Aggregations and Marginalization

of GARCH and Stochastic Volatility Models [*]

Nour Meddahi [†], Eric Renault [‡]

First Draft March 1996

This Draft November 1996

[*]The authors thank Feike Drost, Rob Engle, Christian Gourieroux, Theo Nijman, Bas Werker, Jean Michel Zakoian, and the participants of the ESEM 1996, Istanbul, the Fourth Workshop on Financial Modeling and Econometric Analysis, Tilburg, December 1996, for their helpful comments. They also acknowledge fruitful discussions during CORE, Malinvaud (CREST-Paris) and University of Montreal seminars. Of course, the authors are the only responsible for the remaining errors.

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

[‡]CREST-INSEE, 15 Bd Gabriel Péri, 92245 Malakoff Cedex, France

Abstract

The GARCH and Stochastic Volatility paradigms are often brought into conflict as two competitive views of the appropriate conditional variance concept: conditional variance given past values of the same series or conditional variance given a larger past information (including possibly unobservable state variables). The main thesis of this paper is that, since in general the econometrician has no idea about something like a structural level of disaggregation, a well-written volatility model should be specified in such a way that one is always allowed to reduce the information set without invalidating the model. To this respect, the debate observable past information (in the GARCH spirit) versus unobservable conditioning information (in the state-space spirit) is irrelevant. We stress in this paper a square-root autoregressive stochastic volatility (SR-SARV) model which remains true to the GARCH paradigm of ARMA dynamics for squared innovations but weakens the GARCH structure in order to obtain required robustness properties with respect to various kinds of aggregation. It is shown that the lack of robustness of the usual GARCH setting is due to two very restrictive assumptions: perfect linear correlation between squared innovations and conditional variance on the one hand, linear relationship between the conditional variance of the future conditional variance and the squared conditional variance on the other hand. By relaxing these assumptions thanks to a state-space setting, we obtain aggregation results without renouncing to the conditional variance concept (and related leverage effects) as it is the case for the recently suggested weak GARCH model which gets aggregation results by replacing conditional expectations by linear projections on symmetric past innovations. Moreover, unlike the weak GARCH literature, we are able to define multivariate models, including higher order dynamics and risk-premiums (in the spirit of GARCH(p,p) and GARCH in mean) and to derive conditional moment restrictions well-suited for statistical inference. Finally, we are able to characterize the exact relationships between our SR-SARV models (including higher order dynamics, leverage effect and in mean effect), usual GARCH models and continuous time stochastic volatility models, so that previous results about aggregation of weak GARCH and continuous time GARCH modeling can be recovered in our framework.

Keywords: GARCH, Stochastic Volatility, SR-SARV, Aggregation, Asset Returns, Diffusion Processes.

1 Introduction

R. Engle (1982) introduced his seminal paper about ARCH modeling via the one-period forecast issue of a random variable y_t: "the forecast of today's value based upon the past information, under standard assumptions, is simply $E[y_t \mid y_{t-1}]$, which depends upon the value of the conditioning variable y_{t-1}. The variance of this one-period forecast is given by $V[y_t \mid y_{t-1}]$. Such an expression recognizes that the conditional forecast variance depends upon past information and may therefore be a random variable."

But, on the outside of the very special case of an univariate Markovian (of order 1) process y_t, there is no reason why to summarize the "past information" by last values y_{t-1} of the dependent variable. More generally, if we assume unbounded memory, forecasts are based on an increasing filtration J_t, $t \in I\!R^+$, of σ-fields such that J_t summarizes the information provided by the observation (until time t) of variables of interest. In other words, if we focus for the moment on the volatility issue in discrete time, we address the issue of modeling the second order dynamics of the martingale difference sequence (m.d.s. hereafter):

$$\varepsilon_t = y_t - E[y_t \mid J_{t-1}].$$

A usual way to specify such a model is to start from the factorization:

$$\varepsilon_t = \sqrt{f_{t-1}} u_t \tag{1.1}$$

where:

$$f_{t-1} \in J_{t-1}$$
$$E[u_t \mid J_{t-1}] = 0, \;\; E[u_t^2 \mid J_{t-1}] = 1. \tag{1.2}$$

It is important to notice that, besides the m.d.s. property, the above factorization does not state any additional assumption but only introduces the notation:

$$f_{t-1} = V[\varepsilon_t \mid J_{t-1}] \tag{1.3}$$

A **volatility model** is then a specification (nonparametric, semiparametric or parametric) of dynamics of the (squared) **volatility process** f_t.

On the other hand, a current criticism against ARCH literature is its apparent lack of any structural foundations, that is of any structural dynamic economic theory explaining the variation of conditional second order moments. Faced with that situation, it is important to propose volatility models that do not violate obvious necessary conditions to have structural interpretations. Among these conditions, robustness with respect to both temporal and contemporaneous aggregation as well as marginalization are fairly crucial since in general situations, the econometrician has no idea about something like a structural level of disaggregation. Therefore we would like specify volatility models for which aggregating or pooling could be innocuous.

Of course, this requirement may complicate statistical inference. The first example of such a conflict between structural interpretation and simple statistical inference is the dynamic specification of higher order conditional moments. It is for instance a current practice to assume that standardized innovations u_t in (1.1) are iid. This is a basic assumption of both semiparametric ARCH modeling à la Engle and González-Rivera (1991) and the general volatility definition of Andersen (1992). This assumption allows one to define estimation methods without taking care of conditional skewness or kurtosis while, in the general setting (1.2), they could matter for efficient estimation (see e.g. Bates and White (1991) and Meddahi and Renault (1995)). But a fundamental contribution of Drost and Nijman (1993) (see Section 3, Example 3, "Strong GARCH are not closed") is precisely to have stressed that the classical ARCH(1) model:

$$\varepsilon_t = \sqrt{\omega + \alpha\varepsilon_{t-1}^2}\, u_t \quad , \quad u_t \;\; iid \;\; N(0,1)$$

is not closed under temporal aggregation since, even if we consider the simplest case of stock variables, lower-frequency rescaled innovations (for instance when ε_t is observed only for odd dates t) involve some non degenerated conditional kurtosis. Therefore it is necessary to extend the usual "strong" GARCH or stochastic volatility models to obtain robustness with respect to various kinds of aggregation. There is nevertheless a general agreement to consider that the first model of volatility dynamics we must have in mind is the simplest AR(1):

$$f_t = \omega + \gamma f_{t-1} + \nu_t \tag{1.4}$$

$$0 < \gamma < 1, \; E[\nu_t] = 0, \; Cov(\nu_t, f_\tau) = 0 \;\forall \tau < t.$$

As a matter of fact, the GARCH(1,1) modeling considers the very special case where:

$$\nu_t = \alpha(\varepsilon_t^2 - f_{t-1}) \tag{1.5}$$

since we then recover from (1.4) the usual GARCH(1,1) representation:

$$f_t = \omega + \alpha\varepsilon_t^2 + \beta f_{t-1}, \; 0 < \alpha + \beta < 1. \tag{1.6}$$

Indeed, it is clear that the choice $\nu_t = \alpha(\varepsilon_t^2 - f_{t-1})$ is allowed since:

$$E[\varepsilon_t^2 \mid J_{t-1}] = f_{t-1} \Rightarrow E[\nu_t \mid J_{t-1}] = 0 \Rightarrow E[\nu_t] = 0 \text{ and } Cov(\nu_t, f_\tau) = 0 \;\forall \tau < t.$$

We get then the concept of semi-strong GARCH(1,1) model as it is defined by Drost and Nijman (1993). Unfortunately, Drost and Nijman (1993) have also shown (see Section 3, Example 4, "Semi-strong GARCH not closed") that even this extended class of GARCH processes is not closed under temporal aggregation.

Then, it turns out that we have to extend even more the class of GARCH(1,1) processes if we want to get some structural interpretations. The main idea of the present paper is to offer an alternative to the Drost and Nijman (1993) extension, which presents in our opinion several advantages.

On the one hand, the Drost and Nijman (1993) idea is to weaken the structure:

$$E[\nu_t \mid J_{t-1}] = 0$$

$$\nu_t = \alpha(\varepsilon_t^2 - f_{t-1}) \text{ and thus } J_t = \sigma(\varepsilon_\tau, \tau \leq t)$$

by maintaining the restriction

$$\nu_t = \alpha(\varepsilon_t^2 - f_{t-1}) \tag{1.7}$$

but assuming only:

$$E[\nu_t] = 0 \text{ and } Cov(\nu_t, f_\tau) = 0 \ \forall \tau < t$$

instead of the non-robust conditional moment restrictions:

$$E[\nu_t \mid J_{t-1}] = 0. \tag{1.8}$$

The advantage of this proposal is to focus on the linear structure of GARCH modeling: the usual GARCH representation:

$$\varepsilon_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta f_{t-1}$$

may then be interpreted as a linear projection on the Hilbert space H_t^s spanned by 1 and $\varepsilon_\tau^2, \tau \leq t$. This is the so-called weak GARCH(1,1) model. The main drawback of this approach is that by relaxing (1.8) into the setting (1.7), one renounces to interpret f_{t-1} as a conditional variance of ε_t! This is a pity for both financial interpretation of volatility (what could be for instance the use of linear projections for option pricing which is, by definition, highly nonlinear!) and statistical inference (without conditional moment restrictions, we lose the consistency of usual Quasi Maximum likelihood or GMM).

On the other hand, our main proposal is the following: we remain true to the conditional moment restrictions

$$E[\nu_t \mid J_{t-1}] = 0$$

but we no longer consider that ν_t is perfectly linearly correlated with ε_t^2 conditionally to J_{t-1} (according to (1.7)); it may involves a separate contemporaneous stochastic component, and in that case, according to the usual terminology (see Andersen (1992)), we say that volatility is stochastic.

In order to compare the weak GARCH modeling with our stochastic volatility approach, it is first important to stress that weak GARCH(1,1) processes which are outside the usual semi-strong GARCH class **do involve stochastic volatility features**. Since we only know that:

$$EL[\varepsilon_{t+1}^2 \mid H_t^s] = \omega + \gamma EL[\varepsilon_t^2 \mid H_{t-1}^s]$$

where EL denotes the linear projection, $V[\varepsilon_{t+1} \mid J_t] = E[\varepsilon_{t+1}^2 \mid J_t]$ will generally involve some nondegenerated random variables like:

$$E[\varepsilon_{t+1}^2 \mid J_t] - EL[\varepsilon_{t+1}^2 \mid H_t^s].$$

In other words, weak GARCH processes and stochastic volatility models are not necessarily inconsistent extensions of the semi-strong GARCH class. The choice between the two approaches has been done in relationship with the objective: financial interpretations, statistical inference, robustness with respect to various kinds of aggregation... But we want to argue here that besides its advantages for financial interpretation and statistical inference, the stochastic volatility approach is even better suited for aggregation issues: while Drost and Nijman (1993) are led to restrict themselves to symmetric GARCH processes, no such restrictions are needed within our stochastic volatility framework. This degree of freedom matters to capture the so-called leverage effect, already well documented in the stock price literature (see Black (1976), Nelson (1991), Bollerslev, Engle and Nelson (1994)). Indeed, we check that if we renounce to capture this leverage effect by adding to our setting some symmetry restrictions à la Drost and Nijman (1993), we are led back to weak GARCH processes. Moreover, while Drost and Werker (1996) suggest to "close the GARCH Gap" by relating continuous time stochastic volatility models with discrete time weak GARCH modeling, the stochastic volatility framework allows us to a more direct relation between continuous time and discrete time models.

The paper is organized as follows.
The sections 2 focuses on the so-called SR-SARV(1) class, that is to say the family of volatility models conformable to (1.8) in the framework (1.1) to (1.4). We detail its relation with both discrete time GARCH(1,1) models (semi-strong or weak) and continuous time stochastic volatility models. Indeed, our SR-SARV(1) class is closed

under temporal aggregation and may be viewed as a discrete time sampling from usual continuous time stochastic volatility models. Moreover, we stress that temporal aggregation of semi-strong GARCH(1,1) processes does create in general stochastic volatility features or time-varying coefficients which are hidden in the weak GARCH representation. In the same way, discretization of univariate heteroskedastic diffusions may create stochastic volatility features. This is the reason why the stochastic volatility model is, in our opinion, the most versatile tool.
The section 3 propose various extensions of the SR-SARV(1) model.

In order to get rid of richer correlation patterns in conditional variances than the simplest AR(1) case, we first introduce in section 3.1 the SR-SARV(p) class which corresponds to GARCH(p,p) as SR-SARV(1) corresponds to GARCH(1,1). However, to remain true to the Markovian paradigm of state variables which is dominating in modern Finance, the order-p structure is obtained by marginalization of a p-dimensional VAR(1) process of state variables. This allows us to extend to higher orders, on the one hand the main result about temporal aggregation and on the other hand, the relationship with continuous time stochastic volatility diffusion models.

Besides, to capture some structural interpretations of asset returns time series models, and particularly of dynamic models of first order and second order conditional moments, we need some versatile statistical structures where conditional expectations, conditional variances and covariances may be combined through linear aggregators to characterize risk premia. This is the main puzzle which motivates the sequel of section 3. More precisely, we want to extend the basic SR-SARV(p) model in order to get on the one hand statistical models able to capture structural restrictions about risk premia dynamics and, on the other hand which remain true to the requirement of robustness w.r.t. temporal aggregation. After a brief discussion of such structural restrictions (section 3.2), we introduce in a given asset risk premium another marginalization of the p-dimensional state variable process already used to characterize the conditional variance dynamics. This is the concept of SR-SARV-M (that is SR-SARV in mean) which as presented in subsection 3.3 adapts to our setting the usual ARCH-M model introduced by Engle, Lilien and Robins (1987). As far as we are concerned by robustness w.r.t. temporal aggregation, the SR-SARV-M model is shown to be not only closed under changes of the sampling frequency but also under discretizations where Ito's lemma may introduce squared volatilities in the drift of the processes. Moreover, it is shown that temporal aggregation of the SR-SARV-M processes should introduce some kind of leverage effect, through the volatility features of unobserved risk premia. Besides the corporate finance-based Black's argument, this could provide an additional explanation for the well-documented evidence of leverage effect in financial time series.

Subsection 3.4 is concerned with SV models with a predictable part, which may be itself generated by a VAR(1) models of several factors. Indeed, the complete multivariate model of factors which may appear in both conditional expectations, variances and covariances of a vector of returns is described in Section 4.

The proofs of the main results are provided in the Appendix.

2 The SR-SARV(1) Class

2.1 GARCH versus Stochastic Volatility

Let us consider a martingale difference sequence (m.d.s. hereafter) ε_t, t=1,2,... adapted to the natural filtration

$$I_t = \sigma(\varepsilon_\tau, \tau \leq t). \tag{2.1}$$

Typically, ε_t could be the log-return of a given asset whose price at time t is denoted by S_t:

$$\varepsilon_t = Log \frac{S_t}{S_{t-1}}. \tag{2.2}$$

If S_t is the value at time t of a given currency in terms of dollar, the martingale difference hypothesis is generally accepted, according to the efficiency of foreign exchange markets. More generally, if we consider that log-return could have non-zero conditional expectation, the framework (2.2) may be extended by considering ε_t as the innovation process of the log-return process (see section 3).

Indeed, we follow here the usual practice (according to the mainstream GARCH and SV literature) of introducing the conditional heteroskedasticity setting at the level of innovations processes. As far as we are concerned with GARCH modeling, the basic idea was to introduce a serial linear correlation pattern through ε_t^2 (see Engle (1982), Bollerslev (1986)) while, by definition of the martingale difference property, ε_t does not involve such correlation:

$$E[\varepsilon_t \mid I_{t-1}] = 0 \tag{2.3}$$

This is surely the main reason for the widespread use of the GARCH(1,1) model. If we denote by h_t the conditional variance process:

$$h_t = E[\varepsilon_t^2 \mid I_{t-1}] \tag{2.4}$$

the usual GARCH(1,1) setting:

$$h_t = w + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1},\ \alpha > 0,\ \beta \geq 0,\ \alpha + \beta < 1, \tag{2.5}$$

may be simply rearranged in:

$$\varepsilon_t^2 = w + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta(\varepsilon_{t-1}^2 - h_{t-1}) + (\varepsilon_t^2 - h_t) \tag{2.6}$$

which stresses the ARMA(1,1) structure [1] of ε_t^2 with innovation process

$$\eta_t = \varepsilon_t^2 - h_t. \tag{2.7}$$

Moreover, let us recall that the ARMA(1,1) representation of ε_t^2 is tantamount to an AR(1) representation of the conditional variance process h_t (see (2.4)/(2.5)):

$$h_t = \omega + (\alpha + \beta)h_{t-1} + \alpha\eta_{t-1}. \tag{2.8}$$

As a matter of fact, we want to stress here that, if it is natural to use the ARMA(1,1) setting to summarize the pattern of serial correlation of ε_t^2, the framework (2.6) is too restrictive since the innovation process η_t is not only serially uncorrelated but is a m.d.s. (due to (2.4)). This restrictive feature involves several shortcomings as a lack of robustness with respect to temporal aggregation, contemporaneous aggregation, marginalization... This is the reason why it would be interesting to remain true to the ARMA pattern of serial correlation but with a general weak white noise innovation. Indeed, the serial uncorrelation assumption (in a linear sense) does not involve the same drawbacks that the martingale difference one. This is the reason why this "weak" concept was used by Drost-Nijman (1993) and Nijman-Sentana (1996) to derive aggregation properties in a GARCH setting.

As announced in the introduction, the main goal of this paper is to define a semi-parametric model of conditional heteroskedasticity which on the one hand corresponds to a weak concept of ARMA process for ε_t^2 (in order to obtain robustness properties) but on the other hand remains true to the same martingale concepts. These concepts matter for at least two reasons: statistical inference and logical relationship with continuous time models of modern Finance. This is the reason why we suggest to forsake the GARCH framework (the exact relation with weak GARCH modeling will be made more precise later) and to define the parameters of interest in the following SV setting:

[1]R. Engle (1995), page xiii, explains with humor why, according to (2.5) and (2.6), "ARCH is autoregressive while GARCH is ARMA", in contradiction to his first intuition that the main drawback of the ARCH specification was that "it appeared to be more a moving-average specification than an autoregression".

Definition 2.1

A stationary squared integrable process $\varepsilon_t,\ t \in \mathbf{N}$, is called a **SR-SARV(1) process** with respect to an increasing filtration J_t, $t \in \mathbf{N}$, if there exists a positive J_t-adapted stationary $AR(1)$ process f_t, $t \in \mathbf{N}$, such that:

$$\varepsilon_t = \sqrt{f_{t-1}} u_t \tag{2.9.a}$$

$$E[u_t \mid J_{t-1}] = 0 \tag{2.9.b}$$

$$E[u_t^2 \mid J_{t-1}] = 1 \tag{2.9.c}$$

Of course, the definition 2.1 is related to the so-called square-root stochastic autoregressive volatility setting introduced by Andersen (1994) (which justifies the notation SR-SARV) since by (2.9), the process f_t can be interpreted as the **conditional variance process**:

$$f_{t-1} = Var[\varepsilon_t \mid J_{t-1}], \tag{2.10}$$

and this process admits by definition an $AR(1)$ representation

$$f_t = \omega + \gamma f_{t-1} + \nu_t \tag{2.11.a}$$

$$E[\nu_t \mid J_{t-1}] = 0 \tag{2.11.b}$$

Since stationarity is assumed, we have implicitly assumed that $\mid \gamma \mid < 1$, and in order to provide simple sufficient conditions of positivity, a maintained assumption hereafter will be

$$\omega > 0 \text{ and } 0 < \gamma < 1. \tag{2.12}$$

This AR(1) representation of the conditional variance process remains true to the basic idea of GARCH modeling (see (2.8)) but is stated in a more general framework since nothing is assumed about the information set (observable or not, contains past returns or not...). The logical relationships between SR-SARV and GARCH will be detailed below. Let us first notice moreover that, by (2.9):

$$E[\varepsilon_t \mid J_{t-1}] = 0.$$

In other words, ε_t will be a m.d.s. adapted to J_t if we set the additional assumption that J_t contains past returns:

$$I_t \subset J_t. \tag{2.13}$$

But we are able to prove that one always may reduce the information sets of the SR-SARV representation, even without taking into account the restriction (2.13):

Proposition 2.1 *If ε_t is a SR-SARV(1) process w.r.t. a filtration J_t and f_t the associated conditional variance process, then, for any subfiltration $J_t^* \subset J_t$, ε_t is a SR-SARV(1) process with respect to J_t^* and the associated conditional variance process f_t^* is given by*

$$f_t^* = E[f_t \mid J_t^*].$$

This Proposition 2.1 stresses the fact that, in contradiction with a common idea, the difference between SV and GARCH is not characterized in terms of "observability" of the volatility process. Of course, one is always allowed to consider an information set which is reduced to a σ-field J_t^* spanned by variables which were observed at time t, but the corresponding SR-SARV(1) representation (with $f_t^* = E[f_t \mid J_t^*]$, $\nu_t^* = f_t^* - \omega - \gamma f_{t-1}^*$) is not necessarily a GARCH(1,1) model, since the AR(1) representation of the conditional variance process is not necessarily linear w.r.t. observables, according to (1.5):

$$\nu_t^* = \alpha(\varepsilon_t^2 - f_{t-1}^*)$$

Indeed, this linearity conditional to J_{t-1}^* with a constant slope coefficient α is characteristic property of GARCH models as stated by Proposition 2.2 below. But we want to stress here, that, even outside of the GARCH class, our class of SV models is not affected by the criticism against Log-normal SV models (see section 2.2 below) rightly formulated by Engle (1994). While, for Log-normal SV models, it is right to claim that the conditional variance given observables "has a form that is not easy to evaluate except possibly by simulation methods", the SR-SARV(1) representation provides a conditional variance process which is AR(1), whatever the information set, including an information filtration defined only by past returns.

