, the conditional expectation operator is defined given the past of f_t and u_t which is different from the previous section. However, in the examples considered in the literature and ours, the process $\{u_t\}$ does not cause $\{f_t\}$. More precisely, we have for any square-integrable function $g(f_t)$:

$$E[g(f_t) \mid f_\tau, u_\tau, \tau \leq t-1] = E[g(f_t) \mid f_\tau, \tau \leq t-1] = E[g(f_t) \mid f_{t-1}].$$

The first equality is the non-causality assumption while the second one is due to the Markovianity of f_t. When p is infinite, (4.3) means that the conditional variance of ε_t is the limit in mean-square of $\sum_{i=0}^{\tilde{p}} a_i E_i(f_{t-1})$ when $\tilde{p} \to +\infty$. Observe also that we do not make restrictive assumptions on u_t. In particular, the distribution of u_t may be a standardized Student or any other distribution with zero mean and unit variance. Therefore, some higher moments of u_t may be infinite. However, we will assume that the fourth moment of u_t is finite when we will consider some properties of the process of ε_t like the covariance structure of their square (heteroskedasticity). Note also that we do not preclude leverage effect (Black, 1976; Nelson, 1991), i.e., the random variables u_t and f_t may be correlated.

Definition 4.2. Terminology: *Let $\{\varepsilon_t\}$ an ESV(p) model with an underlying Markovian state variable $\{f_t\}$ with corresponding eigenfunction $E_i(f_t)$, $i \in \mathbf{N}$. If the name of the eigenfunction is known, for instance Hermite polynomials, then we give the same name for the volatility model, for instance Hermite SV model, HSV(p).*

This means that, for instance, when the eigenfunctions are the Jacobi polynomials, we say that ε_t is the Jacobi SV model of order p, JSV(p).

4.1.2 Moments

We now compute the moments of the ESV model. We compute the first four moments, the covariance structures of the squared residual and conditional moment restrictions. We start by computing the marginal moments. Observe that most of the comments made in the HSV and LSV models hold for the ESV ones. Therefore, we will not make then again.

Proposition 4.1 Marginal moments of ESV models. *Consider $\{\varepsilon_t, t \in \mathbf{N}\}$ a discrete-time ESV(p) model. Then the mean and variance of ε_t are given by:*

$$E[\varepsilon_t] = 0;\ Var[\varepsilon_t^2] = a_0. \tag{4.5}$$

When the fourth moment of u_t is finite, the fourth moment and kurtosis of ε_t are given by

$$E[\varepsilon_t^4] = E[u_t^4] \sum_{i=0}^{p} a_i^2; \quad Kurt[\varepsilon_t] = Kurt[u_t] \left(1 + \frac{\sum_{i=1}^{p} a_i^2}{a_0^2}\right). \tag{4.6}$$

Proposition 4.2 Covariance structures of ESV models. *Consider $\{\varepsilon_t, t \in \mathbf{N}\}$ a discrete-time ESV(p) model. Assume that the fourth moment of u_t is finite and that there is no leverage effect, i.e. u_t and f_t are independent. Then, $\forall j > 0$, we have:*

$$Cov[\varepsilon_t^2, \varepsilon_{t-j}^2] = \sum_{i=1}^{p} a_i^2 \lambda_i^j. \tag{4.7}$$

Observe that we assume that there is no leverage effect. However, if one assume there is, one has to specify the dependence between f_t and u_t and, then, take into account this specification in the computation of $Cov[\varepsilon_t^2, \varepsilon_{t-j}^2]$. For instance, consider the Hermite SV case and assume the usual leverage effect specification, i.e.

$$(u_t, v_t) \ \text{ i.i.d. } \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left[\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right]\right) \tag{4.8}$$

where v_t is the standardized innovation of f_t, i.e.

$$f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2}\, v_t. \tag{4.9}$$

Then, we can compute explicitly the covariance structure of the squared process ε_t^2.

Proposition 4.3 Covariance structures of HSV models under leverage effect. *Consider $\{\varepsilon_t, t \in \mathbf{N}\}$ a discrete-time HSV(p) model and assume that the leverage effect specification is given by (4.8). Then, $\forall j > 0$:*

$$Cov[\varepsilon_t^2, \varepsilon_{t-j}^2] = (1-\rho^2)\sum_{i=1}^{p} a_i^2\gamma^{ij} + \rho^2 a_0 a_2 (1-\gamma^2)\gamma^2\sqrt{2} +$$

$$\rho^2 \sum_{i=1}^{p} a_i\gamma^{ij}\left[(1-\gamma^2)\sqrt{(i-1)i}\gamma^{-2}a_{i-2} + (1+2i(1-\gamma^2))a_i + (1-\gamma^2)\sqrt{(i+1)(i+2)}\gamma^2 a_{i+2}\right] \tag{4.10}$$

with the convention that $a_i = 0$ if $i < 0$ or $i > p$.

When there is no leverage effect, i.e. $\rho = 0$, the first element in the right part of (4.10) corresponds exactly to (4.7). Therefore the last two terms in the right part of (4.10) are due to the leverage effect. Following Meddahi and Renault (1996, 2000), we can also characterize the leverage effect in terms of observable restrictions. More precisely, these authors show that the presence of leverage effect implies that $Cov[\varepsilon_t, \varepsilon_{t+1}^2]$ is nonzero. As usual in volatility models, one has to derive the model on σ_t in order to compute moments that involve the standard deviation σ_t and not the variance σ_t^2. This is feasible in exponential models (e.g. EGARCH and log-normal SV) but not in GARCH. This is also our case. Hence, if one has the decomposition of σ_t in terms of eigenfunctions (see Section 2), then one can compute $Cov[\varepsilon_t, \varepsilon_{t+1}^2]$.

Let us now consider the dynamics of the squared residual process. The variance definition means that

$$E[\varepsilon_t^2 \mid \varepsilon_\tau, f_\tau, \tau \leq t-1] = \sum_{i=0}^{p} a_i E_i(f_{t-1}). \tag{4.11}$$

Since each eigenfunction is an AR(1), (4.11) implies, under some condition, that ε_t^2 is an ARMA(p,p). In fact, we have a more restricted observable implications:

Proposition 4.4 Observable conditional moments restrictions of ESV models. *Consider $\{\varepsilon_t, t \in \mathbf{N}\}$ a discrete-time HSV(p) model. Then we have:*

$$E[\varepsilon_t \mid \varepsilon_\tau, \tau \leq t-1] = 0. \tag{4.12}$$

If p is finite, then we have:

$$E[\prod_{i=1}^{p}(1-\lambda_i L)[\varepsilon_t^2 - a_0] \mid \varepsilon_\tau, \tau \leq t-p-1] = 0. \tag{4.13}$$

As a consequence, if the fourth moment of ε_t is finite, ε_t^2 is an ARMA(p,p) model with autoregressive coefficients λ_1, λ_2,...,λ_p.

When the fourth moment of ε_t is finite, then (4.13) implies that the process w_t defined by

$$w_t \equiv \prod_{i=1}^{p}(1-\gamma^i L)[\varepsilon_t^2 - a_0] \tag{4.14}$$

is a weak moving average of order p, MA(p), and, hence, ε_t^2 is a weak ARMA(p,p) and ε_t is a weak GARCH(p,p).

4.1.3 Temporal aggregation and asymptotic behavior of the kurtosis of aggregated process

In order to study the temporal aggregation of the Eigenfunction SV models, Meddahi (2001-b) relaxes the main restrictive assumption of ESV models that is not closed under temporal aggregation, i.e. the i.i.d. assumption on u_t. Then he shows that semiparametric ESV(p) class of models is closed under temporal aggregation.

