Temporal Aggregation of Volatility Models*

Nour Meddahi † ‡ and **Eric Renault** §

January, 2002

Abstract

In this paper, we consider temporal aggregation of volatility models. We introduce semiparametric volatility models, termed square-root stochastic autoregressive volatility (SR-SARV), which are characterized by autoregressive dynamics of the stochastic variance. Our class encompasses the usual GARCH models and various asymmetric GARCH models. Moreover, our stochastic volatility models are characterized by multiperiod conditional moment restrictions in terms of observables. The SR-SARV class is a natural extension of the class of weak GARCH models. This extension has four advantages: i) we do not assume that fourth moments are finite; ii) we allow for asymmetries (skewness, leverage effect) that are excluded from weak GARCH models; iii) we derive conditional moment restrictions; iv) our framework allows us to study temporal aggregation of IGARCH models.

Keywords: GARCH, stochastic volatility, state-space, SR-SARV, temporal aggregation, asset returns, diffusion processes.

JEL Classification: C22, C43, C50, C51.

---

*This is a revision of a part of Meddahi and Renault (1996), "Aggregations and Marginalization of GARCH and Stochastic Volatility Models." Some other results of this manuscript are now included in two companion papers, Meddahi (2002) and Meddahi and Renault (2002b) entitled "Temporal and Cross-sectional Aggregations of Volatility in Mean Models" and "Conditioning Information in Volatility Models" respectively. The authors thank Torben Andersen, Bryan Campbell, Marine Carrasco, Ramdan Dridi, Feike Drost, Jean-Marie Dufour, Ola Elarian, Rob Engle, Jean-Pierre Florens, René Garcia, Ramazan Gençay, Christian Gouriéroux, Stéphane Gregoir, Joanna Jasiak, Tom McCurdy, Theo Nijman, Enrique Sentana, Neil Shephard, Jean-Michel Zakoian, a referee and the Editor, Bas Werker, and the participants of the Econometric society meetings at Istanbul (1996) and Pasadena (1997), the Fourth Workshop on Financial Modeling and Econometric Analysis, Tilburg, December 1996, for their helpful comments. They also acknowledge fruitful discussions during seminars at CEMFI, CORE, CREST, North Carolina (Triangle seminar), Montréal, Oxford. The authors are the only responsible for the remaining errors. The first author acknowledges FCAR, MITACS and IFM2 for financial support.

†Corresponding author.

‡Département de sciences économiques, CRDE, CIRANO, Université de Montréal and CEPR. Address: C.P. 6128, succursale Centre-ville, Montréal (Québec), H3C 3J7, Canada. E-mail: nour.meddahi@umontreal.ca.

§Département de sciences économiques, CRDE, CIRANO, Université de Montréal and CREST-Insee. E-mail: eric.renault@umontreal.ca.

Résumé

Dans cet article, nous considérons l'agrégation temporelle des modèles de volatilité. Nous introduisons une classe de modèles de volatilité semi-paramétrique dénommée SR-SARV et caractérisée par une variance stochastique ayant une dynamique autorégressive. Notre classe contient les modèles GARCH usuels ainsi que plusieurs variantes asymétriques. De plus, nos modèles à volatilité stochastique sont caractérisés par des moments conditionnels observables et à plusieurs horizons. La classe des modèles SR-SARV est une généralisation naturelle des modèles GARCH faibles. Notre extension présente quatre avantages : i) nous ne supposons pas que le moment d'ordre quatre est fini; ii) nous permettons des asymétries (de type skewness et effet de levier) qui sont exclues par les modèles GARCH faibles; iii) nous dérivons des restrictions sur des moments conditionnels utiles pour l'inférence non linéaire; iv) notre cadre de travail nous permet d'étudier l'agrégation temporelle des modèles IGARCH.

Mot clés : GARCH, volatilité stochastique, éspace-état, SR-SARV, agrégation temporelle, rendements d'actifs, processus de diffusion.

1 Introduction

Prices of financial assets, such as stocks, bonds, and currencies, are available at many frequencies from intradaily to annual. When modeling volatility of the returns on such assets, issues related to the effect of temporal aggregation and the choice of the observation frequency arise naturally. Basically, two modeling strategies can be considered: the model can be specified for the observable frequency by implicitly assuming that it is the correct model for this frequency (an assumption which is testable), or the model can be specified at a high frequency, say continuous time, where the implications for a lower frequency are subsequently derived. Typically, models from the ARCH[1] family belong to the first class, while models in Drost and Nijman (1993) and Hansen and Scheinkman (1995) stem from the second strategy.[2] In general, we say that a model is closed under temporal aggregation if the model keeps the same structure, with possibly different parameter values, for any data frequency.

Drost and Nijman (1993) consider temporal aggregation of volatility models. They show that the usual GARCH models of Bollerslev (1986) are not closed under temporal aggregation. The main reason is that such models imply that the squared residual process is a semi-strong ARMA (i.e., an ARMA process for which the innovations form a martingale difference sequence), which is not closed under temporal aggregation. The ARMA literature teaches us that weak ARMA models, where the innovations are serially uncorrelated (weak white noise), are closed under temporal aggregation. Therefore, Drost and Nijman (1993) introduce the class of weak GARCH models which are characterized by a weak ARMA structure of the squared innovations and show that this class is closed under temporal aggregation.

However, weak GARCH models have several limitations. First, since weak GARCH models are characterized by a weak ARMA structure of the squared innovations, Drost and Nijman (1993) assume that the fourth moment of the innovations is finite. This seems to be empirically violated by several financial time series, especially when observed at a high frequency.[3] Secondly, in the weak GARCH setting, linear projections instead of conditional expectations are considered. This is an important drawback if the conditional variance is considered to be the relevant measure of risk. It is also a limitation for statistical purposes since asymptotic properties of inference procedures like QMLE are usually based on conditional moments. Indeed, in a Monte Carlo study we show clearly that QMLE

[1]ARCH models were introduced by Engle (1982) and extended by Bollerslev (1986) to GARCH. For a review of the ARCH literature, see, e.g., Bollerslev, Engle and Nelson (1994).

[2]Hansen and Scheinkman (1995) consider continuous time stochastic differential equations and derive moment restrictions for a given data frequency. Duffie and Glynn (1997) extend this to observations sampled at random times. Nelson bridges the gap between discrete time ARCH models and continuous time models by taking an approximating, filtering, or smoothing approach: Nelson (1990, 1992, 1996), Nelson and Foster (1994).

[3]Recently, Davis and Mikosch (1998) show that for an ARCH(1) of Engle (1982) with infinite fourth moment, the standard estimator of the correlation between ε_t^2 and it lags converges to a random variable.

is not consistent for temporally aggregated GARCH models.[4] Finally, for temporal aggregation of flow variables (e.g., returns), Drost and Nijman (1993) have to exclude asymmetries such as skewed innovations and leverage effects (Black, 1976, Nelson, 1991).

In the present paper, we propose a new class of volatility models which is closed under temporal aggregation and which avoids the limitations of the weak GARCH class. We follow the main idea of Drost and Nijman (1993) by considering an ARMA structure for the squared innovations. However, our approach is based on linear state-space modeling, that is, according to financial terminology, stochastic volatility (SV) modeling.[5] We consider the Square-Root Stochastic Autoregressive Volatility (SR-SARV) models which are characterized by AR dynamics for the conditional variance process. Special ARCH-type examples of SR-SARV include ARCH of Engle (1982), GARCH of Bollerslev (1986), and the asymmetric GARCH models of Glosten, Jagannathan and Runkle (1989), Engle and Ng (1993). Moreover, even if the variance is stochastic, we can still base inference on some conditional moment restrictions which involve only observables. When the fourth moment of the innovations is finite, these moment restrictions imply that the squared innovations process is an ARMA process. Besides, we prove that any symmetric SR-SARV model with finite fourth moment is weak GARCH. Hence, weak GARCH are SV processes rather than standard GARCH and our results generalize those of Drost and Nijman (1993) and of Drost and Werker (1996). Finally, our framework allows us to study temporal aggregation of Integrated GARCH (IGARCH).

Several models in the literature share the property of autoregression of the variance: GARCH models, structural GARCH models of Harvey, Ruiz, and Sentana (1992), SV models of Barndorff-Nielsen and Shephard (2001), and the SR-SARV models of Andersen (1994).[6] Our class of models is closely related to the Andersen (1994) SR-SARV and we adopt his terminology. However, while Andersen (1994) specifies a parametric setting, we take a semiparametric point of view avoiding parametric assumptions on the probability distributions.[7]

Since Akaike (1974), it is well-known that there is an equivalence between **weak** ARMA and **weak** state-space models. In particular, given an ARMA process with finite variance, we can find a state-space model, generally not unique, such that the restrictions implied on the observables are the same for both models. In Meddahi and Renault (2002a), we extend this result to semi-strong models. However, there is no equivalence between semi-strong ARMA models and semi-strong state-space models. More precisely, we show that semi-strong ARMA models admit a particular semi-strong

[4]This is an important difference with Drost and Nijman (1992) who report simulation results which suggest that the QMLE of temporally aggregated GARCH is consistent or has a very small bias. Our results are different from theirs because we aggregate over a much longer period and we take empirically more relevant low frequency parameters.

[5]See Ghysels, Harvey and Renault (1996) and Shephard (1996) for a review.

[6]Several multivariate models in factor GARCH literature also share this property: Diebold and Nerlove (1989), Engle, Ng and Rothschild (1990), King, Sentana and Wadhwani (1994).

[7]Besides specific Gaussian models, distributional assumptions are generally not closed under temporal aggregation.

state-space representation, but the latter amounts only to some multiperiod conditional moment restrictions which are less restrictive than the moment conditions implied by a semi-strong ARMA model. For instance, consider an ARMA(1,1) process z_t. We show that z_t admits a semi-strong state representation if and only if there exist ω and γ such that $E[z_t - \omega - \gamma z_{t-1} \mid z_\tau, \tau \leq t-2] = 0$.[8] It turns out that these weakened multiperiod conditional moment restrictions are closed under temporal aggregation. In other words, this **particular state-space representation** of a semi-strong ARMA(1,1) model is robust to temporal aggregation while semi-strong ARMA models in general are not. Multiperiod conditional moment restrictions are very useful for inference and are introduced in Hansen (1985); see Hansen and Singleton (1996) for a review. When the variance of z_t is finite, these restrictions imply that z_t is a weak and not necessarily a semi-strong ARMA: it is in between.

Starting from the SR-SARV(1) model characterized by AR(1) dynamics of the conditional variance process, we propose several extensions. In the spirit of GARCH (p,p) modeling, we introduce the SR-SARV(p) model: the variance process is the sum of the components (marginalization) of a positive multivariate VAR(1) of dimension p. GARCH(p,p) models are special examples of SR-SARV(p). When fourth moments are finite, the squared innovations process is an ARMA(p,p). In continuous time, all this leads up to consider a SV model in which the variance is a marginalization of a vector of dimension p, that is a multi-factor model for the variance (e.g., Heston, 1993; Duffie and Kan, 1996).[9] The exact discretization of such models is SR-SARV(p), hence the process of squared innovations fulfills the above mentioned multiperiod moment restrictions.

