2004s-37

Conditionally Heteroskedastic Factor Models: Identification and Instrumental Variables Estimation¤

Catherine Doz, Eric Renault

Série Scientifique
Scientific Series

Montréal
Juin 2004

CIRANO
Centre interuniversitaire de recherche en analyse des organisations

ISSN 1198-8177

Conditionally Heteroskedastic Factor Models: Identification and Instrumental Variables Estimation*

Catherine Doz†, *Eric Renault*‡

Résumé / *Abstract*

Cet article propose un cadre semi-paramétrique adapté à la modélisation de l'hétéroscédasticité conditionnelle multivariée. Nous montrons d'abord qu'un modèle factoriel à volatilité stochastique ne peut pas être identifié seulement à partir de la structure de variance conditionnelle des rendements, sauf si l'on impose des restrictions importantes au support de la loi de probabilité des facteurs latents. Nous proposons ensuite des restrictions alternatives permettant d'identifier le modèle de volatilité multivariée. Ces restrictions portent soit sur les moments d'ordre supérieur, soit sur une spécification de la prime de risque fondée sur un prix constant du risque des facteurs. Dans les deux cas, l'identification du modèle est obtenue à partir de restrictions sur les moments conditionnels, ce qui permet l'estimation par variables instrumentales. Une étape préliminaire de détermination du nombre de facteurs et d'identification de portefeuilles représentatifs est proposée. Elle est fondée sur une séquence de tests de sur-identification qui englobe les tests de caractéristiques communes d'Engle et Kozicki (1993).

Mots clés : évaluation d'actifs financiers, volatilité stochastique, modèles conditionnels à facteurs, hétéroscédasticité conditionnelle multivariée, caractéristiques communes, restrictions de moments conditionnels avec retards, Méthode des Moments Généralisés.

This paper provides a semiparametric framework for modelling multivariate conditional heteroskedasticity. First, we show that stochastic volatility factor models with possibly cross-correlated disturbances cannot be identified from returns conditional variance structure only, except when strong restrictions on the support of the probability distribution of latent factors volatility are maintained. Second, we provide an alternative way to maintain identifying restrictions through either higher order moments or through a specification of risk premiums based on constant prices of factor risks. In both cases, identification is obtained with conditional moment restrictions which pave the way for instrumental variables estimation and inference. A preliminary step of determination of the number of factors and identification of mimicking portfolios is proposed through a sequence of GMM overidentification tests which encompass Engle and Kozicki (1993) tests for common features.

Keywords: *asset pricing, stochastic volatility, conditional factor models, multivariate conditional heteroskedasticity, common features, multiperiod conditional moment restrictions, Generalized Method of Moments.*

---

* We would like to thank Manuel Arellano, Prosper Dovonon, Nour Meddahi and Enrique Sentana for helpful comments and suggestions on earlier versions of this paper. This joint work has been made possible by cross-invitations at CIREQ and THEMA which are gratefully aknowledged.

† University Cergy-Pontoise, THEMA, France, email: Catherine.Doz@eco.u-cergy.fr.

‡ CIRANO and CIREQ, University of Montreal, email: Eric.Renault@umontreal.ca.

1 Introduction

Estimation of large multivariate conditionally heteroskedastic models is notoriously challenging, requiring strong assumptions to make such estimation feasible. While several hundred parameters can be necessary to capture joint GARCH kind of dynamics of ten asset returns, more structure is needed to get a parsimonious characterization of the joint conditional covariance matrix.

Among the possible structures, the common factors model is quite popular for at least two reasons. First, as emphasized by Diebold and Nerlove (1989) and King, Sentana and Wadhwani (1994), they are well-suited to capture commonality in the conditional variance movements of the returns (regardless of correlation), as all asset prices react to the arrival of new information. In other words, common factors may represent news which are common among all asset prices. A second advantage of factor models is that they automatically guarantee a positive semidefinite conditional covariance matrix for returns, once we ensure that the conditional covariance matrix of the factors is itself positive semidefinite. A maintained assumption in this paper is that the common factors represent conditionally orthogonal influences, which implies that the factors conditional covariance matrix is diagonal (see Sentana (1998) for more general factor models, which are called oblique).

To summarize, we consider in the whole paper a vector y_{t+1} of n asset returns, observed at time $t+1$, which can be decomposed as :

$$y_{t+1} = \mu + \Lambda f_{t+1} + u_{t+1} \tag{1.1}$$

where f_{t+1} is a $K \times 1$ vector of unobserved common factors, Λ is the $n \times K$ matrix of associated factor loadings and u_{t+1} is a $n \times 1$ vector of idiosyncratic terms. This decomposition will help to characterize the conditional covariance matrix of returns $\Sigma_t = Var\left(y_{t+1} \left| J_t \right.\right)$ given some information set J_t that contains the past values: $y_\tau, \tau \leq t$ and $f_\tau, \tau \leq t$.

The characterization of Σ_t from (1.1) rests upon three basic assumptions. First, as already mentionned, we assume that $D_t = Var\left(f_{t+1} \left| J_t \right.\right)$ is a diagonal matrix. Second, factor loadings are interpreted as conditional betas coefficients of returns on factors, that is :

$$Cov\left(f_{t+1}, u_{t+1} \left| J_t \right.\right) = 0.$$

Third, some identifying assumptions must be maintained about the residual covariance matrix $\Omega_t = Var\left(u_{t+1} \left| J_t \right.\right)$ in order to keep the interpretation of residual shocks u_{t+1} as idiosyncratic ones. Based on those assumptions, the covariance factor structure which is the focus of interest of this paper will be characterized as:

$$\Sigma_t = \Lambda D_t \Lambda' + \Omega_t \tag{1.2}$$

In such a dynamic framework, the concept of idiosyncracy may actually be understood in two different ways. First, extending to a dynamic setting the Ross's (1976) initial intuition, we may adopt a conditional factor analysis approach by assuming that idiosyncratic shocks are conditionaly orthogonal, that is Ω_t is a diagonal matrix. Then parsimony is reached by the fact that only $(n + K)$ independent univariate conditionally heteroskedastic processes have to be specified: the K common factors processes and the n idiosyncratic shocks. This is the approach which has been followed by both Diebold and Nerlove (1989) and King, Sentana and Waddhwani (1994), even though the former assume in addition that Ω_t is a constant matrix. Irrespective of the assumption about the dynamics of Ω_t, the maintained assumption of diagonality allows one to resort at least to identification tools of conventional factor analysis estimation.

However, diagonality of Ω_t may be thought as a too restrictive assumption, particularly because it is not preserved by portfolio formation. This is the reason why Chamberlain and Rotschild (1983) introduced the concept of approximate factor structures, in which the idiosyncratic terms may be correlated, but only up to a certain degree. Since a versatile dynamic extension of the concept of approximate factor structure is still not much developped in the conditional heteroskedasticity literature (see however Sentana (2004)), we focus here on another concept of idiosyncracy, defined in line with common features which have been introduced by Engle and Kozicki (1993). More precisely, we assume that the residual covariance matrix Ω_t is a possibly nondiagonal constant positive definite matrix Ω:

$$\Sigma_t = \Lambda D_t \Lambda' + \Omega \tag{1.3}$$

The maintained assumption of the factor structure (1.3) is akin to see the K common factors as the K sources of conditional heteroskedasticity which should explain the commonality in the conditional variance movements of the returns. By contrast, the common features, defined from the $(n - K)$-dimensional orthogonal space of the range of Λ, characterize the vectorial space of conditionally homoskedastic portfolio returns.

Moreover, it is worth emphasizing that almost all the identification and estimation strategies put forward in this paper could be easily extended to a more general context with a time-varying matrix Ω_t, insofar as only the diagonal coefficients would be allowed to be time-varying. In other words, one could add to the idiosyncratic shocks u_{t+1} of our model (1.1)/(1.3) some orthogonal idiosyncratic shocks v_{t+1} with a conditional covariance matrix conformable to King, Sentana and Wadhwani (1994) conditional factor analysis model.