Definition 4.3. Semiparametric ESV model: *A stationary squared integrable process* $\{\varepsilon_t\}$ *is called a semiparametric ESV(p) w.r.t. an increasing filtration* J_t *with an underlying state variable process* $\{f_t\}$ *if:*

i) the process ε_t *is adapted w.r.t.* J_t*, that is* $I_t \subset J_t$ *where* $I_t = \sigma(\varepsilon_\tau, \tau \leq t)$*;*

ii) ε_t *is a martingale difference sequence w.r.t.* J_{t-1}*, i.e.* $E[\varepsilon_t \mid J_{t-1}] = 0$*;*

iii) the conditional variance process σ_{t-1}^2 *of* ε_t *given* J_{t-1} *is given by*

$$\sigma_{t-1}^2 = \sum_{i=0}^{p} a_i E_i(f_{t-1}) \quad \text{where} \tag{4.15}$$

where the sequence a_i *is such that* $\sum_{i=0}^{p} a_i^2 < \infty$ *and* $E_i(f_t)$*,* $i \in \mathbf{N}$*, are the eigenfunction (with corresponding eigenvalues* λ_i*) of the conditional expectation operator associated to the state variable* f_t *given* J_{t-1}*, i.e.*

$$E[E_i(f_t) \mid J_{t-1}] = \lambda_i E_i(f_{t-1}). \tag{4.16}$$

Proposition 4.5 Temporal aggregation of semiparametric ESV models
Let ε_t *a semiparametric ESV(p) model w.r.t. an increasing filtration* J_t *where* $\{f_t\}$ *is the underlying state variable process. Define for a given integer* m *and real numbers* $w_1, ..., w_m$ *the process* $\{\varepsilon_{tm}^{(m)}, t \in \mathbf{N}\}$ *by* $\varepsilon_{tm}^{(m)} = \sum_{i=0}^{m-1} w_i \varepsilon_{tm-i}$*. Then* $\{\varepsilon_{tm}^{(m)}, t \in \mathbf{N}\}$ *is a semiparametric ESV(p) model w.r.t.* $J_{tm}^{(m)} = \sigma(\varepsilon_{\tau m}^{(m)}, f_\tau, \tau \leq t)$ *with the same underlying state variable process, i.e.* $f_{\tau m}^{(m)} = f_{\tau m}$*. The eigenfunctions are the same but the corresponding eigenvalues are* $\lambda_i^{(m)} = \lambda_i^m$*.*

We now derive the fourth moment of the aggregated process and its asymptotic behavior:

Proposition 4.6 Asymptotic behavior of the kurtosis of aggregated process *Consider* $\{\varepsilon_t, t \in \mathbf{N}\}$ *a discrete-time ESV(p) model. Assume that the fourth moment of* u_t *is finite, the third moment of* u_t *is zero and that there is no leverage effect. Let* m *be an integer and consider the aggregated process* $\{\varepsilon_{\tau m}^{(m)}\}$ *defined by* $\varepsilon_{\tau m}^{(m)} \equiv \sum_{i=0}^{m-1} \varepsilon_{tm-i}$*. Then the kurtosis of* $\varepsilon_{\tau m}^{(m)}$ *is given by*

$$Kurt(\varepsilon_{\tau m}^{(m)}) - 3 = \frac{1}{m}(kurt(\varepsilon_t) - 3) + \frac{6}{a_0^2} \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{(1-\lambda_i)^2} \frac{\lambda_i^m - m\lambda_i + m - 1}{m^2}. \tag{4.17}$$

Moreover, when $m \longrightarrow +\infty$*, we have*

$$Kurt(\varepsilon_{\tau m}^{(m)}) - 3 = \frac{1}{m}\left((kurt(\varepsilon_t) - 3) + \frac{6}{a_0^2} \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{(1-\lambda_i)^2}\right) + \mathrm{o}(\frac{1}{m}). \tag{4.18}$$

4.2 The continuous time Eigenfunction Stochastic Volatility models

Definition 4.3. Continuous time ESV model: *A continuous time process $\{y_t\}$ is called an Eigenfunction SV model of order p, ESV(p), with an underlying diffusion process $\{f_t\}$ if:*

$$dy_t = mdt + \sigma_t \, [\sqrt{1-\rho^2} dW_t^{(1)} + \rho dW_t^{(2)}], \quad \text{with} \tag{4.19}$$

$$\sigma_t^2 = \sum_{i=0}^{p} a_i E_i(f_t), \quad \text{where} \tag{4.20}$$

$$\sum_{i=0}^{p} a_i^2 < \infty, \tag{4.21}$$

$E_i(f_t)$ are the eigenfunctions (with corresponding eigenvalues $(-\delta_i)$) of the infinitesimal generator associated to the stationary process f_t characterized by

$$df_t = \mu(f_t) + \sigma(f_t) dW_t^{(2)}, \tag{4.22}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are two independent standard Brownian processes.

Again, as for the discrete time ESV model, we say that a continuous time SV model is an ESV(p) one if the instantaneous variance process σ_t^2 is a linear combination of the first p eigenfunctions of the infinitesimal generator associated to the diffusion process f_t. All the remarks made in the discrete time SV model hold here. Observe that here we specify explicitly the leverage effect via ρ. In the following section, we show how this model may be extended to incorporate time-varying leverage effect. The last section also considers a time-varying drift model where the drift is also a linear combination of the eigenfunctions $E_i(f_t)$. Finally, we adopt the same terminology as in discrete time. For instance, when the eigenfunctions $E_i(f_t)$ are the Jacobi polynomials, we say that y_t is a Jacobi SV model.

Unconditional and conditional moments of these models are derived in Meddahi (2001-a). As for Hermite and Laguerre SV models, exact discretization of ESV models leads to semiparametric ESV models (see Meddahi, 2001-b):

Proposition 4.7 Exact discretization. *Consider $\{y_t\}$ a continuous time ESV(p) model where $\{f_t\}$ is the underlying state variable. Then, for any sampling interval h, the associated discrete time process $\varepsilon_{th}^{(h)} \equiv y_{th} - y_{(t-1)h}$ is a semiparametric ESV(p) model w.r.t. $J_{th}^{(h)} = \sigma(\varepsilon_{\tau h}^{(h)}, f_{\tau h}, \tau \leq t)$ with the same underlying state variable process, i.e. $f_{\tau h}^{(h)} = f_{\tau h}$. More precisely, we have:*

$$\sigma_{(t-1)}^{(h)2} \equiv Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}] = a_0 h E_0(f_{(t-1)h}^{(h)}) + \sum_{i=1}^{p} \frac{a_i(1-\exp(-\delta_i h))}{\delta_i} E_i(f_{(t-1)h}^{(h)}). \tag{4.23}$$

4.3 Multifactor model

Instead of specifying that the variance process depends on one factor, we can assume that it depends on several ones. For simplicity, we consider the two factor case, i.e. we assume that

$$\sigma_t^2 = F(f_{1,t}, f_{2,t}) \tag{4.24}$$

where $f_{1,t}$ and $f_{2,t}$ are two independent stochastic processes[35] characterized by the SDE

$$df_{i,t} = \mu_i(f_{i,t})dt + \sigma_i(f_{i,t})dW_{i,t}. \tag{4.25}$$

Therefore, to specify the variance process, one has to define the function $F(.,.)$. The literature considers in general additive functions. More precisely, the usual assumption is that

$$\sigma_t^2 = f_{1,t} + f_{2,t}$$

where the processes $f_{i,t}$ are affine processes while the variance is assumed to be

$$\sigma_t^2 = \exp(f_{1,t} + f_{2,t})$$

where the processes $f_{i,t}$ are Ornstein-Uhlenbeck processes.[36] It is important to understand that such specifications are considered for simplicity. However, it is not clear that this simple approach provides models that fit the data.