Let us notice that until now, we have not explicitly assumed that the information set J_t contains at least the past returns. This is however a natural assumption which facilitates several interpretations below: relation with ARCH models, leverage effect... This is the reason why this assumption will be maintained in all the rest of the paper and, except in case of ambiguity, the terminology "SR-SARV(1)" without specifying the benchmark filtration will be used for cases where this filtration J_t contains the "returns filtration" I_t. Let us notice that, since f_t is J_t-adapted by definition, we then have for any SR-SARV(1) model:

$$J_t^0 \subset J_t \tag{2.14.a}$$

where

$$J_t^0 = \sigma(\varepsilon_\tau, f_\tau, \tau \leq t) = \sigma(u_\tau, \nu_\tau, \tau \leq t). \tag{2.14.b}$$

is the natural filtration. We are then able to state:

Proposition 2.2 *The class of semi-strong GARCH(1,1) processes ε_t (defined by (2.3), (2.4) and (2.5) with a conditional variance process h_t) coincides with the subclass of the SR-SARV(1) processes (defined by (2.9) with a squared volatility process f_t) which verifies:*

- *First, ε_t^2 and f_t are conditionally perfectly positively correlated given J_{t-1} (conditional linear correlation equal to 1).*
- *Second, the ratio $\frac{Var[f_t|J_{t-1}]}{Var[\varepsilon_t^2|J_{t-1}]}$ is constant and smaller or equal to γ^2.*

In this case:

$$h_{t+1} = f_t,\ \gamma = \alpha+\beta \text{ and } J_t = \sigma(\varepsilon_\tau, f_\tau, \tau \le t) = \sigma(\varepsilon_\tau, \tau \le t) \text{ with } \alpha = [\frac{Var[f_t \mid J_{t-1}]}{Var[\varepsilon_t^2 \mid J_{t-1}]}]^{1/2}.$$

The first restriction is related to the common idea that ARCH models correspond to the degenerate case where there are no exogenous source of randomness in the conditional variance dynamics. As discussed above through reduction (by Proposition 2.1) of information sets, the invocated degeneracy corresponds to GARCH only if it is a perfect linear correlation.

Moreover the second restriction introduced by Proposition 2.2 is less known even though it was already coined by Nelson and Foster (1994) (pages 21-22): "most commonly-used ARCH models effectively assume that "the variance of the variance" rises linearly with σ_t^4", that is $Var[f_t \mid J_{t-1}]$ is proportional to $Var[\varepsilon_t^2 \mid J_{t-1}]$, which is itself proportional to f_{t-1}^2 (that is σ_t^4) in case of strong GARCH (that is iid standardized innovations u_t). In other words, the semi-strong or strong GARCH setting implies non-trivial restrictions on the conditional kurtosis dynamics; this last remark was the source of a Drost and Nijman (1993) counterexample of lack of robustness w.r.t. temporal aggregation.

As far as we are concerned by the first restriction, it provides some insight on the ability of the various models to capture the so-called leverage effect phenomenon. This effect, first stressed by Black (1976), refers to the well-documented evidence that, for various asset prices, bad news and good news of the same amplitude have not the same effect on subsequent volatilities. Actually, it appears that, according to the theory of the levered firm, stock price volatilities raise relatively more after bad news than after good ones. In any case, a versatile stochastic volatility model should be able to

capture asymmetric responses of the return $\varepsilon_t = \sqrt{f_{t-1}} u_t$ to some shocks in previous information J_{t-1} or contemporaneous volatility $\sqrt{f_t}$.
Two types of asymmetry could be imagined:

- Either, the conditional skewness of u_t given J_{t-1} is non-zero.
- Or, the conditional correlation of u_t with f_t (given J_{t-1}) is non-zero.

According to the notations (2.9) and (2.11) we then define:

Definition 2.2
Leverage effect occurs as soon as one of the two following properties is fulfilled:

$$E[u_t^3 \mid J_{t-1}] \neq 0 \tag{2.15.a}$$

or

$$E[u_t \nu_t \mid J_{t-1}] \neq 0. \tag{2.15.b}$$

Let us notice that we have defined leverage effect w.r.t. conditional probability distributions, which appears more conformable to the idea of responses to shocks. However, unconditional leverage effect in terms of observables ($E\varepsilon_t^3 \neq 0$ or $E\varepsilon_t \varepsilon_{t+1}^2 \neq 0$) implies conditional leverage effect.

Let us consider now processes which, as semi-strong GARCH(1,1) fulfill the first restriction of Proposition 2.2. In this case, $f_t = a_{t-1}\varepsilon_t^2 + b_{t-1}$, with a_{t-1}, $b_{t-1} \in J_{t-1}$, so that:

$$E[u_t \nu_t \mid J_{t-1}] = E[u_t f_t \mid J_{t-1}] = a_{t-1} E[u_t \varepsilon_t^2 \mid J_{t-1}] = a_{t-1} f_{t-1} E[u_t^3 \mid J_{t-1}]$$

so that:

Proposition 2.3 *For any SR-SARV(1) process which fulfills the first restriction of Proposition 2.2 (for instance a semi-strong GARCH(1,1) process), the two conditions of leverage effect (2.15.a) and (2.15.b) are equivalent.*

It is worthwhile to notice that, according to Proposition 2.3 and in contradiction with a common idea, usual GARCH(1,1) process may involve some kind of leverage effect in case of asymmetric innovations, that is:

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

Let us notice that such a leverage effect is not spurious since we are able to check more generally that a reduction of information in the sense of Proposition 2.1 cannot introduce spurious leverage effect:

Proposition 2.4 *If ε_t is a SR-SARV(1) process w.r.t. a filtration J_t, than, for any subfiltration J_t^* such that $I_t \subset J_t^* \subset J_t$, we have with obvious notations:*

$$E[u_t^3 \mid J_{t-1}] = 0 \Longrightarrow E[u_t^{*3} \mid J_{t-1}^*] = 0$$
$$E[u_t \nu_t \mid J_{t-1}] = 0 \Longrightarrow E[u_t^* \nu_t^* \mid J_{t-1}^*] = 0.$$

Actually, if for instance we observe that a GARCH(1,1) process is such that $E[\varepsilon_t^3 \mid I_{t-1}] \neq 0$, we are able to claim that any associated SR-SARV(1) modeling of ε_t should involve usual leverage effect in the sense of (2.15.b). This is the reason why we do consider the conditional skewness of ε_t as a genuine occurrence of leverage effect.

2.2 Statistical Issues

Since Taylor (1986) seminal work, one observes a burgeoning literature about stochastic volatility models such that the terminology is still not well-established (see Ghysels, Harvey and Renault (1996) for a survey). In order to place our SR-SARV(1) concept in relation to the available literature, several properties have to be emphasized:

• First, even in the general case where it is not J_t-adapted, the process ε_t fulfills the m.d.s. property:

$$E[\varepsilon_t \mid J_{t-1}] = 0.$$

This would not be the case if we considered, as Taylor (1994) the so-called "contemporaneous autoregressive random variance model" $\varepsilon_t = \sqrt{f_t} u_t$. Let us recall that this does not prevent us to consider occurrences of leverage effect.

• This leverage effect, or more generally the empirical evidence of asymmetry in the relationship return/volatility has led D. Nelson (1991) to propose the exponential GARCH or EGARCH as an alternative to the usual GARCH setting. On the other hand, the log-normal stochastic variance model introduced by Taylor (1986):

$$Log f_t = \omega + \gamma Log f_{t-1} + \nu_t, \tag{2.16}$$

$$\nu_t \ iid \ N(0, \sigma^2).$$

is the natural SV analogue of EGARCH models. It is the most popular SV model since Harvey, Ruiz and Shephard (1994) have popularized it by exploiting its linear state space form. In particular, the exponential form of EGARCH and (2.16) simplifies inference since non-positive variances are automatically excluded. However, as noticed by Engle (1995, page xiii), "it has the drawback that forecasts of variance require a numerical simulation or at least a distributional assumption which is not the case for linear models". Indeed, as already stressed, our SR-SARV(1) models preserves the linear AR(1) structure of the conditional variance process emphasized by R. Engle as a distinctive feature of GARCH processes. This is the reason why we are able to forecast variances in a **semiparametric** framework, without distributional assumptions on the error terms. Besides robustness, the distribution-free framework is crucial to get convenient properties of aggregation. Moreover, we have shown that, in contradiction with a common idea, leverage effect can be captured in linear settings.

As far as linearity is concerned, we know that the AR(1) representation of the conditional variance process corresponds to an ARMA(1,1) representation of ε_t^2. As already explained, the usual semi-strong GARCH setting assumes that the innovation process of the ARMA(1,1) model is a m.d.s. while the weak GARCH setting only involves white noises in a weak (second order) sense. The point we want to stress here, is that, by introducing the SR-SARV(1) modeling (2.9), we remain true to an ARMA(1,1) modeling for ε_t^2, but with some additional restrictions w.r.t. a weak concept.

Proposition 2.5 *If ε_t admits a SR-SARV(1) representation, then ε_t^2 is a weak stationary ARMA(1,1) process:*

$$\varepsilon_t^2 - \omega - \gamma\varepsilon_{t-1}^2 = \omega_t \tag{2.17}$$

where ω_t is a MA(1) process such that:

$$E[\omega_t \mid J_{t-2}] = 0 \tag{2.18}$$

Moreover, ω_t admits a MA(1) representation $\omega_t = \eta_t - \beta\eta_{t-1}$, η_t white noise, $\beta < 1$.

It is worthwhile to notice that:

$$J_t \supset I_t = \sigma(\varepsilon_\tau, \tau \leq t) \supset \sigma(\varepsilon_\tau^2, \tau \leq t) = \sigma(\omega_\tau, \tau \leq t)$$

In particular, the property (2.18) is much more powerful than the usual definition of a weak ARMA(1,1) process:

$$Cov(\omega_t, \omega_{t-h}) = 0 \quad \forall \;\; h \;\geq\; 2$$

Beside its interpretation close to a martingale difference property, the property (2.18) may be useful for statistical inference through GMM (see Drost, Meddahi and Renault (1996)) because it implies the following observable conditional moment restrictions:

$$E[\varepsilon_t^2 - \omega - \gamma\varepsilon_{t-1}^2 \mid I_{t-2}] = 0 \tag{2.19}$$

Indeed, (2.19) is an example of application of a general class of lagged conditional moment restrictions studied by Hansen and Singleton (1996).

This may produce estimates of the parameters ω and γ of interest for the SV dynamics (2.9) much more accurate than unconditional moments based estimates of the weak ARMA model of ε_t^2. On the other hand, the usual m.d.s. restriction (2.4) on the innovation process of ε_t^2 (which of course implies (2.19)) is too restrictive because it corresponds to the concept of semi-strong GARCH(1,1) model (see Drost and Nijman (1993)) which is not robust w.r.t. temporal aggregation. In the opposite, we shall see in the following subsection 2.3 that the SR-SARV(1) model (2.9) is robust w.r.t. temporal aggregation.

To summarize, the SR-SARV(1) concept provides sufficient restrictions to remain true to some martingale concepts of the semi-strong GARCH class ($E[\varepsilon_t \mid I_{t-1}] = 0$ and $E[\omega_t \mid I_{t-2}] = 0$) but a sufficiently larger class (indeed strictly larger than semi-strong GARCH class) to ensure robustness w.r.t. temporal aggregation.

As already announced, Drost and Nijman (1993) have proposed an other "weakening" of the semi-strong GARCH concept through the so-called weak GARCH property. The main idea is to ensure robustness w.r.t. temporal aggregation by using only **linear** concepts of information sets. More precisely, let us consider the Hilbert space H_{t-1} spanned by the constant, ε_τ and ε_τ^2, $\tau < t$. Drost and Nijman's (1993) weak GARCH definition is:

$$EL[\varepsilon_t \mid H_{t-1}] = 0 \tag{2.20.a}$$

$$EL[\varepsilon_t^2 \mid H_{t-1}] = \omega + \alpha\varepsilon_{t-1}^2 + \beta EL[\varepsilon_{t-1}^2 \mid H_{t-2}] \tag{2.20.b}$$

$$0 < \alpha + \beta < 1 \tag{2.20.c}$$

which is conformable to the usual definition (2.3), (2.4), (2.5) but with usual concepts of conditional expectation replaced by the concept $EL[. \mid H_{t-1}]$ of linear projection on the Hilbert space H_{t-1}. With such a definition, the weak GARCH class clearly encompasses the semi-strong one. Indeed, one may interpret the weak GARCH concept proposed by Drost and Nijman (1993) as the following:

Proposition 2.6 *If ε_t is a weak GARCH(1,1) process, than ε_t^2 is a weak stationary ARMA(1,1) process.*
Conversely, if ε_t^2 is a weak stationary ARMA(1,1) process:

$$\varepsilon_t^2 - \omega - \gamma\varepsilon_{t-1}^2 = \eta_t - \beta\eta_{t-1} \tag{2.21}$$

with η_t weak white noise , ε_t is a weak GARCH(1,1) if and only if:

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

In this case, the representations (2.20) and (2.21) are related by parameters ω, β and $\gamma = \alpha + \beta$. Moreover, $EL[\varepsilon_t^2 \mid H_{t-1}]$ coincides with the linear optimal prediction $EL[\varepsilon_t^2 \mid H_{t-1}^s]$ where H_{t-1}^s is the Hilbert space spanned by the constant and $\varepsilon_\tau^2, \tau < t$.

Of course, the Proposition 2.6 is stated with the maintained assumption (2.3) which implies (2.20.a). Moreover, it is worth noting that (2.22) has something in common with (2.18) since (2.18) means that:

$$E[\eta_t - \beta\eta_{t-1} \mid \varepsilon_\tau, \nu_\tau, \tau < t-1] = 0$$

which in turn implies:

$$Cov(\eta_t - \beta\eta_{t-1}, \varepsilon_\tau) = 0, \ \forall \tau < t-1.$$

Unfortunately, to obtain their temporal aggregation result, Drost and Nijman (1993) are led to restrict themselves to the class of "symmetric weak GARCH processes". But we want to argue here that, insofar as one needs to impose symmetry, the weak GARCH concept is not the convenient one to extend the semi-strong GARCH class since it precludes a leverage effect phenomenon which can be captured by general GARCH models.

Indeed, as soon as the following symmetry assumption is maintained :

$$(\varepsilon_t, \varepsilon_{t'}) \text{ and } (-\varepsilon_t, -\varepsilon_{t'}) \text{ have the same probability distribution for any } (t, t'). \tag{2.23}$$

the weakest notion of leverage effect is precluded in a weak GARCH framework [2] since (2.23) implies that:

$$E[EL[\varepsilon_{t+1}^2 \mid H_t]\varepsilon_t] = 0$$

because:

$$EL[\varepsilon_{t+1}^2 \mid H_t] = EL[\varepsilon_{t+1}^2 \mid H_t^s]$$

and

$$E[\varepsilon_\tau^2 \varepsilon_t] = 0, \ \forall \tau.$$

Moreover, as soon as the above leverage effects are precluded, the weak GARCH concept provides no more structure than the SR-SARV(1) one since it can be proved that:

Proposition 2.7 *If ε_t is a SR-SARV(1) process defined by (2.9-11) with the additional restrictions:*

$$E[u_t^3 \mid J_{t-1}] = 0 \tag{2.24.a}$$

$$E[u_t \nu_t \mid J_{t-1}] = 0 \tag{2.24.b}$$

then ε_t is a weak GARCH(1,1) process.

[2] In a weak GARCH framework, the leverage effect should be characterized by the (unconditional) linear correlation between ε_t and $EL[\varepsilon_{t+1}^2 \mid H_t]$.

To summarize, we claim that the general SR-SARV(1) concept is the convenient structure in order to:

- First, capture leverage effects.
- Second, extend the semi-strong GARCH(1,1) class.
- Third, allow temporal aggregation.

On the opposite, the weak GARCH(1,1) class provides the last two properties only by excluding the first one. Moreover, we shall show in subsection 2.3 below that, not only the SR-SARV(1) representation, but also the restrictions (2.24) which relate this class to the weak GARCH one, are robust w.r.t. temporal aggregation.
Moreover, it is worthwhile to notice that the symmetry assumption used by Drost and Nijman (1993) is even stronger than (2.23) since it assumes that :

For any positive integer h, for any h-uplet $(a_k)_{1\leq k\leq h} \in \{-1,1\}^h$,

$$(\varepsilon_{t+k})_{1\leq k\leq h} \textit{ and } (a_k\varepsilon_{t+k})_{1\leq k\leq h} \textit{ have the same probability distribution.} \tag{2.25}$$

Amazingly, if we are ready to maintain (2.25) in a semi-strong GARCH framework, we can always state a degenerated SR-SARV representation of our process

$$\varepsilon_t = \sqrt{f_{t-1}}u_t$$

with $f_{t-1} = \varepsilon_t^2$ and u_t is the sign of ε_t. Indeed, we are able to check that $E[u_t \mid J_{t-1}] = 0$ since, in this case, J_{t-1} is generated by $\varepsilon_\tau,\ \tau < t$, and ε_t^2, and thanks to (2.25), u_t and ε_t^2 are conditionally independent, given $\varepsilon_\tau,\ \tau < t$.

2.3 Temporal aggregation of SR-SARV(1) process:

As suggested by (2.2), we have in mind the interpretation of ε_t as a continuously compounded rate of return over the period $[t-1,t]$. Since the unit of time of the sampling interval is to a large extent arbitrary, we would surely want the SV model defined by equations (2.9) to be closed under temporal aggregation. As rates of return are flow variables, closeness under temporal aggregation means that for any integer m:

$$\varepsilon_{tm}^{(m)} = Log\frac{S_{tm}}{S_{tm-m}} = \sum_{k=0}^{m-1}\varepsilon_{tm-k}$$

is again conformable to a model of the type (2.9) with suitably adapted parameter values.

However, to encompass too the case of stock (or prices) variables we consider more generally:

$$\varepsilon_{tm}^{(m)} = \sum_{k=0}^{m-1} a_{km}\varepsilon_{tm-k} \tag{2.26}$$

Typically, for stock variables observed at the dates m, 2m, 3m, ..., tm:

$$a_{0m} = 1 \quad and \quad a_{km} = 0 \ \ \forall k > 0$$

while, for flow variables:

$$a_{km} = 1 \ \ \forall k = 0, 1, .., m-1.$$

We are then able to state:

Proposition 2.8 *Let ε_t a SR-SARV(1) process w.r.t. J_t and a conditional variance process f_t with innovation ν_t. If for a given natural integer m, we consider a filtration $J_{km}^{(m)}$, $k \in \mathbf{N}$, such that for any k:*

$$\varepsilon_{km}^{(m)} \in J_{km}^{(m)} \subset J_{km},$$

then *$\varepsilon_{tm}^{(m)}$ is a SR-SARV(1) process w.r.t. $J_{tm}^{(m)}$, with a conditional variance process $f_{tm}^{(m)}$:*

$$f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}].$$

We have:

$$\varepsilon_{tm}^{(m)} = \sqrt{f_{tm-m}^{(m)}}u_{tm}^{(m)}$$

$$f_{tm}^{(m)} = \omega^{(m)} + \gamma^{(m)} f_{tm-m}^{(m)} + \nu_{tm}^{(m)}$$

and

$$\gamma^{(m)} = \gamma^m$$

$$\omega^{(m)} = a^{(m)}\omega\frac{1-\gamma^m}{1-\gamma} + b^{(m)}(1-\gamma^m)$$

where

$$a^{(m)} = \sum_{k=0}^{m-1} a_{km}^2 \gamma^{m-k-1}$$

$$b^{(m)} = \frac{\omega}{1-\gamma}\sum_{k=0}^{m-1} a_{km}^2 (1-\gamma^{m-k-1})$$

To give the intuition of the proof of Proposition 2.8, it is worthwhile to notice that $f^{(m)}_{tm-m}$ is the optimal prediction, given $J^{(m)}_{tm-m}$ available at time $tm-m$, of the aggregated squared volatility $\sum_{i=0}^{m-1} a^2_{im} f_{tm-i-1}$. Therefore, taking into account the Markovian feature of f_t, we see that:

$$f^{(m)}_{tm} = a^{(m)} E[f_{tm} \mid J^{(m)}_{tm}] + b^{(m)}. \tag{2.27}$$

To a certain extent, (2.27) is counterintuitive since it shows that, although volatility of returns is aggregated as a flow variable ($\sum_{i=0}^{m-1} a^2_{im} f_{tm-i-1}$ with $a_{im} = 1, \ \forall i$), its informational content appears as a stock variable, at least in the case of low-frequency observation scheme of the stock type ($f_{tm} \in J^{(m)}_{tm}$). For instance, in this case, to price at time tm an option written on the asset of price S, the volatility $f^{(m)}_{tm}$ corresponding to the low frequency data has the same informational content that the volatility f_{tm} of high frequency data [3].