For sake of clarity, we prefer to focus here on the main contribution of this paper, that is instrumental variables (IV) identification and estimation of conditionally heteroskedastic factor models defined through the concept of common features, that is the mere fact that conditional heteroskedasticity is *a priori* limited to a restricted number of directions. This is a non-trivial issue for the following reason. While the search for common features may allow to identify the range of the matrix Λ of factor loadings, it does not protect against the following lack of identification. Roughly speaking, even when common factors are normalized by the maintained assumption:

$$ED_t = Var(f_{t+1}) = Id_K \tag{1.4}$$

where Id_K is the identity matrix of size K, it is always possible to transfer somme constant variance from factors to idiosyncratic terms through a convenient rescaling of the factor loadings. This degree of indetermination is clearly not innocuous for asset pricing and dynamic risk management as well. It turns out that this difficult identification issue has been overlooked in the literature until now since it may be solved by chance thanks to additional parametric assumptions. For instance, the maintained assumption of joint conditional normality of the idiosyncratic shocks allows Harvey, Ruiz and Shephard (1994) to propose QML consistent estimation of a model similar to (1.3) while, with GARCH factors, Fiorentini, Sentana and Shephard (2003) are even able to propose a likelihood-based estimation procedure.

We argue however that identification of the factor structure (1.3) with as little as possible additional assumptions is important for financial econometrics. Typically, both asset pricing and risk management issues are tightly related to two different features of asset returns' conditional probability distribution : conditional heteroskedasticity on the one hand, conditional tail behaviour on the other hand. It is then fairly important to be able to disentangle these two issues, that is to propose inference procedures about conditional variance dynamics whose validity is not contingent to some joint assumptions about the tail behaviour. This is typically the spirit IV procedures proposed here. We want to identify separately both common factors conditional heteroskedasticity dynamics and idiosyncratic variance by using, as far as possible, only observable conditional moment restrictions about the first two joint moments of asset returns.

Our main results are the following. First, we show that the required identification is much easier to meet when the common factors risk is priced, and a parametric model of price of factors risk is available. The general intuition is that the resulting risk premiums that show up in expected returns make the conditional variances of latent factors almost observable. To see this, let us consider a linear model of factor risk premiums :

$$E\left(f_{t+1}\left|J_t\right.\right) = \tau d_t$$

where the $K \times 1$ vector d_t stackles together the diagonal coefficients of the matrix D_t. Then the regression model (1.1) provides now two sets of conditional moment restrictions :

$$\begin{cases} E\left(y_{t+1}\left|J_t\right.\right) = \mu + \Lambda\tau d_t \\ Var\left(y_{t+1}\left|J_t\right.\right) = \Lambda D_t \Lambda' + \Omega \end{cases} \quad (1.5)$$

Considering these two sets of moment restrictions jointly will protect us against the aforementioned possibilities of variance transfer between D_t and Ω.

Full identification of the matrix Λ of factor loadings and also of the idiosyncratic covariance matrix Ω, is much more involved when one cannot take advantage of a non-zero price τ of factor risks. The solution put forward in this paper rests upon an additional model of joint conditional kurtosis of returns. Even though we consider as a pity to resort to such higher order moments joint assumptions, it is much less restrictive than usual parametric assumptions about the asset returns' joint probability distribution. Moreover, our chosen specification nests the most usual models for volatility factors, like strong GARCH (Diebold and Nerlove (1989)), affine diffusion stochastic volatility factors (Heston (1993), Duffie, Pan and Singleton (2000), Meddahi and Renault (2004) or Ornstein-Uhlenbeck like Levy volatility processes (Barndorff-Nielsen and Shephard (2001)).

In this paper, we assume that probability distributions of both common factors and disturbances are conditionally symmetric, which facilitates the characterization of conditional kurtosis of returns. Extensions which accomodate skewness or leverage effect are considered in a companion paper (Dovonon, Doz and Renault (2004)).

The paper is organized as follows. We first discuss (section 2) the general identification issue of the factor loadings and the idiosyncratic covariance matrix in the general setting (1.3). We put forward the aforementionned possible transfer in variance and characterize its effect on identification. A byproduct of this is that, concerning the volatility dynamics of common factors, only the volatility persistence parameters of stochastic volatility (SV) factors can be identified, while more specific factor structures like GARCH are not testable. Instrumental variables estimation and identification with zero factor risk premiums is presented in section 3 while the issue of linear beta models of risk premium is addressed in section 4. In both cases, a sequential procedure is proposed to identify the number K of factors as well as K mimicking portfolios in a preliminary inference step. Section 5 concludes and proposes some possible extensions. The main proofs are gathered in the appendix.

2 Model identification:

2.1 Identification of the factor loadings:

In this section, we address the identification issue of a factor model for conditional variance:

$$\Sigma_t = \Lambda \; D_t \; \Lambda' + \Omega \tag{2.1}$$

where Σ_t is the conditional covariance matrix of a vector y_{t+1} of n observed random variables:

$$\Sigma_t = Var\left(y_{t+1} \left| J_t \right.\right) \tag{2.2}$$

and J_t is a nondecreasing filtration which defines the relevant conditioning information. In particular, y_t is J_t adapted.

Of course, when writing the factor model (2.1), one has in mind a conditional regression of y_{t+1} on some factors f_{t+1}, given the information J_t:

$$\begin{cases} y_{t+1} = \mu(J_t) + \Lambda f_{t+1} + u_{t+1} \\ E\left(u_{t+1} \left| J_t \right.\right) = 0 \\ E\left(f_{t+1} \left| J_t \right.\right) = 0 \\ Cov\left(u_{t+1}, f_{t+1} \left| J_t \right.\right) = 0 \\ Var\left(f_{t+1} \left| J_t \right.\right) = D_t \end{cases} \tag{2.3}$$

With the additional assumption that the conditional idiosyncratic variance is constant:

$$Var\left(u_{t+1} \left| J_t \right.\right) = \Omega \tag{2.4}$$

the conditional regression model (2.3) implies (2.1)[1].

Two necessary identification conditions of the factor loadings Λ in (2.1) are well known:

[1]The decomposition (2.1) is actually equivalent to the regression model (2.3)/(2.4)with a possibly singular residual covariance matrix Ω (see Gourieroux, Monfort and Renault (1991))

- First, Λ must be a matrix of full column rank, say of rank K. If it was not the case, because for instance the K^{th} column would be a linear combination of the first $(K-1)$ columns, then the factorial representation (2.1) could be rewritten by using only as factor loadings the first $(K-1)$ columns of Λ.

- Second, the matrix D_t must be normalized, for instance by assuming that:

$$ED_t = Id_K \tag{2.5}$$

  identity matrix of size K.

A less well-known necessary identification condition, already pointed out [2] by Fiorentini and Sentana (2001) is the following:

Proposition 2.1 *If some diagonal coefficient σ_{kt}^2 of D_t is positively lower bounded:*

$$\sigma_{kt}^2 \geq \underline{\sigma}_k^2 > 0 \quad a.s. \tag{2.6}$$

then the decomposition

$$\Sigma_t = \Lambda \ D_t \ \Lambda' + \Omega \tag{2.7}$$

can also be written:

$$\Sigma_t = \tilde{\Lambda} \ \tilde{D}_t \ \tilde{\Lambda}' + \tilde{\Omega} \tag{2.8}$$

with

$$\tilde{\Lambda} \neq \Lambda, \tilde{\Omega} \neq \Omega, \tilde{\Omega} \gg \Omega. \tag{2.9}$$

The interpretation of this result is clear: if the conditional variance σ_{kt}^2 of a factor is positively lower bounded, it is always possible to transfer a constant part of it into the residual variance matrix [3]. Therefore, the two contributions cannot be separately identified.

The following proposition confirms this interpretation. We consider without loss of generality observable variables y_{t+1} and latent factors f_{t+1} of zero conditional expectation given J_t and we focus, for sake of notational simplicity, on the case of a single factor model.

[2]The identification results presented in this section are tightly related, although not equivalent or redundant, with some of Fiorentini and Sentana (2001). In order to be self-contained, we provide some autonomous proofs.

[3]Of course, such a transfer would be precluded if we assume that the residual covariance matrix Ω is diagonal. This is another way to get identifiability of SV factor models, see e.g. Fiorentini and Sentana (2001).