In contrast, there is a general theory about basis expansion of general function. In particular if one considers a Taylor expansion of $F(.,.)$, than this says that a general formulation is

$$\sigma_t^2 = \sum_{0 \leq i,j \leq \infty} a_{i,j} f_{1,t}^i f_{2,t}^j.$$

Again, rather than considering expansion in terms of monomials, we will consider the expansion in terms of eigenfunctions. More precisely we assume that

$$\sigma_t^2 = \sum_{0 \leq i,j \leq p} a_{i,j} E_{1,i}(f_{1,t}) E_{2,j}(f_{2,t}) \quad \text{where} \quad \sum_{0 \leq i,j \leq p} a_{i,j}^2 < \infty$$

where $E_{1,i}(.)$ and $E_{2,j}$ are the eigenfunctions of the variables $f_{1,t}$ and $f_{2,t}$ respectively with corresponding eigenvalues $\lambda_{1,i}$ and $\lambda_{2,j}$. Observe that we do not assume that the processes $f_{1,t}$ and $f_{2,t}$ have the same eigenfunctions. In particular, $f_{1,t}$ may be a normalized Ornstein-Uhlenbeck process while $f_{2,t}$ is a square-root one.[37]

This approach has two important advantages. The first one is that multiplicative models are useful for generating fat tails (see Section 2). In particular, they can be interpreted as subordinated processes which is a usual solution to generate fat tails (Clark, 1973).[38] Interestingly, Chernov et al. (2001) found that empirically, multiplicative models outperform the other ones. The second advantage of our approach is that the dynamics of the variance process given as a linear combination of the cross product of the eigenfunctions are very simple. The reason is the following. Consider two independent processes $y_{1,t}$ and $y_{2,t}$ such that

$$E[y_{i,t} \mid J_{t-1}] = \rho_i y_{i,t-1}$$

[35] The independence assumption is usually made in the literature. It is not clear if this assumption is supported by the data.

[36] In their positive Ornstein-Uhlenbeck SV model, Barndorff-Nielsen and Shephard (2001) consider also multifactor model where the variance is assumed to be the sum of independent positive processes.

[37] In a different context, namely terme structure modeling, Ahn, Dittmar, Gallant and Gao (2000) consider a multifactor model where some factors are Ornstein-Uhlenbeck processes and others are affine.

[38] See Carrasco, Hansen and Chen (1999) for a general study of the impact of time deformation on the dependence of a process.

i.e., each process $y_{i,t}$ is a centered AR(1) process. Define y_t by $y_t \equiv y_{1,t}y_{2,t}$. Then,

$$E[y_t \mid J_{t-1}] = E[y_{1,t} \mid J_{t-1}]E[y_{2,t} \mid J_{t-1}] = \rho_1 y_{1,t-1}\rho_2 y_{2,t-1} = \rho_1\rho_2 y_{t-1},$$

i.e. y_t is a centered AR(1) process with autoregressive coefficient equal to $\rho_1\rho_2$. Therefore, by applying this result to $E_{i,j}(f_t)$ defined by

$$E_{i,j}(f_t) \equiv E_{1,i}(f_{1,t})E_{2,j}(f_{2,t}) \quad \text{where} \quad f_t \equiv (f_{1,t}, f_{2,t})', \tag{4.26}$$

one gets

$$E[E_{i,j}(f_{t+h}) \mid J_t] = \exp(-\delta_{i,j}h)E_{i,j}(f_t) \quad \text{where} \quad \delta_{i,j} = \delta_{1,i} + \delta_{2,j}. \tag{4.27}$$

In other words, the variance process σ_t^2 is a linear combination of centered AR(1) processes. Observe that

$$E[E_{i,j}(f_t)E_{k,l}(f_t)] = E[E_{1,i}(f_{1,t})E_{1,k}(f_{1,t})]E[E_{2,j}(f_{2,t})E_{2,l}(f_{2,t})] = \delta_{ik}\delta_{jl}.$$

Hence if $(i,j) \neq (k,l)$, then $E[E_{i,j}(f_t)E_{k,l}(f_t)] = 0$ while $E[E_{i,j}(f_t)E_{i,j}(f_t)] = 1$, i.e. $E_{i,j}$ are orthogonal and their variance is one. As a consequence, the functions $E_{i,j}$ are the eigenfunctions of the conditional expectation operator associated to the state variable f_t. Hence, all the results provided in the previous sections still hold.[39]

4.4 Related literature

In this section, we will review the literature related to our approach. Consider the eigenfunctions approach for time series modeling. As we already mentioned, this approach is pioneered in the econometric and financial literature by Hansen and Scheinkman (1995). This paper deals with identification of diffusion processes through the their infinitesimal generator and derives observable moment restrictions based on both the marginal and conditional distributions of the diffusion. Then, Hansen, Scheinkman and Touzi (1997) consider the identification of a scalar diffusion through the first eigenfunction of the infinitesimal generator. After these two papers, Chen, Hansen and Scheinkman (1998), Darolles, Florens and Gouriéroux (1998) and Darolles, Florens and Renault (1998), highlighted the importance of the eigenfunctions autoregressive dynamics for time series modeling purposes. In particular the nonlinear principal components terminology appears in these papers. While Darolles, Florens and Renault (1998) gives a condition ensuring that the spectrum of the conditional expectation operator is discrete, Chen, Hansen and Scheinkman (1998) and Darolles, Florens and Gouriéroux (1998) consider the nonparametric extraction of the first eigenfunction. Besides, Chen, Hansen and Scheinkman (2000) consider a new approach for continuous time modeling by specifying the marginal distribution and the diffusion (matrix) term. This approach is very appealing since economists are in general mostly concerned about the long run, i.e. the marginal distribution. In this setup, Chen, Hansen and Scheinkman (2000) derives also the principal components. As mentioned in the introduction, our work was motivated by these papers. However we consider a parametric

[39] See Chen, Hansen and Scheinkman (2000) for a more general approach of principal components modeling in the multivariate case.

model and SV framework. Finally, Kessler and Sorensen (1999) consider the estimation of various univariate scalar diffusion by using the eigenfunctions of the infinitesimal generator of the diffusion. The estimation is based on the estimating functions method.

An important advantage of the eigenfunction approach is that the dynamics of the state variable driving the volatility process are not related to the relationship between these two variables. Therefore we reduce the link between the marginal and conditional distributions of the returns. This is in line with the approach of Chen, Hansen and Scheinkman (2000). Barndorff-Nielsen and Shephard (2001) also consider a continuous time SV model where the process driving the volatility process is a positive Levy process that has this nice feature.

We will now review the literature in both volatility in interest rates modeling that implicitly consider some particular eigenfunctions models. We already show that the lognormal and affine models are particular example of our approach. Some papers consider that the variance process is the square of a Gaussian AR(1) process. This is the case of Robinson and Zaffaroni (1998) in discrete time and Stein and Stein (1991) and Ho, Perraudin and Sorensen (1996) in continuous time. The same approach was also considered for modeling interest rates by Constantinides (1991), Ahn, Dittmar and Gallant (2000) and Leippold and Wu (1999). It means that these models consider a Hermite model of order 2. However by taking the square of an AR(1) process, they consider a restricted Hermite model. Gallant, Hsu and Tauchen (1999) and Chernov et al. (2001) consider a model where the variance process is the exponential of the GARCH diffusion process of Nelson (1990-b) and Wong (1964). As we already mention, the spectrum of this process is mixed or continuous. Therefore, when the variance process is square-integrable, we can encompass such models by considering an integral instead of a sum in the definition of the variance process. Interestingly, Chen, Hansen and Carrasco (1999), show that when the spectrum of a scalar diffusion is continuous, there exist a (non-linear) function of this diffusion such that its spectral density is unbounded around zero, which looks like a long-memory property. As a consequence, this is the case of the GARCH process when its spectrum is continuous. In other words, in this case, the return process has a long memory stochastic volatility. This may explain the long-memory founded in Gallant, Hsu and Tauchen (1999).[40] Andersen (1994) considers briefly Taylor expansion of volatility models but did not study them. Finally, Robinson (2001) considers Hermite polynomials expansions to study the memory of volatility models.[41]

Of course, econometricians are familiar with expansions around Hermite polynomials. This is the case in the edgeworth expansion but also in the seminonparametric estimation of Gallant and Tauchen (1989) and in the maximum-likelihood approximation method of Ait-Sahalia (2000). This is different from our case for two reasons. We assume that we consider the true model and not an expansion of the volatility. Moreover, the choice of polynomials and more generally of eigenfunctions depends on the underlying state variable which is not the case of these works.

[40] An alternative approach to generate models that look like as long-memory volatility ones is to consider a two component (or factor) model as in Engle and Lee (1999). This is the multifactor case previously considered.