Finally, we consider temporal aggregation of IGARCH models. In this case, we consider the ISR-SARV class where we relax the assumption of integrability of the variance process while maintaining the stationarity assumption. We show that this class is closed under temporal aggregation.

The rest of the paper is organized as follows. We introduce in Section 2 the SR-SARV(p) model in discrete and continuous time. We start by showing that exact discretization of continuous time SR-SARV models are discrete time SR-SARV models. Then, we show that the discrete time SR-SARV(p) model is closed under temporal aggregation. After that, we derive multiperiod conditional moment restrictions fulfilled by the squared innovations process. We also characterize the relations between SR-SARV, semi-strong GARCH, weak GARCH, and ARMA representations for squared innovations. Section 3 focuses more specifically on the SR-SARV(1) model. In particular, we characterize the SR-SARV(1) models that are semi-strong GARCH(1,1) and we discuss asymmetry issues (leverage effect and skewness). We also consider temporal aggregation of IGARCH models and show by Monte Carlo that the Gaussian QMLE is not consistent for temporally aggregated GARCH models. We

[8] This restriction is less restrictive than saying that the innovation process of z_t is a martingale difference sequence.

[9] Heston (1993) considers a SV model where the volatility is a Constant Elasticity of Variance (CEV) process introduced by Cox (1975). They are characterized by a linear drift and popular in finance for their nonnegativity.

conclude in the last section while all the proofs of the results are provided in the Appendix.

2 SR-SARV(p) model

In this section we introduce the Square-Root Stochastic Autoregressive Volatility model of order p (SR-SARV(p)) in discrete and continuous time. This model involves a state-space representation of order p for the squared (innovation) process. We prove that the continuous time and discrete time models are consistent by showing that the exact discretization of a continuous time SR-SARV(p) model is a discrete time SR-SARV(p) model. This result suggests that the discrete time model is closed under temporal aggregation and, hence, we prove it. Then we derive observable restrictions of our model. These multiperiod conditional moment restrictions involve p lags and hold for the squared process. When the fourth moment of the process is finite, it ensures an ARMA structure for the squared innovation process which is intermediate between weak and semi-strong. Finally we recall the definitions of semi-strong GARCH and weak GARCH and their links with the ARMA structure of the squared innovations.

2.1 The model

2.1.1 Discrete time SR-SARV(p) model

Definition 2.1. Discrete time SR-SARV(p) model: *A stationary square-integrable process* $\{\varepsilon_t, t \in \mathbf{Z}\}$ *is called a* **SR-SARV(p) process** *with respect to a filtration* $J_t, t \in \mathbf{Z}$*, if:*
i) ε_t *is a martingale difference sequence w.r.t.* J_{t-1}*, that is* $E[\varepsilon_t \mid J_{t-1}] = 0$*;*
ii) the conditional variance process f_t *of* ε_{t+1} *given* J_t *is a marginalization of a stationary* J_t*-adapted VAR(1) of dimension p:*

$$f_t \equiv Var[\varepsilon_{t+1} \mid J_t] = e' F_t, \tag{2.1}$$

$$F_t = \Omega + \Gamma F_{t-1} + V_t, \quad \text{with } E[V_t \mid J_{t-1}] = 0, \tag{2.2}$$

where $e \in I\!R^p$*,* $\Omega \in I\!R^p$ *and the eigenvalues of* Γ *have modulus smaller than one.*

Observe that the SR-SARV process is defined for a given information set J_t. The information J_t contains at least the minimal natural filtration associated to the process ε_t and denoted I_t, that is:

$$I_t = \sigma(\varepsilon_\tau, \tau \leq t). \tag{2.3}$$

In particular, J_t may contain macroeconomic variables, information about other assets and markets, the volume of transactions, the spread, the order book and so on.[10] Indeed, we assume that the econometrician observes I_t but not necessarily J_t, even if the economic agent may do. Thus, when

[10]Note also that $\sigma(\varepsilon_\tau, f_\tau, \tau \leq t) \subset J_t$ since the process f_t is adapted w.r.t. J_t.

$I_t \neq J_t$, the model is a Stochastic Volatility (SV) model since the conditional variance process is a function of possibly latent variables.

The process of interest ε_t is assumed to be a martingale difference sequence w.r.t the large information J_t and therefore w.r.t. I_t. Typically, ε_t could be the log-return of a given asset with a price at time t denoted by S_t: $\varepsilon_t = Log(S_t/S_{t-1})$. This assumption of m.d.s. is widespread in financial economics and related to the notion of informational efficiency of asset markets. However, we do not preclude predictable log-returns; in this case, our ε_t should be interpreted as the innovation process (see Meddahi and Renault, 1996).

The model can be interpreted through a state-space representation of ε_t^2 since

$$\varepsilon_t^2 = e'F_{t-1} + (\varepsilon_t^2 - E[\varepsilon_t^2 \mid J_{t-1}]). \tag{2.4}$$

Here, (2.4) is the measurement equation while (2.2) is the transition equation. This state-space representation is convenient for both temporal aggregation and inference purposes. It is implicitly assumed that the process $e'F_t$ is non negative. A sufficient but not necessary condition for this is that all the components of e and F_t are nonnegative.

Note that in contrast to the weak GARCH case, we do not assume that the fourth moment of ε_t is finite. We only assume the integrability of the conditional variance process and, hence, the finiteness of the second moment. Moreover, leverage effect, that is a nonzero correlation between ε_t and f_t, is not precluded.

This model is related to Andersen's (1994) SR-SARV and indeed we adopt his terminology. However, Andersen (1994) considers a fully parametric model[11] by specifying the complete distribution of the process $(\varepsilon_t, F_t')'$ and precludes any leverage effect. The temporal aggregation requirement prevents us from completely specifying the probability distributions. Distributional assumptions or homo-conditional moments restrictions (homoskewness, homokurtosis) are generally not closed under temporal aggregation (see below). Actually, we do not maintain any assumption about the leverage effect or about the high order moments of ε_t (third, fourth...) and V_t.[12] To summarize, we consider a semiparametric SV model.

The SR-SARV class of models nests some well known examples. We list below some of them.

[11] Andersen (1994) considers the general class of SARV models where a function of the conditional variance process is a polynomial of an AR(1) Markov process. When this function is the square-root, Andersen (1994) calls it Square-Root (SR) SARV while he terms Exponential SARV when this function is the exponential one, corresponding to the Taylor (1986) and Harvey, Ruiz and Shephard (1994) lognormal SV model.

[12] Andersen (1994) considers only one factor, so his model is related to a SR-SARV(1). However, he defined the volatility process as a function of a polynomial, say of degree p, of an AR(1) state-variable K_t. Thus, it is a marginalization of the vector $(K_t, K_t^2, ..., K_t^p)'$ which is indeed a VAR(1) of size p. In other words, Andersen (1994) considers implicitly a particular SARV(p) model.

• **Example 1: GARCH(1,1).** This model introduced by Bollerslev (1986) and extended by Engle and Ng (1993) is given by

$$f_t = \omega_1 + \alpha(\varepsilon_t + \lambda)^2 + \beta f_{t-1} \quad \text{and} \quad J_t = I_t. \tag{2.5}$$

Bollerslev's (1986) model corresponds to $\lambda = 0$. It was extended by Engle and Ng (1993) in order to capture the leverage effect. Obviously, we have

$$f_t = \omega + \gamma f_{t-1} + v_t \text{ where } \omega = \omega_1 + \alpha\lambda^2,\ \gamma = \alpha + \beta,\ v_t = \alpha(\varepsilon_t^2 - f_t) + 2\alpha\lambda\varepsilon_t \text{ and } E[v_t \mid J_{t-1}] = 0. \tag{2.6}$$

Since $J_t = I_t$, this model is an ARCH-type instead of an SV one.[13] Moreover, ε_t is a weak GARCH(1,1) when it is assumed in addition that fourth moment of ε_t is finite and the leverage effect is ruled out ($\lambda = 0$).

• **Example 2: Quadratic process.** Consider $\{z_t\}$ an Gaussian AR(1) process given by

$$z_t = \lambda z_{t-1} + e_t \quad \text{where } \mid \lambda \mid < 1 \text{ and } e_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2),$$

and define f_t by $f_t \equiv z_t^2$. Then, we have

$$f_t = \omega + \gamma f_{t-1} + v_t \text{ with } \omega = \sigma^2,\ \gamma = \lambda^2,\ v_t = 2\lambda z_{t-1} e_t,\ E[e_t \mid J_{t-1}] = 0 \text{ where } J_t = \sigma(\varepsilon_\tau, z_\tau, \tau \leq t).$$

• **Example 3: Positive AR(1) process.** Assume that $\{f_t\}$ is defined by

$$f_t = \omega + \gamma f_{t-1} + v_t \quad \text{where} \quad v_t \text{ is } i.i.d.\ D(0, \sigma^2) \tag{2.7}$$

with a lower bounded support for v_t, that is, there exists a real a such that $v_t \geq a$ almost surely. Then the process f_t is nonnegative when $\omega + a \geq 0$.[14] Such processes are considered by Barndorff-Nielsen, Jensen and Sorensen (1998). In this case, ε_t is a SR-SARV(1) w.r.t. to the information $J_t = \sigma(\varepsilon_\tau, f_\tau, \tau \leq t)$.

We now consider continuous time stochastic volatility models which are popular in finance due to their positivity. The exact discretization of these processes is a discrete time SR-SARV(p) process.

2.1.2 Continuous time SR-SARV(p) model

Definition 2.2. Continuous time SR-SARV(p) model: *A continuous time stationary process $\{y_t, t \in I\!R\}$ is called a* **SR-SARV(p)** *process with respect to a filtration $J_t, t \in I\!R$, if and only if there exists a p-variate process F_t^c such that y_t is the stationary solution of*

$$d\begin{pmatrix} y_t \\ F_t^c \end{pmatrix} = \begin{pmatrix} 0 \\ K(\Theta - F_t^c) \end{pmatrix} dt + R_t\, dW_t, \tag{2.8}$$

[13] We give additional ARCH-type examples in the following section.

[14] The reason is that $f_t = \sum_{i=0}^{\infty} \gamma^i (\omega + v_t)$.

where W_t is a $(p+1)$-variate standard Wiener process adapted w.r.t J_t, K is a $p \times p$ positive stable matrix[15] *and R_t is a $(p+1) \times (p+1)$ lower triangular matrix, such that the coefficient $r_{11,t}$ is the square-root of $r_{11,t}^2 \equiv \sigma_t^2 = e' F_t^c$, with $e \in I\!R_+^p$.*

The instantaneous conditional variance of (y_t, F_t^c) given J_t is $R_t R_t'$. The matrix R_t is lower triangular,[16] therefore the conditional variance of y_t given J_t is $r_{11,t}^2$. In other words, we follow the main idea of the discrete time SR-SARV(p) model, that is the conditional variance is a marginalization of a p dimensional VAR(1) positive process F_t^c. Note that as for the discrete time model, we have a semiparametric SV model since we do not define completely the matrix R_t. In particular, we allow for a leverage effect. Of course, the matrix R_t has to fulfill conditions ensuring existence and uniqueness of a stationary solution of the SDE (2.8). For instance, this is consistent with the Duffie and Kan (1996) setting of a multivariate square-root process such that each coefficient of $R_t R_t'$ is of the form $(1, F_t^{c'})\tilde{e}$ with $\tilde{e} \in I\!R^{p+1}$.[17] Again, we do not ruled out the leverage effect, i.e. the first component the Brownian process W_t may be correlated with the other components. Finally, note that the framework allows for models where there are additional factors in R_t.