Proposition 2.2 *If there is one factor f_{t+1} such that:*

$$y_{t+1} = \lambda f_{t+1} + u_{t+1} \tag{2.10}$$

$$\begin{cases} Cov\left(u_{t+1}, f_{t+1}\left|J_t\right.\right) = 0 \\ Var\left(u_{t+1}\left|J_t\right.\right) = \Omega \qquad \textit{positive definite} \\ E\left(Var\left(f_{t+1}\left|J_t\right.\right)\right) = 1 \end{cases} \tag{2.11}$$

and

$$\sigma_t^2 = Var\left(f_{t+1}\left|J_t\right.\right) \geq \underline{\sigma}^2 > 0 \quad a.s. \tag{2.12}$$

then there is another factor $\tilde{f}_{t+1}$ such that

$$y_{t+1} = \tilde{\lambda}\tilde{f}_{t+1} + \tilde{u}_{t+1} \tag{2.13}$$

$$\begin{cases} Cov\left(\tilde{u}_{t+1}, \tilde{f}_{t+1}\left|J_t\right.\right) = 0 \\ Var\left(\tilde{u}_{t+1}\left|J_t\right.\right) = \tilde{\Omega} \gg \Omega \\ E\left(Var\left(\tilde{f}_{t+1}\left|J_t\right.\right)\right) = 1 \end{cases} \tag{2.14}$$

with

$$\tilde{\sigma}_t^2 = Var\left(\tilde{f}_{t+1}\left|J_t\right.\right) = \frac{1}{1-\underline{\sigma}^2}\left(\sigma_t^2 - \underline{\sigma}^2\right) \tag{2.15}$$

$$\tilde{\lambda} = (1-\underline{\sigma}^2)\lambda \textit{ and } \tilde{\Omega} = \Omega + \underline{\sigma}^2\lambda\lambda'$$

Another necessary condition for identification of decomposition (2.1) is of course the presence of conditional heteroskedasticity in each factor $f_{kt}, k = 1, \cdots, K$, that is the maintained assumption that the K diagonal coefficients σ_{kt}^2 of D_t are non degenerate random variables. This assumption is actually sufficient to identify the number K of factors:

Proposition 2.3 *If $\Lambda\ D_t\ \Lambda'\ + \Omega = L\Delta_t L' + W$ with:*

$$\Lambda\left(n \times K\right) \textit{ matrix of rank } K$$
$$L\left(n \times J\right) \textit{ matrix of rank } J$$
$$ED_t = Id_K,\ E\Delta_t = Id_J$$

and if D_t and Δ_t are diagonal matrices whose diagonal coefficients are non degenerate random variables, then: $K = J$ and the ranges of matrices Λ and L coincide.

Proposition 2.3 leads us to define a K SV factors model by the following conditions:

$$\begin{cases} \Sigma_t = \Lambda D_t \Lambda' + \Omega, \\ \Lambda\,(n \times K) \text{ matrix of rank } K, \\ \Omega\,(n \times n) \text{ positive definite matrix,} \\ D_t = Diag(d_t) \\ \text{where } d_t = \left(\sigma^2_{kt}\right)_{1\leq k\leq K} \text{ is a vector of } K \text{ positive random variables,} \\ \text{of expectation unity and such that } Var(d_t) \text{ is a nonsingular matrix} \end{cases} \tag{2.16}$$

Note that (2.16) strengthens the assumptions of Proposition 2.3 by considering not only that the K random variables $\sigma^2_{kt}, k = 1, \cdots, K$ are not degenerate but also that no linear combination of them is degenerate. This stronger assumption is actually needed to be sure to identify the K columns of the matrix Λ of factor loadings up to permutations and multiplication by arbitrary scalar numbers:

Proposition 2.4 *If Σ_t admits two factor decompositions:*

$$\Sigma_t = \Lambda\ D_t\ \Lambda'\ + \Omega = L\Delta_t L' + W$$

which are both conformable to (2.16), then:

$$L = \Lambda\Delta Q$$

for some diagonal matrix Δ and some permutation matrix Q.

Notice that postmultiplication of the matrix Λ of factor loadings by a diagonal matrix is akin to rescale each column of Λ according to the intuition put forward by proposition 2.2. The necessary identification condition of Proposition 2.1 is actually sufficient too:

Proposition 2.5 *If Σ_t admits two factor decompositions:*

$$\Sigma_t = \Lambda\ D_t\ \Lambda'\ + \Omega = L\Delta_t L' + W$$

which are both conformable to (2.16) with:

$$\begin{aligned} D_t &= diag\left(\sigma^2_{1t}, \cdots, \sigma^2_{Kt}\right) \\ \Delta_t &= diag\left(\tilde{\sigma}^2_{1t}, \cdots, \tilde{\sigma}^2_{Kt}\right) \end{aligned}$$

and for $k = 1, \cdots K$:

$$\Pr\left(\sigma_{kt}^2 \geq \sigma^2\right) < 1 \textit{ and } \Pr\left(\tilde{\sigma}_{kt}^2 \geq \sigma^2\right) < 1,$$

for any positive number σ*, then:*

$$\Omega = W \textit{ and } L = \Lambda\Delta Q$$

for some permutation matrix Q *and some diagonal matrix* Δ *the diagonal coefficients of which are all (+1) or (-1).*

Then, we do get identification of factor loadings up to sign and permutation of the factors insofar as we "minimize" the conditional variance of each factor f_{kt} by considering that

$$\sigma_{kt}^2 \geq \underline{\sigma}_k^2 \text{ a.s. } \Longrightarrow \underline{\sigma}_k^2 = 0. \tag{2.17}$$

According to Engle (2002) general discussion of non-negative processes, condition (2.17) means that there is a positive probability that the conditional factor variance is equal to zero, or arbitrarily close to zero.

Of course, as already mentioned, one can avoid identification condition (2.17) by either imposing more structure on the residual covariance matrix Ω or by maintaining additional assumptions about higher order moments. While the first route will not be considered in this paper, the second one will be followed in the third part of section 3. We first discuss in subsection 2.2 below the implications of condition (2.17) in terms of model specification for the conditional variance processes of the K factors.

2.2 GARCH or SV Factors ?

For sake of notational simplicity, we only consider in this section the case of a one factor model. But everything can easily be extended to the K factors case.

The GARCH factor model, as first introduced by Diebold and Nerlove (1989), specifies the latent factor f_t as a GARCH (1,1):

$$\begin{cases} \sigma_t^2 = (1-\gamma) + \alpha f_t^2 + (\gamma - \alpha)\,\sigma_{t-1}^2 \\ 0 < \alpha \leq \gamma < 1 \end{cases} \tag{2.18}$$

Note that the intercept $(1-\gamma)$ has been chosen to fulfill the restriction $E\sigma_t^2 = 1$. However, an obvious implication of the GARCH specification is that the identification condition (2.17) is not fulfilled. We have identically:

$$\sigma_t^2 \geq \frac{1-\gamma}{1-\gamma+\alpha} = \underline{\sigma}^2 > 0 \tag{2.19}$$

Therefore, a number of latent volatility factors $\tilde{f}_t$ are observationally equivalent to the GARCH factor f_t. For instance, a volatility factor $\tilde{f}_t$ associated to a conditional variance process $\tilde{\sigma}_t^2$ defined, according to proposition 2.2, by:

$$\tilde{\sigma}_t^2 = \frac{\sigma_t^2 - \underline{\sigma}^2}{1 - \underline{\sigma}^2} \tag{2.20}$$

cannot be in general GARCH (1,1) since the lower bound of $\tilde{\sigma}_t^2$, when equal to zero, cannot be conformable to a condition like (2.18). In other words, GARCH (1,1) structures of volatility factors cannot be fully identified from the only observation of returns volatility dynamics. The only dynamic features of factors that can be identified in this context are the ones which are invariant with respect to transformations like (2.20). This leads us to focus on the autoregressive dynamics of the conditional variance process σ_t^2 obviously implied by the GARCH structure:[4]

$$E\left[\sigma_t^2 \left| \sigma_\tau, \tau < t \right.\right] = 1 - \gamma + \gamma \sigma_{t-1}^2 \tag{2.21}$$

Following Andersen (1994) and Meddahi and Renault (2004), we define more generally:

Definition 2.6 *A scalar process $\{f_t, t \in \mathbf{Z}\}$ is SR-SARV (1) (Square Root Stochastic Autoregressive Volatility of order 1) with respect to a filtration J_t if:*

$$\begin{aligned}
&E\left[f_{t+1} \left| J_t \right.\right] = 0 \quad , \quad E\left[f_{t+1}^2 \left| J_t \right.\right] = \sigma_t^2 \\
&E\left[\sigma_{t+1}^2 \left| J_t \right.\right] = 1 - \gamma + \gamma \sigma_t^2 \\
&0 < \gamma < 1.
\end{aligned}$$

Section 3 will confirm that the persistence parameter γ is identifiable from the return volatility dynamics. Of course, this does not preclude to maintain a factor GARCH (1,1) assumption and to try to separately identify the GARCH parameters α and $(\gamma - \alpha)$ within this framework. The following result shows that this is actually possible, at least when maintaining some additional assumptions about higher order moments:

[4] While we focus in this paper on GARCH (1,1) factors and associated AR(1) volatility dynamics, most of the results could easily be extended to higher orders.