[41] In a different perspective, Ghysels, Gouriéroux and Jasiak (1998) and Pitt and Walker (1999) consider latent variable models where the loading function is nonlinear. This function was specified such that the marginal distribution of the observable process is fixed, say Gamma. In this case, the dynamics of the observed process are complicated and not characterized. Therefore, this approach is different from ours.

5 Concluding remarks

In this paper, we consider a flexible approach for volatility modeling. We follow the main idea of the SV literature, that is by specifying the volatility as a function of a state variable. Rather than specifying the variance as being equal to a specific function of the state variable, we assume that it is a linear combination of the eigenfunctions of the conditional expectation operator associated to the state variable. Special examples are the log-normal and square root SV models. When the state variable is Gaussian (resp square-root), the eigenfunctions are the Hermite (resp Laguerre) polynomials. This flexible approach has many advantages. First of all, it allows us to generate fat tails. Moreover, the structure of the squared returns are ARMA and hence simple to estimate. For the discrete time model, we derive conditional moment restrictions based on the state variable as well as on the observations. For the continuous time model, unconditional and conditional moments of the discrete time returns are derived in a companion paper, Meddahi (2001-a). We also consider the long run properties of our model and study the variance of the variance. Finally, we extend our model to the multifactor case.

Several other extensions of our model may be considered. For instance, we can incorporate a time varying mean or drift, denoted m_t. In particular, a flexible approach is obtained by assuming

$$m_t = \sum_{i=0}^{p} m_i E_i(f_t) \quad \text{where} \quad \sum_{i=0}^{p} m_i^2 < \infty.$$

As a consequence, we allow the mean/drift to be a (square-integrable) nonlinear function of the variance process.

We can also incorporate jumps in the continuous time model. The intensity (and/or size) of the jump is then assumed to be a linear combination of the eigenfunctions. Instead of assuming that (a component of) f_t is continuous, we can assume that it is a continuous time Markov chain, that is f_t takes a finite number of values. It turns out that in this case, we have also the general theory about the eigenfunctions. Moreover, when the Markov chain is time reversible, the eigenvalues are real numbers.

A usual assumption made in the SV literature is that the leverage effect is constant. However, empirical evidence based on options data suggest that the leverage effect is time varying. This leads Garcia, Luger and Renault (2001) to consider a discrete time model where the leverage effect is time varying. Observe also that if one considers a multifactor model for the variance process where the Brownian motions associated to the volatility factors are correlated to the stock Brownian motion, then this implies a time varying leverage effect. This is the case of Jones (2000) and Chernov et al. (2001). An alternative approach is to specify directly a stochastic process for the correlation parameter. Consider the one factor case considered in Section 4 and assume that the instantaneous correlation between the Brownian motion processes $W_t^{(1)}$ and $W_t^{(2)}$ is not constant but time-varying and given by $\rho_t = G(z_t)$, where z_t is a state variable and G a function which values are in $[-1, 1]$. The state variable z_t may be the same as the factor governing the volatility, i.e. f_t, or another variable. The process z_t may also be a Markov Chain in continuous time. In this case, the set of values of ρ_t is finite. Observe that in this case the correlation is also a Markov chain and, hence, a mean-reverting process. A different approach to have mean reverting correlation with continuous state variable is obtained by assuming that the correlation is characterized by a Jacobi diffusion:

$$d\rho_t = \frac{1}{2}[(\beta+1)(1-\rho_t) - (\alpha+1)(1+\rho_t)]dt + \sqrt{1-\rho_t^2}dW_t^{(3)}$$

where the Brownian motion process $W_t^{(3)}$ is assumed to be independent with $(W_t^{(1)}, W_t^{(2)})'$.

Finally, while all the previous models were considered in an SV setting, the same approach may be considered for ARCH-type models. More precisely, consider the process ε_t defined by

$$\varepsilon_t = \sigma_{t-1} u_t$$

where the process u_t is assumed to be i.i.d. with some distribution $D(0,1)$ and assume that σ_{t-1} is adapted to the information $\sigma(\varepsilon_\tau, \tau \leq t-1)$. In other words, we consider an ARCH-type model. Assume that the distribution function of u_t is $F(u,\theta)$ where F is a known function and θ an unknown parameter. Then, define the process v_t by $v_t = \Phi^{-1}(F(u_t,\theta))$ where Φ is the distribution function of the $\mathcal{N}(0,1)$ random variable. Let γ a real number, $\mid \gamma \mid < 1$, and define the processes f_t and σ_{t-1} by

$$f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2} v_t \quad \text{and}$$

$$\sigma_{t-1}^2 = \sum_{i=0}^{p} a_i H_i(f_t) \quad \text{where} \quad \sum_{i=0}^{p} a_i^2 < \infty.$$

Then the process ε_t is an ARCH-type model. Moreover it has the flexibility of the eigenfunctions volatility models studied in the previous sections. We call this model the Hermite ARCH model. A special example is the Gaussian exponential ARCH model of Nelson (1991). Observe that this model automatically incorporates the leverage effect when the odd Hermite polynomials are included in the variance decomposition.

Table 1. Decomposition of the variance process: discrete time model.

i	a_i	w_i	cum_i	γ^i	a_i	w_i	cum_i	γ^i
0	.558	-	-	-	.349	-	-	-
1	.422	.742	.742	.978	.219	.816	.816	.980
2	.226	.212	.953	.956	.098	.161	.977	.961
3	.098	.041	.994	.935	.036	.022	.998	.942
4	.038	.006	.999	.914	.011	.002	.999	.923
5	.013	6.7e-04	1.00	.894	3.1e-04	1.6e-04	1.00	.904
6	3.8e-3	6.3e-05	1.00	.874	7.9e-04	1.1e-05	1.00	.886
7	1.1e-03	5.2e-06	1.00	.855	1.9e-04	6.1e-07	1.00	.868
8	2.9e-04	3.7e-07	1.00	.836	4.2e-05	3.1e-08	1.00	.851
9	7.5e-05	2.4e-08	1.00	.817	8.8e-06	1.3e-09	1.00	.834
10	1.8e-05	1.4e-09	1.00	.799	1.8e-06	5.2e-11	1.00	.817

Note. In the discrete time log-normal model $y_t = \sigma_{t-1} u_t$ with $\log(\sigma_t^2) = \omega + \gamma \log \sigma_{t-1}^2 + \sigma_v v_t$, where u_t and v_t are independent and $\mathcal{N}(0,1)$, we define f_t by $f_t = \dfrac{\log(\sigma_t^2) - \mu}{\sigma}$ where $\mu = \frac{\omega}{1-\gamma}$ and $\sigma^2 = \frac{\sigma_v^2}{1-\gamma^2}$. σ_t^2 is decomposed in terms of Hermite polynomials: $\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t)$ where $a_i = \exp(\mu + \frac{\sigma^2}{2}) \frac{\sigma^i}{\sqrt{i!}}$. $w_i = a_i^2 / (\sum_{j=1}^{\infty} a_j^2)$ is the relative weight of each polynomila in the variance decomposition while $cum_i = \sum_{j=1}^{i} w_j$ is their cumulative weight. γ^i is the autoregressive parameter of each Hermite polynomial $H_i(f_t)$. We use the empirical results of Kim, Shephard et Chib (1998) for Pound-US$ and Yen-US$ exchange rates returns. The empirical estimates of $(\omega, \gamma, \sigma_v)$ are respectively (-.019,.978,.158) and (-.025,.980,.125).