The continuous time SR-SARV class nests several well known models:

Example 4: CEV and GARCH diffusion models. Consider the one factor model where σ_t^2 is given by:[18]

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \delta(\sigma_t^2)^\lambda dW_{2,t}, \text{ with } 1/2 \le \lambda \le 1. \tag{2.9}$$

This is the CEV process introduced by Cox (1975). When $\lambda = 1/2$, the CEV process becomes the square-root model considered by Heston (1993) while the case $\lambda = 1$ corresponds to the GARCH diffusion model of Nelson (1990). Notice that the Brownian motion processes $W_{1,t}$ and $W_{2,t}$ may be perfectly correlated, which leads to a GARCH-type model.[19]

Example 5: Quadratic model. As for Example 2, assume that σ_t^2 is the square of a driftless Ornstein-Uhlenbeck process z_t, i.e. $\sigma_t^2 = z_t^2$ where

$$dz_t = -k_1 z_t dt + \sigma dW_{2,t}.$$

Then, by using Ito's Lemma, we get

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + r_{22,t} dW_{2,t} \text{ where } k = 2k_1, \theta = \frac{\sigma^2}{k} \text{ and } r_{22,t} = 2\sigma z_t. \tag{2.10}$$

[15] This means that the eigenvalues of K have positive real parts (see Horn and Johnson, 1994, Chapter 2). Indeed, a usual assumption, see e.g. Bergstrom (1990), page 53, is that the eigenvalues of K are distinct. Therefore K is diagonalisable, i.e. there exists a matrix H such that $HKH^{-1} = Diag(\lambda_1, ..., \lambda_p) \equiv \Lambda$. As a consequence, for $u > 0$, $He^{-uK}H^{-1} = e^{-u\Lambda} = Diag(e^{-u\lambda_1}, ..., e^{-u\lambda_p})$ with $e^Z = \sum_{i=0}^{\infty} \frac{Z^i}{i!}$. The positivity of the real parts of the eigenvalues K ensures the existence of e^{-uK} $\forall u > 0$.

[16] This Gramm-Schmidt normalization rule is standard and can be maintained without loss of generality.

[17] See Duffie and Kan (1996) for sufficient conditions of existence of a stationary solution of (2.8) in this case.

[18] Since there is only one factor, we change the notations by taking $F_t^c \equiv \sigma_t^2$, $W_t = (W_{1,t}, W_{2,t})'$.

[19] This model is considered by Heston and Nandi (2000) with $\lambda = 1/2$.

Example 6: Positive Ornstein-Uhlenbeck process. Recently, Barndorff-Nielsen and Shephard (2001) have considered a new class of continuous time stochastic volatility models, termed positive OU processes, where σ_t^2 is given by

$$\sigma_t^2 = e^{-\lambda t}\sigma_0^2 + \int_0^t e^{-\lambda(t-s)} dz(\lambda s) \tag{2.11}$$

where $\sigma_0^2 = \int_{-\infty}^0 e^s dz(s)$, $\{z_t, t \in I\!R\}$ is an integrable homogenous Lévy process with positive increments and λ a positive number. Note the similarity with Example 3. Actually, positive Lévy processes and CEV processes are two ways to get positive autoregressive processes that are appropriate for SR-SARV.

The consistency between the notions of continuous time SR-SARV(p) and discrete time SR-SARV(p) is ensured by the following result:

Proposition 2.1 Exact discretization of continuous time SR-SARV(p)

Let $\{y_t, t \in I\!R\}$ be a continuous time SR-SARV(p) process with a corresponding factor process $\{F_t^c, t \in I\!R\}$. Assume that the second moment of F_t^c is finite. Then, for any sampling interval h, the associated discrete time process $\varepsilon_{th}^{(h)} = y_{th} - y_{(t-1)h}$, $t \in \mathbf{Z}$, is a SR-SARV(p) process w.r.t. $J_{th}^{(h)}$, $J_{th}^{(h)} = \sigma(\varepsilon_{\tau h}^{(h)}, F_{\tau h}^c, \tau \leq t, \tau \in \mathbf{Z})$. The corresponding conditional variance process $f_{(t-1)h}^{(h)} \equiv Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}]$ is given by $f_{th}^{(h)} = e' F_{th}^{(h)}$ with $F_{th}^{(h)} = A^{(h)} F_{th}^c + B^{(h)}$, where $A^{(h)} = K^{-1}(Id - e^{-Kh})$ and $B^{(h)} = (hId - A^{(h)})\Theta$.

Proposition 2.1 maintains the assumption of square-integrability of F_t^c to ensure that its exact discretization is a VAR(1) process. This is why CEV processes with $\lambda > 1$ are excluded. When $\lambda < 1$, the integrability assumption is fulfilled while the additional restriction $\sigma^2/2k < 1$ is needed for the GARCH diffusion model of Nelson (1990) ($\lambda = 1$).

The previous result suggests that the SR-SARV(p) class is closed under temporal aggregation. This is the main focus of interest of the paper and the purpose of the next subsection.

2.2 Temporal aggregation of SR-SARV(p) models

Let us consider a process $\{\varepsilon_t, t \in \mathbf{Z}\}$ and the aggregated process $\{\varepsilon_{tm}^{(m)}, t \in \mathbf{Z}\}$ defined by

$$\varepsilon_{tm}^{(m)} = \sum_{k=0}^{m-1} a_k \varepsilon_{tm-k}, \tag{2.12}$$

with $m \in \mathbf{N}^*$, $a = (a_0, a_1, .., a_{m-1})' \in I\!R^m$. Temporal aggregation of stock variables observed at the dates m, 2m, 3m,.., Tm, corresponds to $a = (1, 0, 0..., 0)'$, while for flow variables $a = (1, .., 1)'$. This latter case is particularly suitable for log-returns.

Proposition 2.2 Temporal aggregation of SR-SARV(p) models

Let ε_t be a SR-SARV(p) process w.r.t. an increasing filtration J_t and a conditional variance process $f_t = e'F_t$. For a given integer m, the process $\varepsilon_{tm}^{(m)}$ defined by (2.12) is a SR-SARV(p) w.r.t. $J_{tm}^{(m)} = \sigma(\varepsilon_{\tau m}^{(m)}, F_{\tau m}, \tau \leq t)$. More precisely, we have:

$$f_{tm-m}^{(m)} \equiv Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}] = e'(A^{(m)} F_{tm-m} + B^{(m)}), \tag{2.13}$$

$$\text{where } A^{(m)} = \sum_{k=0}^{m-1} a_k^2 \Gamma^{m-k-1} \text{ and } B^{(m)} = (\sum_{k=0}^{m-1} a_k^2 (\sum_{i=0}^{m-k-2} \Gamma^i))\Omega. \tag{2.14}$$

Assume that $e'A^{(m)} \neq 0$, then $f_{tm}^{(m)} = e^{(m)\prime} F_{tm}^{(m)}$ with

$$e^{(m)} = A^{(m)\prime} e, \quad F_{tm}^{(m)} = F_{tm} + e^{(m)} (e^{(m)\prime} e^{(m)})^{-1} e' B^{(m)}. \tag{2.15}$$

As well, $F_{tm}^{(m)}$ is a VAR(1) process with an autoregressive matrix $\Gamma^{(m)}$ given by

$$\Gamma^{(m)} = \Gamma^m. \tag{2.16}$$

In other words, the assumption that the conditional variance is a marginalization of a VAR(1) process of dimension p is robust to temporal aggregation. The intuition of this result is the following. Consider the initial process ε_t with the information J_t at high frequency and define the process at low frequency $\varepsilon_{tm}^{(m)}$ by (2.12). Define $f_{tm}^{(m)}$ as the conditional variance of $\varepsilon_{(t+1)m}^{(m)}$ given the information at **high frequency** J_{tm} (first part of 2.13). This information is generally not observable either by the agent or by the econometrician and thus the variance is stochastic. But by something like a Markovian property,[20] the conditional variance $f_{tm}^{(m)}$ is a function of F_{tm}. By the linearity of the model, this function is indeed affine (second part of (2.13)). Define the information at low frequency by $J_{tm}^{(m)} \equiv \sigma(\varepsilon_{\tau m}^{(m)}, F_{\tau m}^{(m)}, \tau \leq t)$. Then $\varepsilon_{tm}^{(m)}$ is still a m.d.s. with respect to $J_{tm}^{(m)}$ since $E[\varepsilon_{tm+m}^{(m)} \mid J_{tm}] = 0$ and $J_{tm}^{(m)} \subset J_{tm}$. Of course, by definition, the conditional variance $f_{tm}^{(m)}$ of $\varepsilon_{tm+m}^{(m)}$ given $J_{tm}^{(m)}$ is positive. Then assuming that $e'A^{(m)} \neq 0$,[21] we can rewrite this conditional variance as a marginalization of a new state variable $F_{tm}^{(m)}$. The latter is a VAR(1) since it is the sum of a VAR(1) and a constant. Thus, $\varepsilon_{tm}^{(m)}$ is a SR-SARV(p) w.r.t. $J_{tm}^{(m)}$. Finally, the autoregressive parameter of the VAR(1) F_{tm}^m is equal to the autoregressive parameter of the high frequency vector F_t to the power m (2.16). It means that the persistence increases exponentially with the frequency. Conversely, conditional heteroskedasticity vanishes when the frequency is low. This corresponds to a well-documented empirical evidence and was pointed out by Diebold (1988), Drost and Nijman (1993) and Drost and Werker (1996).

[20] If one has in mind an underlying continuous time representation like (2.8), the low frequency process $(y_{tm}^{(m)}, F_{tm}^{(m)})$ is Markovian. More generally, our setting ensures that the conditional variance $f_{tm}^{(m)}$ depends on past information only through F_{tm}.

[21] The equality $e'A^{(m)} = 0$ would mean that the process $\varepsilon_{tm}^{(m)}$ is homoskedastic which is a degenerate SR-SARV model. In other words, temporal aggregation would cancel the volatility effect.

Temporal aggregation of conditionally heteroskedastic models was already considered by Drost and Nijman (1993) and lead to the weak GARCH paradigm while the links between continuous time SV models and weak GARCH were put forward by Drost and Werker (1996). In the next subsection, we will recap these results and characterize the links between weak GARCH and SR-SARV models.

2.3 Observable restrictions

2.3.1 Multiperiod conditional moment restrictions

The SR-SARV is defined w.r.t. an increasing filtration J_t, which may not be observable by the economic agent or the econometrician. However, following Meddahi and Renault (2002a), the state-space representation of ε_t^2 allows us to derive conditional moments fulfilled by the observable process ε_t given the minimal information $I_t = \sigma(\varepsilon_\tau, \tau \leq t)$. These restrictions are multiperiod ones of order p.