Proposition 2.7 *If y_{t+1} is described by two different one-factor GARCH (1,1) models:*

$$\begin{cases} y_{t+1} = \lambda f_{t+1} + u_{t+1} = \tilde{\lambda} \tilde{f}_{t+1} + \tilde{u}_{t+1} \\ Cov\left(f_{t+1}, u_{t+1} \left| J_t \right.\right) = Cov\left(\tilde{f}_{t+1}, \tilde{u}_{t+1} \left| J_t \right.\right) = 0 \\ Var\left(f_{t+1} \left| J_t \right.\right) = \sigma_t^2,\ Var\left(\tilde{f}_{t+1} \left| J_t \right.\right) = \tilde{\sigma}_t^2 \\ Var\left(u_{t+1} \left| J_t \right.\right) = \Omega,\ Var\left(\tilde{u}_{t+1} \left| J_t \right.\right) = \tilde{\Omega} \\ E\sigma_t^2 = E\tilde{\sigma}_t^2 = 1 \end{cases}$$

where f_{t+1} and $\tilde{f}_{t+1}$ are both GARCH (1,1) processes with constant conditional kurtosis, then: $\sigma_t^2 = \tilde{\sigma}_t^2$.

If in addition, the two conditional kurtosis coefficients coincide, then: $f_{t+1}^2 = \tilde{f}_{t+1}^2$.

Note that the assumption of constant conditional kurtosis is implied by the strong GARCH property as defined by Drost and Nijman (1993), that is by the i.i.d. property of standardized innovations f_{t+1}/σ_t. When the conditional probability distribution of f_{t+1}/σ_t is given, for instance when it is supposed to be gaussian, the factor f_{t+1} is identified up to a sign. Since such assumptions are generally maintained for any kind of parametric inference about GARCH or SV type models, the identification issue about latent GARCH factors has been overlooked in the literature. An alternative identifying assumption within the latent GARCH(1,1) framework is to maintain the ARCH(1) specification, that is $\alpha = \gamma$ in (2.18) (see eg. Dellaportas, Giakoumatos and Politis(1999) and Diebold and Nerlove (1989)). For given γ, this value of α is actually the one which minimizes the lower bound of variance $\underline{\sigma}^2$ in (2.19).

Since the focus of interest of this paper is statistical identification and inference about joint volatility dynamics of a vector of returns with as little as possible additional assumptions about higher order moments, we focus on general latent SV factors rather than on latent GARCH factors. Moreover, the convenient identification result of proposition 2.7 is not specific to GARCH factors. We will actually be able to show in section 3 that the maintained assumption of fixed conditional kurtosis is sufficient to hedge against the identification problem of section 2 in a general framework of SR-SARV(1) factors with quadratic variance of the conditional variance.

3 Model with constant risk premiums: identification and IV estimation

We consider in this section a vector y_{t+1} of n asset returns with constant conditional expectation $\mu = E\left(y_{t+1}|J_t\right)$. Then, the factor structure must be identified only from information about the conditional covariance matrix and possibly higher order conditional moments. We first discuss a statistical procedure of determination of the number K of factors. A byproduct of this procedure is the identification of a subset of K asset returns the conditional heteroskedasticity of which does involve the K factors.

Then, we are able, from the semiparametric model of the conditional variance matrix with K SR-SARV(1) factors, to perform efficient IV estimation of the K coefficients of volatility persistence, the range of the matrix Λ of factor loadings and the residual covariance matrix Ω up to $\frac{K(K+1)}{2}$ degrees of freedom.

These degrees of freedom correspond to the possible transfer of constant variance from some linear combinations of factors to the idiosyncratic terms. To get rid of them, we propose to add higher order conditional moment restrictions that are tightly related to a conditional multivariate kurtosis model for the vector of returns. Then, we are able to fully identify the matrices Λ and Ω and to perform IV estimation of their coefficients.

3.1 Determination of the number K of factors

According to the general definitions of section 2, we consider that the conditional heteroskedasticity of the vector y_{t+1} of asset returns is characterized by a K SV factor model if there exist K positive stochastic processes σ_{kt}^2, $k = 1, \cdots, K$, such that:

$$\left\{\begin{array}{l}
E\left(y_{t+1}\left|J_t\right.\right) = \mu \\
E\left(\sigma_{kt+1}^2\left|J_t\right.\right) = (1-\gamma_k) + \gamma_k\ \sigma_{kt}^2, \quad 0 < \gamma_k < 1 \\
\text{and} \\
Var\left(y_{t+1}\left|J_t\right.\right) = \Lambda D_t \Lambda' + \Omega \\
\text{with} \\
\Lambda\left(n \times K\right) \text{ matrix of rank } K, \\
\Omega\left(n \times n\right) \text{ positive definite matrix}, \\
\text{and } D_t = Diag\left(\sigma_{1t}^2, \cdots, \sigma_{Kt}^2\right)
\end{array}\right. \tag{3.1}$$

Statistical inference about model (3.1) must be based on some information set I_t available at time t to the econometrician. Typically, I_t is a sub σ-algebra of J_t containing at least the past and current observations of returns:

$$\sigma\left[y_\tau, \tau \leq t\right] \subset I_t \subset J_t.$$

In this section and in all the rest of the paper as well, the symbols E_t, V_t, Cov_t respectively denote conditional expectation, variance and covariance "given available information at time t". These notations do no longer make explicit the distinction between the theoretical information set J_t and the econometrician information set $I_t \subset J_t$.

Actually, this distinction may be omitted insofar as the moment conditions considered for inference are about conditional moments of future values of the process y_t which is I_t - adapted.

From proposition 2.3, the number K of factors in model (3.1) is well identified. Its statistical determination will be performed through a sequential testing procedure. The sequences of hypotheses are defined for $k = 0, 1, \cdots, n-1$ by:

H_{0k}: The number of factors is k and

H_k: The number of factors is larger or equal to k.

We want to test H_{0k} against $H_{k+1} = H_k - H_{0k}$ and we consider the following sequences of tests:

(i) Test of H_{00}, that is test of joint conditional homoskedasticity of the vector y_{t+1} of asset returns. Of course, if H_{00} is accepted, there is no need to look for any factor.

(ii) Otherwise, H_{01} is tested against H_2. If H_{01} is accepted, the procedure stops and we accept the hypothesis: y_{t+1} is governed by a one-factor model. If H_{01} is rejected, we test H_{02} against H_3, and so on.

The first step is standard. It should be based on the $\frac{n(n+1)}{2}$ conditional moment restrictions

$$vech\left[E_t\left(y_{t+1}y'_{t+1} - C\right)\right] = 0$$

where C is an unknown positive symmetric matrix.

A simpler and more natural procedure would be to consider only the diagonal restrictions:

$$E_t\left(y^2_{it+1} - c_{ii}\right) = 0, i = 1, \cdots, n.$$

Note however that, by contrast with more common univariate tests of homoskedasticity, it is important to check the orthogonality of $\left(y^2_{it+1} - \omega_{ii}\right)$ not only with lagged squared values $y^2_{i\tau}, \tau \leq t$

of return i but also with lagged squared values $y_{j\tau}^2, \tau \leq t, j \neq i$, of other returns and possibly lagged cross products $y_{i\tau}y_{j\tau}, \tau \leq t$.

The test of H_{0k} against H_{k+1} for $k \geq 1$ must be performed by taking into account that H_{0k-1} has been rejected in the previous step. In other words, we know that the number of factors is larger than $(k-1)$ (hypothesis H_k) and we wonder whether it is exactly k (hypothesis H_{0k}). In this case, $K = k$ and it is possible to select k rows of the matrix Λ of factor loadings such that the corresponding submatrix $\bar{\Lambda}$ of Λ is a square non singular matrix of size k. In terms of decomposition of the vector of returns y_{t+1}, this property can be characterized in the following way :

Proposition 3.1 *Under the hypothesis H_k that the number K of factors is greater or equal to k, if $\bar{y}_{t+1}$ denotes k selected components of the vector y_{t+1} and $\bar{\bar{y}}_{t+1}$ denotes the $(n-k)$ remaining components, the following conditions are equivalent :*

i) the matrix $\bar{\Lambda}$ of factor loadings associated to $\bar{y}_{t+1}$ is a squared non singular matrix of size $k = K$

ii) there exists a matrix B such that $\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}$ is conditionally homoskedastic.