As far as we are concerned with weak GARCH processes which are SR-SARV (which is the case for semi-strong GARCH), our Proposition 2.6 extends the results of Drost and Nijman's (1993) section 3 about temporal aggregation of GARCH(1,1).
In a first case, we consider stock variables and obtain:

$$\gamma^{(m)} = \gamma^m$$

$$\omega^{(m)} = \omega \frac{1-\gamma^m}{1-\gamma}$$

which in GARCH notations provides Drost and Nijman's (1993) equation (9):

$$\alpha^{(m)} + \beta^{(m)} = (\alpha+\beta)^m$$

$$\omega^{(m)} = \omega \frac{1-(\alpha+\beta)^m}{1-(\alpha+\beta)}.$$

In the second case of flow variables, we have

$$\gamma^{(m)} = \gamma^m$$

[3] (2.27) proves that, in the particular case $f_{tm} \in J^{(m)}_{tm}$, if we consider the more flexible SV representation: $\varepsilon_t = \sqrt{a f_{t-1} + b}\ u_t$, the aggregation result may be written without changing the f process. In particular if $a = 1$ and $b = 0$: $\varepsilon^{(m)}_{tm} = \sqrt{a^{(m)} f_{tm-m} + b^{(m)}}\ u^{(m)}_{tm}$ admits f too as leading process of its volatility process. Of course, the degree of freedom added by the scalar a and b implies a lack of identifiability for the coefficients ω and γ of the AR representation. This is the reason why we have preferred to maintain the usual identifiability restriction $(a, b) = (1, 0)$ which implies some rescaling of f for a given aggregation schedule.

$$\omega^{(m)} = m\omega \frac{1-\gamma^m}{1-\gamma} \tag{2.28}$$

which in GARCH notations provides Drost and Nijman's (1993) equation (13):

$$\alpha^{(m)} + \beta^{(m)} = (\alpha+\beta)^m$$

$$\omega^{(m)} = m\omega \frac{1-(\alpha+\beta)^m}{1-(\alpha+\beta)}. \tag{2.29}$$

Indeed, as it is clear from the SARV setting, the parameter of interest is $\gamma = \alpha + \beta$; in particular, if $f_{tm} \in J_{tm}^{(m)}$, it specifies the weight of the exponential smoothing of the stochastic volatility innovations since one can easily deduce in this case from (2.27) and (2.11) that:

$$\nu_{tm}^{(m)} = a^{(m)} \sum_{k=0}^{m-1} \gamma^k \nu_{tm-k} \tag{2.30}$$

In other words, as it already well-known in the GARCH literature (see Drost and Nijman (1993)), $\gamma^{(m)} = (\alpha+\beta)^m$ characterizes the persistence of shocks in the volatility process. Therefore, if the stochastic volatility feature appears to be still significant at low frequency data (γ^m significant for large m), it is likely to be highly persistent for high frequency data (γ close to 1).

As far as we are concerned with leverage effect, the decompositions (2.26) and (2.30) allow us to characterize its invariance w.r.t. temporal aggregation:

- First, the **unconditional leverage effect** $E[\varepsilon_t \nu_t] \neq 0$ occurs at the lowest frequency if and only if it occurs at the highest.

- Second, if **conditional leverage effect** occurs at the lowest frequency:

  $$E[\varepsilon_{tm}^{(m)} \nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] \neq 0$$

  it necessarily occurs at the highest:

  $$E[\varepsilon_{tm} \nu_{tm} \mid J_{tm-1}] \neq 0.$$

  The converse is true if above conditional covariances are constant.

More generally, we have introduced in (2.24) two restrictions of conditional symmetry which are equivalent in the semi-strong GARCH case and allow us to consider the class of "symmetric" (in the sense of (2.24)) SR-SARV(1) processes as a subclass of the weak GARCH processes. In order to show that our aggregation result (for SR-SARV(1) process) extends the Drost and Nijman's (1993) one (for this subclass of weak GARCH processes), we have to check that the restrictions (2.24) themselves are robust w.r.t. temporal aggregation. This is the following result:

Proposition 2.9 *With the assumptions and notations of Proposition 2.8. If:*

$$E[u_t^3 \mid J_{t-1}] = E[u_t \nu_t \mid J_{t-1}] = 0$$

Then for any $m \geq 1$:

$$E[(u_{tm}^{(m)})^3 \mid J_{tm-m}^{(m)}] = E[u_{tm} \nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = 0.$$

2.4 Continuous time SR-SARV modeling:

According to the martingale difference framework for log-returns, we consider here a general continuous time stochastic volatility model with zero-drift for the asset price process:

$$dLogS_t = \sigma_t \; dW_t \tag{2.31}$$

$$d\sigma_t = \gamma_t \; dt + \delta_t \; dW_t^\sigma$$

$$Cov(dW_t, dW_t^\sigma) = \rho_t \; dt$$

where (W_t, W_t^σ) is a bivariate Wiener process and the stochastic processes σ_t, γ_t, δ_t and ρ_t are $J_t^c = \sigma(W_\tau, W_\tau^\sigma, \tau \leq t)$ adapted where the superscript c means that the available information is now defined in **c**ontinuous time. To ensure that σ_t is a nonnegative process one typically follows either one of two strategies: (1) considering a diffusion with a linear drift for $\log\sigma_t^2$ or (2) describing σ_t^2 as a CEV process (or Constant Elasticity of Variance process following Cox (1975) and Cox and Ross (1976)) [4]. The former is frequently encountered in the option pricing literature (see e.g. Wiggins (1987)) and is also clearly related to Nelson (1991), who introduced EGARCH, and to the log-normal SARV model of Taylor (1986 and 1994). The second modeling strategy involves CEV processes which can be written as

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \delta(\sigma_t^2)^\lambda \; dW_t^\sigma \tag{2.32}$$

where $\frac{1}{2} \leq \lambda \leq 1$ ensures that σ_t^2 is a stationary process with nonnegative values. Equation (2.32) can be viewed as the continuous time analogue of the discrete time SR-SARV(1) class of models presented above as it is confirmed by the exact discretization

[4] Occasionally one encounters specifications which do not ensure nonnegativity of the σ_t process. For the sake of computational simplicity some authors for instance have considered Ornstein-Uhlenbeck processes for σ_t or σ_t^2 (see e.g. Stein and Stein (1991)).

results of continuous time SV models stated below. Here, as in the previous section, it will be tempting to draw comparisons with the GARCH class of models, in particular the GARCH-diffusion processes proposed by Drost and Werker (1996) in line with the temporal aggregation of weak GARCH processes.

Indeed, one should note that the CEV process in (2.32) implies a SR-SARV model in discrete time for σ_t^2, namely:

$$\sigma_{t+h}^2 = \theta(1 - e^{-kh}) + e^{-kh}\sigma_t^2 + e^{-kh}\int_t^{t+h} e^{k(u-t)}\delta(\sigma_u^2)^\lambda dW_u^\sigma \tag{2.33}$$

We are then able to prove:

Proposition 2.10 *When the continuous time stochastic process S_t is conformable to (2.31), for any sampling interval h, the associated discrete time process $Log\frac{S_{th}}{S_{(t-1)h}}$, $t \in \mathbf{N}$, is a SR-SARV(1) process w.r.t. $J_{th}^{(h)}$, $J_{th}^{(h)} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, \sigma_{\tau h}^2, \tau \leq t, \tau \in \mathbf{N})$.*

In other words, from the diffusion (2.31) and (2.32), we obtain the class of discrete time SR-SARV(1) which is closed under temporal aggregation, as discussed in the previous section.

As already announced, we have so built a class of SR-SARV(1) processes in discrete time which automatically fulfill the positivity requirement of the volatility process, thanks to the well-suited dynamics of the underlying continuous-time process.

The relation between the continuous time parameters k, θ, and the discrete time parameters $\omega^{(h)}$ and $\gamma^{(h)}$ is the following:

$$\gamma^{(h)} = e^{-kh} \tag{2.34.a}$$
$$\omega^{(h)} = h\theta(1 - e^{-kh}) \tag{2.34.b}$$

The relation between continuous time and discrete time volatility process is similar to (2.27):

$$f_{th} = a^{(h)}\sigma_{th}^2 + b^{(h)} \tag{2.35}$$

with

$$a^{(h)} = \frac{1 - e^{-kh}}{k}$$
$$b^{(h)} = \theta\ (h - \frac{1 - e^{-kh}}{k}).$$

Since the log-return are of course considered here as flow variables, (2.34) has to be seen as a generalization of (2.28) with:

$$\gamma = \gamma^{(1)} = e^{-k}$$

$$\omega = \theta(1-\gamma)$$

where θ is the unconditional variance, that is the expectation of σ_t^2.

If we use the GARCH notations (2.29), it is clear that (2.34) is a generalization of the temporal aggregation result of Drost and Werker (1996), which closes "the GARCH Gap" by interpreting discrete time sampling in a continuous time SV model of the type (2.31), (2.32) as a weak GARCH process. Moreover, contrarily to Drost and Werker (1996), we do not exclude the possibility of leverage effect (no restrictions are considered with respect to the correlation process ρ_t). Moreover, as explained in the previous subsections, we prefer the SR-SARV representation which provides an explicit characterization of innovations in variance (which is not the case for the weak GARCH modeling). As it is shown in the proof of Proposition 2.10, we have here the following innovation process:

$$\nu_{th}^{(h)} = \frac{(1-e^{-kh})}{k}\delta e^{-kh}\int_{(t-1)h}^{th} e^{k(u-(t-1)h)}(\sigma_u^2)^{\lambda}\, dW_u^{\sigma} \tag{2.36}$$

The exponential smoothing formula (2.30) is clearly implied by (2.36). Therefore, (2.36) allows one to state same conclusions about leverage effect; generically, the occurrence of this effect at the highest frequency (that is in continuous time) is tantamount to the occurrence of it at lower frequencies.

3 Univariate SV models for asset returns:

We propose in this section various extensions of the basic SR-SARV(1) model in order to capture some well-documented evidence about asset returns.

A first evidence is that the patterns of the sample autocorrelations for the squared process are not always conformable to the theoretical pattern of an ARMA(1,1) process. This is the reason why there are more and more findings in the empirical GARCH literature of GARCH models of higher orders, or even of long memory GARCH models, to reproduce some stylized empirical regularities like for instance a sample autocorrelation function which "decreases very fast at the beginning, and then decreases very very slowly and remains significantly positive" (quoted from Ding and Granger (1996)). The fact that "the (1,1) order specification fails to account for the variety of dynamic patterns in many time series" has also recently been stressed by Diebold and Lopez (1995) and by their discussant Steigerwald (1995). However, a major drawback of GARCH processes of higher order for asset returns (that is of ARMA process of higher order for squared asset returns) is that they do not generally remain true to the Markovian (of order 1) property of the volatility process usual in Finance; in particular they may lead to differentiable underlying continuous time process for asset prices, which is inconsistent with the fundamental no free-lunch assumption (existence of an equivalent martingale measure conformable to arbitrage pricing theory à la Harrison et Kreps (1979)). This is the reason why we will suggest in this section a particular class of ARMA(p,q) representations of the volatility process (namely ARMA(p,p-1)) which can be interpreted as marginalization of a VAR(1) process of p state variables. In other words, according to a classical asset pricing methodology, "under the umbrella of the Harrison-Kreps model" (quoted from Duffie (1992)), the Markovian property is maintained at the level of a latent multivariate process of state variables, including the case of a continuous time multifactor representation.

This idea is first exploited in subsection 3.1 to define the SR-SARV(p) class which corresponds to a weak concept of GARCH(p,p) models (that is a weak concept of ARMA(p,p) models for squared returns) as well as the SR-SARV(1) class corresponds to weak GARCH(1,1) model. We prove in particular a temporal aggregation property which extends the Drost and Nijman (1993) result for weak GARCH as it was already extended for weak GARCH(1,1) in section 2. The relationship with both semi-strong GARCH(p,p) and continuous time multifactor models are also characterized.

The main goal of the rest of section 3 is to capture in our SV framework the basic idea of GARCH-M modeling: the conditional variance process may enter in the conditional mean through the so-called risk premium. We first briefly discuss in subsection 3.2 some lessons of economic theory about the trade-off between mean of returns and their variance. We propose in section 3.3 various SR-SARV-M specifications which are conformable to these economic ideas and mimic the usual GARCH-M specification. Moreover, our SR-SARV-M models are closed under temporal aggregation while GARCH-M are not for **three** reasons:

- First, usual GARCH specifications are not closed under temporal aggregation, except if we weaken their structure à la Drost and Nijman (1993).
- Second, there is no aggregation theory available for GARCH-M models, even in a weak sense, since the proposed risk premium patterns are clearly not robust w.r.t. temporal aggregation.
- Third, as already emphasized by Proposition 2.2, "commonly-used ARCH models effectively assume that the variance of the variance rises linearly with σ_t^4" (see Nelson and Foster (1994)) which is inconsistent with a joint aggregation of risk premia and conditional variance (see Proposition 3.7 below).

3.1 The SR-SARV(p) models

Definition 3.1 A stationary squared integrable process ε_t is called SR-SARV(p) w.r.t. an increasing filtration J_t, $t \in \mathbf{N}$, if there exists a p-dimensional J_t-adapted stationary VAR(1) process F_t, $t \in \mathbf{N}$, with nonnegative components, such that:

$$\varepsilon_t = \sqrt{f_{t-1}} u_t \tag{3.1.a}$$
$$E[u_t \mid J_{t-1}] = 0 \tag{3.1.b}$$
$$E[u_t^2 \mid J_{t-1}] = 1 \tag{3.1.c}$$
$$f_t = e' F_t \tag{3.1.d}$$

with e is a vector of $\mathbb{R}^p$ with nonnegative components.

It is clear that definition 3.1 extends the definition 2.1 to a higher order dimensional vector F_t of state variables. This process admits by definition a VAR(1) representation:

$$F_t = \Omega + \Gamma F_{t-1} + V_t \tag{3.2.a}$$
$$E[V_t \mid J_{t-1}] = 0 \tag{3.2.b}$$

while $f_t = e' F_t$ is still interpreted as the conditional variance process:

$$f_{t-1} = Var[\varepsilon_t \mid J_{t-1}]. \tag{3.3}$$

In other words, the conditional variance process is now characterized as a linear combination (defined by e) of state variables; the vector e will be called hereafter the variance marginalization vector.

Since stationarity of F_t is assumed, we have implicitly assumed that the eigenvalues of Γ have a modulus smaller than 1, and in order to provide simple sufficient conditions of positivity, a maintained assumption hereafter will be the nonnegativity of the coefficients of Ω.

One can lead here the same discussion about information sets as in the comment of definition 2.1. In the same way, one may state:

Proposition 3.1 *If ε_t is a SR-SARV(p) process w.r.t. a filtration J_t, with F_t the associated p-dimensional state variable and e the corresponding variance marginalization vector, then, for any subfiltration $J_t^* \subset J_t$, ε_t is a SR-SARV(p) process with respect to J_t^*, and the associated p-dimensional state variable F_t^* is given by*

$$F_t^* = E[F_t \mid J_t^*]$$

with the same variance marginalization vector.

As previously announced, the volatility process f_t is no longer Markovian but follows an ARMA(p,p-1) process (as a marginalization of a VAR(1) of size p, see e.g. Lutkepohl (1991)). We are then able to prove the following generalization of Proposition 2.5 which shows that for $p > 1$, the squared innovation ε_t^2 will appear as an ARMA of higher order and therefore reproduce a richer class of autocorrelation patterns:

Proposition 3.2 *If ε_t admits a SR-SARV(p) representation, then ε_t^2 is a weak stationary ARMA(p,p) process:*

$$\varepsilon_t^2 - \omega - \sum_{i=1}^{p} \gamma_i \varepsilon_{t-i}^2 = \omega_t \tag{3.4}$$

where ω_t is an MA(p) process such that

$$E[w_t \mid J_{t-p-1}] = 0 \tag{3.5}$$

We denote by $\omega_t = \eta_t - \sum_{i=1}^{p} \beta_i \eta_{t-i}$, η_t white noise, $\sum_{i=1}^{p} \beta_i < 1$, the MA(p) representation of ω_t. It is then straightforward to deduce from this representation something like a weak GARCH(p,p) representation for ε_t:

$$EL[\varepsilon_t^2 \mid H_{t-1}^s] = h_t \tag{3.6.a}$$
$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{3.6.b}$$

where, according to the notations of Proposition 2.4, H_{t-1}^s is the Hilbert space spanned by the constant and ε_τ^2, $\tau < t$.
Moreover:

$$\gamma_i = \alpha_i + \beta_i, \quad i = 1, 2..p$$

and then (see proof of Proposition 3.2):

$$\sum_{i=1}^{p} \gamma_i = 1 - det(Id - \Gamma)$$

is the convenient measure of persistence of conditional heteroskedasticity if p is the minimal order of the ARMA(r,r) representations (r≥p) of ε_t^2 [5].

However, as already stressed in section 2, a property like (3.5) is much more powerful that the usual definition of weak ARMA(p,p) (and corresponding notion of weak GARCH(p,p)) which ensures only zero linear correlations. Its interpretation is close to

[5]Of course, if the marginalization $f_t = e'F_t$ creates some degeneracies in the dynamics (that is some common roots in the ARMA representation (3.4)), the measure of persistence should be reevaluated.

a martingale difference property, which may allow statistical inference through GMM from the following observable moment restrictions:

$$E[\varepsilon_t^2 - \omega - \sum_{i=1}^{p} \gamma_i \varepsilon_{t-i}^2 \mid I_{t-p-1}] = 0 \tag{3.7}$$

Like to (2.19), (3.7) belongs to the class of lagged moment restrictions studied by Hansen and Singleton (1996). For the SV framework, see Drost, Meddahi and Renault (1996). Moreover, the interest of restrictions (3.5) is that they are consistent with the requirement of closeness under temporal aggregation. This explain that the SR-SARV(p) class itself is closed under temporal aggregation:

Proposition 3.3 *Let ε_t a SR-SARV(p) process w.r.t. J_t, with the associated p-dimensional state variable F_t and marginalization vector e. If for a given natural integer m, we consider a filtration $J_{km}^{(m)}$, $k \in \mathbf{N}$, such that for any k:*

$$\varepsilon_{km}^{(m)} \in J_{km}^{(m)} \subset J_{km},$$

where

$$\varepsilon_{tm}^{(m)} = \sum_{k=0}^{m-1} a_{km} \varepsilon_{tm-k}, \tag{3.8}$$

then *$\varepsilon_{tm}^{(m)}$, $t \in \mathbf{N}$, is a SR-SARV(p) process w.r.t. $J_{tm}^{(m)}$, with a corresponding p-dimensional state variable vector $F_{tm}^{(m)}$ and a marginalization vector $e^{(m)}$ defined by:*

$$e^{(m)} = e$$

$$F_{tm}^{(m)} = A^{(m)} E[F_{tm} \mid J_{tm}^{(m)}] + B^{(m)}$$

where

$$A^{(m)} = \sum_{k=0}^{m-1} a_{km}^2 \Gamma^{m-k-1}$$

$$B^{(m)} = (\sum_{k=0}^{m-1} a_{km}^2 (\sum_{i=0}^{m-k-2} \Gamma^i))\Omega$$

We have:

$$\varepsilon_{tm}^{(m)} = \sqrt{f_{tm-m}^{(m)}} u_{tm}^{(m)}$$

$$f_{tm}^{(m)} = e' F_{tm}^{(m)}$$

$$F_{tm}^{(m)} = \Omega^{(m)} + \Gamma^{(m)} F_{tm-m}^{(m)} + V_{tm}^{(m)}$$

with

$$\Gamma^{(m)} = \Gamma^m$$

$$\Omega^{(m)} = A^{(m)} (Id - \Gamma^m)(Id - \Gamma)^{-1} \Omega + (Id - \Gamma^m) B^{(m)}.$$

In other words, the SR-SARV(p) class is on the one hand better suited to statistical inference than the weak GARCH(p,p) class (thanks to (3.7)) and on the other hand shares with the weak GARCH class a temporal aggregation property (see Drost and Nijman (1993)) for the weak GARCH result). Moreover, we have a generalization of the result (2.27):

$$F_{tm}^{(m)} = A^{(m)} E[F_{tm} \mid J_{tm}^{(m)}] + B^{(m)} \tag{3.9}$$

which proves that the temporal aggregation does not change the state variables process, up to a convenient rescaling.

Indeed, our SR-SARV(p) representation shares with the weak GARCH(p,p) one the idea of ARMA(p,p) dynamics for squared innovations. But, by stressing the underlying VAR(1) process of state variables F_t, we get, as already announced, a direct relationship with usual continuous time multifactor modeling in Finance. This is the issue addressed by Proposition 3.3 below. To the best of 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: diffusion approximation / filtering à la Nelson (1990) / Nelson and Foster (1994) or closing the GARCH gap à la Drost and Werker (1996).

Definition 3.2
A continuous time stochastic volatility model (with zero-drift) for the asset price S_t with p volatility factors F_t^c is defined by:

$$d\begin{pmatrix} LogS_t \\ F_t^c \end{pmatrix} = \begin{pmatrix} 0 \\ K(\Theta - F_t^c) \end{pmatrix} dt + N(Diag(1, F_t^{c'}))^{1/2} dW_t \tag{3.10}$$

where W_t is a (p+1)-univariate standard Wiener process and $Diag(x)$ is the diagonal matrix whose diagonal coefficients are defined by the coefficients of the row vector x.

Proposition 3.4 below is then the multifactor generalization of Proposition 2.10:

Proposition 3.4 *When the continuous time stochastic process S_t is conformable to (3.10), for any sampling interval h, the associated discrete time process $Log\frac{S_{th}}{S_{(t-1)h}}$, $t \in \mathbf{N}$, is a SR-SARV(p) process w.r.t. $J_{th}^{(h)}$, $J_{th}^{(h)} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, F_{\tau h}^c, \tau \leq t, \tau \in \mathbf{N})$.*

In other words, from the diffusion (3.10), we obtain the class of discrete time SR-SARV(p) which is closed under temporal aggregation, as discussed above. The relation between the continuous time parameters K, Θ and the discrete time parameters $\Gamma^{(h)}$, $\Omega^{(h)}$ is the following :

$$\Gamma^{(h)} = e^{-Kh} \tag{3.11.a}$$

$$\Omega^{(h)} = (Id - e^{-Kh})\Theta h \tag{3.11.b}$$

To see this, it is useful to understand that the state variable process $F_t^{(h)}$ in discrete time is one-to-one linearly related to the state process F_t^c in continuous time:

$$F_t^{(h)} = A^{(h)} F_t^c + B^{(h)} \tag{3.12}$$

with

$$A^{(h)} = \int_{(t-1)h}^{th} e^{-K(u-(t-1)h)}\ du = K^{-1}(Id - e^{-kh})$$

$$B^{(h)} = \int_{(t-1)h}^{th} [(Id - e^{-K(u-(t-1)h)}]\Theta\ du = [h\ Id - K^{-1}(Id - e^{-kh})]\Theta.$$

The equations (3.11)/(3.12) extend respectively (2.34)/(2.35). We obtain in particular the discrete time representation (3.1) for $h = 1$.