Table 2. Decomposition of the variance process: continuous time model.

i	a_i	w_i	cum_i	γ^i	a_i	w_i	cum_i	γ^i
0	.396	-	-	-	0.551	-	-	-
1	.133	.945	.945	.994	.384	.778	.778	.987
2	.032	.053	.998	.988	.189	.189	.966	.974
3	.006	.002	.999	.982	.076	.031	.996	.960
4	.001	5.6e-05	1.00	.975	.027	.004	.999	.947
5	1.5e-04	1.2e-06	1.00	.969	.009	.0004	1.00	.934
6	2.1e-05	2.3e-08	1.00	.963	.003	2.8e-05	1.00	.922
7	2.7e-06	3.8e-10	1.00	.957	6.1e-04	1.9e-06	1.00	.909
8	3.1e-07	5.4e-12	1.00	.952	2.0e-04	1.2e-07	1.00	.897
9	3.6e-08	6.8e-14	1.00	.946	3.5e-05	6.5e-09	1.00	.885
10	3.8e-09	7.6e-16	1.00	.939	7.8e-06	3.1e-10	1.00	.873

Note. In the continuous time log-normal model $dy_t = \sigma_t \left[\sqrt{1-\rho^2} dW_t^{(1)} + \rho dW_t^{(2)}\right]$, $d\log(\sigma_t^2) = k[\theta - \log(\sigma_t^2)]dt + \sigma dW_t^{(2)}$, we define f_t by $f_t \equiv \frac{\sqrt{2k}}{\sigma}(\log \sigma_t^2 - \theta)$. σ_t^2 is decomposed in terms of Hermite polynomials: $\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t)$ where $a_i = \exp(\theta + \frac{\sigma^2}{4k}) \frac{(\sigma/\sqrt{2k})^i}{\sqrt{i!}}$. $w_i = a_i^2 / (\sum_{j=1}^{\infty} a_j^2)$ is the relative weight of each polynomila in the variance decomposition while $cum_i = \sum_{j=1}^{i} w_j$ is their cumulative weight. γ^i is the autoregressive parameter of each Hermite polynomial $H_i(f_t)$ for a daily frequency. We use the empirical results of Andersen, Benzoni and Lund (2001). The empirical estimates of $(\theta, k, \sigma, \rho)$ are respectively (-.984,.0062,.038,0) and (-.838,.0136,.115,-.575).

Table 3. GMM estimation of HSV(2) model

Data	a_0	a_2	γ	Jtest
Pound-US$	0.496	0.606	0.982	0.432
	(0.057)	(0.158)	(0.023)	
Yen-US$	0.338	0.266	0.953	1.516
	(0.030)	(0.053)	(0.042)	

Note. GMM Esimates and standard errors (in parenthese) of the HSV(2) model $y_t = \sigma_{t-1} u_t$ where $\sigma_{t-1}^2 = a_0 + a_2 H_2(f_{t-1})$ with $f_t = \gamma f_{t-1} + \sqrt{1-\gamma^2} v_t$, u_t and v_t are independent and $\mathcal{N}(0,1)$. The moments used in the inference are $E[y_t^2]$, $E[y_t^4]$, $Cov(y_t^2, y_{t-i}^2)$ for $i = 15$ and 20. Newey-West (1987) procedure with 10 lags is used. Jtest is the overidentification test.

Table 4. Implied moments by the different models

-	Data	HSV(2)	log-n1	log-n2
Pound-US$				
Var$[\varepsilon_t]$	.505	.496	.558	.488
$E[\varepsilon_t^4]$	2.01	1.84	1.65	1.69
Kurtosis$[\varepsilon_t]$	7.87	7.48	5.32	7.11
γ	-	.982	.978	.978
Yen-US$				
Var$[\varepsilon_t]$	.352	.338	.349	.338
$E[\varepsilon_t^4]$	.612	.556	.543	.554
Kurtosis$[\varepsilon_t]$	4.97	4.86	4.45	4.86
γ	-	.954	.980	.918

Note. Implied moments (variance, fourth moment and kurtosis) by the differente models. For the log-normal model, we use Kim, Shephard and Chib (1998) estimators (log-n1) and GMM estimators (log-n2) from Table 5.

Table 5. GMM estimation of log-normal model

Data	μ	γ	σ	Jtest
Pound-US$	-1.150	0.978	0.929	0.425
	(0.107)	(0.034)	(0.121)	
Yen-US$	-1.326	0.918	0.694	1.548
	(0.094)	(0.078)	(0.088)	

Note. GMM Esimates and standard errors (in parenthese) of the log-normal model $y_t = \sigma_{t-1} u_t$ where $\log(\sigma_t^2) = \omega + \gamma \log \sigma_{t-1}^2 + \sigma_v v_t$, where u_t and v_t are independent and $\mathcal{N}(0,1)$. We estimate (μ, γ, σ) where $\mu = \frac{\omega}{1-\gamma}$ and $\sigma^2 = \frac{\sigma_v^2}{1-\gamma^2}$. The moments used in the inference are $E[y_t^2]$, $E[y_t^4]$, $Cov(y_t^2, y_{t-i}^2)$ for $i = 15$ and 20. Newey-West (1987) procedure with 10 lags is used. Jtest is the overidentification test.

Table 6. Implied sixth standardized moment

Data	2	4	6	8	log-n1	log-n2
Pound-US: 13.1	10.7	32.3	168.5	1073	5.4	12.7
Yen-US: 3.6	4.3	10.1	46.5	288.6	3.3	4.3

Note. The standardized sixth moment is defined by $E[y_t^6]/(15E[y_t^2]^3)$. We report this moment for HSV models where $\sigma_t^2 = a_0 + a_p H_p(f_t)$ for $p = 2, 4, 6$, and 8 by using the estimates of Table 3, and for the log-normal model by using Kim, Shephard and Chib (1998) estimates (log-n1) and GMM ones (log-n2) from Table 5.

References

Ahn, D.G., R.F. Dittmar and A.R. Gallant (2000), "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, forthcoming.

Ahn, D.G., R.F. Dittmar, A.R. Gallant and B. Gao (2000), "Purebred or Hybrid? Reproducing the Volatility in Term Structure Dynamics," unpublished manuscript, University of North Carolina.

Ait-Sahalia, Y. (2000), "Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, forthcoming.

Andersen, T.G. (1992), "Volatility", Discussion paper, Northwestern University.

Andersen, T.G. (1994), "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance*, 4, 75-102.

Andersen, T.G., L. Benzoni and J. Lund (2001), " An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, forthcoming.

Andersen, T.G., H.J. Chung and B.E. Sorensen (1999), "Efficient Method of Moment Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61-87.

Andersen, T.G. and B.E. Sorensen (1996), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics*, 14, 328-352.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3-30.

Bakshi, G., C. Cao and Z. Chen (1997), "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003-2049.

Barkovec, M. and C. Kluppelberg (2000), "Extremal Behaviour of Diffusion Models in Finance," Working Paper, Munich University of Technology.

Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian OU based Models and some of their uses in Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167-241.

Bates, D.S. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *The Review of Financial Studies*, 9, 69-107.

Black, F. (1976), "Studies in Stock Price Volatility Changes," *Proceedings of the 1976 Business Meeting of the Business and Statistics Section, American Statistical Association*, 177-181.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1987), " A Conditional Heteroskedastic Time Series Model For Speculative Prices and Rates of Returns," *Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T. and R. Engle (1993), "Common Persistence in Conditional Variance," *Econometrica*, 61, 166-187.

Bollerslev, T., R. Engle and D.B. Nelson (1994), "ARCH Models," *Handbook of Econometrics*, Vol IV.

Carrasco, M., L.P. Hansen and X. Chen (1999), "Time Deformation and Dependence," unpublished manuscript, University of Chicago.

Chen, X., L.P. Hansen and M. Carrasco (1999), "Nonlinearity and Temporal Dependence," unpublished manuscript, University of Chicago.

Chen, X., L.P. Hansen and J. Scheinkman (1998), "Shape-Preserving Estimation of Diffusions," unpublished manuscript, University of Chicago.

Chen, X., L.P. Hansen and J. Scheinkman (2000), "Principal Components and the Long Run," unpublished manuscript, London School of Economics and University of Chicago.

Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (1999), "A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation," Working Paper, University of North Carolina.

Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (2001), "Alternative Models for Stock Price Dynamics," Working Paper, University of North Carolina.

Chib, S., F. Nardari and N. Shephard (2000), "Markov Chain Monte Carlo Methods for Generalized Stochastic Volatility Models," *Journal of Econometrics*, forthcoming.

Chihara, T.S. (1978), *An Introduction to Orthogonal Polynomials*, Gordon and Breach Science Publishers.

Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135-156.

Conley, T., L.P. Hansen and W.F. Liu (1997), "Bootstraping the Long Run," *Macroeconomic Dynamics,* 1, 279-311.