Moreover, when these conditions are fulfilled, B is necessarily the matrix $\bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$, where $\bar{\bar{\Lambda}}$ denotes the matrix of factor loadings associated to the subvector $\bar{\bar{y}}_{t+1}$ of returns.

This suggests to test H_{0k} against H_{k+1} for $k \geq 1$ through the overidentification test of the conditional moment restrictions:

$$vech\left[E_t\left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)' - C\right]\right] = 0 \tag{3.2}$$

for unknown matrices B and C.

However, (3.2) does not encompass all the information provided by the factor model. To see this, note that:

$$\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)' = \left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)\bar{\bar{y}}'_{t+1} - \left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)\bar{y}'_{t+1}B' \tag{3.3}$$

and that under the null of $k = K$ factors, both terms of this difference have a constant conditional expectation. In other words, efficient inference about H_{0k} must be based on the following extended set of conditional moment restrictions:

$$vec\left[E_t\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)y'_{t+1} - D\right] = 0 \tag{3.4}$$

for unknown matrices B and D.

Note however that from (3.3), the lower part $\bar{\bar{D}}$ of the matrix D can be seen as:

$$\bar{\bar{D}} = C + \bar{D}B'$$

so that, for a given B, the matrix D specification only involves $K(n-K) + \frac{(n-K)(n-K+1)}{2}$ free parameters corresponding to the specification of, first, the upper part $\bar{D}$ of D and, second, the symmetric matrix C of size $(n-K)$.

To fully realize the important difference between (3.2) and (3.4) several remarks are in order. First, by contrast with (3.2), (3.4) is linear with respect to the unknown parameters, which is of course more convenient for computation and statistical inference. Second, nonlinearity of (3.2) is even more detrimental than usual here since, following the terminology of Arellano, Hansen and Sentana (1999), (3.2) is a case where identification is guaranted (by proposition 3.1) even though there is a first-order lack of identification. To see this, note that, by (3.4), the Jacobian matrix of (3.2) with respect to B is constant, and thus, the rank condition for joint identification of B and C fails. As already shown by Sargan (1983), this may produce non-standard asymptotic probability distributions for some parameter estimates. For testing for common features, Engle and Kozicki (1993) compute an overidentification test statistic after concentrating with respect to B, that is replacing C by $C(B) = \frac{1}{T}\sum_{t=1}^{T}\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)'$.

The focus of interest of this section is a test of the factor structure (3.4), which is a submodel of the common features model. We are actually able to show that the rank condition for GMM inference about B and D is fulfilled in the case of (3.4) for a convenient choice of instrumental variables:

Proposition 3.2 *When conditions of proposition 3.1 are fulfilled, if z_t is an I_t-adapted real valued stochastic process such that:*

$$Cov(z_t, \sigma_{kt}^2) \neq 0 \text{ for } k = 1, \cdots, K$$

and if $\Phi(B, D)$ denotes the vector:

$$\Phi(B, D) = E\left[\begin{bmatrix} 1 \\ z_t \end{bmatrix} \otimes vec\left[E_t\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)y'_{t+1} - D\right]\right]$$

then the jacobian matrix :

$$\frac{\partial \Phi(B, D)}{\partial\left[(vec\ B)', (vec\ D)'\right]}$$

is of full column rank.

There is however an additional issue for using overidentification tests of conditional moment restrictions (3.4) in order to define a sequential testing procedure of hypotheses $H_{0k}, k \geq 1$. Since the test of H_{00} is a standard test of multivariate conditional homoskedasticity, we only have to define the test of $H_{0k}, k \geq 1$ when H_{0k-1} has been rejected in the previous step. Let us first stress that, since we consider a sequence of hypotheses in an increasing order, there is no hope to control the overall size of the test by an argument of independence of two consecutive test statistics under the null. This issue is common when determining the order of a time series model, like order of an ARMA or of a GARCH process. Moreover, as shown by proposition 3.1, the identification condition needed for GMM inference about H_{0k} takes crucially advantage of the fact that H_{0k-1} has been rejected in the previous step.

An additional difficulty here is that H_{0k} is actually defined as a union of hypotheses $H_{0k}(\bar{y})$, each of them being characterized by standard conditional moment restrictions. Let us write :

$$H_{0k} = \bigcup_{\bar{y} \in S_k} H_{0k}(\bar{y})$$

where S_k denotes the set of all subvector $\bar{y}_{t+1}$ of y_{t+1} of dimension k and $H_{0k}(\bar{y})$ is defined by the conditional moment restrictions (3.4):

$$vec\ E_t \left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1} \right) y'_{t+1} - D \right] = 0$$

For a given choice of a vector z_t of H I_t -adapted instruments, such conditional moment restrictions are usually tested through their unconditional consequence:

$$\bar{H}_{0k}(\bar{y}) : E\left[z_t \otimes Vec\left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1} \right) y'_{t+1} - D \right] \right] = 0$$

Let us denote by $S_T(\bar{y})$ the Hansen $J-$Test statistic to test $\bar{H}_{0k}(\bar{y})$. Since we consider that H_{0k-1} is wrong, we maintain conditions of propositions 3.1 and 3.2 for $K = k$ and then, we know from Hansen (1982) that the test of $H_{0k}(\bar{y})$ defined by:

$$\text{Reject } H_{0k}(\bar{y}) \Leftrightarrow S_T(\bar{y}) > \mathcal{X}^2_{1-\alpha} \left[Hn(n-k) - 2k(n-k) - \frac{(n-k)(n-k+1)}{2} \right]$$

is asymptotically of level α if $\mathcal{X}^2_{1-\alpha}(n)$ denotes the $(1-\alpha)$ quantile of a $\mathcal{X}^2(n)$.

For a given choice of $\bar{y}$ in S_k, this suggests to adapt the following rule to test for H_{0k}:

$$\text{Reject } H_{0k} \Leftrightarrow S_T(\bar{y}) > \mathcal{X}^2_{1-\alpha} \left[Hn(n-k) - 2k(n-k) - \frac{(n-k)(n-k+1)}{2} \right] \tag{3.5}$$

Unfortunately, such a testing procedure for H_{0k} will intuitively suffer from severe size distortions (with respect to the nominal size α) since H_{0k} will be rejected on the basis that we only think that a particular hypothesis $H_{0k}(\bar{y})$ is wrong. This is actually far to imply that H_{0k} itself is wrong. However, nobody would like to test H_{0k} by rejecting it only when inequality (3.5) is fulfilled for any choice of $\bar{y}$ in S_k. This would produce a test for H_{0k} much too conservative since the probability of the intersection of two events like (3.5), for two different choices of $\bar{y}$, may be as small as the square of α.

Therefore, our proposal will be to test H_{0k} through (3.5) [5] , but for a convenient choice of $\bar{y}$. The trick amounts to pre-select, among the possible $\bar{y} \in S_k$, the one which is the "most likely" to capture conditional heteroskedasticity of the n returns, that is to fulfill $H_{0k}(\bar{y})$. Then, if (3.5) is nevertheless fulfilled by such an $\bar{y}$, it makes sense to consider that other choices of $\bar{y}$ would a fortiori violate $H_{0k}(\bar{y})$ and then that H_{0k} should be rejected. In other words, rejection on H_{0k} through this strategy should not lead to an effective size much larger than α: the probability of (3.5) under H_{0k} is not much larger than α, that is the probability of (3.5) under $H_{0k}(\bar{y})$, since we do think that H_{0k} cannot be fulfilled without $H_{0k}(\bar{y})$ itself being fulfilled.