If we use the GARCH notations (3.6), we see that we have "closed the GARCH gap" à la Drost and Werker (1996) for GARCH of higher orders. Once more, we stress that the SR-SARV(p) structure is richer than the weak GARCH(p,p) one. On the other hand, any semi-strong GARCH(p,p) can be seen as a particular SR-SARV(p) process, according to Proposition 3.5 below:

Proposition 3.5 *If ε_t admits a semi-strong GARCH(p,p) representation then it is a SR-SARV(p) w.r.t. J_t, $J_t = I_t = \sigma(\varepsilon_\tau, \tau \leq t)$.*

Of course, the factor process F_t of size p exhibited in Proposition 3.5 is highly degenerated since it depends of one source of randomness. More generally, it is important to keep in mind that a given ARMA(p,p-1) volatility process f_t may be represented as the marginalization of a lot of VAR(1) p-dimensional factor process F_t, and in particular degenerated VAR(1) processes.

However, for asset pricing purposes, the modeling we suggest here raises at least two new issues:

• First, when the market for a given financial asset of price S_t involves incompleteness, it is not always sufficient to introduce one state variable σ_t to fully describe the relevant uncertainty. Indeed, even the joint process (S_t, σ_t) may be not Markovian and the relevant Markovian representation (S_t, F_t) for asset pricing involve more state variables. Moreover, this case appears to be realistic, according to the widespread finding of asset returns dynamics which correspond to GARCH of higher orders.

• Second, the definition 3.2 introduces a multivariate concept of leverage effect which could be identified with discrete time data through the correlation structure between underlying u_t and V_t. Let us notice that we could consider in (3.10) even more general structures (for instance positive exponent other than 1/2 and different for all element in the diagonal matrix of diffusion coefficients of the state variables F_t^c) since we only need to ensure positivity and stationarity of the process F_t^c.

3.2 Time-varying risk premia

A large variety of dynamic asset pricing models is now available to explain how the dynamics of the risk premia in asset returns may be related to stochastic volatility dynamics by structural relations involving fixed parameters like risk aversion, discount factor, elasticity of intertemporal substitution... Such structural relations are often deduced from Euler equations corresponding to an intertemporal optimization program of a representative agent. Among these models, the so-called consumption based CAPM (see Lucas (1978)) may be considered as a template. It provides the following Euler equation:

$$1 = \beta E_t[r_{t+1}(\frac{c_{t+1}}{c_t})^{-a}] \tag{3.13}$$

where β is a discount factor, a is a relative risk aversion parameter, c_t is the time t consumption and r_{t+1} is the return on a given asset over period $[t, t+1]$. E_t denotes the conditional expectation given available information at time t. Without going into details about it, we may deduce from (3.13) and an assumption of joint conditional log-normality the following usual relation (see for instance Hansen and Singleton (1983)):

$$-Log\beta = m_{rt} - am_{ct} + \frac{1}{2}(\sigma_{rt}^2 + a^2\sigma_{ct}^2 - 2a\sigma_{rct}) \tag{3.14}$$

where, with obvious notations, m_{rt}, m_{ct}, σ_{rt}^2, σ_{ct}^2 and σ_{rct} denote respectively conditional expectations, conditional variances and covariances.

The main issue we want to stress here is that, through such Euler equation, the conditional expectation of a log-return may appear like a linear combination of:

- First, conditional expectation of some macroeconomic factors like aggregate consumption. As usual, when risk aversion is present ($a \neq 0$), predictability of asset returns is not inconsistent with market efficiency.

- Second, conditional variances of the return itself and of other factors.

- Third, conditional covariances between this return and other factors.

Of course, for several models (see for instance the basic CAPM), some factors are themselves asset returns.

In other words, some structural interpretations of asset returns time series models, and particularly of dynamic models of first order and second order conditional moments, imply that we have at our disposal versatile statistical structures where conditional expectations, conditional variances and covariances may be combined through linear aggregators. This is the main puzzle which motivates the sequel of the paper. More precisely, we want to extend the basic SR-SARV(p) model in order to get on the one hand statistical models able to capture structural restrictions like (3.14) about risk premia dynamics and, on the other hand which remain true to the requirement of robustness w.r.t. temporal aggregation.

According to (3.14), we have first to introduce in a given asset risk premium an affine function of its conditional variance. This is the concept of SR-SARV-M (that is SR-SARV in mean) which as presented in subsection 3.3 adapts to our setting the usual ARCH-M model introduced by Engle, Lilien and Robins (1987). As far as we are concerned by robustness w.r.t. temporal aggregation, the SR-SARV-M model is shown to be not only closed under changes of the sampling frequency but also under discretizations where Ito's lemma may introduce squared volatilities in the drift of the processes.

3.3 SR-SARV-M models

The main purpose of this subsection is to extend the SR-SARV(p) model in order to describe the dynamics of a time series y_t, $t = 1, 2...$ of log-returns which is not a martingale difference sequence, due to a risk premium linear w.r.t. conditional variance.

Indeed such a linearity must be seen as a marginalization of the state variable vector F_t whose variance marginalization $e'F_t$ defines the conditional variance process. In order to be consistent with our temporal aggregation requirement, it is important to allow these two marginalizations to be different, that is to consider a process:

$$y_t = c + d'F_{t-1} + \varepsilon_t$$

$$E[\varepsilon_t \mid J_{t-1}] = 0$$

$$Var[\varepsilon_t \mid J_{t-1}] = e'F_{t-1}$$

where d (resp e) is the risk premium (resp variance) marginalization vector, and these two marginalization vectors may not be collinear. If one imagines for instance that the sampling frequency is divided by two, one has to consider the conditional variance of

y_{t+1} given a sub-σ field $J_{t-1}^{(2)}$ of J_{t-1}. It is then clear that:

• On the one hand, the conditional variance $Var[y_{t+1} \mid J_{t-1}^{(2)}]$ will involve, not only the volatility due to the innovation terms ε_{t+1} (and eventually ε_t in the flow case), but also the conditional variance of the high-frequency risk premium $d'F_t$. In other words, the temporal aggregation mixes the two marginalization vectors d and e in the conditional variance process. This is the reason why a concept which would require collinearity between d and e should not be robust with respect to temporal aggregation.

• On the other hand, the mere fact that the conditional variance of the high frequency risk premium enters the conditional variance of the low frequency process implies that we need a SV specification of the state variable process F_t itself.

This is the reason why we propose the following definition:

Definition 3.3

A stationary squared integrable process y_t is called SR-SARV(p) in mean (SR-SARV(p)-M hereafter) w.r.t. an increasing filtration J_t, $t \in \mathbf{N}$, if there exists a p-dimensional J_t-adapted stationary VAR(1) process F_t, $t \in \mathbf{N}$, with nonnegative components, such that:

$$y_t = c + d'F_{t-1} + \varepsilon_t \tag{3.15.a}$$

$$\varepsilon_t = \sqrt{f_{t-1}}u_t \tag{3.15.b}$$

$$f_t = e'(1, F_t')' \tag{3.15.c}$$

$$E[u_t \mid J_{t-1}] = 0 \tag{3.15.d}$$

$$E[u_t^2 \mid J_{t-1}] = 1 \tag{3.15.e}$$

$$Vech(Var[\begin{pmatrix} y_t \\ F_t \end{pmatrix} \mid J_{t-1}]) = R + SF_{t-1} \tag{3.15.f}$$

where d is a vector of size p, e a vector of size $p+1$ with nonnegative components, $R = (r_i)$ is a vector of size $\frac{(p+1)(p+2)}{2}$ and $S = (s_{ij})$ is a matrix of size $\frac{(p+1)(p+2)}{2} \times p$ such that:

$$(r_1, (s_{1i})'_{1\leq i\leq p})' = e. \tag{3.15.g}$$

It is clear that definition 3.3 extends the definition 3.1, which, roughly speaking, corresponds to the case $c = 0$ and $d = 0$. But, as explained above with regard to the temporal aggregation requirement, we need the additional restriction (3.15.f) about $Var[F_t \mid J_{t-1}]$ and $Cov(y_t, F_t \mid J_{t-1})$. Indeed, let us notice that, taking into account (3.15.g), the north-west equality of (3.15.f) is nothing but the definition of e as the variance marginalization vector of the SR-SARV(p) process ε_t[6]. Moreover, following

[6]The variance marginalization vector e is defined here by $f_t = e'(1, F_t')'$ instead of $f_t = e'F_t$ in the definition 3.1. Indeed, this only corresponds to a change in the intercept of the VAR(1) process F_t of state variables.

a widespread tradition (see e.g. Bollerslev, Engle and Wooldridge (1988), Engle and Kroner (1995)) we represent in (3.15.f) linear functions of F_{t-1} taking values in the space of symmetric positive definite matrices by the notation $Vech$ which does not ensure by itself the positivity requirement. Of course, positivity could be ensured by a representation à la BEKK (Baba, Engle, Kraft and Kroner quoted by Engle and Kroner (1995)):

$$Var[\left(\begin{array}{c} y_t \\ F_t \end{array} \right) \mid J_{t-1}] = H + \Lambda \ Diag \ (F'_{t-1}) \ \Lambda' \tag{3.16}$$

which reinforces the assumption (3.15.f). Indeed, we shall introduce below (see Proposition 3.9) a continuous time process whose instantaneous variance is of the type (3.16), so that positivity is automatically fulfilled for any frequency of discrete time sampling.

To summarize, one way to understand the above discrete time modeling is to have in mind an underlying continuous time model where both the "variance of the variance" and the (conditional multivariate) leverage effect have to rise linearly with the conditional variance. As noticed by Nelson and Foster (1994) (see Proposition 3.7 below), this is inconsistent with usual GARCH modeling. But Nelson and Foster (1994) (pages 21-22) themselves with an optimal filtering point of view have stressed the continuous time foundation of such an assumption.

By definition, the state variable process F_t admits a VAR(1) representation

$$F_t = \Omega + \Gamma F_{t-1} + V_t$$

$$E[V_t \mid J_{t-1}] = 0$$

where Γ has eigenvalues of modulus smaller than 1. Let us notice that with a multivariate notion of SR-SARV(p) process (see section 4 below for a precise definition), (3.15.f) means that $(\varepsilon_t, V'_t)'$ is a SR-SARV(p) process w.r.t. J_t.

As far as we are concerned by reducing the information sets along the lines of Propositions 2.1 and 3.1, it is clear that any time-varying risk premium conformable to a structural model like Euler equations will generally become stochastic w.r.t. a reduced information set. This is the reason why we get only a "weak" version of the reduced information result:

Proposition 3.6 *If y_t is a SR-SARV(p)-M process w.r.t. a filtration J_t, with F_t the associated p-dimensional state variable, then, for any subfiltration $J_t^* \subset J_t$, y_t is a SR-SARV(p)-M process w.r.t. J_t^* as soon as F_t is J_t^*-adapted.*

When applied to a general SR-SARV(p) process, this result is weaker than Proposition 3.1 since we have assumed that:

$$F_t^* = F_t \in J_t^*$$

instead of

$$F_t^* = E[F_t \mid J_t^*].$$

This additional requirement is necessary to ensure that the risk premium $c + d'F_{t-1}$ belongs to J_{t-1}^*. Such a requirement is not innocuous since it may prevent us to maintain an assumption of "observability" of the state variables, that is to say a modeling w.r.t. an information filtration defined only by past returns and other data available for the econometrician[7] (volumes, durations, other asset price series..). Indeed, as already noticed by Pagan and Ullah (1988), Glosten, Jagannathan and Runkle (1993) and by King, Sentana and Wadhwani (1994), "the relations between risk premia and conditional variances are sensitive to differential information between agents and econometricians" (quoted from King, Sentana and Wadhwani (1994)). In other words, our framework is not inconsistent with structural models of risk premia which belong, by definition, to the information sets of the agents. We only claim that there are cases where the information set of the economic agents are larger than the one of econometricians. Option pricing models à la Hull and White (1987) are typical examples where a stochastic volatility process may belong to the agent's information set (for instance because it is one-to-one related to quoted option prices) but is a latent (unobserved) process for the econometrician.

Of course, one could imagine to reduce the information set to past returns. As already stressed in Proposition 2.2, this does not necessarily lead to a GARCH representation since the latter maintains an additional assumption of perfect linear correlation between conditional variance and squared returns. The point we want to stress here is that such a perfect linear correlation is generally inconsistent with restriction (3.15.f). In other words, even though GARCH models are particular case of SR-SARV processes, they cannot be considered as innovation process of our SR-SARV(p)-M models. Therefore, the setting of definition 3.3 does not contain the so-called "GARCH in mean" processes à la Engle, Lilien and Robins (1987). This is the price to pay to define a class which is robust w.r.t. temporal aggregation. The following proposition provides a counterexample:

Proposition 3.7 *Let ε_t be a strong-ARCH(1) process, with conditional variance (given I_{t-1}) denoted by h_t, and $u_t = \frac{\varepsilon_t}{\sqrt{h_t}}$ standard gaussian white noise. Then, if there was*

[7] We are grateful to R. Engle for having drawn our attention on this issue.

a state variable F_t of size p conformable to the definition 3.3 w.r.t. the filtration I_t ($f_{t-1} = h_t = e'(1, F'_{t-1})' \in I_{t-1}$), we should have, for any positive integer k, a marginalization e_k ($e_k \in I\!R^{p+1}$) such that $h_t^k = e'_k(1, F'_{t-1})'$.

In other words, we are not able in general to define a state variable vector F_t which ensures that a given ARCH-M model

$$y_t = f(h_t) + \varepsilon_t$$

$$\varepsilon_t \text{ ARCH}$$

falls in the category of SR-SARV(p)-M. The main reason for this inconsistency is the restriction (3.15.f) which plays an essential role for the temporal aggregation result below for SR-SARV(p)-M. Moreover, according to the Engle requirement (already quoted in subsection 2.2) that linearity is needed to allow one to compute forecasts of variance without distributional assumptions, (3.15.f) is needed to compute something like $Var[y_{t+1} \mid J_{t-1}]$ without assumptions about the conditional higher order moments of y_t (for instance conditional kurtosis should appear through the conditional variance of the risk premium $d'F_t$ given J_{t-1}).

As far as we are concerned by temporal aggregation, we want to extend the arguments of Proposition 2.8 with respect to aggregates:

$$y_{tm}^{(m)} = \sum_{k=0}^{m-1} a_{km} y_{tm-k} \tag{3.17}$$

where y is a SR-SARV(p)-M. We are then able to prove:

Proposition 3.8 *Let y_t a SR-SARV(p)-M process w.r.t. J_t, with the associated p-dimensional state variable F_t and marginalization vectors d and e. If for a given natural integer m, we consider a filtration $J_{km}^{(m)}$, $k \in \mathbf{N}$, such that for any k:*

$$y_{km}^{(m)} \in J_{km}^{(m)} \subset J_{km} \text{ and } F_{km} \in J_{km}^{(m)}$$

then *$y_{tm}^{(m)}$, $t \in \mathbf{N}$ is a SR-SARV(p)-M process w.r.t. $J_{tm}^{(m)}$, with a corresponding p-dimensional state variable vector $F_{tm}^{(m)}$ and a risk premium (resp variance) marginalization vector $d^{(m)}$ (resp $e^{(m)}$):*

$$y_{tm}^{(m)} = c^{(m)} + d^{(m)\prime} F_{tm-m} + \varepsilon_{tm}^{(m)}$$

$$\varepsilon_{tm}^{(m)} = \sqrt{e'(1, F'_{tm-m})'} u_{tm}^{(m)}$$

$$F_{tm} = \Omega^{(m)} + \Gamma^{(m)} F_{tm-m} + V_{tm}^{(m)}$$

$$\Gamma^{(m)} = \Gamma^m$$

To understand the Proposition 3.8, it is worthwhile to have in mind the same intuition as in (3.9) with the additional requirement $F_{tm} \in J_{tm}^{(m)}$. Indeed, we have now

$$F_{tm}^{(m)} = F_{tm}.$$

Since as a difference with proposition 3.3 the matrices $A^{(m)}$ and $B^{(m)}$ do no longer incorporate the computation of an expected aggregated squared volatility (see subsection 2.3 for more explicit formulas in the simplest case), one has to change the variance marginalization vector: $e^{(m)} \neq e$.

On the other hand, due to the already explained more complicated effects of temporal aggregation on risk premia, the risk premium marginalization vector $d^{(m)}$ does not admit a so simple expression. However, we have of course:

$$d = 0 \Rightarrow d^{(m)} = 0$$

since temporal aggregation cannot introduce "in mean" variance effects. Besides, it turns out that if "in mean" effects are present, they will generally introduce spurious leverage effects by temporal aggregation since it can be shown (see Appendix) that:

$$\varepsilon_{tm}^{(m)} = \sum_{k=1}^{m-1} L_{km} V_{tm-k} + \sum_{k=0}^{m-1} a_{km} \varepsilon_{tm-k}$$

where

$$L_{km} = d' \sum_{i=0}^{k-1} a_{im} \Gamma^{k-i-1}$$

such as the leverage effect will be introduced by the terms $L_{km} V_{tm-k}$ as soon as $d \neq 0$. This general phenomenon will be illustrated below by an example in continuous time.

As in subsection 3.1, it is interesting to notice that our temporal aggregation result can be extended to close the gap between continuous time and discrete time modeling, by stating a direct relationship between our discrete time SR-SARV(p)-M framework and usual multifactor diffusion models.

More precisely, we can consider an asset price process for which risk premium is linear w.r.t. the factors of volatility:

$$d\left(\begin{array}{c} LogS_t \\ F_t^c \end{array} \right) = \left(\begin{array}{c} \mu + \psi' F_t^c \\ K(\Theta - F_t^c) \end{array} \right) dt + N(Diag(1, F_t^{c'}))^{1/2} dW_t \tag{3.18}$$

Let us notice that we had focused in the previous section on the particular case $\mu = 0$ and $\psi = 0$. Moreover, the precise structure of the diffusion matrix $[Diag(1, F_t^{c'})]^{1/2}$ does matter now since we want to introduce conditional variances and covariances which are linear w.r.t. F_t^c (as already noticed, these restrictions were to a large extent useless in the pure SR-SARV(p) setting). We are then able to state:

Proposition 3.9 *When the continuous time stochastic process S_t is conformable to (3.18), for any sampling interval h, the associated discrete time process $Log\frac{S_{th}}{S_{(t-1)h}}$, $t \in \mathbf{N}$, is a SR-SARV(p)-M process w.r.t. $J_{th}^{(h)}$, $J_{th}^{(h)} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, F_{\tau h}^c, \tau \leq t, \tau \in \mathbf{N})$.*

As for (3.15.f), the crucial assumption for this temporal aggregation result is the linear structure of the instantaneous variance matrix $NDiag(1, F_t^{c'})N$ w.r.t. F_t^c. This type of multivariate square root process was already emphasized by Duffie and Kan (1996), and by Frachot and Lesne (1993) as necessary and sufficient to get linear factorial representation of the term structure of interest rates. It is quite amazing to observe that this class of processes is also well-suited for linear aggregation.

To give more insight on the resulting leverage effect, it is worthwhile to detail the case of one factor ($F_t^c = \sigma_t^2$) which follows a usual square root process:

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \delta\sigma_t \; dW_t^\sigma \tag{3.19.a}$$

and can be identified with the instantaneous volatility of:

$$\frac{dS_t}{S_t} = (\mu + \psi\sigma_t^2) \; dt \; + \; \sigma_t \; dW_t. \tag{3.19.b}$$

If for instance we assume a constant leverage effect

$$Cov(dW_t, dW_t^\sigma) = \rho \; dt \tag{3.19.c}$$

the model (3.19) is conformable to the general setting (3.18). This model is widely used in the option pricing literature (see Bates and Pennachi (1990), Gennotte and Marsh (1993) and Heston (1993)). Moreover, it is important to notice that the "**in mean**" effect introduces a feature which is new w.r.t. all the previous results of this paper. Until now, we were always able to claim (see for instance Proposition 2.9) that if leverage effect appears at a low frequency, it is necessarily occurs at the highest. Unfortunately, this is no longer true with SR-SARV(p)-M models since, due to mean effect, temporal aggregation can creates a spurious leverage effect. To see this, we can compute one-period returns by integrating (3.18.b):

$$Log\frac{S_{t+1}}{S_t} = \int_t^{t+1} [\mu + (\psi - 1/2)\sigma_u^2]du \; + \; \int_t^{t+1} \sigma_u dW_u. \tag{3.20}$$

Indeed, there is "in mean" effect as soon as $\psi \neq 1/2$. In this case, we observe that the term $(\psi - 1/2)\sigma_u^2$ introduces the path of W_u^σ, $t \leq u \leq t+1$, in the innovation of $Log\frac{S_{t+1}}{S_t}$. As a consequence, it produces **automatically** a correlation between this innovation and the one of the volatility process in discrete time. In other words, as soon as $\psi \neq 1/2$, it turns out that, even $\rho = 0$, a leverage effect occurs which is **spurious** because due to the risk premium $\psi\sigma_u^2$. Moreover, it is amazing to observe that if the risk premium is not too large ($\psi < 1/2$), the resulting leverage effect will be automatically in the usual sense (negative correlation). This could provide a theoretical explanation for widespread empirical finding as stressed for instance by French, Schwert and Stambaugh (1987): "Our longer sample period and more inclusive market index support Black's conclusion: leverage is probably not the sole explanation for the negative relation between stock returns and volatility."