Proposition 2.3 SR-SARV and multiperiod restrictions
Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ be a stationnary process. It admits a SR-SARV(p) representation w.r.t. an increasing filtration J_t if and only if there exist $p+1$ reals ω, γ_1,..,γ_p, such that the roots of $1 - \sum_{i=1}^{p} \gamma_i L^i$ are outside the unit circle and

$$E[\varepsilon_t^2 - \omega - \sum_{i=1}^{p} \gamma_i \varepsilon_{t-i}^2 \mid \varepsilon_\tau, \tau \leq t-p-1] = 0. \tag{2.17}$$

Therefore, when the fourth moment of ε_t is finite, ε_t^2 is an ARMA(p,p) defined by (2.17), that is an ARMA property which is intermediate between weak and semi-strong. The (semi-strong) ARMA structure was the main idea of the ARCH models introduced by Engle (1982) and generalized by Bollerslev (1986). Indeed, the clustering effect in financial data that these models account for is directly related to the ARMA structure of the squared residuals.

For temporal aggregation purposes, Drost and Nijman (1993) introduce the weak GARCH models where the squared residuals process is a weak ARMA. Following the Drost and Nijman (1993) terminology, we precisely define below the various concepts and show how they are nested.

2.3.2 GARCH(p,q)

Definition 2.3. GARCH(p,q): *Let a stationary process $\{\varepsilon_t, t \in \mathbf{Z}\}$ and define the processes $\{h_t, u_t, t \in \mathbf{Z}\}$ by the stationary solution of*

$$B(L)h_t = \omega + A(L)\varepsilon_t^2 \tag{2.18}$$

and $u_t = \varepsilon_t/\sqrt{h_t}$, with $A(L) = \sum_{i=1}^{q} \alpha_i L^i$, $B(L) = 1 - \sum_{i=1}^{p} \beta_i L^i$ where the roots of $B(L) - A(L)$ and $B(L)$ are assumed to be different and outside the unit circle. We say that:

i) ε_t *is a strong GARCH(p,q) if the process* u_t *is i.i.d.* $D(0,1)$*;*

ii) ε_t *is a semi-strong GARCH(p,q) if the process* u_t *is such that*

$$E[u_t \mid \varepsilon_\tau, \tau \leq t-1] = 0 \quad \text{and} \quad Var[u_t \mid \varepsilon_\tau, \tau \leq t-1] = 1; \tag{2.19}$$

iii) ε_t *is a weak GARCH(p,q) if*

$$EL[\varepsilon_t \mid H_{t-1}] = 0 \quad \text{and} \quad EL[\varepsilon_t^2 \mid H_{t-1}] = h_t, \tag{2.20}$$

where $EL[x_t \mid H_{t-1}]$ *denotes the best linear predictor of* x_t *in the Hilbert space,* H_{t-1}*, spanned by* $\{1, \varepsilon_\tau, \varepsilon_\tau^2, \tau \leq t-1\}$*, that is*

$$E[(x_t - EL[x_t \mid H_{t-1}])\varepsilon_{t-i}^r] = 0 \quad \text{for} \quad i \geq 1 \quad \text{and} \quad r = 0,1,2. \tag{2.21}$$

Note that in the strong and semi-strong cases, we do not assume that the fourth moment is finite while the weak GARCH setting requires this assumption.

Proposition 2.4 Semi-strong GARCH and ARMA

Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ *be a m.d.s. (*$E[\varepsilon_t \mid \varepsilon_\tau, \tau \leq t-1] = 0$*). It is a semi-strong GARCH(p,q) if and only if* ε_t^2 *is a semi-strong ARMA(max*$\{p,q\}$*,p) with an innovation process which is a m.d.s. w.r.t.* I_t*.*

Note that the innovation process of the squared process is assumed to be a m.d.s. w.r.t. I_t and not only w.r.t. $\tilde{I}_t = \sigma(\varepsilon_\tau^2, \tau \leq t))$ since the conditional variance process is defined given I_t (and not $\tilde{I}_t$). Bollerslev (1988) already pointed out that the squared values of a strong GARCH(p,q) have a semi-strong ARMA(max$\{p,q\}$,p) structure. Note that strong GARCH implies only semi-strong ARMA: when ε_t^2/h_t is i.i.d., the ARMA process ε_t^2 should in general be conditionally heteroskedastic.

Since ε_t^2 is a semi-strong ARMA, it fulfills a multiperiod conditional moment restriction of order max(p,q).[22] Therefore, Proposition 2.3 implies that ε_t admits a SR-SARV(max$\{p,q\}$) representation.

Corollary 3.1 Semi-strong GARCH and SR-SARV

Let $\{\varepsilon_t\}$ *be a semi-strong GARCH(p,q). Then* $\{\varepsilon_t\}$ *is a SR-SARV(max*$\{p,q\}$*) w.r.t.* I_t*.*

Note that Corollary 3.1 and Proposition 2.1 put together provide a continuous time model, the SR-SARV(p) one, which is consistent with GARCH(p,p) in discrete time. To our knowledge, the relationship between GARCH(p,p) modeling of higher order ($p > 1$) and continuous time stochastic volatility models was not clearly stated before in the literature, whatever the approach of diffusion approximating (Nelson, 1990), filtering (Nelson and Foster, 1994) or closing the GARCH Gap (Drost and Werker, 1996). Finally, the temporal aggregation of a GARCH model is a SR-SARV model. In

[22] More precisely, a semi-strong ARMA($\tilde{q},\tilde{p}$) implies a multiperiod conditional moment restrictions of order equal to max$\{\tilde{q},\tilde{p}\}$. Thus a semi-strong ARMA(max$\{p,q\}$,p) implies a multiperiod restriction of order max$\{p,q\}$.

other words, to close the class of GARCH processes, we have to plug it into the stochastic volatility class of models. This is not a surprising result since we know that semi-strong ARMA processes are not closed under temporal aggregation.

In the next section, we give additional insights as to why GARCH models are not robust to temporal aggregation. Drost and Nijman (1993) already focused on this weakness of standard GARCH models. They give examples of strong and semi-strong GARCH which are not closed under temporal aggregation. Then, they introduce the weak GARCH model where the squared residuals are weak ARMA in order to benefit from the temporal aggregation of the weak ARMA structure.

Proposition 2.5 Weak GARCH and ARMA

Let H^s_{t-1} the Hilbert space spanned by $\{1, \varepsilon^2_\tau, \tau \leq t-1\}$ and $\{h^s_t, \eta_t\}$ the processes defined by $h^s_t = EL[\varepsilon^2_t \mid H^s_{t-1}]$ and $\eta_t = \varepsilon^2_t - h^s_t$. If ε_t is a weak GARCH(p,q) process, then $h_t = h^s_t$ a.s. and, hence, ε^2_t is a weak stationary ARMA($max\{p,q\}$,p) process and

$$Cov(\eta_t, \varepsilon_\tau) = 0, \ \forall \tau < t. \tag{2.22}$$

Conversely, if ε^2_t is a weak stationary ARMA(q,p) process and (2.22) holds, then ε_t is a weak GARCH(p,q).

Thus, the weak GARCH property is slightly more restrictive than the weak ARMA assumption for the squared residuals. In particular, (2.22) is like a symmetry assumption, which is implied by the maintained m.d.s. condition for ε_t when assuming semi-strong GARCH. In fact, Drost and Nijman (1993) take a "coherent" definition in the sense that they project both the residual and its square onto the same space H_{t-1}. However, using the ARMA structure of the squared residuals is the main idea of weak GARCH.[23] As we can already see, the class of weak ARMA processes strictly contains the class of ARMA models with a semi-strong state-space representation and finite variance. Therefore, weak GARCH processes are in fact **Stochastic Volatility** models, i.e., Drost and Nijman (1993) plug also the class of GARCH models into the SV one.[24]

However, to show that weak GARCH class is closed under temporal aggregation for flow variables, Drost and Nijman (1993) maintain at least one of the following symmetry assumptions:

$$\forall h \in \mathbf{N}^*, \forall (a_k)_{1\leq k\leq h} \in \{-1,1\}^h, (\varepsilon_{t+k})_{1\leq k\leq h} \ = \ (a_k\varepsilon_{t+k})_{1\leq k\leq h} \text{ in distribution, or} \tag{2.23}$$

$$\forall 0 \leq i \leq j \ E[\varepsilon_t\varepsilon_{t-i}\varepsilon_{t-j}] = 0 \text{ and } \forall \ 0 \leq i \leq j \leq k, i \neq 0 \text{ or } j \neq k \ E[\varepsilon_t\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-k}] = 0. \tag{2.24}$$

[23] When Nijman and Sentana (1996) and Drost and Weker (1996) prove respectively that a marginalization of a multivariate GARCH and that the discretization of (2.8) for p=1 under (2.9) are weak GARCH, they only deal with the ARMA property of squared residuals.

[24] See the following section where we establish the exact links between SR-SARV and weak GARCH.

Such symmetry restrictions are indeed quite restrictive both on theoretical and empirical grounds. They preclude two types of asymmetry which appear relevant for financial data. First, even in the strong GARCH setting, the probability distribution of the standardized innovations $\varepsilon_t/\sqrt{h_t}$ may be skewed. Second, since the weak GARCH models are SV ones (outside the standard GARCH class), another type of asymmetry (termed the leverage effect by Black, 1976, and popularized by Nelson, 1991) may matter. A clear distinction between these two types of asymmetric behavior of a general SR-SARV(1) process will be made in Section 3 below. Equivalently, the leverage effect can be introduced in the continuous time setting by allowing the volatility matrix R_t to be non-diagonal, unlike the case considered by Drost and Werker (1996) and Andersen and Bollerslev (1998). Finally, note that our results concerning temporal aggregation and exact discretization are consistent with those of Drost and Nijman (1993) and Drost and Werker (1996).[25] In particular, the restrictions on the persistence parameters are the same ($\Gamma^{(m)} = \Gamma^m$).[26]

3 SR-SARV(1)

3.1 SR-SARV(1) and GARCH(1,1)

The GARCH(1,1) model is nowadays dominant w.r.t. any other ARCH or GARCH type model in the empirical finance literature. We discuss in more detail its relationship with the general SR-SARV(1). In the previous section, we proved that a semi-strong GARCH(p,q) is also a SR-SARV(max$\{p, q\}$). We first characterize those SR-SARV(1) processes which are also semi-strong GARCH(1,1). The general notations of Definition 2.1 are adapted in the case $p = 1$ by $f_t = F_t$ and $\gamma = \Gamma$ with $|\gamma| < 1$.

Proposition 3.1 Semi-strong GARCH(1,1) and SR-SARV(1)
Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ be a SR-SARV(1) process with a conditional variance process f_t and a positive persistence parameter γ. Then, ε_t is a semi-strong GARCH(1,1) with $\alpha > 0$ and $\beta \geq 0$ if and only if: i) ε_t^2 and f_t are conditionally perfectly linearly and positively correlated given J_{t-1}; ii) the ratio $Var[f_t \mid J_{t-1}]/Var[\varepsilon_t^2 \mid J_{t-1}]$ is constant and smaller or equal to γ^2. In this case: $h_{t+1} = f_t$, $J_t = I_t$ and $\beta = \gamma - \alpha$ with $\alpha = \sqrt{Var[f_t \mid J_{t-1}]/Var[\varepsilon_t^2 \mid J_{t-1}]}$.