The preselection of a convenient $\bar{y}$ can be based on the fact that, when testing H_{0k},we had previously rejected H_{0k-1}, on the basis that some preselected $\tilde{\bar{y}} \in S_{k-1}$ did not fulfill $H_{0k-1}(\tilde{\bar{y}})$ defined through a decomposition $\left(\tilde{\bar{y}}', \tilde{\bar{\bar{y}}}'\right)'$ of returns and a corresponding matrix $\tilde{B}$. It will then be natural, after selection of $\tilde{\bar{y}}$, to build $\bar{y}$ by adding to $\tilde{\bar{y}}$ a well-chosen return $\tilde{\bar{\bar{y}}}_{i_0}$, component of the vector $\tilde{\bar{\bar{y}}}$ of $(n-k+1)$ remaining returns, and to choose $\tilde{\bar{\bar{y}}}_{i_0}$, which is "the most responsible" for the rejection of H_{0k-1}.

In other words, after concluding that H_{00} is rejected, the first one-dimensional $\bar{y}$ considered will be the return y_{i_0} the conditional heteroskedasticity of which is maximum, in terms of conditional variance coefficient of variation. Thus, we choose i_0 as a solution of:

$$\max_{1 \le i \le n} \frac{Var\left[Var_t\left(y_{it+1}\right)\right]}{Var\left(y_{it+1}\right)} = \frac{\lambda_i^2 Var \sigma_t^2}{\lambda_i^2 + \omega_{ii}}$$

Similarly when H_{0k-1} has been rejected, one will choose to add to the previous set $\tilde{\bar{y}}$ of $(k-1)$ returns, a k^{th} return the index i_0 of which is choosen as a solution of:

$$\max_i \frac{Var\left[Var_t\left(\tilde{\bar{\bar{y}}}_{it+1} - \tilde{B}_i' \tilde{\bar{y}}_{t+1}\right)\right]}{Var\left(\tilde{\bar{\bar{y}}}_{it+1} - \tilde{B}_i' \tilde{\bar{y}}_{t+1}\right)}$$

[5] In finite samples the number of degrees of freedom in (3.5) may be too large for reliable size and power properties. It may then be relevant to focus only on a subset of moment conditions (3.4).

$y_{i_0 t+1}$ is the return the conditional heteroskedacticity of which is the least captured by the mimicking portfolios $\tilde{B}\tilde{\bar{y}}_{t+1}$.

As already announced, this sequential procedure is not fully closed in terms of controlling statistical risk. But, as usual with model choice strategies, it must be seen as only a preliminary exploratory analysis to perform before the comprehensive statistical strategy of subsections 3.2 and 3.3.

3.2 Estimation of spanning factor loadings

The focus of interest of this subsection is efficient IV estimation of the K factors SV-model (3.1). In other words, the unknown parameters of interest are:

- First, the coefficients of the matrix Λ of factor loadings and the residual covariance matrix Ω.

- Second, the K persistence parameters $\gamma_k, k = 1 \cdots K$ of the volatility factors.

- Third, the coefficients of the conditional mean vector $\mu = E_t y_{t+1}$.

Note that since we do not give in this subsection any statistical content to the identification assumptions of propositions 2.4 and 2.5, only the range of Λ is identified (see proposition 2.3). Equivalently, since the unconditional covariance matrix $\Sigma = \Lambda\Lambda' + \Omega$ is of course identifiable, some lack of identification in the residual covariance matrix Ω is implied by the lack of identification of Λ, as exhibited in proposition 2.2. More precisely:

(i) Λ is identified up to a right multiplication by an arbitrary non singular matrix M of size K: Λ and ΛM are observationally equivalent.

(ii) Ω is identified up to an arbitrary symmetric positive definite matrix M of size K: $\Omega = \Sigma - \Lambda\Lambda'$ and $\Omega = \Sigma - \Lambda A \Lambda'$ (see $A = MM'$) are observationally equivalent.

Therefore, we are able to identify in particular the position [6] of K rows of Λ which defines a nonsingular matrix $\bar{\Lambda}$.

The determination of such a position is actually a byproduct of the testing procedure defined in subsection 3.1 through the set of conditional moment restrictions:

$$E_t \left[\left(\bar{\bar{y}}_{t-1} - B\bar{y}_{t+1} \right) y'_{t+1} - D \right] = 0$$

By assuming without loss of generality that $\bar{y}_{t+1}$ corresponds to the first K rows of y_{t+1}, a

[6]Such a position is invariant by right multiplication of Λ by an arbitrary non singular matrix M of size K.

natural identifiable parameterization of Λ is then:

$$\Lambda = \begin{bmatrix} Id_K \\ B \end{bmatrix} \tag{3.6}$$

In other words, we choose $M = \bar{\Lambda}^{-1}$.

This choice also identifies $\Omega_{11} = \Sigma_{11} - Id_K$, and then fixes the required $\frac{K(K+1)}{2}$ coefficients to identify Ω. Ω_{12} and Ω_{22} are then respectively identified from the unconditional moments:

$$Cov\left[\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}, \bar{y}_{t+1}\right] = \Omega_{21} - B\Omega_{11}$$

and

$$Cov\left[\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}, \bar{\bar{y}}_{t+1}\right] = \Omega_{22} - B\Omega_{12}.$$

We are so able to state the set of conditional moment restrictions which allows efficient estimation of $\Lambda, \Omega, \mu, \gamma_k, k = 1, \cdots K$ in the model (3.1) with the normalization rule (3.6). If we denote $\Omega_{1.} = [\Omega_{11} \quad \Omega_{12}]$ and $\Omega_{2.} = [\Omega_{21} \quad \Omega_{22}]$ this set of conditional moment restrictions is the following one:

Proposition 3.3 *In the K SV factor model with factor loadings: $\Lambda = [Id_K \quad B']'$, and idiosyncratic covariance matrix $\Omega = [\Omega_{1.}' \quad \Omega_{2.}']'$ the parameters μ, Λ, Ω and $\gamma_k, k = 1 \cdots K$ are characterized by:*

$$E_t\left[y_{t+1} - \mu\right] = 0 \tag{3.7}$$

$$vec\ E_t\left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right) y_{t+1}'\right] \quad = \quad vec\left[\left(\bar{\bar{\mu}} - B\bar{\mu}\right)\mu' + \Omega_{2.} - B\Omega_{1.}\right] \tag{3.8}$$

$$vechE_{t-K}\left[\prod_{k=1}^{K}\left(1 - \gamma_k L\right)\left(y_{t+1}y_{t+1}' - \Lambda\Lambda' - \Omega - \mu\mu'\right)\right] = 0 \tag{3.9}$$

To understand the role of proposition 3.3 within the general issue of inference on SV-factors models, two remarks are in order.

First, efficient IV estimation through conditional moment restrictions (3.7), (3.8) and (3.9) is supposed to be performed in a second stage, after the testing strategy of subsection 3.1 has been applied. In particular, the number K of factors and the selection $\bar{y}_{t+1}$ of K mimicking portfolios (see $\bar{\Lambda} = Id_K$) are considered as already known. Note that a first stage estimation of the matrix B of conditional beta coefficients of other returns $\bar{\bar{y}}_{t+1}$with respect to the K factors should also be a byproduct of the first stage testing strategy. However, the joint use of conditional moment restrictions (3.7), (3.8) and (3.9) should provide more efficient estimators of B and the other parameters of interest as well. Second, multi-period conditional moment restrictions as (3.9) have already been put forward by Meddahi and Renault (2004) for estimation of SR-SARV models. The issue of optimal instruments for such moment conditions is addressed by Hansen, Heaton and Ogaki (1988) and Hansen and Singleton (1996).To see what is at play in these moment restrictions, it is worth noting that, for all pair (y_{it+1}, y_{jt+1}) of returns, we can write with obvious notations:

$$y_{it+1}y_{jt+1} = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk} f_{kt+1}^2 + \omega_{ij} + \mu_i\mu_j + v_{t+1}^{(i,j)} \text{ with } E_t\left[v_{t+1}^{(i,j)}\right] = 0.$$

Therefore:

$$\begin{aligned}(1-\gamma_1 L)\left(y_{it+1}y_{jt+1} - \omega_{ij} - \mu_i\mu_j\right) &= \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk}\left[\left(f_{kt+1}^2 - 1\right) - \gamma_1\left(f_{kt}^2 - 1\right)\right] \\ &+ (1-\gamma_1)\sum_{k=1}^{K} \lambda_{ik}\lambda_{jk} + v_{t+1}^{(i,j)} - \gamma_1 v_t^{(i,j)}.\end{aligned}$$

But, by definition:

$$E_{t-1}\left[\left(f_{1t+1}^2 - 1\right) - \gamma_1\left(f_{1t}^2 - 1\right)\right] = E_{t-1}\left[\sigma_{1t}^2 - \gamma_1\sigma_{1t-1}^2 - (1-\gamma_1)\right] = 0. \tag{3.10}$$

In other words, the first factor volatility dynamics are annihilated by filtering by the lag polynomial $1 - \gamma_1 L$, up to a moving average of order one effect (we consider only expectation at time $(t-1)$ of $f_{1t+1}^2 - \gamma_1 f_{1t}^2$). By iterating this argument with consecutive filtering by $(1-\gamma_k L), \gamma_k = 1 \cdots K$, we clearly get conditional moment restrictions (3.9). Note moreover that for efficient estimation, the cross restrictions about $y_{it+1}y_{jt+1}, i \neq j$, are intuitively as informative as the diagonal restrictions about squared returns y_{it+1}^2.