3.4 SV models with a predictable component

The previous section has stressed the difficulty of introducing common features in risk premium and conditional variance processes. This difficulty has even led us to brush away the GARCH processes as innovations processes.

This is no longer the case if we restrict the predicable part to be linear function of lagged endogenous variables. The general multivariate setting which integrates the two types of predictable components will be presented in section 4. Let us just briefly stress in this subsection some specific features of univariate AR(1) processes with SR-SARV(p) innovations.

Definition 3.4

A stationary squared integrable process y_t is called AR(1) with SR-SARV(p) errors w.r.t. an increasing filtration J_t, $t \in \mathbf{N}$, if:

$$y_t = c + \rho y_{t-1} + \varepsilon_t \tag{3.21}$$

$$| \rho |< 1$$

$$\varepsilon_t \text{ SR} - \text{SARV(p) w.r.t. } J_t.$$

Since, on the one hand the AR(1) structure is robust w.r.t. information reduction or temporal aggregation, and on the other hand (3.21) does not restrict the dynamics of the innovation process more than the general SR-SARV(p) setting, all the results of section 3.1 can easily be extended to the framework of definition 3.4. For example, AR(1) process with semi-strong GARCH(p,p) innovations are particular cases of (3.21).

To be more precise, we just detail below some results about temporal aggregation. To extend the arguments of Proposition 3.3, we should be interested on aggregates:

$$y_{tm}^{(m)} = \sum_{k=0}^{m-1} a_{km} y_{tm-k}$$

where y is AR(1) with SR-SARV(p) innovations. In this case, the temporal aggregation of the AR(1) structure would lead to ARMA(1,1) in the general case (see e.g. Drost and Nijman (1993)). To limit ourselves to a simpler case, easy to interpret (see below interpretations in terms of interest rates), we consider the case of stock variables:

$$y_{tm}^{(m)} = y_{tm}, \; t = 1, 2, ...$$

Proposition 3.10 *Let $y_t = c + \rho y_{t-1} + \varepsilon_t$, where ε_t is a SR-SARV(p) w.r.t. an increasing filtration J_t. If for a given natural integer m, we consider a subfiltration $J_{km}^{(m)}$, $k \in \mathbf{N}$, such that for any k:*

$$y_{km}^{(m)} = y_{km} \in J_{km}^{(m)} \subset J_{km}$$

then, *$y_{km}^{(m)}$, $t \in \mathbf{N}$, is an AR(1) process with SR-SARV(p) innovation process $\varepsilon_{tm}^{(m)}$:*

$$y_{tm}^{(m)} = c^{(m)} + \rho^{(m)} y_{tm-m}^{(m)} + \varepsilon_{tm}^{(m)}$$

with

$$\begin{aligned} c^{(m)} &= c\tfrac{1-\rho^m}{1-\rho} \\ \rho^{(m)} &= \rho^m \\ \varepsilon_{tm}^{(m)} &= \textstyle\sum_{k=0}^{m-1} \rho^k \varepsilon_{tm-k} \end{aligned}$$

Let us notice that the properties of the SR-SARV(p) process $\varepsilon_{tm}^{(m)}$ can be deduced from those of ε_t by applying the general results of Proposition 3.3 with $a_{km} = \rho^k$.

As in previous sections, the temporal aggregation result allows us to close the gap with a classical continuous time model:

Proposition 3.11 *Let us consider a generalized square-root process y_t defined by:*

$$d\begin{pmatrix} y_t \\ F_t^c \end{pmatrix} = \begin{pmatrix} k(\theta - y_t) \\ K(\Theta - F_t^c) \end{pmatrix} dt + N(Diag(1, F_t^{c'}))^{1/2} dW_t. \tag{3.22}$$

Then, for any sampling interval h, the associated discrete time process y_{th}, $t \in \mathbf{N}$, is AR(1) with SR-SARV(p) innovations w.r.t. $J_{th}^{(h)}$, $J_{th}^{(h)} = \sigma(y_{\tau h}, F_{\tau h}^c, \tau \leq t, \tau \in \mathbf{N})$, $t \in \mathbf{N}$:

$$y_{th} = \theta(1 - e^{-kh}) + e^{-kh} y_{(t-1)h} + \varepsilon_{th}^{(h)}$$

where $\varepsilon_{th}^{(h)}$ is conformable to the process characterized by Proposition 3.4.

An interesting particular case is the degenerated one, where the only factor of the SR-SARV process is y_t itself. We are then led to consider the **square-root process** popularized by Cox, Ingersoll and Ross (1985) when y_t is a short term rate:

$$dy_t = k(\theta - y_t)dt + \sqrt{ay_t + b}\, dW_t. \tag{3.23}$$

In this case, the discrete time representation may be written:

$$y_{th} = \theta(1 - e^{-kh}) + e^{-kh} y_{(t-1)h} + \varepsilon_{th}^{(h)}$$

with

$$\varepsilon_{th}^{(h)} = e^{-kh} \int_{(t-1)h}^{th} e^{k(u-(t-1)h)} \sqrt{a y_u + b}\, dW_u$$

and

$$Var[\varepsilon_{th}^{(h)} \mid y_{\tau h}, \tau < t] = a^{(h)} y_{(t-1)h} + b^{(h)}$$

where:

$$a^{(h)} = a \frac{1 - e^{-2kh}}{k}$$
$$b^{(h)} = (a\theta + b) \frac{1 - e^{-2kh}}{2k} - a e^{-kh} \frac{1 - e^{-kh}}{k}.$$

It is amazing to notice that we have in this case something like AR(1) process with ARCH(1) innovations, except that the conditional heteroskedasticity is characterized by linear combinations of past values of the process itself rather than squared innovations[8]. Indeed, this is not surprising since, as already stressed, the general reduction information result (in line of Proposition 3.1) may be applied in this setting. In other words, one can always reduced the information set to past observables, which opens the door to ARCH-type models.

As far as we are concerned by temporal aggregation of interest rates models, it is worth noting that if y_t is a continuously compounded "short" term interest rate and we divide by m the frequency of data recording, we generally observe, not only the short term interest rate y_{tm}, $t \in \mathbf{N}$, but also a longer term interest rate which is a flow-type aggregate of short term ones:

$$y_{tm}^{(m)} = \frac{1}{m} \sum_{k=0}^{m-1} y_{tm-k}.$$

We are then able to complete the Proposition 3.10 (which may be applied to the underlying continuous time model (3.23)) by considering the general case

$$y_{tm}^{(m)} = \sum_{k=0}^{m-1} a_{km} y_{tm-k}. \tag{3.24}$$

Proposition 3.12 *Let $y_t = c + \rho y_{t-1} + \varepsilon_t$, where ε_t is a SR-SARV(p) w.r.t. an increasing filtration J_t. If, for a given natural integer m, we consider a filtration $J_{km}^{(m)}$, $k \in \mathbf{N}$, such that for any k:*

$$\{y_{tm}^{(m)}, y_{tm}\} \subset J_{km}^{(m)} \subset J_{km}$$

[8]We thank Feike Drost to have drawn our attention on this example.

where $y_{tm}^{(m)}$ *is defined by (3.24),* **then***:*

$$y_{tm}^{(m)} = c^{(m)} + \rho^{(m)} y_{tm-m} + \varepsilon_{tm}^{(m)}$$

with

$$c^{(m)} = c\frac{1}{1-\rho}\sum_{k=0}^{m-1} a_{km}(1-\rho^{m-k})$$

$$\rho^{(m)} = \sum_{k=0}^{m-1} a_{km}\rho^{m-k}$$

$$\varepsilon_{tm}^{(m)} \text{ SR} - \text{SARV(p) process w.r.t. } J_{km}^{(m)}.$$

4 Multivariate Case

We first provide in the subsection 4.1 below some natural multivariate generalizations of the previous concepts of SR-SARV(p) and SR-SARV(p)-M processes. We do not detail the extended statements of the previous results (projection, temporal aggregation, relationships with GARCH and diffusion models) since these results could generally be easily extended at the price of cumbersome formulas. We prefer to stress in subsection 4.2 below the specific issues of the multivariate analysis, that is contemporaneous aggregation and marginalization.

4.1 The general setting

We first extend to a multivariate m.d.s. Y_t the definition 3.1:

Definition 4.1 A stationary second order process Y_t of size n, is called a multivariate SR-SARV(p) w.r.t. an increasing filtration J_t, $t \in \mathbf{N}$, if there exists a p-dimensional J_t-adapted stationary VAR(1) process F_t, $t \in \mathbf{N}$, with nonnegative components, such that:

$$Y_t = (G_{t-1})^{\frac{1}{2}} U_t \tag{4.1.a}$$

$$E[U_t \mid J_{t-1}] = 0 \tag{4.1.b}$$

$$Var[U_t \mid J_{t-1}] = Id_n \tag{4.1.c}$$

$$Vech G_t = R + S\ F_t \tag{4.1.d}$$

where $Y_t \in I\!\!R^n$, $F_t \in I\!\!R^p$, $U_t \in I\!\!R^n$, $R \in I\!\!R^{\frac{n(p+1)}{2}}$, S is a matrix of size $\frac{n(n+1)}{2} \times p$ and G_t is a process of symmetric positive definite matrices of size n.

It is clear that definition 4.1 extends the definition 3.1 to a multivariate setting. Indeed, if $n = 1$, $G_t = R + S\ F_t$ is a scalar process which can be rewritten $G_t = e'\tilde{F}_t$ where $e = S'$ and $\tilde{F}_t$ is a VAR(1) process of size p which can for instance be defined from F_t by:

$$\tilde{F}_t = \frac{R}{s_1}(1, 0...0)' + F_t$$

if the first coefficient s_1 of S is nonzero. In any case, the VAR(1) representation of F_t and $\tilde{F}_t$ differ only by the intercept [9].

Up to this slight change of notations, all the results of section 3.1 can easily be extended to this setting. Even if we do not detail it for briefness, we want particularly stress here the temporal aggregation result for a multivariate SR-SARV(p) process, since, to the best of our knowledge, no such results are available in the literature for multivariate conditional heteroskedasticity. Indeed, the weak GARCH concept has not been extended to a multivariate setting.

As far as one is concerned by the relationship between definition 4.1 and multivariate GARCH, one can notice some similarity between our definition 4.1 and the multivariate Generalized ARCH models as described by Engle and Kroner (1995). It turns out that, like us, Engle and Kroner (1995) introduce first the so-called "vec representation" before claiming that the BEKK representation is "a new parameterization that easily imposes (the positivity) restrictions and that eliminates very few if any interesting models allowed by the vec representation". As already announced, we are able to write a BEKK type representation:

$$G_{t-1} = H + \Lambda \; Diag(F'_{t-1})\Lambda' \tag{4.2}$$

which ensures positivity. Moreover, a continuous time setting may be built with such a variance representation (the multivariate extension of definition 3.2 is straightforward). On the other hand, for temporal aggregation purposes, we have chosen here to stress the VAR(1) representation of a state variables vector. This VAR(1) representation is not well-suited with respect to the BEKK parameterization since there is no simpler way to incorporate VAR(1) dynamics

$$F_t = \Omega + \Gamma F_{t-1} + V_t \tag{4.3}$$

in the BEKK representation (4.2). Of course, as noticed by Engle and Kroner (1995), almost all Vec representation can be rewritten in a BEKK form:

$$G_{t-1} = H + \sum_{k=1}^{K} \Lambda_k \; Diag(F'_{t-1})\Lambda'_k \tag{4.4}$$

for sufficiently large K. As already explained, we have preferred here for various reasons (in particular, the relation with modern continuous time finance) to focus on

[9]Indeed, this choice of parameterization was already encountered in the ARCH literature. When one considers a ARCH(1) model: $\varepsilon_t = \sqrt{h_t}u_t$, $h_t = \omega + \alpha\varepsilon_{t-1}^2$, various authors (see for instance Broze and Gouriéroux (1993) and Newey and Steigerwald (1995)) prefer the parameterization $\varepsilon_t = \sigma\sqrt{\tilde{h}_t}u_t$, $\tilde{h}_t = 1 + \tilde{\alpha}\varepsilon_{t-1}^2$.

p-dimensional VAR(1) process of state variables. Let us recall that it may represent higher order GARCH(p,p) models (as it was detailed in the univariate setting, see proposition 3.5) while BEKK representation of GARCH(p,p) model à la Engle and Kroner (1995) necessitates a large number of parameters. Indeed, one additional advantage of our representation is that the p state variables play the role of **factors** in **both** the transversal and longitudinal dimensions: they summarize not only the cross-covariances but also the higher order dynamics. Our approach follows an old tradition of multivariate time series analysis where ARMA processes are represented by Markovian state space models: "If one accepts the notion of a vector variable, then it is natural to think that in a sufficiently full description of a given physical situation the process would be Markov" (quoted from Whittle (1990), page 18).

Finally, it is worthwhile to notice that the definition 4.1 allows for stochastic dynamics even at the level of cross-correlations; in other words, we capture more general multivariate dynamics than the restricted multivariate GARCH model of Bollerslev (1990) or the log-normal multivariate stochastic volatility model of Harvey, Ruiz and Shephard (1994) which maintain an assumption of constant cross-correlations. However, it is clear that such a degree of freedom could be introduced in the log-normal SV model in the same way that here by considering for instance:

$$G_t = A\ exp(Diag(F_t'))\ A'.$$

Let us now consider the multivariate extension of the SV in mean definition 3.3:

Definition 4.2 A stationary second order process Y_t of size n, is called a multivariate SR-SARV(p)-M w.r.t. an increasing filtration J_t, $t \in \mathbf{N}$, if there exists a p-dimensional J_t-adapted stationary VAR(1) process F_t, $t \in \mathbf{N}$, with nonnegative components, such that:

$$Y_t = C + DF_{t-1} + \varepsilon_t \tag{4.5.a}$$

$$\varepsilon_t = (G_{t-1})^{\frac{1}{2}} U_t \tag{4.5.b}$$

$$E[U_t \mid J_{t-1}] = 0 \tag{4.5.c}$$

$$Var[U_t \mid J_{t-1}] = Id_n \tag{4.5.d}$$

$$Vech(Var[\begin{pmatrix} Y_t \\ F_t \end{pmatrix} \mid J_{t-1}]) = R + S\ F_{t-1} \tag{4.5.e}$$

where $Y_t \in I\!R^n$, $F_t \in I\!R^p$, $U_t \in I\!R^n$, $C \in I\!R^n$, $R \in I\!R^{\frac{n(p+1)}{2}}$, D is a matrix of size $n \times p$, S is a matrix of size $\frac{n(n+1)}{2} \times p$ and G_t is a process of symmetric positive definite matrices of size n.

Let us notice that, for sake of notational simplicity, we have not introduced in definition 4.2 the most general setting with both risk premia and predictable components in the spirit of definition 3.4. Both the previous definitions and the previous results (projection, temporal aggregation, relation with continuous time) could be easily extended at the price of cumbersome formulas. This is the reason why we have chosen to focus in this section 4 on the specific features of the multivariate analysis, that is contemporaneous aggregation and marginalization.

4.2 Marginalization and Contemporaneous aggregation

As already stressed by Nijman and Sentana (1996), little attention has been paid in the literature until now on first the relation between a multivariate model of conditional heteroskedasticity and the implied univariate models for the components (the so-called **marginalization** issue) and second the effect of taking linear combination of univariate GARCH models (the so-called **contemporaneous aggregation** issue). Nijman and Sentana (1996) emphasize some contexts of Financial Econometrics where contemporaneous aggregation is a crucial issue:

• First, if one considers "the (log) returns in the Deutsche mark/US dollar exchange rate , the US dollar/Japanese yen rate, and the Deutsche mark/Japanese yen rate. As the returns on the third exchange rate are simply the sum of the returns on the first two exchange rates, the GARCH models for these exchange rates implicitly specify a model for the third exchange rate as well."

• Second, in the same spirit, "the relation between the models for (the) individual stocks and the one for the portfolio" should imply some robustness of conditional heteroskedasticity models with respect to contemporaneous aggregation.

But, Nijman and Sentana (1996) observe that "the parametric structure of the commonly used GARCH models is lost by taking linear combinations or by marginalizing". This is the reason why they prove that linear combinations and marginalizations are of the weak GARCH type.

However, their seminal work raises at least two issues:

• On the one hand, as already stressed in section 2, the weak GARCH concept suffers from a lack of structure and one would like to get more statistical properties about the scalar processes obtained by marginalization or contemporaneous aggregation.

• On the other hand, it is a pity that Nijman and Sentana (1996) do not obtain a robustness result à la Drost and Nijman (1993). Indeed, they have not been able to define a multivariate weak GARCH concept and to prove that the resulting class of processes is invariant by linear transformations.

This is the reason why we believe that it is useful to prove a general invariance result for the class of multivariate SR-SARV(p) processes. This is the following proposition 4.1.

Proposition 4.1 *If Y_t is a multivariate (of size n) SR-SARV(p) (resp SR-SARV(p)-M) process w.r.t. an increasing filtration J_t, $t \in \mathrm{N}$, with the associated p-dimensional state variable F_t, while A and B are given matrices of respective sizes $q \times 1$ and $q \times n$,* **then** *$B\,Y_t$ (resp $A+B\,Y_t$) is a multivariate SR-SARV(p) (resp SR-SARV(p)-M) process w.r.t. J_t with the same vector F_t of state variables.*

Of course, Proposition 4.1 is a generalization of the Nijman and Sentana result, due to the relationships between SR-SARV, semi-strong GARCH and weak GARCH already described in section 2. Let us notice moreover that our general multivariate linear setting may be incorporated without additional difficulty within simultaneous equations systems (including exogenous variables) à la Engle and Kroner (1995). In the same way, structural ARCH à la Harvey, Ruiz and Sentana (1992) and King, Sentana and Wadhwani (1994) as well as ARCH factor models à la Diebold and Nerlove (1989) can be seen as particular cases of our setting.

5 Conclusion

We have proposed in this paper a new concept of semiparametric stochastic volatility model which appears to be the good framework for structural interpretations of times series models with conditional heteroskedasticity. Actually, if one wants to consider time series of conditionally heteroskedastic asset returns, there was no framework available until now to capture in the same setting temporal aggregation or portfolios of these returns. On the one hand, it is well known that the usual GARCH setting is not robust with respect to temporal and contemporaneous aggregations. On the other hand, the only robust setting already suggested in the literature, that is the Drost and Nijman (1993) weak GARCH one, suffers from several drawbacks:

- It renounces to the concept of conditional variance, which is a pity for both financial interpretation and statistical inference.
- It does not admit multivariate or "in mean" versions and cannot capture the well documented leverage effect.

The SR-SARV setting proposed here overcomes these difficulties because it extends the usual GARCH class without losing the essential above properties and the nice intuition of ARMA representation of squared innovations. Moreover, all the results of the weak GARCH literature (Drost and Nijman (1993), Drost and Werker (1996), Nijman and Sentana (1996)) are shown to be particular cases of our general results since, roughly speaking, if leverage effect is precluded, our SR-SARV processes are weak GARCH.

Moreover, we are even able to give stochastic volatility type representations of GARCH(p,p)

The only loss with respect to GARCH is the introduction of an unobserved stochastic volatility process which obliges one to think in a state space form. But, perhaps one of the main contribution of this paper is to stress that volatility models are always specified with respect to a given information set and that, according to the mainstream asset pricing tradition, the econometrician is always allowed to reduce the information set without invalidating the model. In particular, one always can reduce the information set up to observables, without necessarily encountering the GARCH setting since it is shown that the usual GARCH(1,1) model is tantamount to two restrictive assumptions:

- First, it assumes perfect **linear** correlation between squared innovations and conditional variance.
- Second, it assumes that the variance of the variance raises linearly with the squared variance (a drawback already pointed out by Nelson and Foster (1994)).

Moreover, we are even able to give stochastic volatility type representations of GARCH(p,p) of higher orders ($p > 1$) through a state (volatility) vector which is a VAR(1) of size p.

The only case where one cannot reduce the information sets without changing the form of the model is the "in mean" case, where risk premium have to be in the information sets of the economic agents. This is conformable to the most recent modeling of asset returns (see e.g. King, Sentana and Wadhwani (1994)) and opens the door for future research on structural multivariate modeling of time varying volatility. Finally, let us recall that the inference issue is not explicitly addressed in this paper even though we provide here the main tool for it: conditional lagged moment restrictions à la Hansen and Singleton (1996). The details of practical implementation in the stochastic volatility setting of Hansen and Singleton (1996) general approach for optimal instruments are discussed in Drost, Meddahi and Renault (1996) paper.

References

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.

Bates D. and G. Pennacchi (1990), "Estimating a Heteroskedastic State Space Model of Asset Prices", Working paper, University of Illinois at Urbana.

Bates C.E. and H. White (1990), "Efficient Instrumental Variables Estimation for Systems of Implicit Heterogeneous Nonlinear Dynamic Equation with Nonspherical Errors", *Dynamic Econometric Modeling*, edited by W. Barnett, E. Berndt and H. White, Cambridge University Press.

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. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach", *Review of Economics and Statistics*, 72, 498-505.

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

Broze, L. and C. Gouriéroux (1993), "Covariance Estimators and Adjusted Pseudo-Maximum Likelihood Method", CORE DP 9313.