Conley, T., L.P. Hansen, E. Luttmer and J. Scheinkman (1997), "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies,* 10, 525-577.

Constantinides, J. (1992), "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies,* 5, 531-552.

Cox, J.C., J. Ingersoll and S. Ross, "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, 385-408.

Darolles, S., J.P. Florens and C. Gouriéroux (1998), "Kernel-Based Nonlinear Canonical Analysis," CREST-INSEE DP 9855.

Darolles, S., J.P. Florens and E. Renault (1998), "Nonlinear Principal Components and Inference on a Conditional Expectation Operator," unpublished manuscript, CREST-INSEE.

Darolles, S., C. Gouriéroux and J. Jasiak (2001), "Compound Autoregressive Models," unpublished manuscript, CREST-INSEE.

Davidian, M. and and R.J. Caroll (1987), "Variance Function Estimation," *Journal of the American Statistical Association*, 82, 1079-1091.

Davydov, D. and V. Linetsky (2000), "Pricing Options on Scalar Diffusions: An Eigenfunction Expansion Approach," unpublished manuscript, Northwestern University.

Diebold, F.X. (1988), *Empirical Modeling of Exchange Rates*, Springer Verlag.

Drost, F.C. and TH.E. Nijman (1993), "Temporal Aggregation of GARCH processes," *Econometrica*, 61, 909-927.

Duffie, D. (1992), *Dynamic Asset Pricing Theory*, Princeton University Press.

Duffie, D. and R. Kan (1996), "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379-406.

Duffie, D., J. Pan and K. Singleton (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1342-1376.

Duffie, D. and K. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Engle, R.F. and G.G.J. Lee (1999), "A Permanent and Transitory Component Model for Stock Return Volatility," in *Cointegration, Causality and Forecasting: a festschrift in Honour of Clive W.J. Granger*, R.F. Engle and H. White eds., Oxford University Press.

Eraker, B., M.S. Johannes and N.G. Polson (1999), "Return Dynamics in Continuous Time with Jumps to Volatility and Returns," Working paper, University of Chicago.

Feller, W. (1951), "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 173-182.

Fiorentini, G. and E. Sentana (1998), "Conditional Means of Time Series Processes and Time Series Processes for Conditional Means," *International Economic Review*, 39, 1101-1118.

Florens, J.P., E. Renault and N. Touzi (1998), "Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes," *Econometric Theory*, 14, 744-769.

Gallant, A.R., C. Hsu and G. Tauchen (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *The Review of Economics and Statistics*, 81, 617-631.

Gallant, A.R., P. Rossi and G. Tauchen (1993), "Nonlinear Dynamic Structures," *Econometrica*, 61, 871-907.

Gallant, A.R. and G. Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091-1120.

Gallant, A.R. and G. Tauchen (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657-681.

Gallant, A.R. and G. Tauchen (1999), "The Relative Efficiency of Methods of Moments Estimators," *Journal of Econometrics*, 92, 149-172.

Garcia, R., R. Luger and E. Renault (2001), "Asymmetric Smiles, Leverage Effects and Structural Parameters," CIRANO DP 2001s-01.

Ghysels, E., A.C. Harvey and E. Renault (1996), "Stochastic Volatility," in: G.S. Maddala and C.R. Rao eds, *in Handbook of Statistics*, vol 14, 119-191.

Ghysels, E., C. Gouriéroux and J. Jasiak (1997), "Stochastic Volatility Duration Models," unpublished manuscript, CREST-Paris.

Giraitis, L., P. Kokoszka and R, Leipus (2000), "Stationary ARCH Models: Dependence structure and Central Limit Theorem," *Econometric Theory*, 16, 3-22.

Gouriéroux C. and J. Jasiak (2001), "Autoregressive Gamma Processes," unpublished manuscript, CREST-INSEE.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, 85-118.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hansen, L.P. (1985), "A Method for Calculating Bounds in the Asymptotic Covariance Matrix of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203-238.

Hansen, L.P, J.C. Heaton and M. Ogaki (1988), "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863-871.

Hansen, L.P. and J. Scheinkman (1995), "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 767-804.

Hansen, L.P., J. Scheinkman and N. Touzi (1998), "Spectral Methods for Identifying Scalar Diffusions," *Journal of Econometrics*, 86, 1-32.

Hansen, L.P. and K.J. Singleton (1996), "Efficient Estimation of Linear Asset-Pricing Models With Moving Average Errors," *Journal of Business and Economic Statistics*, 14, 53-68.

Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247-264.

Heston, S.L. (1993), "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-344.

Ho, M.S., W.R.M. Perraudin and B.E. Sorensen (1996), "A Continuous Time Arbitrage Pricing Model with Stochastic Volatility and Jumps," *Journal of Business, Economics and Statistics*, 14, 31-43.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics,* 12, 371-389.

Jacquier, E., N.G. Polson and P.E. Rossi (1999), "Stochastic Volatility: Univariate and Multivariate Extensionss," CIRANO DP 99s-26.

Johnson, N.L, S. Kotz and N. Balakrishnan (1994), *Continuous Univariate Distributions*, Vol 1, Wiley.

Jones, C.S. (2000), "The Dynamics of Stochastic Volatility," Working paper, Rochester University.

Kessler, M. and M. Sorensen (1999), "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Processes," *Bernoulli*, 5, 299-314.

Kim, S., N. Shephard and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

Leippold, M. and L. Wu (1999), "Asset Pricing Under the Quadratic Class," unpublished manuscript, University of St. Gallen and Fordham University.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.

Meddahi, N. (2001-a), "Moments of Continuous Time Stochastic Volatility Models," in preparation.

Meddahi, N. (2001-b), "Temporal Aggregation of Nonlinear Volatility Models," in preparation.

Meddahi, N. and E. Renault (1996), "Aggregation and Marginalization of GARCH and Stochastic Volatility Models," GREMAQ DP 96.30.433, Univesité de Toulouse and CRDE DP 3597, Université de Montréal.

Meddahi, N. and E. Renault (2000), "Temporal Aggregation of Volatility Models," CIRANO DP 2000s-22.

Melino, A. and S.M. Turnbull (1990), "Pricing Foreing Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239-265.

Nelson, D.B. (1988), "The Time-Series Behavior of Stock Market Volatility and Returns," unpublished Ph.D Thesis, MIT.

Nelson, D.B. (1990-a), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318-334.

Nelson, D.B. (1990-b), "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7-39.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

Nelson, D.B. (1996), "A Note on the Normalized Errors in ARCH and Stochastic Volatility Models," *Econometric Theory*, 12, 113-128.

Nelson, D.B. and D.P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH models," *Econometrica*, 62, 1-41.

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Pitt, M.K. and S.G. Walker (1999), "Construction of Stationary Time Series with Application to Volatility Models," unpublished manuscript, Imperial College, London.

Robinson, P.M. (2001), "The Memory of Stochastic Volatility Models," *Journal of Econometrics*, 101, 195-218.

Robinson, P.M. and P. Zaffaroni (1998), "Nonlinear Time Series with Long Memory: a Model for Stochastic Volatility," *Journal of Statistical Planning and Inference*, 68, 359-371.

Schoutens, W. (2000), *Stochastic Processes and Orthogonal Polynomials*, Lecture Notes in Statistics 146, Springer-Verlag.

Schwert, G.W. (1989), "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44, 1115-1154.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (Eds), *Time Series Models in Econometrics, Finance and Other Fields,* pp. 1-67, London, Chapman and Hall.

Stein, E.M. and J. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach", *Review of Financial Studies*, 4, 727-752.

Szego , G. (1975), *Orthogonal Polynomials*, American Mathematical Society Publishers.

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

Wiggins, J.B. (1987), "Options Values under Stochastic Volatility: Theory and Empirical Estimates", *Journal of Financial Economics*, 19, 351-372.

Wong, E. (1964), "The Construction of a Class of Stationary Markoff Processes," in Belleman, R., (Ed.), Sixteenth Symposium in Applied Mathematics, Stochastic Processes in Mathematical Physics and Engineering, American Mathematical Society, Providence, 264-276.