For all practical purposes, the first condition implies that f_t is adapted to I_t, i.e. there are no exogenous sources of randomness in the conditional variance. Actually, this is the case for the GARCH(1,1) model and also for more general ARCH-type models as ones listed below in (3.1), (3.2) (3.3), (3.4) and (3.5). But perfect linear conditional correlation between f_t and ε_t^2 is a specification of the GARCH model. The second condition is less known even though it was already coined by Nelson and Foster (1994). They observed that the most commonly used ARCH models assume that

[25] Nevertheless, Drost and Werker (1993) consider only the one factor case.
[26] For more details, see Meddahi and Renault (1996).

the variance of the variance rises linearly with the square of the variance, which is the main drawback of GARCH models in approximating SV models in continuous time. Thus, the semi-strong GARCH setting imposes nontrivial restrictions on the dynamics of the conditional kurtosis.

Nelson (1991) stressed that one limitation of GARCH models is that only the magnitude and not the sign of unanticipated excess returns affects the conditional variance. Therefore, alternative asymmetric GARCH models have been introduced in the literature. For instance, Glosten, Jagannathan and Runkle (1989, GJR) introduce a model based on a GARCH formulation but accounting for the sign of the past residuals. More generally, asymmetric models have been studied and compared by Engle and Ng (1993) who consider the following models:

$$\mathrm{GJR}: \mathrm{h_t} = \omega_1 + \alpha\varepsilon_{\mathrm{t}-1}^2 + \beta \mathrm{h_{t-1}} + \lambda\ \mathrm{S_{t-1}}\varepsilon_{\mathrm{t}-1}^2, \text{ where } \mathrm{S_t} = 1 \text{ if } \varepsilon_{\mathrm{t}} < 0,\ \mathrm{S_t} = 0 \text{ otherwise} \tag{3.1}$$

$$\text{Asymmetric GARCH}: \ h_t = \omega_1 + \alpha(\varepsilon_{t-1} + \lambda)^2 + \beta h_{t-1}; \tag{3.2}$$

$$\text{Nonlinear Asymmetric GARCH}: \ h_t = \omega_1 + \alpha(\varepsilon_{t-1} + \lambda\sqrt{h_{t-1}})^2 + \beta h_{t-1}; \tag{3.3}$$

$$\text{VGARCH}: \ h_t = \omega_1 + \alpha(\varepsilon_{t-1}/\sqrt{h_{t-1}} + \lambda)^2 + \beta h_{t-1}; \tag{3.4}$$

Let us also consider a related model considered by Heston and Nandi (1999):

$$\text{Heston and Nandi}: \ h_t = \omega_1 + \alpha(\varepsilon_{t-1}/\sqrt{h_{t-1}} - \lambda\sqrt{h_{t-1}})^2 + \beta h_{t-1}. \tag{3.5}$$

Actually, we show that all these models are in the SR-SARV(1) class.[27]

Proposition 3.2 Asymmetric GARCH and SR-SARV(1)

Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ be a m.d.s. and define h_t the conditional variance of ε_t, i.e. $h_t \equiv Var[\varepsilon_t \mid \varepsilon_\tau, \tau \leq t-1]$. Assume that h_t is given by (3.2), (3.3), (3.4), or by (3.5), then ε_t is a SR-SARV(1) process. If $u_t = \varepsilon_t/\sqrt{h_t}$ is i.i.d., then the GJR model defined by (3.1) is also a SR-SARV(1) process.

3.2 SR-SARV(1) and weak GARCH(1,1)

We will now focus on the relationships between SR-SARV and weak GARCH. As already mentioned, Drost and Nijman (1993) prove the temporal aggregation property of symmetric weak GARCH (assuming (2.23) or (2.24)) which excludes the leverage effect and all the asymmetric models considered in Proposition 3.2 We specify two kinds of asymmetries for the SR-SARV model:

Definition 3.1. Leverage effect and skewness: *Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ be a SR-SARV(1) process w.r.t. a filtration J_t with corresponding processes $\{f_t, u_t, \nu_t\}$ where $f_t = \omega + \gamma f_{t-1} + \nu_t$ and $u_t = \varepsilon_t/\sqrt{f_{t-1}}$. We say that:*

[27]Finally, Drost (1993) shows that symmetric QARCH of Sentana (1995) are weak GARCH. Indeed, it is easy to show that any QARCH is a SR-SARV model. This is also the case of the HARCH model of Muller and al. (1997) since this model is a restricted QARCH.

i) ε_t *does not possess a (conditional) leverage effect w.r.t.* J_t *if and only if*

$$E[u_t \nu_t \mid J_{t-1}] = 0 \text{ or equivalently } E[\varepsilon_t \varepsilon_{t+1}^2 \mid J_{t-1}] = 0; \tag{3.6}$$

ii) ε_t *does not possess a (conditional) skewness w.r.t.* J_t *if and only if*

$$E[u_t^3 \mid J_{t-1}] = 0 \text{ or equivalently } E[\varepsilon_t^3 \mid J_{t-1}] = 0. \tag{3.7}$$

We show in the appendix that the two conditions of (3.6) (and (3.7)) are equivalent. Now we can show that a SR-SARV model without leverage effect and skewness is a weak GARCH.

Proposition 3.3 Weak GARCH(1,1) and SR-SARV(1)
If ε_t *is a SR-SARV(1) process with finite fourth moment and without leverage effect or skewness, that is if (3.6) and (3.7) hold, then* ε_t *is a weak GARCH(1,1) process.*

Therefore, there is no major difference between symmetric weak GARCH and symmetric SR-SARV. However, we do not prove an equivalence result, and it is clear that the class of symmetric weak GARCH is larger than one of symmetric SR-SARV. Indeed, one can interpret the weak GARCH model as a SV model, but not endowed with a sufficiently rich specification for statistical inference and economic interpretation. In addition, we have proved in Section 2 that this weakness is not needed to close the GARCH gap with continuous time as in Drost and Werker (1996). In a sense, by introducing the SR-SARV, we have restricted the weak GARCH models by adding useful restrictions for financial and statistical interpretations. Furthermore, SR-SARV allows for asymmetries like the leverage effect and skewness. Indeed, the corresponding symmetry assumptions are closed under temporal aggregation.

Proposition 3.4 Temporal aggregation, leverage effect and skewness
Let $\{\varepsilon_t, t \in \mathbf{Z}\}$ *be a SR-SARV process w.r.t. an increasing filtration* J_t *with corresponding processes* $\{f_t, u_t, \nu_t, t \in \mathbf{Z}\}$. *Define* $\varepsilon_{tm}^{(m)}$ *by (2.12) and the corresponding SR-SARV(1) representation of Proposition 3.2,* $J_{tm}^{(m)}$, $\{f_{tm}^{(m)}, u_{tm}^{(m)}, \nu_{tm}^{(m)},\}$. *Then the symmetric SR-SARV class is closed under temporal aggregation. More precisely, we have:*

$$E[u_t \nu_t \mid J_{t-1}] = 0 \Longrightarrow E[u_{tm}^{(m)} \nu_{tm}^{(m)} \mid J_{(t-1)m}^{(m)}] = 0, \text{ and} \tag{3.8}$$

$$E[u_t \nu_t \mid J_{t-1}] = E[u_t^3 \mid J_{t-1}] = 0 \Longrightarrow E[u_{tm}^{(m)} \nu_{tm}^{(m)} \mid J_{(t-1)m}^{(m)}] = E[(u_{tm}^{(m)})^3 \mid J_{(t-1)m}^{(m)}] = 0. \tag{3.9}$$

This proposition means that our results are tightly related to those of Drost and Nijman (1993) and Drost and Werker (1996), since symmetric SR-SARV are weak GARCH and are closed under temporal aggregation. Besides, the relationship between parameters at various frequencies, already stressed by these authors (particularly the persistence parameter) are maintained in our SR-SARV setting.

Furthermore, a symmetry assumption about the standardized innovation cannot be maintained for various frequencies without precluding leverage effect as well. It is easy to see that when leverage effect is present, the symmetry condition (3.7) is not robust to temporal aggregation. Therefore, when one observes significant skewness at a low frequency, it may be due either to genuine skewness or to leverage effect at higher frequency, while the presence of leverage effect at a low frequency implies the same feature at higher frequencies.

Proposition 3.5 Observable restrictions of leverage effect and skewness

Let ε_t be a SR-SARV(1) w.r.t. a filtration J_t.

i) If ε_t is without leverage effect ((3.6) holds), then

$$E[\varepsilon_t \varepsilon_{t+1}^2 \mid I_{t-1}] = 0. \tag{3.10}$$

ii) If ε_t is without skewness ((3.7) holds), then

$$E[\varepsilon_t^3 \mid I_{t-1}] = 0. \tag{3.11}$$

Therefore we can derive moments restrictions based on observable data which can be used to test the absence of leverage effect or skewness. Moreover, usual GARCH allows for a leverage effect as soon as there is skewness since the conditions (3.6) and (3.7) are equivalent in this case. Indeed, in the introduction of his EGARCH paper, Nelson (1991) explicitly mentions that **symmetric** GARCH models do not take into account the leverage effect.

3.3 Temporal aggregation of IGARCH(1,1) models

Until now, we have considered temporal aggregation of volatility models with integrable volatility. However, some empirical evidence supports the Integrated GARCH model introduced by Engle and Bollerslev (1986). This process is not second-order stationary, since the second moment is infinite. This evidence is even more pronounced for high frequency data (5 and 10 minutes returns); see for instance Andersen and Bollerslev (1997a) and Gençay et al. (1998). While the second moment of the residuals is not finite, the notion of conditional variance is valid since the squared residual process is nonnegative and hence its possibly infinite conditional expectation is well defined. Moreover, we know that the GARCH(1,1) process is strictly stationary when $E[ln(\beta + \alpha u_t^2)] < 0$ (and $\omega > 0$) with i.i.d. standardized residuals (see Nelson, 1990). This condition is ensured when $\alpha + \beta = 1$,[28] that is for IGARCH(1,1). Therefore we can extend our notion of SR-SARV to nest the IGARCH class.

[28]By Jensen's inequality, we have $E[ln(\beta + \alpha u_t^2)] < lnE[\beta + \alpha u_t^2] = ln(\alpha + \beta) = 0$.

Definition 5.1. Integrated SR-SARV(1) model: *A strictly stationary process* $\{\varepsilon_t, t \in \mathbf{Z}\}$ *is called an* **ISR-SARV(1) process** *w.r.t.* J_t *if:*

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

ii) the conditional variance process f_{t-1} *of* ε_t *given* J_{t-1} *is such that:*

$$E[f_t \mid J_{t-1}] = \omega + f_{t-1}. \tag{3.12}$$

Obviously an IGARCH(1,1) is an ISR-SARV(1). Note that strict stationarity is not important for modeling purposes since we can remove it in the definition of an ISR-SARV. However, it is useful for inference. We now consider temporal aggregation of ISR-SARV:

Proposition 3.6 Temporal aggregation of ISR-SARV(1)

Let ε_t *be an ISR-SARV(1) process w.r.t. an increasing filtration* J_t *and a conditional variance process* f_{t-1}*. The process* $\varepsilon_{tm}^{(m)}$ *defined by* $\varepsilon_{tm}^{(m)} \equiv \sum_{k=0}^{m-1} a_k \varepsilon_{tm-k}$ *is an ISR-SARV(1) w.r.t.* $J_{tm}^{(m)} = \sigma(\varepsilon_{\tau m}^{(m)}, f_{\tau m}, \tau \le t)$.