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

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

Cox, J.C. and S. Ross (1976), "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics*, 3, 145-166.

Diebold, F.X. and J.A. Lopez (1995), "Modeling Volatility Dynamics", in Macroeconometrics, 427-466, edited by K.V. Hoover, Kluwer Academic Publishers.

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.

Ding, Z. and C.W.J. Granger (1996), "Modeling Volatility Persistence of Speculative Returns: A new Approach", *Journal of Econometrics*, 73, 185-215.

Drost, F.C., N. Meddahi and E. Renault (1996), "Instrumental Variables Estimation of Stochastic Volatility Models", in progress.

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. (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.

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

Engle, R. (1994), commentary on Jacquier-Polson-Rossi's "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, 12, 395-396.

Engle, R. (1995), *ARCH, Selected Readings*, in Advanced Texts in Econometrics, Oxford University Press.

Engle, R. and G. González-Rivera (1991), "Semiparametric GARCH models", *Journal of Business and Economic Statistics*, 9, 345-359.

Engle, R. and Kroner F.K (1995), "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122-150.

Engle, R., D.M. Lilien and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Models", *Econometrica*, 55, 391-407.

Frachot, A. and J.P. Lesne (1993), "Econometrics of Linear Factor Models of Interest Rates", mimeo, Banque de France.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), "Expected Stock Returns and Volatility", *Journal of Financial Economics*, 19, 3-29.

Gennotte, G. and T.A. Marsh (1993), "Variations in Economic Uncertainty and Risk Premiums on Capital Assets", *European Economic Review*, 37, 1021-1041.

Ghysels, E., A.C. Harvey and E. Renault (1995), "Stochastic Volatility", *in Handbook of Statistics*, vol 14, forthcoming.

Glosten, L.R., R. Jagannathan and D. Runkle (1993), "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. and K.J. Singleton (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns", *Journal of Political Economy*, 91, 249-265.

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.

Harrison, J.M. and D.M. Kreps (1979), "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381-408.

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.

Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities", *The Journal of Finance*, Vol XLII, 281-300.

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

Lucas, R.E. Jr (1978), "Asset Prices in an Exchange Economy", *Econometrica*, 46, 1429-1445.

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

Meddahi, N. and E. Renault (1995), "Linear Statistical Inference for ARCH-Type Processes", Discussion paper, GREMAQ, Univesité de Toulouse.

Newey, W.K. and D. Steigerwald (1995), "Consistency of Quasi-Maximum Likelihood Estimators for Models with Conditional Heteroscedasticity", mimeo, University of Santa Barbara.

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

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.

Pagan, A. and A. Ullah (1988), "The Econometric Analysis of Models with Risk Terms", *Journal of Applied Econometrics*, 3, 87-105.

Steigerwald, D.G. (1995), commentary on Diebold-Lopez's "Modeling Volatility Dynamics", in Macroeconometrics, 467-472, edited by K.V. Hoover, Kluwer Academic Publishers.

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

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

Taylor, S.J. (1994), "Modeling Stochastic Volatility: A Review and Comparative Study", *Mathematical Finance*, 4, 183-204.

Whittle, P. (1990), "Statistical Inference in Time Series", in Time Series and Statistics, The New Palgrave.

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

APPENDIX

Proof of Proposition 2.1

Let $u_t^* = \frac{\varepsilon_t}{\sqrt{f_{t-1}^*}}$. We have:

$E[u_t^* \mid J_{t-1}^*] = \frac{1}{\sqrt{f_{t-1}^*}} E[E[\varepsilon_t \mid J_{t-1}] \mid J_{t-1}^*] = 0,$

$E[(u_t^*)^2 \mid J_{t-1}^*] = \frac{1}{f_{t-1}^*} E[E[(\varepsilon_t)^2 \mid J_{t-1}] \mid J_{t-1}^*] = \frac{1}{f_{t-1}^*} E[f_{t-1} \mid J_{t-1}^*] = 1.$

Proof of Proposition 2.2

Let us consider ε_t a GARCH(1,1) defined by (2.3-5). 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, u_t is conformable to (2.9.b) and (2.9.c)

$E[u_t \mid I_{t-1}] = 0$ and $E[u_t^2 \mid I_{t-1}] = 1$

while f_t is an I_t-adapted AR(1) process since, by (2.5):

$$f_t = \omega + \alpha\varepsilon_t^2 + \beta f_{t-1} = \omega + \alpha f_{t-1} u_t^2 + \beta f_{t-1} = \omega + (\alpha + \beta) f_{t-1} + \nu_t$$

which provides the AR(1) representation with an innovation process:

$$\nu_t = \alpha f_{t-1}(u_t^2 - 1)$$

since:

$$E[u_t^2 \mid I_{t-1}] = 1 \Longrightarrow E[\nu_t \mid I_{t-1}] = 0.$$

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 $\alpha > 0$). A fortiori this is 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 $\alpha^2 \leq \gamma^2 = (\alpha + \beta)^2$.

Conversely, let us now consider a SR-SARV(1) process ε_t which fulfills the two restrictions of Proposition 2.2. 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:

$$f_t = \alpha f_t + b_t$$

and, taking conditional expectation $E[\bullet \mid J_{t-1}]$:

$$E[f_t \mid J_{t-1}] = \alpha f_{t-1} + b_t.$$

By identification with the AR(1) representation (see (2.11.a)) of f_t, we conclude that:

$b_t = \omega + \beta f_{t-1}$ where $\beta = \gamma - \alpha \geq 0$.

Thus: $\qquad f_t = \omega + \alpha\varepsilon_t^2 + \beta f_{t-1}$,

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 2.3

Let ε_t be a SR-SARV(1) process which fulfills the first restriction of Proposition 2.2. Then:

$$f_t = a_t\varepsilon_t^2 + b_t,\ \ a_t, b_t \in J_{t-1},$$

and:

$$E[f_t \mid J_{t-1}] = a_t E[\varepsilon_t^2 \mid J_{t-1}] + b_t = a_t f_{t-1} + b_t.$$

Therefore, by difference:

$$\nu_t = a_t(\varepsilon_t^2 - f_{t-1})$$

and:

$$E[u_t\nu_t \mid J_{t-1}] = a_t E[u_t\varepsilon_t^2 \mid J_{t-1}] - a_t f_{t-1} E[u_t \mid J_{t-1}] = a_t f_{t-1} E[u_t^3 \mid J_{t-1}],$$

which proves that:

$$E[u_t\nu_t \mid J_{t-1}] \neq 0 \Longleftrightarrow E[u_t^3 \mid J_{t-1}] \neq 0.$$

Proof of Proposition 2.4

$$\varepsilon_t = \sqrt{f_{t-1}}u_t = \sqrt{f_{t-1}^*}u_t^*.$$

Then: $E[\varepsilon_t^3 \mid J_{t-1}] = f_{t-1}^{3/2} E[u_t^3 \mid J_{t-1}]$, and

$$E[\varepsilon_t^3 \mid J_{t-1}] \neq 0 \Longleftrightarrow E[u_t^3 \mid J_{t-1}] \neq 0.$$

For the same reason,

$$E[\varepsilon_t^3 \mid J_{t-1}^*] \neq 0 \Longleftrightarrow E[(u_t^*)^3 \mid J_{t-1}^*] \neq 0.$$

Since, by iterated projections:

$$E[\varepsilon_t^3 \mid J_{t-1}] = 0 \Longrightarrow E[\varepsilon_t^3 \mid J_{t-1}^*] = 0,$$

we can conclude that:

$$E[u_t^3 \mid J_{t-1}] = 0 \Longrightarrow E[(u_t^*)^3 \mid J_{t-1}^*] = 0.$$

On the other hand:

$$E[u_t \nu_t \mid J_{t-1}] = E[u_t f_t \mid J_{t-1}] = \frac{1}{\sqrt{f_{t-1}}} E[\varepsilon_t f_t \mid J_{t-1}].$$

Thus:

$$E[u_t \nu_t \mid J_{t-1}] = 0 \Longleftrightarrow E[\varepsilon_t f_t \mid J_{t-1}] = 0,$$

and for the same reason

$$E[u_t^* \nu_t^* \mid J_{t-1}^*] = 0 \Longleftrightarrow E[\varepsilon_t f_t^* \mid J_{t-1}^*] = 0.$$

But:

$$E[\varepsilon_t f_t^* \mid J_{t-1}^*] = E[\varepsilon_t E[f_t \mid J_t^*] \mid J_{t-1}^*] = E[E[\varepsilon_t f_t \mid J_t^*] \mid J_{t-1}^*] = E[\varepsilon_t f_t \mid J_{t-1}^*]$$

since $\varepsilon_t \in I_t \subset J_t^*$ and $J_{t-1}^* \subset J_t^*$. But, by iterated projections:

$$E[\varepsilon_t f_t \mid J_{t-1}] = 0 \Longrightarrow E[\varepsilon_t f_t \mid J_{t-1}^*] = 0,$$

and therefore

$$E[u_t \nu_t \mid J_{t-1}] = 0 \Longrightarrow E[u_t^* \nu_t^* \mid J_{t-1}^*] = 0.$$

Proof of Proposition 2.5

$$\varepsilon_t^2 = f_{t-1} u_t^2 = \omega u_t^2 + \gamma f_{t-2} u_t^2 + \nu_{t-1} u_t^2.$$

Therefore, if we define: $\omega_t = \varepsilon_t^2 - \omega - \gamma \varepsilon_{t-1}^2$, we have:

$$\omega_t = \omega(u_t^2 - 1) + \gamma f_{t-2}(u_t^2 - u_{t-1}^2) + \nu_{t-1} u_t^2.$$

It is clear that: $E[\omega_t \mid J_{t-2}] = 0$
since:

$$E[u_t^2 \mid J_{t-1}] = 1$$
$$E[u_t^2 - u_{t-1}^2 \mid J_{t-2}] = 1 - 1 = 0$$
$$E[\nu_{t-1} u_t^2 \mid J_{t-2}] = E[\nu_{t-1} E[u_t^2 \mid J_{t-1}] \mid J_{t-2}] = E[\nu_{t-1} \mid J_{t-2}] = 0.$$

Therefore, since: $\sigma(\omega_\tau, \tau \leq t-2) = \sigma(\varepsilon_\tau^2, \tau \leq t-2) \subset J_{t-2}$, we have:

$$Cov(\omega_t, \omega_{t-h}) = 0 \;\forall\; h \;\geq\; 2.$$

Thanks to our stationarity assumption, we are able to conclude that ω_t is MA(1), that is $\omega_t = \eta_t - \beta\eta_{t-1}$, where η_t is a white noise.

Proof of Proposition 2.6

Let us denote by η_t the innovation of the process ε_t^2:

$$\varepsilon_t^2 - \eta_t = EL[\varepsilon_t^2 \mid 1, \varepsilon_\tau^2, \tau < t].$$

We want to show that ε_t is a weak GARCH if and only if:

$$Cov(\eta_t, \varepsilon_\tau) = 0,\ \forall \tau < t.$$

But, these equalities are tantamount to claim that

$$\varepsilon_t^2 - EL[\varepsilon_t^2 \mid 1, \varepsilon_\tau^2, \tau < t]$$

is orthogonal, not only to $1, \varepsilon_\tau^2, \tau < t$, but also to $\varepsilon_\tau, \tau < t$, that is:

$$EL[\varepsilon_t^2 \mid 1, \varepsilon_\tau^2, \tau < t] = EL[\varepsilon_t^2 \mid H_{t-1}].$$

In other words:

$$Cov(\eta_t, \varepsilon_\tau) = 0,\ \forall \tau < t \Longleftrightarrow \varepsilon_t^2 - \eta_t = EL[\varepsilon_t^2 \mid H_{t-1}] \;\; \forall t.$$

Thanks to this characterization, we first check that (2.22) implies that ε_t is a weak GARCH. Let us define: $\alpha = \gamma - \beta$.

We want to prove that (2.22) implies that:

$$EL[\varepsilon_t^2 \mid H_{t-1}] = \omega + \alpha\varepsilon_{t-1}^2 + \beta EL[\varepsilon_{t-1}^2 \mid H_{t-2}]$$

that is, according to the above characterization, that:

$$\varepsilon_t^2 - \eta_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta(\varepsilon_{t-1}^2 - \eta_{t-1}).$$

This results straightforwardly from the ARMA representation of ε_t^2.

Conversely, let us assume that ε_t is a weak GARCH, that is:

$$EL[\varepsilon_t^2 \mid H_{t-1}] = \omega + \alpha\varepsilon_{t-1}^2 + \beta EL[\varepsilon_{t-1}^2 \mid H_{t-2}]$$

for a given α, with $0 < \alpha + \beta < 1$. We want to prove that ε_t^2 is an ARMA(1,1) process conformable to (2.21) and (2.22).

Let us define $\tilde{\eta}_t = \varepsilon_t^2 - EL[\varepsilon_t^2 \mid H_{t-1}]$. By definition:

$$Cov(\tilde{\eta}_t, \varepsilon_\tau) = 0,\ \forall \tau < t.$$

Therefore, it is sufficient to prove that $\tilde{\eta}_t = \eta_t$. But by definition of the weak GARCH representation, we have:

$$\varepsilon_t^2 - \tilde{\eta}_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta(\varepsilon_{t-1}^2 - \tilde{\eta}_{t-1})$$

that is:

$$\varepsilon_t^2 - (\alpha+\beta)\varepsilon_{t-1}^2 = \omega + \tilde{\eta}_t - \beta\tilde{\eta}_{t-1}.$$

Moreover, thanks to the assumed stationarity of ε_t, we know that $\tilde{\eta}_t$ is a white noise, and by (2.20.c), $0 < \alpha + \beta < 1$.

We are then able to conclude, thanks to the unicity of the ARMA representation of the stationary process ε_t^2 that:

$\tilde{\eta}_t = \eta_t$ and $\alpha + \beta = \gamma$.

This achieves the proof of Proposition 2.6.

Proof of Proposition 2.7

The SR-SARV(1) property implies, by Proposition 2.5, that ε_t^2 is an ARMA(1,1) process. Therefore, by Proposition 2.6, 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, $\tau \le t$, 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):

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

- 1st case: $t' > t$: Then:

$$E[\varepsilon_{t'}\varepsilon_t^2 \mid J_{t-1}] = E[\varepsilon_t^2 E[\varepsilon_{t'} \mid J_{t'-1}] \mid J_{t-1}] = 0$$

by (2.9.b).

- 2nd case: $t' = t$: 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 (2.24.a).

• 3rd case: $t' < t$: Then

$$E[\varepsilon_{t'}\varepsilon_t^2 \mid J_{t'-1}] = E[\varepsilon_{t'} f_{t-1} E[u_t^2 \mid J_{t-1}] \mid J_{t'-1}] = E[\varepsilon_{t'} f_{t-1} \mid J_{t'-1}]$$

by to (2.9.c).

Since f_t is an AR(1) (by (2.11.a)), 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 $i \geq t - t'$

$$E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = \nu_{t-1-i} E[\varepsilon_{t'} \mid J_{t'-1}]] = 0$$

by (2.9.b).

If $i = t - t' - 1$

$$E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = E[\nu_{t'}\varepsilon_{t'} \mid J_{t'-1}] = \sqrt{f_{t'-1}} E[u_{t'}\nu_{t'} \mid J_{t'-1}] = 0$$

by (2.24.b).

And, if $i < t - t' - 1$

$$E[\nu_{t-1-i}\varepsilon_{t'} \mid J_{t'-1}] = E[\varepsilon_{t'} E[\nu_{t-1-i} \mid J_{t-i-2}] \mid J_{t'-1}] = 0$$

by (2.11.b).

Hence, $E[\varepsilon_{t'} f_{t-1} \mid J_{t'-1}] = 0$, which achieves the proof of Proposition 2.7.

Proof of Proposition 2.8

Let

$$f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[(\varepsilon_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}]$$

where $\varepsilon_{tm}^{(m)}$ is defined by (2.26). We have:

$$(\varepsilon_{tm}^{(m)})^2 = \sum_{k=0}^{m-1} a_{km}^2 \varepsilon_{tm-k}^2 + 2 \sum_{0 \leq i < j \leq m-1} a_{im} a_{jm} \varepsilon_{tm-i} \varepsilon_{tm-j}$$

$\forall (i,j),\ 0 \leq i < j \leq m-1$, we have by (2.9.b):

$$E[a_{im} a_{jm} \varepsilon_{tm-i} \varepsilon_{tm-j} \mid J_{tm-m}^{(m)}] = E[a_{im} a_{jm} \varepsilon_{tm-j} E[\varepsilon_{tm-i} \mid J_{tm-i-1}] \mid J_{tm-m}^{(m)}] = 0$$

and by (2.9.c), for $0 \leq k \leq m-1$:

$$E[\varepsilon_{tm-k}^2 \mid J_{tm-m}^{(m)}] = E[f_{tm-k-1} E[u_{tm-k}^2 \mid J_{tm-k-1}] \mid J_{tm-m}^{(m)}] = E[f_{tm-k-1} \mid J_{tm-m}^{(m)}]$$

Hence:

$$f_{tm-m}^{(m)} = E[\sum_{k=0}^{m-1} a_{km}^2 \varepsilon_{tm-k}^2 \mid J_{tm-m}^{(m)}] = E[\sum_{k=0}^{m-1} a_{km}^2 f_{tm-k-1} \mid J_{tm-m}^{(m)}]$$

By (2.11.a), we have:

$$\forall k, \quad f_{tm-k-1} = \omega \frac{1-\gamma^{m-k-1}}{1-\gamma} + \gamma^{m-k-1} f_{tm-m} + \sum_{i=0}^{m-k-2} \gamma^i \nu_{tm-k-1-i}, \tag{A.2}$$

so

$$E[f_{tm-k-1} \mid J_{tm-m}^{(m)}] = \omega \frac{1-\gamma^{m-k-1}}{1-\gamma} + \gamma^{m-k-1} E[f_{tm-m} \mid J_{tm-m}^{(m)}].$$

Hence

$$f_{tm-m}^{(m)} = a^{(m)} E[f_{tm-m} \mid J_{tm-m}^{(m)}] + b^{(m)}$$

where

$$a^{(m)} = \sum_{k=0}^{m-1} a_{km}^2 \gamma^{m-k-1}$$

$$b^{(m)} = \frac{\omega}{1-\gamma} \sum_{k=0}^{m-1} a_{km}^2 (1-\gamma^{m-k-1}).$$

Since

$$f_{tm} = \omega \frac{1-\gamma^m}{1-\gamma} + \gamma^m f_{tm-m} + \sum_{k=0}^{m-1} \gamma^k \nu_{tm-k},$$

we have

$$E[f_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = a^{(m)} E[E[f_{tm} \mid J_{tm}^{(m)}] \mid J_{tm-m}^{(m)}] + b^{(m)}$$

$$= a^{(m)} \omega \frac{1-\gamma^m}{1-\gamma} + b^{(m)} + \gamma^m a^{(m)} E[f_{tm-m} \mid J_{tm-m}^{(m)}] + a^{(m)} E[(\sum_{k=0}^{m-1} \gamma^k \nu_{tm-k}) \mid J_{tm-m}^{(m)}]$$

so

$$E[f_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = a^{(m)} \omega \frac{1-\gamma^m}{1-\gamma} + b^{(m)} + \gamma^m (f_{tm-m}^{(m)} - b^{(m)}) + a^{(m)} E[(\sum_{k=0}^{m-1} \gamma^k \nu_{tm-k}) \mid J_{tm-m}^{(m)}]$$

$$= w^{(m)} + \gamma^{(m)} f_{tm-m}^{(m)}$$

where

$$\gamma^{(m)} = \gamma^m \tag{A.3}$$

$$\omega^{(m)} = a^{(m)} \omega \frac{1-\gamma^m}{1-\gamma} + b^{(m)} (1-\gamma^m) \tag{A.4}$$

$$\nu_{tm}^{(m)} = f_{tm}^{(m)} - \omega^{(m)} - \gamma^{(m)} f_{tm-m}^{(m)}. \tag{A.5}$$

Let

$$u_{tm}^{(m)} = \frac{\varepsilon_{tm}^{(m)}}{\sqrt{f_{tm-m}^{(m)}}}$$

We must show that $u_{tm}^{(m)}$ and $\nu_{tm}^{(m)}$ are conformable to the restrictions (2.9.b), (2.9.c) and (2.11.b).

• restriction (2.9.b):

$$E[u_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = \frac{1}{\sqrt{f_{tm-m}^{(m)}}} E[\sum_{k=0}^{m-1} a_{km} E[\varepsilon_{tm-k} \mid J_{tm-k-1}] \mid J_{tm-m}^{(m)}] = 0.$$

• restriction (2.9.c):

$$E[(u_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}] = \frac{1}{f_{tm-m}^{(m)}} E[(\varepsilon_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}] = 1.$$

• restriction (2.11.b):

$$E[\nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[f_{tm}^{(m)} - \omega^{(m)} - \gamma^{(m)} f_{tm-m}^{(m)} \mid J_{tm-m}^{(m)}] = E[f_{tm}^{(m)} \mid J_{tm-m}^{(m)}] - \omega^{(m)} - \gamma^{(m)} f_{tm-m}^{(m)} = 0.$$

This achieves the Proof of Proposition 2.8.

Proof of Proposition 2.9

We should prove that $u_{tm}^{(m)}$ and $\nu_{tm}^{(m)}$ are conformable to the restrictions (2.24.a) and (2.24.b).