3.3 Identification through higher order moments

To introduce the main idea, let us first consider a one-factor model:

$$y_{t+1} = \mu + \lambda f_{t+1} + u_{t+1}.$$

For identification and inference, we only used so far the conditional moment restrictions produced by the first two conditional moments:

$$\begin{cases} E_t(y_{t+1}) = \mu \\ E_t(y_{t+1}y_{t+1}') = \lambda\lambda'\sigma_t^2 + \Omega + \mu\mu'. \end{cases}$$

Although the process σ_t^2 is not observed, the maintained assumption of AR(1) dynamics for this process has allowed us to identify the parameter of interest up to a scale factor in λ (and a corresponding degree of freedom in Ω).

Instead of trying to control this degree of freedom by an additional assumption about the support of σ_t^2 as in section 2, we will prefer to generalize to SV-factor processes the approach of proposition 2.7: when the standardized factor f_{t+1}/σ_t is assumed to be i.i.d., or at least with constant conditional kurtosis, this may preclude the transfer of variance between common factors and residual variance and thus may allow identification. Following the general approach of this section, identification and estimation issues will be addressed simultaneously through conditional moment restrictions.

For this purpose, a well-suited assumption is akin to impose a VARMA (1,1) structure for the pair (f_t^2, f_t^4). While volatility persistence was estimated in previous subsection thanks to multilag conditional moment restrictions (3.9) corresponding to an ARMA (1,1) structure for f_t^2 (see (3.10)), we add now the second VARMA (1,1) equation for f_t^4:

Assumption 3.4 *There exists a, b, c such that*

$$E_{t-1}\left[f_{t+1}^4 - a - bf_t^2 - cf_t^4\right] = 0 \tag{3.11}$$

Note that, jointly with (3.10), (3.11) precisely means that

$$\begin{bmatrix} f_{t+1}^2 \\ f_{t+1}^4 \end{bmatrix} - \begin{bmatrix} \gamma & 0 \\ b & c \end{bmatrix} \begin{bmatrix} f_t^2 \\ f_t^4 \end{bmatrix}$$

is uncorrelated with any function of I_{t-1}, which implies a fortiori vectorial $MA(1)$ dynamics.

In order to see to what extent this subsection 3.3 generalizes the result of proposition 2.7, it is worth revisiting assumption 3.4 in the case of a factor with constant conditional kurtosis:

$$E_t f_{t+1}^4 = \kappa \sigma_t^4 \tag{3.12}$$

Then:

$$\begin{aligned} E_{t-1} f_{t+1}^4 &= \kappa E_{t-1} \sigma_t^4 \\ &= \kappa \left[V_{t-1}\sigma_t^2 + \left(E_{t-1}\sigma_t^2\right)^2\right] \\ &= \kappa \left[V_{t-1}\sigma_t^2 + (1-\gamma)^2 + 2\gamma(1-\gamma)\sigma_{t-1}^2 + \gamma^2\sigma_{t-1}^4\right] \end{aligned}$$

which shows that:

$$
\begin{aligned}
E_{t-1}f_{t+1}^4 - bE_{t-1}f_t^2 - cE_{t-1}f_t^4 &= \left(\gamma^2 - c\right)\kappa\sigma_{t-1}^4 + \left[2\gamma\left(1-\gamma\right)\kappa - b\right]\sigma_{t-1}^2 \\
&\quad +\kappa\left(1-\gamma\right)^2 + \kappa V_{t-1}\sigma_t^2.
\end{aligned}
$$

By identification with (3.11) we show that:

Proposition 3.5 *In the case of constant conditional factor kurtosis:*

$$E_t f_{t+1}^4 = \kappa\sigma_t^4$$

assumption 3.4 is equivalent to the property:

$V_{t-1}(\sigma_t^2)$ *is a quadratic function of* σ_{t-1}^2.

In this case,

$$V_{t-1}\left(\sigma_t^2\right) = \frac{a}{\kappa} - \left(1-\gamma\right)^2 + \left[\frac{b}{\kappa} - 2\gamma\left(1-\gamma\right)\right]\sigma_{t-1}^2 + \left(c-\gamma^2\right)\sigma_{t-1}^4$$

Proposition 3.5 shows that, when a, b, c are free parameters, $V_{t-1}(\sigma_t^2)$ is a general quadratic function of σ_{t-1}^2. Note however that the normalization condition $Ef_t^2 = 1$ implies, by Cauchy-Schwartz inequality: $Ef_t^4 > 1$, that is, with non negative a and b:

$$c < 1 \text{ and } a + b + c > 1.$$

Proposition 2.7 was dealing with the case of a GARCH (1,1) factor with constant conditional kurtosis:

$$\sigma_t^2 = \omega + \alpha f_t^2 + \left(\gamma - \alpha\right)\sigma_{t-1}^2.$$

Then:

$$V_{t-1}\sigma_t^2 = \alpha^2 V_{t-1}f_t^2 = \alpha^2(\kappa - 1)\sigma_{t-1}^4$$

The general quadratic specification of $V_{t-1}\sigma_t^2$, in the context of SR-SARV(1) processes, is much more general than the GARCH (1,1) case since it nests in particular:

- First, affine processes of conditional variance as considered by Heston (1993), Duffie, Pan and Singleton (2000) and Meddahi and Renault (2004). Then $V_{t-1}\sigma_t^2$ is affine with respect to σ_{t-1}^2.

- Second, Ornstein-Uhlenbeck like Levy-processes of conditional variance as introduced by Barndorff-Nielsen and Shephard (2001). Then $V_{t-1}\sigma_t^2$ is time invariant.

Assumption 3.4 will ensure full identification of the factor loadings and of the residual covariance matrix of SV factor models (with general SR-SARV(1) factors) by allowing to consider joint dynamics of conditional variance and conditional kurtosis of asset returns. For sake of notational simplicity, we will assume in all the rest of this subsection that $E_t y_{t+1} = \mu = 0$. At the cost of tedious notation, a free parameter μ would not be difficult to introduce in all the formulas. We also maintain, in the whole following, assumption 3.6 below:

Assumption 3.6 *$f_{t+1}, f_{t+1}^2, f_{t+1}^3$ are conditionally uncorrelated with any polynomial function of the u_{it+1} 's of degree smaller than four.*

Let us then consider the conditional kurtosis of a particular asset return i:

$$\begin{aligned} E_t y_{it+1}^4 &= E_t\left[(\lambda_i f_{t+1} + u_{it+1})^4\right] \\ &= \lambda_i^4 E_t f_{t+1}^4 + 6\lambda_i^2 \omega_{ii} E_t f_{t+1}^2 + E_t u_{it+1}^4 \end{aligned} \tag{3.13}$$

It is then clear that, if we get rid of the residual dynamics, the VARMA (1,1) structure of $\left(f_{t+1}^2, f_{t+1}^4\right)$ will allow us to write conditional moment restrictions and thus to be able to identify separately λ_i^2 and ω_{ii} (from λ_i^4 and $\lambda_i^2\omega_{ii}$).

Since we consider in the whole paper that error terms do not feature any conditional heteroskedasticity, it is fairly natural to discard any residual dynamics, even at higher orders:

$$E_t\left[u_{it+1}^4\right] = \kappa_{ii}\omega_{ii}^2 \tag{3.14}$$

where κ_{ii} denotes the conditional kurtosis coefficient of the error term u_{it+1}.