As a consequence, a temporally aggregated IGARCH process is also an integrated process but of SV type.[29] Empirically, the IGARCH model is rejected at low frequencies, e.g. monthly. Therefore by the aggregation result, one should conclude that the model at high frequency is not an integrated one. A potential explanation of this is long memory in the volatility. For instance, Bollerslev and Mikkelsen (1998) (resp Comte and Renault, 1998) show via a Monte Carlo study that when the true model is FIGARCH (resp long memory continuous time SV), estimation of a GARCH model by QMLE suggests an IGARCH model. Temporal aggregation of long memory volatility models is beyond the scope of this paper; see Andersen and Bollerslev (1997b).

3.4 QMLE is not consistent for temporally aggregated GARCH(1,1) models

The main objective of this section is to show that the Gaussian QMLE is not consistent for aggregated strong GARCH processes. Drost and Nijman (1992) provided some Monte Carlo results that suggested that the QMLE is consistent or at least have a small bias. More precisely, they generated several strong GARCH(1,1) processes, then aggregated them. Note that these authors consider a very long sample size (80000) after aggregation in order to study the consistency of the QMLE. Then, these authors estimated the weak GARCH(1,1) model by using the QML method and concluded that this method work well. In contrast, Nijman and Sentana (1996) showed in a Monte Carlo study, with the same sample size, that the QMLE is not consistent when one aggregates two independent GARCH processes. Therefore, the Monte Carlo results of Drost and Nijman (1992) are puzzling.

[29] Engle and Bollerslev (1986) consider temporal aggregation of IGARCH model with $\omega = 0$ which is not, however, a strictly stationary process. Moreover, the variance process converges a.s. to a constant (Nelson, 1991).

Francq and Zakoian (2000) consider also the estimation of weak GARCH models. In particular, they propose a Yule-Walker based two-step method to estimate weak GARCH models. Note however that the definition of weak GARCH adopted by Francq and Zakoian (2000) is different from one of Drost and Nijman (1993). The difference is that Francq and Zakoian (2000) defined the variance process h_t as the best linear projection of ε_t^2 given the Hilbert space generated by $\{1, \varepsilon_\tau^2, \tau \leq t-1\}$ and not $\{1, \varepsilon_\tau, \varepsilon_\tau^2, \tau \leq t-1\}$. Indeed, as pointed out by Francq and Zakoian (2000), their weak GARCH models are not closed under temporal aggregation. Finally, it is worth noting that Francq and Zakoian (2000) study the consitency of the Gaussian QMLE of aggregated strong GARCH models in a Monte Carlo experiment. They reach the same conclusion as Drost and Nijman (1992).
We consider a Gaussian GARCH model at high frequency, that is

$$y_t = \mu^1 + \varepsilon_t = \mu^1 + \sqrt{h_t} u_t, \quad \text{with} \quad h_t = \omega^1 + \alpha^1 \varepsilon_{t-1}^2 + \beta^1 h_{t-1} \tag{3.1}$$

where u_t is i.i.d. $\mathcal{N}(0,1)$ with $(\mu^1, \omega^1, \alpha^1, \beta^1, \gamma^1) = (0, 2.8\text{E-}06, .0225, .9770, .9995)$ where $\gamma^1 = \alpha^1 + \beta^1$. We choose these parameters such that after aggregation as flow over m periods with m=400, we obtain a weak GARCH model with the coefficients $(\mu^0, \omega^0, \alpha^0, \beta^0) = (0, 0.4, 0.206, 0.594, 0.8)$. The persistence parameter at the high frequency, γ^1, is conformable to the empirical study of Andersen and Bollerslev (1997a). α^1 and β^1 are chosen such that after temporal aggregation, α^0 and β^0 are close to those of a specification considered by Nijman and Sentana (1996).

The aggregated model we consider is very different from ones considered by Drost and Nijman (1992) (and Francq and Zakoian, 2000, as well). The first difference is that we aggregate more than Drost and Nijman (1992) since we consider $m = 400$ while these authors considered $m = 2, 4, 8, 16$. Moreover, the implied parameters after aggregation are more realistic in our case. For instance, the volatility persistence parameter is .663 with m=8 and .44 with m=16, while we consider persistence parameters equal to .8.

We follow Drost and Nijman (1992) and Nijman and Sentana (1996) by considering very long samples. We consider sample sizes equal to 80000 and 150000. For the first sample, our results provided in Table 1 are based on 100 replications while they are based on 50 replications for the second sample size. From Table 1, it is clear that the QMLE is not consistent for temporally aggregated GARCH models. Therefore, one has to consider a consistent method as the one developed in Francq and Zakoian (2000) or a method based on the multiperiod conditional moment restrictions.[30]

[30] See the previous version of this paper, Meddahi and Renault (2000).

4 Conclusion

In this paper, we have considered temporal aggregation of volatility models. We introduce a semiparametric class of volatility models termed square-root stochastic autoregressive volatility (SR-SARV) characterized by an autoregressive dynamic of the stochastic variance. Our class encompasses the usual GARCH models of Bollerslev (1986), the asymmetric GARCH models of Glosten, Jagannathan and Runkle (1989) and Engle and Ng (1993). Moreover, even if the volatility is stochastic, that is may involve a second source of randomness, the considered models are characterized by observable multiperiod conditional moment restrictions (Hansen, 1985). The SR-SARV class is a natural extension of the weak GARCH models of Drost and Nijman (1993) in discrete time and Drost and Werker (1996) in continuous time. The SR-SARV class extends the weak GARCH class since it does not assume that the fourth moment is finite and, moreover, allows for asymmetries (skewness, leverage effects). On the other hand, it provides a statistical structure which remains true to the concept of conditional variance, and maintains the validity of conditional moment restrictions, which are useful for inference. Finally we also consider temporal aggregation of IGARCH models.

References

Akaike, H. (1974), "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," Ann. Inst. Stat. Math., 26, 363-387.

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

Andersen, T.G. and T. Bollerslev (1997a), "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 4, 115-158.

Andersen, T.G. and T. Bollerslev (1997b), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52, 975-1005.

Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.

Barndorff-Nielsen, O.E., Jensen J.L. and M. Sorensen (1998), "Some Stationary Processes in Discrete and Continuous Time," *Advanced Applied Probability*, 30, 989-1007.

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.

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.

Bergstrom, A.R. (1990), *Continuous Time Econometric Modelling*, Oxford University Press.

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

Bollerslev, T. (1988), "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Processes," *Journal of Times Series Analysis*, 9, 121-131.

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

Bollerslev, T. and H.L. Mikkelsen (1996), "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151-184.

Comte, F. and E. Renault (1998), "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291-323.

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

Cox, J.C. (1975), "Notes on Option Pricing I: Constant Elasticity of Variance Diffusions," Discussion Paper, Stanford University.

Davis, R.A. and T. Mikosch (1998), "The Sample Autocorrelations of Heavy-Tailed Processes with Application to ARCH," *Annals of Statistics*, 26, 2049-2080.

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

Diebold, F.X. and M. Nerlove (1989), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent ARCH Model," *Journal of Applied Econometrics*, 4, 1-21.

Drost, F.C. (1993), "Temporal Aggregation of Time-Series," in J. Kaehler and P. Kugler eds, *Econometric Analysis of Financial Markets*, 11-21.

Drost, F.C. and TH.E. Nijman (1992), "Temporal Aggregation of GARCH processes," Tilburg University DP 9240.

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

Drost, F.C. and B.J.M. Werker (1996), "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31-58.

Duffie, D. and P. Glynn (1997), "Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals," unpublished paper, Stanford University.

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

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

Engle, R.F. and T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews,* 5, 1-50.

Engle, R.F., and V.K. Ng (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749-1778.

Engle, R.F., V.K. Ng and M. Rothschild (1990), "Asset Pricing with a Factor ARCH Structure: Empirical estimates for Treasury Bills," *Journal of Econometrics*, 45, 213-237.

Francq, C. and J.M. Zakoian (2000), "Estimating Weak GARCH Representations," *Econometric Theory*, 16, 692-728.

Gençay, R., G. Ballocchi, M. Dacorogna, R. Olsen and P. Pictet (1998), "Real-Time Trading Models and Statistical Properties of Foreign Exchange Rates," unpublished paper, University of Windsor.

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.

Glosten, L.R., R. Jagannathan and D. Runkle (1989), "On the Relation between the Expected Value of the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, Vol XLVIII, 1779-1801.

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. 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 E. Sentana (1992), "Unobserved Component time series models with ARCH disturbances," *Journal of Econometrics*, 52, 129-157.

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.

Heston, S.L. and S. Nandi (2000), "A Closed-Form GARCH Option Pricing Model," *Review of Financial Studies*, 13, 585-625.

Horn, R.A. and C.R. Johnson (1994), "Topics in Matrix Analysis," Cambridge University Press.

King, M., E. Sentana and S. Wadhwani (1994), "Volatility and Links between National Stock Markets," *Econometrica*, 62, 901-933.

Lutkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer Verlag.

Meddahi, N. (2002), "Temporal and Cross-sectional Aggregations of Volatility in Mean 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.

Meddahi, N. and E. Renault (2002a), "State Space and Multiperiod ARMA Models," in preparation.

Meddahi, N. and E. Renault (2002b), "Conditioning Information in Volatility Models," in preparation.

Muller, U.A., M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Pictet and J.E. von Weizsacker (1997), "Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components," *Journal of Empirical Finance*, 4, 213-239.

Nelson, D.B. (1990), "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. (1992), "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61-90.

Nelson, D.B. (1996), "Asymptotically Optimal Smoothing with ARCH Models," *Econometrica*, 64, 561-573.

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

Nijman, TH. and E. Sentana (1996), "Marginalization and Contemporaneous Aggregation of Multivariate GARCH Processes," *Journal of Econometrics*, 71, 71-87.

Sentana, E. (1995), "Quadratic ARCH Models," *Review of Economic Studies*, 62, 639-661.

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.

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

Table 1. QML estimation of temporally aggregated GARCH models

Sample size	$\mu^0 = 0$	$\omega^0 = .4$	$\alpha^0 = .206$	$\beta^0 = .504$	$\gamma^0 = .8$
80000	0.0004368 (0.005546)	0.4629 (0.02318)	0.1963 (0.006935)	0.5688 (0.01547)	0.7651 (0.01219)
150000	0.000246 (0.003856)	0.4618 (0.01238)	0.1959 (0.004394)	0.5695 (0.008015)	0.7654 (0.006655)

NOTE. The reported statistics are based on 100 replications for the first sample size and 50 replications for the second one. For each cell, the first number shows the mean and the second the standard deviation (in parentheses).

APPENDIX

Proof of Proposition 2.1. From (2.8), we have $dy_t = \sqrt{e'F_t^c}dW_{1t}$ where W_{1t} is the first component of W_t. Therefore $\varepsilon_{th}^{(h)} = \int_{th-h}^{th} \sqrt{e'F_u^c}dW_{1u}$ and $f_{(t-1)h}^{(h)} \equiv Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}] = E[\int_{th-h}^{th} e'F_u^c du \mid J_{(t-1)h}^{(h)}] = e' \int_{th-h}^{th} E[F_u^c \mid J_{(t-1)h}^{(h)}]du$.