Restriction (2.24.a):

$$E[(u_{tm}^{(m)})^3 \mid J_{tm-m}^{(m)}] = \frac{1}{(f_{tm-m}^{(m)})^{\frac{3}{2}}} E[\sum_{0 \le i,j,k \le m-1} a_{im} a_{jm} a_{km} \varepsilon_{tm-i} \varepsilon_{tm-j} \varepsilon_{tm-k} \mid J_{tm-m}^{(m)}].$$

Let (i, j, k) as $i \le j \le k \le m - 1$.

• If $i < j \le k$, 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.1).

So we have:

$$E[(u_{tm}^{(m)})^3 \mid J_{tm-m}^{(m)}] = 0.$$

Restriction (2.24.b):

$$E[u_{tm}^{(m)}\nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = \frac{a^{(m)}}{\sqrt{f_{tm-m}^{(m)}}}E[\sum_{0\leq i,j\leq m-1} a_{im}\gamma^j\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}^{(m)}].$$

But (see third case of the proof of Proposition 2.7), we already know that (2.24) implies that:

$$E[\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}] = 0$$

for $i, j = 0, 1, ..., m-1$.
A fortiori:

$$E[\varepsilon_{tm-i}\nu_{tm-j} \mid J_{tm-m}^{(m)}] = 0$$

Hence, we have:

$$E[u_{tm}^{(m)}\nu_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = 0.$$

This achieves the Proof of Proposition 2.9.

Proof of Proposition 2.10

Let

$$\varepsilon_{th}^{(h)} = log\frac{S_{th}}{S_{(t-1)h}}, \quad t \in \mathbf{N},$$

and

$$f_{(t-1)h}^{(h)} = Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}] = E[(\varepsilon_{th}^{(h)})^2 \mid J_{(t-1)h}^{(h)}]$$

with

$$J_{(t-1)h}^{(h)} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, \sigma_{\tau h}^2, \tau < t, \tau \in \mathbf{N}).$$

We have:

$$\varepsilon_{th}^{(h)} = \int_{(t-1)h}^{th} \sigma_u \; dW_u$$

$$f_{(t-1)h}^{(h)} = E[(\int_{(t-1)h}^{th} \sigma_u \; dW_u)^2 \mid J_{(t-1)h}^{(h)}] = \int_{(t-1)h}^{th} E[\sigma_u^2 \mid J_{(t-1)h}^{(h)}] \; du$$

By using (2.33), we obtain:

$$E[\sigma_u^2 \mid J_{(t-1)h}^{(h)}] = \theta + e^{-k(u-(t-1)h)}(\sigma_{(t-1)h}^2 - \theta)$$

Hence:

$$f_{(t-1)h}^{(h)} = \int_{(t-1)h}^{th}[(\theta + e^{-k(u-(t-1)h)}(\sigma_{(t-1)h}^2 \; - \theta)]du$$

$$= \theta h + (\sigma_{(t-1)h}^2 - \theta)\frac{1-e^{-kh}}{k} = a^{(h)}\sigma_{(t-1)h}^2 \; + b^{(h)}$$

with

$$a^{(h)} = \frac{1 - e^{-kh}}{k}$$

$$b^{(h)} = \theta\ (h - \frac{1 - e^{-kh}}{k})$$

Using again (2.33), we obtain:

$$\sigma^2_{th} = \theta(1 - e^{-kh}) + e^{-kh}\sigma^2_{(t-1)h} + \delta e^{-kh}\int_{(t-1)h}^{th} e^{k(u-(t-1)h)}(\sigma^2_u)^\lambda\ dW^\sigma_u$$

Hence

$$f^{(h)}_{th} = \omega^{(h)} + \gamma^{(h)} f^{(h)}_{(t-1)h} + \nu^{(h)}_{th}$$

with

$$\omega^{(h)} = h\theta(1 - e^{-kh})$$

$$\gamma^{(h)} = e^{-kh}$$

$$\nu^{(h)}_{th} = \frac{(1 - e^{-kh})}{k}\delta e^{-kh}\int_{(t-1)h}^{th} e^{k(u-(t-1)h)}(\sigma^2_u)^\lambda\ dW^\sigma_u$$

Let

$$u^{(h)}_{th} = \frac{\varepsilon_{th}}{\sqrt{f_{(t-1)h}}}.$$

We must show that $u^{(h)}_{th}$ and $\nu^{(h)}_{th}$ are conformable to (2.9.b), (2.9.c) and (2.11.b).
Let $J_{(t-1)h} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, \sigma_{\tau h}, \tau \leq t-1, \tau \in I\!R)$.

• Restriction (2.9.b): since $J^{(h)}_{th} \subset J_{th}$, we have

$$E[u^{(h)}_{(t+1)h} \mid J^{(h)}_{th}] = \frac{1}{\sqrt{f^{(h)}_{th}}} E[E[\int_{th}^{(t+1)h} \sigma_u dW_u \mid J_{th}] \mid J^{(h)}_{th}] = 0.$$

• Restriction (2.9.c):

$$E[(u^{(h)}_{(t+1)h})^2 \mid J^{(h)}_{th}] = \frac{1}{f^{(h)}_{th}} E[E[(\int_{th}^{(t+1)h} \sigma_u dW_u)^2 \mid J_{th}] \mid J^{(h)}_{th}] = 1.$$

• Restriction (2.11.b):

$$E[\nu^{(h)}_{(t+1)h} \mid J^{(h)}_{th}] = \frac{1 - e^{-kh}}{k}\delta E[E[\int_{th}^{(t+1)h} e^{k(u-th)}(\sigma^2_u)^\lambda dW^\sigma_u \mid J_{th}] \mid J^{(h)}_{th}] = 0.$$

This achieves the Proof of Proposition 2.10.

Proof of Proposition 3.1

Similar to the poof of Proposition 2.1.

Proof of Proposition 3.2

We have:

$$\varepsilon_t^2 = f_{t-1}u_t^2 = f_{t-1} + v_t$$

with

$$v_t = f_{t-1}(u_t^2 - 1).$$

Note that $E[v_t \mid J_{t-1}] = 0$. 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: $Deg(Det(Id - \Gamma L)) \leq p$ and $Deg(e'(Id - \Gamma L)^*) \leq p - 1$ where $Deg(.)$ is the maximal degree of the lag polynomials, coefficients of the matrix. So f_t is an ARMA(p,p-1).

Let $z_t = e'(Id - \Gamma L)^*V_t$ and $\omega = e'(Id - \Gamma)^*\Omega$. We have:

$$Det(Id - \Gamma L)\varepsilon_t^2 = Det(Id - \Gamma L)(f_{t-1} + v_t) = \omega + z_{t-1} + Det(I - \Gamma L)v_t = \omega + \omega_t.$$

Since $Deg((Id - \Gamma L)^*) \leq p - 1$ and $E[V_t \mid J_{t-1}] = 0$, we have

$$E[(Id - \Gamma L)^*V_{t-1} \mid J_{t-p-1}] = 0.$$

Since $Deg(Det(Id - \Gamma L)) \leq p$ and $E[v_t \mid J_{t-1}] = 0$, we have

$$E[Det(Id - \Gamma L)v_t \mid J_{t-p-1}] = 0.$$

So, we have:

$$E[\omega_t \mid J_{t-p-1}] = e'E[(Id - \Gamma L)^*V_{t-1} \mid J_{t-p-1}] + E[Det(Id - \Gamma L)v_t \mid J_{t-p-1}] = 0$$

and we conclude that ω_t is an MA(p) and that ε_t^2 is an ARMA(p,p).

Let

$$Det(Id - \Gamma L) = \sum_{i=0}^{p} r_i L^i.$$

We have

$$r_0 = Det(Id - \Gamma\ 0) = Det(Id) = 1.$$

and

$$\varepsilon_t^2 + \sum_{i=1}^{p} r_i \varepsilon_{t-i}^2 = \omega + \omega_t.$$

Let $\gamma_i = -r_i, \ i = 1, 2..p$. We have:

$$\varepsilon_t^2 - \omega - \sum_{i=1}^{p} \gamma_i \varepsilon_{t-i}^2 = \omega_t$$

The persistence parameter is

$$\sum_{i=1}^{p} \gamma_i = -\sum_{i=1}^{p} r_i = 1 - \sum_{i=0}^{p} r_i = 1 - Det(Id - \Gamma\ 1) = 1 - Det(Id - \Gamma).$$

Proof of Proposition 3.3

This proof is similar to the proof of Proposition 2.8.
Let $f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[(\varepsilon_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}]$. We have:

$$f_{tm-m}^{(m)} = E[\sum_{k=0}^{m-1} a_{km}^2 f_{tm-k-1} \mid J_{tm-m}^{(m)}] = e' E[\sum_{k=0}^{m-1} a_{km}^2 F_{tm-k-1} \mid J_{tm-m}^{(m)}]$$

$$= e'(A^{(m)} E[F_{tm-m} \mid J_{tm-m}^{(m)}] + B^{(m)})$$

with

$$A^{(m)} = \sum_{k=0}^{m-1} a_{km}^2 \Gamma^{m-k-1}$$

$$B^{(m)} = (\sum_{k=0}^{m-1} a_{km}^2 (\sum_{i=0}^{m-k-2} \Gamma^i))\Omega$$

Let

$$F_{tm-m}^{(m)} = A^{(m)} E[F_{tm-m} \mid J_{tm-m}^{(m)}] + B^{(m)}$$

We have:

$$F_{tm}^{(m)} = \Omega^{(m)} + \Gamma^{(m)} F_{tm-m}^{(m)} + V_{tm}^{(m)}$$

with

$$\Gamma^{(m)} = \Gamma^m$$

$$\Omega^{(m)} = A^{(m)} (\sum_{k=0}^{m-1} \Gamma^k)\Omega + (Id - \Gamma^m) B^{(m)}$$

To conclude, we have to show that:

$$u_{tm}^{(m)} = \frac{\varepsilon_{tm}^{(m)}}{\sqrt{f_{tm-m}^{(m)}}} \quad \text{and} \quad V_{tm}^{(m)} = F_{tm}^{(m)} - \Omega^{(m)} - \Gamma^{(m)} F_{tm-m}^{(m)}$$

are conformable to the restrictions (3.1) and (3.2). This is exactly the same proof that for the proof of Proposition 2.8.

Proof of Proposition 3.4

This proof is similar to the proof of Proposition 2.10.

Let

$$\varepsilon_{th}^{(h)} = log\frac{S_{th}}{S_{(t-1)h}}$$

and

$$f_{(t-1)h}^{(h)} = Var[\varepsilon_{th}^{(h)} \mid J_{(t-1)h}^{(h)}] = E[(\varepsilon_{th}^{(h)})^2 \mid J_{(t-1)h}^{(h)}]$$

with

$$J_{(t-1)h}^{(h)} = \sigma(Log\frac{S_{\tau h}}{S_{(\tau-1)h}}, F_{\tau h}^c, \tau \leq t, \tau \in \mathbf{N}).$$

The multivariate analog of (2.33) is:

$$F_{t+h}^c = (Id - e^{-Kh})\Theta \ + \ e^{-Kh}F_t^c \ + \ e^{-Kh}\int_t^{t+h} e^{K(u-t)}M_{22}Diag(1, F_u^{c'})^{1/2}dW_u \quad \text{(A.6)}$$

where M_{22} is the $(p \times (p+1))$ matrix defined by $M_{22} = (0, I_p)N$.

Let us define $N = (n_{ij})_{1\leq i\leq p+1, 1\leq j\leq p+1}$ and $e = (n_{1i}^2)_{2\leq i\leq p+1}$. We assume that $n_{11} = 0$ to have the convenient representation. This only corresponds to a change in the intercept of the vector F_t^c.

We have:

$$\varepsilon_{th}^{(h)} = \int_{(t-1)h}^{th} (1, 0..0)NDiag(1, F_u^{c'})^{1/2}\ dW_u$$

and

$$(1, 0..0)NDiag(1, F_u^{c'})N'(1, 0..0)' = e'F_u^c$$

Hence:

$$f_{(t-1)h}^{(h)} = E[(\int_{(t-1)h}^{th} (1, 0..0)NDiag(1, F_u^{c'})^{1/2}\ dW_u)^2 \mid J_{(t-1)h}^{(h)}]$$

$$= \int_{(t-1)h}^{th} E[e'F_u^c \mid J_{(t-1)h}^{(h)}]\ du = e'\int_{(t-1)h}^{th} E[F_u^c \mid J_{(t-1)h}^{(h)}]\ du$$

By (A.6), we have:

$$f_{(t-1)h}^{(h)} = e'\int_{(t-1)h}^{th} [(Id - e^{-K(u-(t-1)h)})\Theta \ + \ e^{-K(u-(t-1)h)}F_{(t-1)h}^c]\ du,$$

that is

$$f_{(t-1)h}^{(h)} = e'F_{(t-1)h}^{(h)}$$

with

$$F_{(t-1)h}^{(h)} = A^{(h)}F_{(t-1)h}^c + B^{(h)}$$

where

$$A^{(h)} = \int_{(t-1)h}^{th} e^{-K(u-(t-1)h)} \ du = K^{-1}(Id - e^{-Kh})$$

$$B^{(h)} = \int_{(t-1)h}^{th} [(Id - e^{-K(u-(t-1)h)}]\Theta \ du = [Id \ h - K^{-1}(Id - e^{-Kh})]\Theta.$$

We have:

$$F_{th}^{(h)} = \Omega^{(h)} + \Gamma^{(h)} F_{(t-1)h}^{(h)} + V_{th}^{(h)}$$

with

$$\Omega^{(h)} = h(Id - e^{-Kh})\Theta$$

$$\Gamma^{(h)} = e^{-Kh}$$

$$V_{th}^{(h)} = K^{-1}(Id - e^{-Kh}) \ e^{-Kh} \int_{(t-1)h}^{th} e^{K(u-(t-1)h)} M_{22} Diag(1, F_u^{c'})^{1/2} dW_u$$

Let

$$u_{th}^{(h)} = \frac{\varepsilon_{th}}{\sqrt{f_{(t-1)h}}}.$$

We must show that $u_{th}^{(h)}$ and $V_{th}^{(h)}$ are conformable to the restrictions of definition 3.1. The proof is similar than for the Proposition 2.10.

Proof of Proposition 3.5

In this proof, we will produce a VAR(1) such that the GARCH(p,p) conditional variance process h_t is a marginalization of this vector.

Let $\varepsilon_t = \sqrt{h_t} u_t$ a GARCH(p,p), with

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

Hence

$$h_t = \omega + \sum_{i=1}^{p} (\alpha_i + \beta_i) h_{t-i} + \sum_{i=1}^{p} \alpha_i h_{t-i} (u_{t-i}^2 - 1)$$

This implies that the process h_t is an ARMA(p,p-1) and its innovation process is $\xi_{t-1} = \alpha_1 h_{t-1}(u_{t-1}^2 - 1)$. Then, there exist coefficients (λ_i, ψ_i), $i = 1, ..p$ of modulus smaller than one and such that:

$$[\prod_{i=1}^{p} (1 - \lambda_i L)](h_t - \mu) = [\prod_{i=1}^{p-1} (1 - \psi_i L)] \xi_{t-1}$$

where $\mu = E[h_t]$ and L is the Lag operator. For the simplicity of the proof, we assume that the coefficients λ_i are different, and are not equal to zero (other cases could be handled in the same way). Let us define for j=1,2..p,:

$$A_j(x) = \prod_{i=1, i \neq j}^{p} (1 - \lambda_i x).$$

For j=1,2..p, $A_j(x)$ is in $\mathbf{C}^{p-1}[X]$, the family of polynomial of degree smaller or equal than $p-1$ and with coefficients in the field of complex numbers $\mathbf{C}$. We have

$$\forall i, i \neq j, \quad A_j(\frac{1}{\lambda_i}) = 0$$

and

$$A_j(\frac{1}{\lambda_j}) \neq 0.$$

Therefore $A_j(x),\ j = 1, ..p$ is a basis of $\mathbf{C}^{p-1}[X]$. $\prod_{i=1}^{p-1}(1-\psi_i x)$ is in $\mathbf{C}^{p-1}[X]$, so there exist $a_j,\ j = 1...p$ such that:

$$\prod_{i=1}^{p-1}(1-\psi_i x) = \sum_{j=1}^{p} a_j A_j(x).$$

Let

$$F_{j,t-1} = \frac{\mu}{p} + (1-\lambda_j L)^{-1} a_j \xi_{t-1}.$$

We have:

$$[\prod_{i=1}^{p}(1-\lambda_i L)](F_{j,t-1} - \frac{\mu}{p}) = A_j(L) a_j \xi_{t-1}.$$

Hence

$$\sum_{j=1}^{p}[\prod_{i=1}^{p}(1-\lambda_i L)(F_{j,t-1} - \frac{\mu}{p})] = \sum_{j=1}^{p} A_j(L) a_j \xi_{t-1},$$

that is

$$[\prod_{i=1}^{p}(1-\lambda_i L)][\sum_{j=1}^{p}(F_{j,t-1} - \frac{\mu}{p})] = [\sum_{j=1}^{p} a_j A_j(L)]\xi_{t-1} = \prod_{i=1}^{p-1}(1-\psi_i L)\xi_{t-1}$$

But the ARMA representation defines an unique second order stationary process, we can conclude that

$$h_t = \sum_{j=1}^{p} F_{j,t-1}$$

that is $h_t = e' F_{t-1}$ with $e' = (1, ..1)$ and $F_t' = (F_{1,t}, ..F_{p,t})$. Moreover, it is clear that F_t is a VAR(1) w.r.t. I_t.

This achieves the proof of Proposition 3.5.

Proof of Proposition 3.6

Since $F_t \in J_t^*$ and $J_t^* \subset J_t$, we have:

$$E[y_t \mid J_{t-1}^*] = c + d' F_{t-1} + E[E[\varepsilon_t \mid J_{t-1}] \mid J_{t-1}^*] = c + d' F_{t-1}.$$

By Proposition 3.1, we know that ε_t is a SR-SARV(p) w.r.t. J_t^*. To complete the proof, we should prove that (y_t, F_t) is conformable to the restriction (3.15.f) w.r.t. J_t^*. We have:

$$
\begin{aligned}
&Vech(Var[\begin{pmatrix} y_t \\ F_t \end{pmatrix} \mid J_{t-1}^*]) \\
&= Vech(E[Var[\begin{pmatrix} y_t \\ F_t \end{pmatrix} \mid J_{t-1}] \mid J_{t-1}^*] + Var[E[\begin{pmatrix} y_t \\ F_t \end{pmatrix} \mid J_{t-1}] \mid J_{t-1}^*]) \\
&= E[R + S\ F_{t-1} \mid J_{t-1}^*] + Vech(Var[\begin{pmatrix} c + d'F_{t-1} \\ \Omega + \Gamma F_{t-1} \end{pmatrix} \mid J_{t-1}^*]) \\
&= R + S\ F_{t-1} + 0 = R + S\ F_{t-1}.
\end{aligned}
$$

This achieves the proof of Proposition 3.6.

Proof of Proposition 3.7

The proof will be performed by induction on k, taking into account that for $k = 1$, h_t^k is by definition a marginalization of $(1, F_{t-1}')'$:

$$h_t = e'(1, F_{t-1}')'.$$

Let as assume that:

$$h_t^r = e_r'(1, F_{t-1}')', \quad for\ r = 1, 2..k.$$

Then

$$Cov(h_t, h_t^k \mid J_{t-2}) = e'Cov((1, F_{t-1}')', (1, F_{t-1}')' \mid J_{t-2})e_k = e'\begin{pmatrix} 0 & 0 \\ 0 & Var[F_{t-1} \mid J_{t-2}] \end{pmatrix} e_k$$

Since, by (3.15.f), $Var[F_{t-1} \mid J_{t-2}]$ has all its coefficients expressed as a marginalization $(1, F_{t-2}')'$, it is then clear that $Cov(h_t, h_t^k \mid J_{t-2})$ is itself a marginalization of $(1, F_{t-2}')'$.

But, by the ARCH representation:

$$h_t = \omega + \alpha\varepsilon_{t-1}^2,$$

we have:

$$h_t^k = \omega^k + \alpha^k \varepsilon_{t-1}^{2k} + \sum_{j=1}^{k-1} C_k^j \omega^j \alpha^{k-j} \varepsilon_{t-1}^{2(k-j)}$$

and thus:

$$
\begin{aligned}
Cov(h_t, h_t^k \mid J_{t-2}) &= \alpha^{k+1} h_{t-1}^{k+1} Cov(u_{t-1}^2, u_{t-1}^{2k} \mid J_{t-2}) \\
&+ \sum_{i=0}^{k} C_k^j \omega^j \alpha^{k+1-j} h_{t-1}^{k+1-j} Cov(u_{t-1}^2, u_{t-1}^{2(k-j)} \mid J_{t-2}).
\end{aligned}
$$

Moreover, since ε_t is a strong ARCH(1) process: $Cov(u_{t-1}^2, u_{t-1}^{2(k-j)} \mid J_{t-2})$ for $j = 0, 1..k-1$, is a fixed real number $\alpha(k-j)$. Taking into account the induction assumption, we get:

$$Cov(h_t, h_t^k \mid J_{t-2}) = \alpha^{k+1} h_{t-1}^{k+1} \alpha(k) + + \sum_{i=0}^{k} C_k^j \omega^j \alpha^{k+1-j} \alpha(k-j) e'_{k+1-j} (1, F'_{t-2})'.$$

Since $Cov(h_t, h_t^k \mid J_{t-2})$ is also a marginalization of $(1, F'_{t-2})'$, we conclude that $\alpha^{k+1} h_{t-1}^{k+1} \alpha(k)$ is such a marginalization too. This proves that h_{t-1}^{k+1} is itself a marginalization of $(1, F'_{t-2})'$ (which achieves the proof by induction) as soon as:

$$\alpha^{k+1} \alpha(k) \neq 0$$

that is, as soon as:

$\alpha \neq 0$ (genuine ARCH effect) and $\alpha(k) = E(u_{t-1}^{2k+2}) - E(u_{t-1}^{2k}) \neq 0$.