Appendix

In the sequel, we will use the following notations

$$I_t = \sigma(\varepsilon_\tau, \tau \leq t) \quad \text{and} \quad J_t = \sigma(\varepsilon_\tau, f_\tau, \tau \leq t). \tag{A.1}$$

Proof of Proposition 4.1. We have:
$E[\varepsilon_t] = E[\sigma_{t-1} E[u_t \mid J_{t-1}]] = 0$;
$E[\varepsilon_t^2] = E[\sigma_{t-1}^2 E[u_t^2 \mid J_{t-1}]] = E[\sigma_{t-1}^2] = \sum_{i=0}^p a_i E[E_i(f_t)] = a_0$;
$E[\varepsilon_t^4] = E[\sigma_{t-1}^4 E[u_t^4 \mid J_{t-1}]] = E[u_t^4]E[\sigma_{t-1}^4] = Kurt[u_t] \sum_{i=0}^p \sum_{j=0}^p a_i a_j E[E_i(f_t)E_j(f_t)] = Kurt[u_t] \sum_{i=0}^p \sum_{j=0}^p a_i a_j \delta_{ij} = Kurt[u_t] \sum_{i=0}^p a_i^2$. Therefore, (4.6) is straithforwardly deduced.□

Proof of Proposition 4.2.
$Cov(\varepsilon_t^2, \varepsilon_{t-j}^2) = E[\varepsilon_t^2 \varepsilon_{t-j}^2] - E[\varepsilon_t^2]^2$
$= E[\sigma_{t-1}^2 \sigma_{t-1-j}^2] - a_0^2$ (since the processes u_t and f_t are independent)
$= \sum_{i=0}^p \sum_{k=0}^p a_i a_k E[E_i(f_{t-1})E_k(f_{t-1-j})] - a_0^2 = \sum_{i=0}^p \sum_{k=0}^p a_i a_k \delta_{ik} \lambda_i^j - a_0^2$
$= \sum_{i=1}^p a_i^2 \lambda_i^j$, i.e. (4.7).□

Proof of Proposition 4.3. We have:
$E[\varepsilon_t^2 \varepsilon_{t-j}^2] = E[\sigma_{t-1}^2 u_t^2 \sigma_{t-j-1}^2 u_{t-j}^2] = E[\sigma_{t-1}^2 \sigma_{t-j-1}^2 u_{t-j}^2] = \sum_{i=0}^p a_i E[E[H_i(f_{t-1}) \mid J_{t-j}]\sigma_{t-j-1}^2 u_{t-j}^2] = \sum_{i=0}^p a_i \gamma^{i(j-1)} E[H_i(f_{t-j})\sigma_{t-j-1}^2 u_{t-j}^2] = \sum_{i=0}^p a_i \gamma^{i(j-1)} E[E[H_i(f_{t-j}) u_{t-j}^2 \mid J_{t-j-1}]\sigma_{t-j-1}^2]$.

We will now compute $E[H_i(f_{t-j})u_{t-j}^2 \mid J_{t-j-1}]$. Define z_t by $u_t = \rho v_t + \sqrt{1-\rho^2} z_t$. The process z_t is i.i.d. $\mathcal{N}(0,1)$ and independent from v_t. Therefore $u_t^2 = \rho^2 v_t + (1-\rho^2) z_t^2 + 2\rho\sqrt{1-\rho^2} v_t z_t$. Hence,
$E[H_i(f_{t+1})u_{t+1}^2 \mid J_t] = \rho^2 E[H_i(f_{t+1})v_{t+1}^2 \mid J_t] + (1-\rho^2) E[H_i(f_{t+1}) \mid J_t]$
$= \rho^2 E[H_i(f_{t+1})v_{t+1}^2 \mid J_t] + (1-\rho^2)\gamma^i H_i(f_t)$. The process $H_i(f_{t+1})v_{t+1}^2$ is square-integrable; hence:
$H_i(f_{t+1})v_{t+1}^2 = \sum_{j=0}^\infty w_j H_j(f_{t+1})$ where $w_j = E[H_i(f_{t+1})H_j(f_{t+1})v_{t+1}^2]$. From the transition equation of f_t (4.9), we have: $v_{t+1}^2 = \dfrac{1}{1-\gamma^2}(f_{t+1}^2 + \gamma^2 f_t^2 - 2\gamma f_{t+1} f_t)$. Therefore
$E[v_{t+1}^2 \mid f_{t+1}] = \dfrac{1}{1-\gamma^2}(f_{t+1}^2 + \gamma^2 E[f_t^2 \mid J_{t+1}] - 2\gamma f_{t+1} E[f_t \mid f_{t+1}])$
$= \dfrac{1}{1-\gamma^2}(f_{t+1}^2 + \gamma^2 E[\sqrt{2}H_2(f_t) + 1 \mid J_{t+1}] - 2\gamma f_{t+1} E[H_1(f_t) \mid f_{t+1}])$
$= \dfrac{1}{1-\gamma^2}(f_{t+1}^2 + \gamma^2(\sqrt{2}\gamma^2 H_2(f_{t+1}) + 1) - 2\gamma f_{t+1}\gamma H_1(f_{t+1})$ since f_t is time-reversible. Thus, $E[v_{t+1}^2 \mid f_{t+1}] = f_{t+1}^2(1-\gamma^2) + \gamma^2$. As a consequence,
$w_j = E[H_i(f_{t+1})H_j(f_{t+1})E[v_{t+1}^2 \mid f_{t+1}]] = (1-\gamma^2)E[H_i(f_{t+1})H_j(f_{t+1})f_{t+1}^2] + \gamma^2 E[H_i(f_{t+1})H_j(f_{t+1})]$.
$= (1-\gamma^2)E[H_i(f_{t+1})H_j(f_{t+1})f_{t+1}^2] + \gamma^2\delta_{ij}$. We compute now w_i. We have:
- for $i=0$: $w_j = (1-\gamma^2)E[H_j(f_{t+1})(\sqrt{2}H_2(f_{t+1})+1)] + \gamma^2\delta_{0j} = (1-\gamma^2)(\sqrt{2}\delta_{2j} + \delta_{0j}) + \gamma^2\delta_{0j} = (1-\gamma^2)\sqrt{2}\delta_{2j} + \delta_{0j}$;

• for $i \geq 1$: for $j \neq 0$, by using the recursive formula (2.6) which is true for $i \geq 1$ with convention $H_{-1} = 0$, we have

$E[H_i(f_{t+1})H_j(f_{t+1})f_{t+1}^2]$

$= E[\{\sqrt{i+1}H_{i+1}(f_{t+1}) + \sqrt{i}H_{i-1}(f_{t+1})\}\{\sqrt{j+1}H_{j+1}(f_{t+1}) + \sqrt{j}H_{j-1}(f_{t+1})\}]$

$= \delta_{ij}(1+2i) + \sqrt{(i-1)i}\delta_{(i-2)j} + \sqrt{(i+1)(i+2)}\delta_{(i+2)j}$. Thus,

$w_j = (1 + 2i(1-\gamma^2))\delta_{ij} + (1-\gamma^2)\sqrt{(i-1)i}\delta_{(i-2)j} + (1-\gamma^2)\sqrt{(i+1)(i+2)}\delta_{(i+2)j}$.

For $j = 0$, $w_0 = (1-\gamma^2)E[H_i(f_{t+1})f_{t+1}^2] + \gamma^2\delta_{i0} = (1-\gamma^2)E[H_i(f_{t+1})(\sqrt{2}H_2(f_{t+1}) + 1)] = (1-\gamma^2)\sqrt{2}\delta_{i2}$.

As a conclusion, in all cases, under the convention $H_{-1} = 0$, we have

$H_i(f_{t+1})v_{t+1}^2$

$= (1-\gamma^2)\sqrt{(i-1)i}H_{i-2}(f_{t+1}) + (1+2i(1-\gamma^2))H_i(f_{t+1}) + (1-\gamma^2)\sqrt{(i+1)(i+2)}H_{i+2}(f_{t+1})$.