Consider the equation (2.8), then we have $dF_t^c = K(\Theta - F_t^c)dt + M_{22}\, R_t dW_t$ where M_{22} is the $p\times(p+1)$ matrix defined by $M_{22} = (0, I_p)$. Therefore, we have

$$\forall h > 0, F_{t+h}^c = (Id - e^{-Kh})\Theta + e^{-Kh}F_t^c + e^{-Kh}\int_t^{t+h} e^{K(u-t)}M_{22}R_t dW_u. \tag{A.1}$$

Hence, $f_{th-h}^{(h)} = e'F_{th-h}^{(h)}$ with
$F_{th-h}^{(h)} = \int_{th-h}^{t} \left\{(Id - e^{-K[u-(th-h]})\Theta + e^{-K[u-(th-h]}F_{th-h}^c\right\} du = A^{(h)}F_{th-h}^c + B^{(h)}$ where $A^{(h)} = K^{-1}(Id - e^{-Kh})$ and $B^{(h)} = (hId - A^{(h)})\Theta$. Since $\{F_{th}^c, t \in \mathbf{Z}\}$ is a VAR(1) due to (A.1) and since $A^{(h)}$ is non singular, $\{F_{th}^{(h)}, t \in \mathbf{Z}\}$ is also a VAR(1) with the same autoregressive matrix than $\{F_{th}^c\}$ that is e^{-Kh}.□

Proof of Proposition 2.2. We show that the points of the Definition 2.1 are fulfilled: i) by definition of $J_{tm}^{(m)}$; ii) we have $J_{tm}^{(m)} \subset J_{tm}$. Hence, $E[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = \sum_{i=0}^{m-1} a_i E[E[\varepsilon_{tm-i} \mid J_{tm-i-1} \mid J_{tm-m}^{(m)}] = 0$, that is $\varepsilon_{tm}^{(m)}$ is a m.d.s. w.r.t. $J_{tm-m}^{(m)}$; iii) we have: $Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[(\varepsilon_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}] = \sum_{i=0}^{m-1} a_i^2 E[\varepsilon_{tm-i}^2 \mid J_{tm-m}^{(m)}] + 2\sum_{0\leq i<j\leq m-1} a_i a_j E[\varepsilon_{tm-i}\varepsilon_{tm-j} \mid J_{tm-m}^{(m)}]$
$= \sum_{i=0}^{m-1} a_i^2 E[E[\varepsilon_{tm-i}^2 \mid J_{tm-i-1}] \mid J_{tm-m}^{(m)}] + 2\sum_{0\leq i<j\leq m-1} a_i a_j E[\varepsilon_{tm-j}E[\varepsilon_{tm-i} \mid J_{tm-i-1}] \mid J_{tm-m}^{(m)}]$
$= E[\sum_{i=0}^{m-1} a_i^2 f_{tm-i-1} \mid J_{tm-m}^{(m)}]$. Since f_t is a marginalization of a the VAR(1) process F_t, one easily gets (see the proof of Proposition 2.3 of Meddahi and Renault, 2002a) $Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[e'(A^{(m)}F_{tm-m} + B^{(m)}) \mid J_{tm-m}^{(m)}]$ where $A^{(m)}$ and $B^{(m)}$ are defined by (2.14). By definition of $J_{tm-m}^{(m)}$, $F_{tm-m}^{(m)}$ is adapted w.r.t. $J_{tm-m}^{(m)}$. Hence, $Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = e'(A^{(m)}F_{tm-m} + B^{(m)}) = e^{(m)'}F_{tm-m}^{(m)}$ where $e^{(m)}$ and $F_{tm-m}^{(m)}$ are defined by (2.15). Besides, $F_{tm-m}^{(m)}$ is a VAR(1) with autoregressive matrix Γ^m.□

Proof of Proposition 2.3. Consider $\{\varepsilon_t, t \in \mathbf{Z}\}$ a SR-SARV(p). Hence $\varepsilon_t^2 = f_{t-1} + \eta_t$. where $\{f_t\}$ admits a state-space representation $\{F_t, \eta_t\}$ w.r.t J_t. We have $F_t = \Omega + \Gamma F_{t-1} + V_t \Rightarrow (Id - \Gamma L)F_t = \Omega + V_t \Rightarrow Det(Id - \Gamma L)F_t = (Id - \Gamma L)^*(\Omega + V_t)$ where L is the Lag Operator, $Det(.)$ is the determinant function and $(Id - \Gamma L)^*$ is the adjoint matrix of $(Id - \Gamma L)$. Hence : $Det(Id - \Gamma L)f_t = Det(Id - \Gamma L)e'F_t = e'(Id - \Gamma)^*\Omega + e'(Id - \Gamma L)^*V_t$ We have: $Deg(e'(Id - \Gamma L)^*) \leq p - 1$ where $Deg(.)$ is the maximal degree of the lag polynomials, coefficients of the matrix. Hence $E[Det(Id - \Gamma L)f_t - e'(Id - \Gamma)^*\Omega \mid J_{t-p}] = 0$. Thus $E[Det(Id - \Gamma L)\varepsilon_{t+1}^2 - e'(Id - \Gamma)^*\Omega \mid J_{t-p}] = 0$ since $\varepsilon_{t+1}^2 = f_t + \eta_{t+1}$ and the (maximal) degree of $Det(Id - \Gamma L)$ is p. Define $a_1, .., a_p$ by $1 - \sum_{i=1}^p a_i L^i = Det(Id - \Gamma L)$ and the real ω by $\omega = e'(Id - \Gamma)^*\Omega$. By definition 2.1, the eigenvalues of Γ are smaller than one in modulus.

Therefore the roots of $1 - \sum_{i=1}^{p} a_i L^i$ are outside the unit circle. Finally, $\sigma(\varepsilon_\tau, \tau \leq t-p) \subset J_{t-p}$. Hence $E[\varepsilon_{t+1}^2 - \omega - \sum_{i=1}^{p} a_i \varepsilon_{t+1-i}^2 \mid z_\tau, \tau \leq t-p] = 0$, that is (2.17).

Conversely, consider a process ε_t such that (2.17). Define F_{t-1} by $F_{t-1} = (E[\varepsilon_{t+p-1}^2 \mid I_{t-1}], E[\varepsilon_{t+p-2}^2 \mid I_{t-1}], ..., E[\varepsilon_t^2 \mid I_{t-1}])'$. Thus $\varepsilon_t^2 = (0,0,...,0,1)F_{t-1} + v_t$ with $E[v_t \mid I_{t-1}] = 0$.

For $i = 2, ..., p$, we have again $E[F_t(i) \mid I_{t-1}] = E[\varepsilon_{t+p+1-i}^2 \mid I_{t-1}] = F_{t-1}(i-1)$.

$E[F_t(1) \mid I_{t-1}] = E[\varepsilon_{t+p}^2 \mid I_{t-1}] = E[(\varepsilon_{t+p}^2 - \omega - \sum_{i=1}^{p} a_i \varepsilon_{t+p-i}^2) + \omega + \sum_{i=1}^{p} a_i \varepsilon_{t+p-i}^2 \mid I_{t-1}]$ $= \omega + \sum_{i=1}^{p} a_i E[\varepsilon_{t+p-i}^2 \mid I_{t-1}] = \omega + \sum_{i=1}^{p} a_i F_{t-1}(i)$. Hence, $E[F_t \mid I_{t-1}] = \Omega + \Gamma F_{t-1}$. As a conclusion, $\{\varepsilon_t^2\}$ has a state space representation $\{F_t, v_t\}$ w.r.t. I_t. On the other hand, ε is a m.d.s. w.r.t. I_t. Thus, ε_t is a SR-SARV(p) w.r.t. I_t.□

Proof of Proposition 2.4. Let ε_t be a semi-strong GARCH(p,q) defined by (2.18) and (2.19). Then $(B(L) - A(L))\varepsilon_t^2 = \omega + B(L)\eta_t$ with $\eta_t = \varepsilon_t^2 - h_t$. By assumption, the roots $B(L) - A(L)$ and $B(L)$ are not common and are outside the unit circle. Finally, η_t is a m.d.s. w.r.t. I_t since $E[\eta_t \mid I_{t-1}] = E[\varepsilon_t^2 \mid I_{t-1}] - h_t = 0$. Conversely, consider a m.d.s. $\{\varepsilon_t\}$ such that $Q(L)\varepsilon_t^2 = \omega + P(L)\eta_t$ where $Q(L) = 1 - \sum_{i=1}^{q} a_i L^i$, $P(L) = 1 - \sum_{i=1}^{p} b_i L^i$, $a_q \neq 0$, $b_p \neq 0$ and $p \leq q$. Assume that $E[\eta_t \mid I_{t-1}] = 0$. Define h_t by $h_t \equiv E[\varepsilon_t^2 \mid I_{t-1}]$. Hence $h_t = \omega + (1 - Q(L))\varepsilon_t^2 + (P(L) - 1)\eta_t$ and $\eta_t = \varepsilon_t^2 - h_t$. Thus $h_t = \omega + (P(L) - Q(L))\varepsilon_t^2 + (1 - P(L))h_t$ and $P(L)h_t = \omega + (P(L) - Q(L))\varepsilon_t^2$. By assumption, the roots of $P(L)$ and $P(L) - (Q(L) - P(L))$, i.e. $Q(L)$, are not common and outside the unit circle. Define u_t by $u_t \equiv \varepsilon_t / \sqrt{h_t}$. We have $E[u_t \mid I_{t-1}] = 0$ since ε_t is a m.d.s.; moreover, $Var[u_t \mid I_{t-1}] = E[\varepsilon_t^2 \mid I_{t-1}]/h_t = 1$, i.e. (2.19).□

Proof of Proposition 2.5. Since $H_t^s \subset H_t$, $EL[h_t \mid H_{t-1}^s] = h_t^s$. But $h_t = \omega/B(1) + B(L)^{-1}A(L)\varepsilon_t^2$ and hence $h_t \in H_{t-1}^s$. Thus $h_t = h_t^s$. Therefore ε_t^2 is a weak ARMA (since $B(L)h_t^s = \omega + A(L)\varepsilon_t^2$) and $cov(\eta_t, \varepsilon_\tau) = 0 \; \forall \tau < t$.

Conversely, assume that ε_t^2 is a weak ARMA and (2.22). We have: $h_t = EL[\varepsilon_t^2 \mid H_{t-1}] = h_t^s + EL[\eta_t \mid H_{t-1}]$. By definition of η_t, $\forall \tau < t$, $cov(\varepsilon_\tau^2, \eta_t) = 0$. Therefore, by combination with (2.22), $\forall z \in H_{t-1}$, $cov(z, \eta_t) = 0$. Thus $EL[\eta_t \mid H_{t-1}]$ and $h_t = h_t^s$ and ε_t is a weak GARCH.□

Proof of Proposition 3.1. Let us consider ε_t a GARCH(1,1). Let $f_{t-1} = h_t = E[\varepsilon_t^2 \mid I_{t-1}]$ and $u_t = \frac{\varepsilon_t}{\sqrt{h_t}}$. By definition, $E[u_t \mid I_{t-1}] = 0$ and $E[u_t^2 \mid I_{t-1}] = 1$. while f_t is an I_t-adapted AR(1) process. with an innovation process: $\nu_t = \alpha f_{t-1}(u_t^2 - 1)$. Then, given I_{t-1}, ε_t^2 and $\nu_t = \alpha f_{t-1}(\frac{\varepsilon_t^2}{f_{t-1}} - 1)$ are conditionally perfectly positively correlated (since $\alpha > 0$). Thus, this is also the case for ε_t^2 and $f_t = \omega + \gamma f_{t-1} + \nu_t$. Moreover: $Var[f_t \mid J_{t-1}] = Var[\nu_t \mid J_{t-1}] = \alpha^2 Var[\varepsilon_t^2 \mid J_{t-1}]$ with $\alpha^2 \leq \gamma^2 = (\alpha + \beta)^2$ since $\beta \geq 0$.