This is the case if $u_t, t \in \mathbf{N}$, are i.i.d $\mathcal{N}(0,1)$ since in this case, we have $E(u_{t-1}^{2k+2}) = (2k+1) E(u_{t-1}^{2k+2})$.

Proof of Proposition 3.8

Let

$$g_{tm-m}^{(m)} = E[y_{tm}^{(m)} \mid J_{tm-m}^{(m)}],$$
$$\varepsilon_{tm}^{(m)} = y_{tm}^{(m)} - E[y_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = y_{tm}^{(m)} - g_{tm-m}^{(m)}$$

and

$$f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = E[(\varepsilon_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}].$$

We have:

$$y_{tm}^{(m)} = c(\sum_{k=0}^{m-1} a_{km}) + (\sum_{k=0}^{m-1} a_{km} d' F_{tm-k-1}) + \sum_{k=0}^{m-1} a_{km} \varepsilon_{tm-k}.$$

Hence:

$$g_{tm-m}^{(m)} = c(\sum_{k=0}^{m-1} a_{km}) + \sum_{k=1}^{m-1} a_{km} d' E[F_{tm-k-1} \mid J_{tm-m}^{(m)}]$$

$$= c(\sum_{k=0}^{m-1} a_{km}) + (\sum_{k=1}^{m-1} a_{km} d' (\sum_{j=0}^{k-m-1} \Gamma^j) \Omega + \Gamma^{k-m} F_{tm-m}) = c^{(m)} + d^{(m)'} F_{tm-m}$$

where

$$c^{(m)} = c(\sum_{k=0}^{m-1} a_{km}) + (\sum_{k=1}^{m-1} a_{km} d' (\sum_{j=0}^{m-k-1} \Gamma^j) \Omega$$

$$d^{(m)} = (\sum_{k=1}^{m-1} a_{km} \Gamma^{k-m})' \ d.$$

We have:

$$\varepsilon_{tm}^{(m)} = y_{tm}^{(m)} - g_{tm-m}^{(m)}$$

$$= d'(\sum_{k=1}^{m-1} a_{km}(\sum_{j=0}^{m-k-1} \Gamma^j V_{tm-k-1-j}) + \sum_{k=0}^{m-1} a_{km}\varepsilon_{tm-k} = \sum_{k=1}^{m-1} L_{km}V_{tm-k} + \sum_{k=0}^{m-1} a_{km}\varepsilon_{tm-k}$$

where

$$L_{km} = d' \sum_{i=0}^{k-1} a_{(k-i-1)m}\Gamma^i$$

Hence:

$$\varepsilon_{tm}^{(m)} = \sum_{k=0}^{m-1} P_{km}\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix}$$

where $P_{km} = (a_{km}, L_{km})$ for $k = 1, ..m$ and $P_{0m} = (a_{0m}, 0)$.

If we define $F_{tm}^{(m)} = F_{tm}$, $F_{tm}^{(m)}$ is a VAR(1) with:

$$F_{tm} = \Omega^{(m)} + \Gamma^{(m)}F_{tm-m} + V_{tm}^{(m)}$$

with

$$\Omega^{(m)} = \sum_{k=0}^{m-1} \Gamma^k\Omega$$

$$\Gamma^{(m)} = \Gamma^m$$

$$V_{tm}^{(m)} = \sum_{k=0}^{m-1} \Gamma^k V_{tm-k} = \sum_{k=0}^{m-1} (0, \Gamma^k)\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix}$$

We will prove that $\begin{pmatrix} \varepsilon_{tm}^{(m)} \\ V_{tm}^{(m)} \end{pmatrix}$ is conformable to (3.15.f). We have:

$$\begin{pmatrix} \varepsilon_{tm}^{(m)} \\ V_{tm}^{(m)} \end{pmatrix} = \sum_{k=0}^{m-1} T_{km}\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix}$$

with $T_{km} = \begin{pmatrix} P_{km} \\ 0, \ \Gamma^k \end{pmatrix}$.

Since $(\varepsilon_{tm-k}, V'_{tm-k})$ are not correlated, we have:

$$Var[\begin{pmatrix} \varepsilon_{tm}^{(m)} \\ V_{tm}^{(m)} \end{pmatrix} \mid J_{tm-m}^{(m)}] = \sum_{k=0}^{m-1} Var[T_{km}\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix} \mid J_{tm-m}^{(m)}] = \sum_{k=0}^{m-1} T_{km}Var[\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix} \mid J_{tm-m}^{(m)}]T'_{km}.$$

Let us recall some properties of the Vec and Vech operators, quoted from Lutkepohl (1991).Let D_* and L_* the duplication and elimination matrices, with the appropriate sizes. We have:

$$Vec(A) = D_*Vech(A),$$
$$Vech(A) = L_*Vec(A),$$
$$Vec(ABC) = (C' \otimes A)Vec(B).$$

Hence

$$Vech(ABC) = L_*(C' \otimes A)Vec(B) = L_*(C' \otimes A)D_*Vech(B).$$

Then we have:

$$Vech(Var[\begin{pmatrix} \varepsilon_{tm}^{(m)} \\ V_{tm}^{(m)} \end{pmatrix} \mid J_{tm-m}^{(m)}]) = \sum_{k=0}^{m-1} Vech(T_{km} Var[\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix} \mid J_{tm-m}^{(m)}] T'_{km})$$

$$= \sum_{k=0}^{m-1} L_*(T_{km} \otimes T_{km}) D_* Vech(Var[\begin{pmatrix} \varepsilon_{tm-k} \\ V_{tm-k} \end{pmatrix} \mid J_{tm-m}^{(m)}])$$

$$= \sum_{k=0}^{m-1} L_*(T_{km} \otimes T_{km}) D_* E[R + SF_{tm-k-1} \mid J_{tm-m}^{(m)}]$$

$$= \sum_{k=0}^{m-1} L_*(T_{km} \otimes T_{km}) D_*[R + S[(\sum_{i=0}^{m-k-1} \Gamma^i)\Omega + \Gamma^{m-k} F_{tm-m}]] = R^{(m)} + S^{(m)} F_{tm-m}$$

where

$$R^{(m)} = \sum_{k=0}^{m-1} L_*(T_{km} \otimes T_{km}) D_*[R + S[(\sum_{i=0}^{m-k-1} \Gamma^i)\Omega]]$$

$$S^{(m)} = \sum_{k=0}^{m-1} L_*(T_{km} \otimes T_{km}) D_* S \Gamma^{m-k}$$

Let $r_1^{(m)}$ the first coefficient of $R^{(m)}$, $e_1^{(m)'} = (S_{1i}^{(m)})_{1 \leq i \leq p}$ and $e^{(m)'} = (1, e_1^{(m)'})'$. We Have:

$$f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = e^{(m)'}(1, F'_{tm-m})'.$$

Let

$$u_{tm}^{(m)} = \frac{\varepsilon_{tm}^{(m)}}{\sqrt{f_{tm-m}^{(m)}}}$$

We must show that $u_{tm}^{(m)}$ is conformable to the restrictions (3.15.d) and (3.15.e).
By construction $E[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = 0$ and $f_{tm-m}^{(m)} = Var[\varepsilon_{tm}^{(m)} \mid J_{tm-m}^{(m)}]$. This implies that $E[u_{tm}^{(m)} \mid J_{tm-m}^{(m)}] = 0$ and $E[(u_{tm}^{(m)})^2 \mid J_{tm-m}^{(m)}] = 1$, that is $u_{tm}^{(m)}$ is conformable to (3.15.d) and (3.15.e).
This achieves the proof of Proposition 3.8.
Proof of Proposition 3.9 Let us first show an exact discretisation formula in a general setting. Let a multivariate process X_t of size n such that:

$$dX_t = G(H - X_t) + N(X_t)dW_t \tag{A.7}$$

We define $Y_t = e^{Gt}(X_t - H)$. By the Ito lemma, we have

$$dY_t = e^{Gt} N(X_t) dW_t$$

Hence

$$Y_{t+\Delta t} = Y_t + \int_t^{t+\Delta t} e^{Gu} N(X_u) dW_u$$

and

$$X_{t+\Delta t} = [Id_n - e^{-G\Delta t}]H + e^{-G\Delta t}X_t + e^{-\Delta t}\int_t^{t+\Delta t} e^{G(u-t)}N(X_u)dW_u \tag{A.8}$$

Let us consider the process defined by (3.18) , we assume, without loss of generality, that $-\psi'\Theta = \mu$[10]. The process defined by (3.18) is conformable to (A.7) with:

$$X_t = (LogS_t, F_t^{c'})', \quad H = (\mu, \Theta')', \quad G = \begin{pmatrix} 0 & -\psi' \\ 0 & K \end{pmatrix}, \quad N(X_t) = NDiag(1, F_t^{c'})^{1/2}.$$

Straightforward calculus show that in this case, for $k \geq 1$,

$$(Gh)^k = \begin{pmatrix} 0 & -h\psi'(Kh)^{k-1} \\ 0 & (Kh)^k \end{pmatrix}$$

and hence

$$e^{Gh} = \begin{pmatrix} 1 & -\psi' K^{-1}(e^{Kh} - Id_p) \\ 0 & e^{Kh} \end{pmatrix}$$

$$e^{-Gh} = \begin{pmatrix} 1 & -\psi' K^{-1}(e^{-Kh} - Id_p) \\ 0 & e^{-Kh} \end{pmatrix}.$$

By (A.8), we obtain:

$$\begin{pmatrix} LogS_{(t+1)h} \\ F^c_{(t+1)h} \end{pmatrix} = [Id_{p+1} - e^{-Gh}]\begin{pmatrix} \mu \\ \Theta \end{pmatrix} + e^{-Gh}\begin{pmatrix} LogS_{th} \\ F^c_{th} \end{pmatrix} + e^{-h}\int_{th}^{(t+1)h} e^{G(u-t)}NDiag(1, F_u^{c'})^{1/2}dW_u.$$

That is:

$$\begin{pmatrix} Log\frac{S_{(t+1)h}}{S_{th}} \\ F^c_{(t+1)h} \end{pmatrix} = \begin{pmatrix} -\psi' K^{-1}(Id_p - e^{-Kh})\Theta \\ (Id_p - e^{-Kh})\Theta \end{pmatrix} + \begin{pmatrix} \psi' K^{-1}(Id_p - e^{-Kh})F^c_{th} \\ e^{-Kh}F^c_{th} \end{pmatrix} + \begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \tag{A.9}$$

where

$$\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} = e^{-h}\int_{th}^{(t+1)h}\begin{pmatrix} 1 & -\psi' K^{-1}(e^{K(u-th)} - Id_p) \\ 0 & e^{K(u-th)} \end{pmatrix}NDiag(1, F_u^{c'})^{1/2}dW_u.$$

The first component of the equation (A.9) gives the SR-SARV(p)-M equation, and the p last components give the dynamic of F^c_{th}, $t \in \mathbb{N}$. As usual,

$$E[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}] = E[E[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^c_{th}] \mid J^{(h)}_{th}] = 0.$$

where $J^c_{th} = \sigma(W_{\tau h}, \tau \leq t, \tau \in \mathbb{R})$. To achieve the proof, we should prove that $Var[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}]$ is conformable to (3.15.f). We have:

[10]This corresponds to a change of the intercept of F^c_t.

$$Vech(Var[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}]) = Vech(e^{-2h} \int_{th}^{(t+1)h} T(u)\ E[Diag(1, F^{c'}_u) \mid J^{(h)}_{th}] T(u)' du$$

where

$$T(u) = \begin{pmatrix} 1 & -\psi' K^{-1}(e^{K(u-th)} - Id_p) \\ 0 & e^{K(u-th)} \end{pmatrix} N$$

Hence

$$Vech(Var[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}]) = e^{-2h} \int_{th}^{(t+1)h} L_*[T(u) \otimes T(u)] D_* Vech[Diag(1, E[F^{c'}_u \mid J^{(h)}_{th}])] du$$

Then, by (A.6), we have:

$$Vech(Var[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}])$$

$$= e^{-2h} \int_{th}^{(t+1)h} L_*[T(u) \otimes T(u)] D_* Vech[Diag(1, (Id_p - e^{-K(u-th)})\Theta + e^{-K(u-th)} F^c_{th})')] du$$

There exist matrices $R(u)$ and $S(u)$ such that:

$$Vech[Diag(1, (Id_p - e^{-K(u-th)})\Theta + e^{-K(u-th)} F^c_{th})')] = R(u) + S(u) F^c_{th}.$$

We conclude that

$$Vech(Var[\begin{pmatrix} \varepsilon^{(h)}_{(t+1)h} \\ V^{(h)}_{(t+1)} \end{pmatrix} \mid J^{(h)}_{th}]) = R^{(h)} + S^{(h)} F^c_{th}$$

with

$$R^{(h)} = e^{-2h} \int_{th}^{(t+1)h} L_*[T(u) \otimes T(u)] D_* R(u) du$$

$$S^{(h)} = e^{-2h} \int_{th}^{(t+1)h} L_*[T(u) \otimes T(u)] D_* S(u) du$$

This achieves the proof of Proposition 3.9.

Proof of Proposition 3.10

Let L be the Lag operator. We have $y^{(m)}_{tm} = y_{tm}$ and

$$(1 - \rho L) y_{tm} = c + \varepsilon_{tm}.$$

Hence

$$(\sum_{i=0}^{m-1} \rho^i L^i)(1 - \rho L) y_{tm} = (1 - \rho^m L^m) y_{tm} = (\sum_{i=0}^{m-1} \rho^i L^i)(c + \varepsilon_{tm}),$$

that is

$$(1-\rho^{(m)}L^m)y_{tm}^{(m)} = c^{(m)} + \varepsilon_{tm}^{(m)}$$

where

$$\rho^{(m)} = \rho^m$$

$$c^{(m)} = (\sum_{i=0}^{m-1}\rho^i)c = \frac{1-\rho^m}{1-\rho}c$$

$$\varepsilon_{tm} = \sum_{i=0}^{m-1}\rho^i\varepsilon_{tm-i}$$

By proposition 3.3, we know that $\varepsilon_{tm}^{(m)}$ (with $a_{km} = \rho^k$) is SR-SARV(p) w.r.t J_{tm}. By the projection theorem 3.1 and since $J_{tm}^{(m)} \subset J_{tm}$, we conclude that that $\varepsilon_{tm}^{(m)}$ is also a SR-SARV(p) w.r.t. $J_{tm}^{(m)}$.

This achieves the proof of Proposition 3.10.

Proof of Proposition 3.11

The process defined by (3.22) is conformable to the one defined by (A.7) with

$$X_t = (y_t, F_t^{c'})', \quad H = (\theta, \Theta')', \quad G = \begin{pmatrix} k & 0 \\ 0 & K \end{pmatrix}, \quad N(X_t) = NDiag(1, F_t^{c'})^{1/2}.$$

By (A.8), we have:

$$\begin{pmatrix} y_{(t+1)h} \\ F_{(t+1)h}^c \end{pmatrix} = \begin{pmatrix} \theta(1-e^{-kh}) \\ (Id_p - e^{-Kh})\Theta \end{pmatrix} + \begin{pmatrix} e^{-kh}y_{th} \\ e^{-Kh}F_{th}^c \end{pmatrix} + \begin{pmatrix} \varepsilon_{(t+1)h}^{(h)} \\ V_{(t+1)}^{(h)} \end{pmatrix}$$

with

$$\begin{pmatrix} \varepsilon_{(t+1)h}^{(h)} \\ V_{(t+1)}^{(h)} \end{pmatrix} = e^{-h}\int_{th}^{(t+1)h} \begin{pmatrix} e^{k(u-th)} & 0 \\ 0 & e^{K(u-th)} \end{pmatrix} NDiag(1, F_u^{c'})^{1/2}dW_u.$$

We have again

$$E[\begin{pmatrix} \varepsilon_{(t+1)h}^{(h)} \\ V_{(t+1)}^{(h)} \end{pmatrix} \mid J_{th}^{(h)}] = E[E[\begin{pmatrix} \varepsilon_{(t+1)h}^{(h)} \\ V_{(t+1)}^{(h)} \end{pmatrix} \mid J_{th}^c] \mid J_{th}^{(h)}] = 0.$$

where $J_{th}^c = \sigma(W_{\tau h}, \tau \leq t, \tau \in I\!R)$. To achieve the proof, we should prove that $Var[\begin{pmatrix} \varepsilon_{(t+1)h}^{(h)} \\ V_{(t+1)}^{(h)} \end{pmatrix} \mid J_{th}^{(h)}]$ is conformable to (3.15.f). It is similar to the proof of Proposition 3.9.

This achieves the proof of Proposition 3.11.

Proof of Proposition 3.12

We have $y_t = c + \rho y_{t-1} + \varepsilon_t$, hence:

$$y_{tm-k} = c\sum_{i=0}^{m-k-1}\rho^i + \rho^{m-k}y_{tm-m} + \sum_{i=0}^{m-k-1}\rho^i\varepsilon_{tm-i} = c\frac{1-\rho^{m-k}}{1-\rho} + \rho^{m-k}y_{tm-m} + \sum_{i=0}^{m-k-1}\rho^i\varepsilon_{tm-i}.$$

This implies that:

$$y_{tm}^{(m)} = \sum_{k=0}^{m-1}a_{km}y_{tm-k} = \frac{c}{1-\rho}\sum_{k=0}^{m-1}a_{km}(1-\rho^{m-k}) + \sum_{k=0}^{m-1}a_{km}(1-\rho^{m-k})y_{tm-m} + \varepsilon_{tm}^{(m)}$$

where $\varepsilon_{tm}^{(m)} = \sum_{k=0}^{m-1}(\sum_{i=0}^{m-k-1}\rho^i\varepsilon_{tm-i})$. By the temporal aggregation theorem 3.3, we know that $\varepsilon_{tm}^{(m)}$ is a SR-SARV(p) w.r.t. J_{tm} and so w.r.t. $J_{tm}^{(m)}$.
This achieves the proof of Proposition 3.12.

Proof of Proposition 4.1

• Let Y_t a multivariate SR-SARV(p) and $Y_t^* = B\ Y_t$ where B is a matrix. Let $G_{t-1}^* = Var[Y_t^* \mid J_{t-1}]$ and $U_t^* = (G_{t-1}^*)^{-\frac{1}{2}}Y_t^*$. We have:

$$Y_t^* = (G_{t-1}^*)^{\frac{1}{2}}U_t^*$$

$$E[U_t^* \mid J_{t-1}] = (G_{t-1}^*)^{-\frac{1}{2}}B(G_{t-1})^{\frac{1}{2}}E[U_t \mid J_{t-1}] = 0$$

$$Var[U_t^* \mid J_{t-1}] = (G_{t-1}^*)^{-\frac{1}{2}}Var[Y_t^* \mid J_{t-1}](G_{t-1}^*)^{-\frac{1}{2}} = Id_q$$

$$Vech(G_t^*) = Vech(Var[Y_{t+1}^* \mid J_t]) = Vech(BVar[Y_{t+1} \mid J_t]B')$$

$$= L_*(B\otimes B)D_*Vech(Var[Y_{t+1} \mid J_t]) = L_*(B\otimes B)D_*Vech(G_t) = R^* + S^*F_t$$

where $R^* = L_*(B\otimes B)D_*R$ and $S^* = L_*(B\otimes B)D_*S$.

• Let Y_t a multivariate SR-SARV(p)-M and $Y_t^* = A + B\ Y_t$ where A and B are matrices. Let $G_{t-1}^* = Var[Y_t^* \mid J_{t-1}]$ and $U_t^* = (G_{t-1}^*)^{-\frac{1}{2}}B\varepsilon_t$. We have:

$$Y_t^* = C^* + D^*F_{t-1} + \varepsilon_t^*$$

where $C^* = A + BC$, $D^* = BD$ and $\varepsilon_t^* = (G_{t-1}^*)^{\frac{1}{2}}U_t^*$. We have:

$$E[U_t^* \mid J_{t-1}] = (G_{t-1}^*)^{-\frac{1}{2}}B(G_{t-1})^{\frac{1}{2}}E[U_t \mid J_{t-1}] = 0$$

$$Var[U_t^* \mid J_{t-1}] = (G_{t-1}^*)^{-\frac{1}{2}}Var[\varepsilon_t^* \mid J_{t-1}](G_{t-1}^*)^{-\frac{1}{2}} = (G_{t-1}^*)^{-\frac{1}{2}}Var[Y_t^* \mid J_{t-1}](G_{t-1}^*)^{-\frac{1}{2}} = Id_q$$

$$Vech(Var[\begin{pmatrix} Y_t^* \\ F_t \end{pmatrix} \mid J_{t-1}]) = Vech(Var[T\begin{pmatrix} Y_t \\ F_t \end{pmatrix} \mid J_{t-1}])$$

where $T = \begin{pmatrix} B & 0 \\ 0 & Ip \end{pmatrix}$. Hence:

$$Vech(Var[\begin{pmatrix} Y_t^* \\ F_t \end{pmatrix} \mid J_{t-1}]) = Vech(TVar[\begin{pmatrix} Y_t \\ F_t \end{pmatrix} \mid J_{t-1}]T')$$

$$= L_*(T \otimes T)D_* Vech(Var[\begin{pmatrix} Y_t \\ F_t \end{pmatrix} \mid J_{t-1}]) = R^* + S^* F_{t-1}$$

where $R^* = L_*(T \otimes T)D_* R$ and $S^* = L_*(T \otimes T)D_* S$.
This achieves the proof of Proposition 4.1.