This implies that

$E[H_i(f_{t-j})v_{t-j}^2 \mid J_{t-j-1}] = (1-\gamma^2)\sqrt{(i-1)i}\gamma^{i-2}H_{i-2}(f_{t-j-1})$

$+(1+2i(1-\gamma^2))\gamma^i H_i(f_{t-j-1}) + (1-\gamma^2)\sqrt{(i+1)(i+2)}\gamma^{i+2}H_{i+2}(f_{t-j-1})$.

Hence, $E[H_i(f_{t-j})u_{t-j}^2 \mid J_{t-j-1}] = (1-\rho^2)\gamma^i H_i(f_{t-j-1}) + \rho^2\left\{(1-\gamma^2)\sqrt{(i-1)i}\gamma^{i-2}H_{i-2}(f_{t-j-1})\right.$ $\left.+(1+2i(1-\gamma^2))\gamma^i H_i(f_{t-j-1}) + (1-\gamma^2)\sqrt{(i+1)(i+2)}\gamma^{i+2}H_{i+2}(f_{t-j-1})\right\}$.

Thus, $E[E[H_i(f_{t-j})u_{t-j}^2 \mid J_{t-j-1}]\sigma_{t-j-1}^2] = (1-\rho^2)\gamma^i a_i + \rho^2\gamma^i\left\{(1-\gamma^2)\sqrt{(i-1)i}\gamma^{-2}a_{i-2}\right.$ $\left.+(1+2i(1-\gamma^2))a_i + (1-\gamma^2)\sqrt{(i+1)(i+2)}\gamma^2 a_{i+2}\right\}$, with the convention $a_i = 0$ if $i < 0$ or $i > p$. Then, (4.10) is deduced.□

Proof of Proposition 4.4. Since $E[\varepsilon_t \mid \varepsilon_\tau, f_\tau, \tau \leq t-1] = 0$, we have $E[\varepsilon_t \mid \varepsilon_\tau, \tau \leq t-1] = 0$, i.e. (4.12).

By (4.11), we get

$$\varepsilon_t^2 = a_0 + \sum_{i=1}^{p} a_i E_i(f_{t-1}) + \eta_t \quad \text{where} \quad E[\eta_t \mid \varepsilon_\tau, f_\tau, \tau \leq t-1] = 0.$$

But each $E_i(f_{t-1})$ is an AR(1) with autoregressive coefficient λ_i. Therefore

$$E_i(f_{t-1}) = \lambda_i E_i(f_{t-2}) + \eta_{i,t-1} \quad \text{where} \quad E[\eta_{i,t-1} \mid \varepsilon_\tau, f_\tau, \tau \leq t-2] = 0.$$

Hence, when p is finite,

$\prod_{i=1}^p(1-\lambda_i L)[\varepsilon_t^2 - a_0] = \sum_{i=1}^p \prod_{j=1, j\neq i}^p (1-\lambda_j L)\eta_{i,t-1} + \prod_{i=1}^p(1-\lambda_i L)\eta_t$.

As a consequence, $E[\prod_{i=1}^p(1-\lambda_i L)[\varepsilon_t^2 - a_0] \mid \varepsilon_\tau, f_\tau, \tau \leq t-p-1] = 0$ and, hence, (4.13).

When the fourth moment of ε_t is finite, the second moment of w_t defined in (refwt) is finite. Besides, (4.13) implies that $E[w_t \mid \varepsilon_\tau, \tau \leq t-p-1] = 0$. Therefore $Cov(w_t, w_{t-h}) = 0$ for all $h > p$. In other words, w_t is a MA(p). Therefore, ε_t^2 is an ARMA(p,p) with autoregressive coefficients λ_1, λ_2,...,λ_p.

Proof of Proposition 4.6. Since ε_t is a m.d.s., we have: $E[\varepsilon_{tm}^{(m)2}] = mE[\varepsilon_t^2] = ma_0$. The no leverage effect assumption implies that:

$E[\varepsilon_{tm}^{(m)4}] = \sum_{i=0}^{m-1} E[\varepsilon_{tm-i}^4] + 6\sum_{0\leq i<j\leq m-1} E[\varepsilon_{tm-i}^2\varepsilon_{tm-j}^2] = mE[\varepsilon_t^4] + 6\sum_{s=1}^{m-1}(m-s)E[\varepsilon_t^2\varepsilon_{t-s}^2]$.

For

$0 < s$: $E[\varepsilon_t^2 \varepsilon_{t-s}^2] = \sum_{0 \leq i,j \leq p} a_i a_j E[E_i(f_{t-1}) E_j(f_{t-s-1})] = \sum_{0 \leq i,j \leq p} a_i a_j \lambda_i^s E[E_i(f_{t-s-1}) E_j(f_{t-s-1})] = \sum_{i=0}^{p} a_i^2 \lambda_i^s$. Thus, $\sum_{s=1}^{m-1} (m-s) E[\varepsilon_t^2 \varepsilon_{t-s}^2] = \sum_{s=1}^{m-1} (m-s) (\sum_{i=0}^{p} a_i^2 \lambda_i^s) = \sum_{i=0}^{p} a_i^2 (\sum_{s=1}^{m-1} (m-s) \lambda_i^s)$

$$= a_0^2 \frac{m(m-1)}{2} + \sum_{i=1}^{p} a_i^2 (\sum_{s=1}^{m-1} (m-s) \lambda_i^s) = a_0^2 \frac{m(m-1)}{2} + \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{(1-\lambda_i)^2} (\lambda_i^m - m\lambda_i + m - 1).$$

Hence, $E[\varepsilon_{tm}^{(m)4}] = mE[\varepsilon_t^4] + 3a_0^2 m(m-1) + 6 \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{(1-\lambda_i)^2} (\lambda_i^m - m\lambda_i + m - 1)$

Therefore,

$kurt(\varepsilon_{tm}^{(m)}) = \frac{E[\varepsilon_{tm}^{(m)4}]}{E[\varepsilon_{tm}^{(m)2}]^2} = \frac{1}{m} \frac{E[\varepsilon_t^4]}{a_0^2} + 3\frac{(m-1)}{m} + \frac{6}{a_0^2} \sum_{i=1}^{p} a_i^2 \frac{\lambda_i}{(1-\lambda_i)^2} \frac{\lambda_i^m - m\lambda_i + m - 1}{m^2}$. As a consequence, we get (4.17).

When m $\longrightarrow +\infty$, we have $\frac{\lambda_i^m - m\lambda_i + m - 1}{m^2} = \frac{1}{m} + o(\frac{1}{m})$. Hence, we get (4.18).□

Lemma A1. Sixth moment *Let* $y_t = \sigma_{t-1} u_t$ *such that* $\sigma_{t-1} = a_0 + a_p H_p(f_t)$. *Then:*

$$\frac{E[y_t^6]}{15E[y_t^2]^3} = \frac{a_0^3 + 3a_0 a_p^2 + a_p^3 [\frac{p!}{(p/2)!}]^3}{a_0^3}.$$

For the log-normal model,

$$\frac{E[y_t^6]}{15E[y_t^2]^3} = \exp(3\sigma^2).$$

Proof of Lemma A1. We have

$E[y_t^6] = E[\sigma_{t-1}^6 u_t^6] = E[\sigma_{t-1}^6] E[u_t^6] = E[\sigma_{t-1}^6] 15$.

HSV model: $\sigma_{t-1}^2 = a_0 + a_p H_p(f_{t-1})$. Thus,

$E[\sigma_{t-1}^6] = a_0^3 + 3a_0^2 a_p E[H_p(f_{t-1})] + 3a_0 a_p^2 E[H_p(f_{t-1})^2] + a_p^3 E[H_p(f_{t-1})^3]$

$= a_0^3 + 3a_0 a_p^2 + a_p^3 E[H_p(f_{t-1})^3]$.

When p is even, $E[H_p(f_{t-1})^3] = [\frac{p!}{(p/2)!}]^3$. Therefore $E[y_t^6] = 15[\frac{p!}{(p/2)!}]^3$.

Log-normal model: $E[\sigma_t^6] = \exp(3\mu + \frac{9\sigma^2}{2})$ while $E[\sigma_t^2] = \exp(\mu + \frac{\sigma^2}{2})$.□