Conversely, let us now consider a SR-SARV(1) process ε_t which fulfills the two restrictions of Proposition 3.1. By the first restriction, we know that: $f_t = a_t \varepsilon_t^2 + b_t$, $a_t, b_t \in J_{t-1}$, with $(Var[f_t \mid J_{t-1}])^{1/2} = a_t (Var[\varepsilon_t^2 \mid J_{t-1}])^{1/2}$.

Thus, by the second restriction, we know that a_t is a positive constant α smaller or equal to γ. Therefore: $f_t = \alpha\varepsilon_t^2 + b_t$ and $E[f_t \mid J_{t-1}] = \alpha f_{t-1} + b_t$. By identification with the AR(1) representation of f_t, we conclude that: $b_t = \omega + \beta f_{t-1}$ where $\beta = \gamma - \alpha \geq 0$. Thus: $f_t = \omega + \alpha\varepsilon_t^2 + \beta f_{t-1}$, which proves that f_t is also I_t-adapted (see $0 < \alpha \leq \gamma < 1$). Then we know by Proposition 2.1 that ε_t is also a SR-SARV(1) process w.r.t. I_t and $f_t = Var[\varepsilon_{t+1} \mid I_t]$. Therefore, with: $h_t = f_{t-1} = Var[\varepsilon_t \mid I_{t-1}]$ we do get the GARCH(1,1) representation: $h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.□

Proof of Proposition 3.2. Define u_t as the standardized residuals ($u_t = \varepsilon_t/\sqrt{h_t}$). Straightforward calculus show that all the models can be rewritten as $h_t = \omega + \gamma h_{t-1} + \nu_{t-1}$ with:

GJR: $\omega = \omega_1$, $\gamma = \alpha + \beta + \lambda S$, $\nu_{t-1} = \alpha(\varepsilon_{t-1}^2 - h_{t-1}) + \lambda(S_{t-1}\varepsilon_{t-1}^2 - Sh_{t-1})$, where $S = E[S_t u_t^2 \mid I_{t-1}]$.

Asymmetric GARCH: $\omega = \omega_1 + \alpha\lambda^2$, $\gamma = \alpha + \beta$, $\nu_{t-1} = \alpha(\varepsilon_{t-1}^2 - h_{t-1}) + 2\alpha\lambda\varepsilon_{t-1}$.

Nonlinear Asymmetric GARCH: $\omega = \omega_1$, $\gamma = \alpha(1 + \lambda^2) + \beta$, $\nu_{t-1} = \alpha h_{t-1}(u_{t-1}^2 - 1 + 2\lambda u_{t-1})$.

VGARCH: $\omega = \omega_1 + \alpha(1 + \lambda^2)$, $\gamma = \beta$, $\nu_{t-1} = \alpha(u_{t-1}^2 - 1 + 2\lambda u_{t-1})$.

Heston-Nandi: $\omega = \omega_1 + \alpha$, $\gamma = \alpha\lambda^2 + \beta$, $\nu_{t-1} = \alpha(u_{t-1}^2 - 1 - 2\lambda\varepsilon_{t-1})$.

By the restrictions $E[\varepsilon_{t-1} \mid I_{t-2}] = E[u_{t-1} \mid I_{t-2}] = 0$, $E[\varepsilon_{t-1}^2 \mid I_{t-2}] = h_{t-1}$ and $E[u_{t-1}^2 \mid I_{t-2}] = 1$, we have $E[\nu_{t-1} \mid I_{t-2}] = 0$, that is ε_t is a SR-SARV(1).□

Proof of the equivalence of the two conditions of (3.6) and (3.7). i) We have: $E[\varepsilon_t\varepsilon_{t+1}^2 \mid J_{t-1}] = \sqrt{f_{t-1}}E[u_t E[\varepsilon_{t+1}^2 \mid J_t] \mid J_{t-1}] = \sqrt{f_{t-1}}E[u_t f_t \mid J_{t-1}] = \sqrt{f_{t-1}}E[u_t(\omega + \gamma f_{t-1} + \nu_t) \mid J_{t-1}] = \sqrt{f_{t-1}}E[u_t\nu_t \mid J_{t-1}]$. Hence $E[u_t\nu_t \mid J_{t-1}] = 0 \iff E[\varepsilon_t\varepsilon_{t+1}^2 \mid J_{t-1}] = 0$ since $f_t \neq 0$ almost surely.

ii) We have: $E[\varepsilon_t^3 \mid J_{t-1}] = (f_{t-1})^{3/2}E[u_t^3 \mid J_{t-1}]$. Hence $E[u_t^3 \mid J_{t-1}] = 0 \iff E[\varepsilon_t^3 \mid J_{t-1}] = 0$.□

Proof of Proposition 3.3. The SR-SARV(1) property implies, by Proposition 2.3, that ε_t^2 fulfill the multiperiod restrictions (2.17) with $p = 1$. Define ω_t by $\omega_t = \varepsilon_t^2 - \omega - \gamma\varepsilon_t^2$. We have $E[\omega_t \mid I_{t-2}] = 0$ and ω_t is a square integrable process since ε_t has a finite fourth moment. Therefore ω_t is a weak MA(1) and hence ε_t^2 is a weak ARMA(1,1). Therefore, by Proposition 2.5, ε_t is a weak GARCH(1,1) if and only if (2.22) is fulfilled. But, since by the ARMA representation of ε_t^2, the Hilbert space H_t^s coincides with the Hilbert space spanned by 1, η_τ, $\tau \leq t$, the condition (2.22) is implied by the following symmetry property of the process ε: $Cov(\varepsilon_{t'}, \varepsilon_t^2) = 0 \; \forall \; t, t'$ that is $E(\varepsilon_{t'}\varepsilon_t^2) = 0 \; \forall \; t, t'$. Thus, we are going to prove this symmetry property. Indeed, we will prove the stronger result (which will be useful in the following):

$$E[\varepsilon_{t'}\varepsilon_t^2 \mid J_\tau] = 0 \; \forall \; t, t' \text{ and } \tau = Min(t, t') - 1 \tag{A.5}$$

If $t' > t$, then $E[\varepsilon_{t'}\varepsilon_t^2 \mid J_{t-1}] = E[\varepsilon_t^2 E[\varepsilon_{t'} \mid J_{t'-1}] \mid J_{t-1}] = 0$ since $\varepsilon_{t'}$ is an m.d.s. w.r.t. $J_{t'-1}$.

If $t' = t$, then $E[\varepsilon_{t'}\varepsilon_t^2 \mid J_{t-1}] = E[\varepsilon_t^3 \mid J_{t-1}] = f_{t-1}^{\frac{3}{2}}E[u_t^3 \mid J_{t-1}] = 0$ by (3.7).

If $t' < t$, then $E[\varepsilon_{t'}\varepsilon_t^2 \mid J_{t'-1}] = E[\varepsilon_{t'}f_{t-1}E[u_t^2 \mid J_{t-1}] \mid J_{t'-1}] = E[\varepsilon_{t'}f_{t-1} \mid J_{t'-1}]$. Since f_t is an AR(1), we have $f_{t-1} = \sum_{i=0}^{\infty}\gamma^i\nu_{t-1-i} + E[f_{t-1}]$. Hence $E[\varepsilon_{t'}f_{t-1} \mid J_{t'-1}] = \sum_{i=0}^{\infty}\gamma^i E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}]$.

But: if $i \geq t-t'$, then $E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = \nu_{t-1-i}E[\varepsilon_{t'} \mid J_{t'-1}]] = 0$; if $i = t-t'-1$, then $E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = E[\nu_{t'}\varepsilon_{t'} \mid J_{t'-1}] = \sqrt{f_{t'-1}}E[u_{t'}\nu_{t'} \mid J_{t'-1}] = 0$ by (3.6); finally, if $i < t-t'-1$, then $E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = E[\varepsilon_{t'}E[\nu_{t-1-i} \mid J_{t-i-2}] \mid J_{t'-1}] = 0$ since ν_t is an m.d.s. w.r.t. J_t. Hence, $E[\varepsilon_{t'}f_{t-1} \mid J_{t'-1}] = 0$, which achieves the proof of Proposition 3.3.□

Proof of Proposition 3.4. $E[u_{tm}^{(m)}\nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = \frac{a^{(m)}}{\sqrt{f_{tm-m}^{(m)}}}E[\sum_{0\leq i,j\leq m-1} a_{im}\gamma^j\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}^{(m)}]$. But (see third case of the proof of Proposition 2.3), (3.6) implies that $E[\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}] = 0$ for $i,j = 0,1,...,m-1$. Thus, $E[\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}^{(m)}] = 0$ and hence $E[u_{tm}^{(m)}\nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = 0$, i.e. (3.8).

$E[(u_{tm}^{(m)})^3 \mid J_{tm-m}^{(m)}] = \frac{1}{(f_{tm-m}^{(m)})^{\frac{3}{2}}}E[\sum_{0\leq i,j,k\leq m-1} a_{im}a_{jm}a_{km}\varepsilon_{tm-i}\varepsilon_{tm-j}\varepsilon_{tm-k} \mid J_{tm-m}^{(m)}]$. Let (i,j,k) as $i \leq j \leq k \leq m-1$. If $i < j \leq k$, then $E[\varepsilon_{tm-i}\varepsilon_{tm-j}\varepsilon_{tm-k} \mid J_{tm-m}^{(m)}] = E[\varepsilon_{tm-j}\varepsilon_{tm-k}E[\varepsilon_{tm-i} \mid J_{tm-i-1}] \mid J_{tm-m}^{(m)}] = 0$. If $i = j = k$, then $E[\varepsilon_{tm-i}\varepsilon_{tm-j}\varepsilon_{tm-k} \mid J_{tm-m}^{(m)}] = E[(f_{tm-i-1})^{\frac{3}{2}}E[(u_{tm-i})^3 \mid J_{tm-i-1}] \mid J_{tm-m}^{(m)}] = 0$. If $i = j < k$, then $E[\varepsilon_{tm-i}\varepsilon_{tm-j}\varepsilon_{tm-k} \mid J_{tm-m}^{(m)}] = E[E[\varepsilon_{tm-k}(\varepsilon_{tm-i})^2 \mid J_{tm-m}] \mid J_{tm-m}^{(m)}] = 0$ by (A.5). So we have: $E[(u_{tm}^{(m)})^3 \mid J_{tm-m}^{(m)}] = 0$.□

Proof of Proposition 3.5. The second part of (3.6) implies (3.10). The second part of (3.7) implies (3.11).□

Proof of Proposition 3.6. This is exactly the same proof as for Proposition 2.2 by taking $\Gamma = 1$.