Then, by writing (3.13) at two consecutive dates and using the law of iterated expectations, we get:

$$E_{t-1} y_{it+1}^4 = \lambda_i^4 E_{t-1} f_{t+1}^4 + 6\lambda_i^2 \omega_{ii} E_{t-1} f_{t+1}^2 + \kappa_{ii}\omega_{ii}^2$$

and

$$E_{t-1} y_{it}^4 = \lambda_i^4 E_{t-1} f_t^4 + 6\lambda_i^2 \omega_{ii} E_{t-1} f_t^2 + \kappa_{ii}\omega_{ii}^2$$

By substracting c times the second equation to the first one and using assumption 3.4, we get:

$$\begin{aligned} E_{t-1}\left(y_{it+1}^4 - c y_{it}^4\right) &= \lambda_i^4 E_{t-1}\left(a + b f_t^2\right) + 6\lambda_i^2\omega_{ii} E_{t-1}\left(f_{t+1}^2 - c f_t^2\right) + \kappa_{ii}\omega_{ii}^2(1-c) \\ &= a\lambda_i^4 + \kappa_{ii}\omega_{ii}^2(1-c) + \left(b\lambda_i^2 - 6c\omega_{ii}\right)\left(E_{t-1}y_{it}^2 - \omega_{ii}\right) \\ &\quad +6\omega_{ii}\left(E_{t-1}y_{it+1}^2 - \omega_{ii}\right) \end{aligned}$$

since $E_{t-1}\left(\lambda_i^2 f_t^2\right) = E_{t-1} y_{it}^2 - \omega_{ii}$.

Therefore, we get the following conditional moment restrictions in terms of observed returns of asset i:

$$E_{t-1}\left[(1-cL)\left(y_{it+1}^4 - 6\omega_{ii} y_{it+1}^2\right) - b\lambda_i^2 y_{it}^2\right] = a\lambda_i^4 + \kappa_{ii}\omega_{ii}^2(1-c) - b\lambda_i^2\omega_{ii} - 6\omega_{ii}^2(1-c) \quad (3.15)$$

To assess the marginal informational content of (3.15) with respect to the SR-SARV condition (3.9), it is worthwhile to rewrite (3.9) as:

$$E_{t-1}\left[y_{it+1}^2 - \gamma y_{it}^2 - (1-\gamma)\left(\omega_{ii} + \lambda_i^2\right)\right] = 0 \quad (3.16)$$

Typically, (3.16) corresponds to a diagonal coefficient of (3.9) which does not allow to disentangle the respective roles of ω_{ii} and λ_i^2 within the unconditional variance $(\omega_{ii} + \lambda_i^2)$ of y_{it+1}. By contrast, assumption 3.4 rewritten as (3.15) ensures identification of:

(i) c as coefficient of y_{it}^4

(ii) ω_{ii} from the coefficient of y_{it+1}^2

(iii) λ_i^2 from the knowledge of ω_{ii} and the unconditional variance of y_{it+1}

(iv) b from the coefficient of y_{it}^2 and the knowledge of c, ω_{ii} and λ_i^2.

However, a and κ_{ii} are not identified separately from (3.15). In the same way as second order dynamics of the vector of returns did not allow us to disentangle the respective contributions of factor and residual volatility inside the return variance, the respective contributions of the idiosyncratic kurtosis κ_{ii} and of the factor kurtosis $Ef_t^4 = \frac{a+b}{1-c}$ (through the free parameter a) cannot be identified from fourth order dynamics of return i.

To summarize, we have shown:

Proposition 3.7 *Let us consider the one SV factor model (3.1) (with $K = 1$), with $E_t y_{t+1} = 0$. If, for some asset i, the idiosyncratic conditional kurtosis is constant ($E_t\left[u_{it+1}^4\right] = \kappa_{ii}\omega_{ii}^2$), the unknown parameters λ, Ω, γ, b, c, and $a\lambda_i^4 + \kappa_{ii}\omega_{ii}^2(1-c)$ are identified by the following set of conditional moment restrictions :*

$$
\begin{aligned}
& E_t\left(y_{t+1}\right) = 0 \\
& vecE_t\left[\left(\bar{\bar{y}}_{t+1} - \bar{\bar{\lambda}}\bar{\lambda}^{-1}\bar{y}_{t+1}\right) y_{t+1}'\right] = vec\left[\Omega_{2.} - \bar{\bar{\lambda}}\bar{\lambda}^{-1}\Omega_{1.}\right] \\
& vechE_{t-1}\left[(1-\gamma L)\left(y_{t+1}y_{t+1}' - \lambda\lambda' - \Omega\right)\right] = 0 \\
& E_{t-1}\left[(1-cL)\left(y_{it+1}^4 - 6\omega_{ii} y_{it+1}^2\right) - b\lambda_i^2 y_{it}^2\right] = a\lambda_i^4 + \kappa_{ii}\omega_{ii}^2(1-c) - b\lambda_i^2\omega_{ii} - 6\omega_{ii}^2(1-c)
\end{aligned}
$$

Of course, for efficient estimation, it may be useful to even maintain an assumption of fixed conditional idiosyncratic multivariate kurtosis, which, jointly with assumption 3.4 and 3.6 is tantamount to an assumption about the conditional multivariate kurtosis of returns :

Proposition 3.8 *Let us consider the one-SV factor model (3.1) (with $K = 1$) with the additional assumptions 3.4, 3.6 and:*

$$E_t y_{t+1} = 0$$
$$E_t \left[u_{t+1} u'_{t+1} \otimes u_{t+1} u'_{t+1} \right] = \Theta,$$

then, if D_n^+ denotes the Moore-Penrose inverse of the duplication matrix of size n [7], *the multivariate conditional kurtosis of y_{t+1} is given by:*

$$E_t \left[\left(vech y_{t+1} y'_{t+1} \right) \left(vech y_{t+1} y'_{t+1} \right)' \right]$$
$$= D_n^+ \left[\begin{array}{c} \lambda\lambda' \otimes \lambda\lambda' E_t \left(f_{t+1}^4 \right) + \Theta + \\ \sigma_t^2 \left[4\lambda\lambda' \otimes \Omega + (vec\lambda\lambda') (vec\Omega)' + (vec\Omega) (vec\lambda\lambda')' \right] \end{array} \right] D_n^{+\prime}$$

Proposition 3.8 allows to write matricial observable moment restrictions about $y_{t+1} y'_{t+1} \otimes y_{t+1} y'_{t+1}$ in the same way as, while focusing only on diagonal coefficients, we deduced (3.15) from the conditional kurtosis of return i. Of course, these conditional moment restrictions must be considered jointly with those of proposition 3.3. However, the normalization condition $\bar{\Lambda} = Id_K$, maintained in proposition 3.3 is now irrelevant since the higher order moment restrictions allow us to fully identify the matrix Λ of factor loadings and not only the matrix $B = \bar{\bar{\Lambda}} \bar{\Lambda}^{-1}$.

To summarize, in the one factor case, the unknown parameters $\lambda, \Omega, \gamma, b$ and c are identified, whereas a and Θ cannot be separately identified because only $a\lambda\lambda' \otimes \lambda\lambda' + (1 - c)\Theta$ is identified. We then obtain the following proposition :

Proposition 3.9 *Under the assumptions of Proposition 3.8, efficient instrumental variables estimation of λ, Ω, γ, b, c, and $a\lambda\lambda' \otimes \lambda\lambda' + (1 - c)\Theta$ can be obtained through the following set of conditional moment restrictions :*

$$\begin{array}{l} E_t \left(y_{t+1} \right) = 0 \\ vec E_t \left[\left(\bar{\bar{y}}_{t+1} - \bar{\bar{\lambda}} \bar{\lambda}^{-1} \bar{y}_{t+1} \right) y'_{t+1} \right] = vec \left[\Omega_{2.} - \bar{\bar{\lambda}} \bar{\lambda}^{-1} \Omega_{1.} \right] \\ vech E_{t-1} \left[(1 - \gamma L) y_{t+1} y'_{t+1} - \lambda\lambda' - \Omega \right] = 0 \end{array}$$

[7] The duplication matrix of size n is the $(n^2, \frac{n(n+1)}{2})$ matrix D_n such that, for any symmetric matrix A of size n, $vecA = D_n vechA$. Then, the Moore-Penrose inverse D_n^+ of D_n satisfies: $D_n^+ vecA = vechA$

$$
\begin{aligned}
& D_n^+ E_{t-1}\left[(1-cL)\,\phi\left(y_{t+1},\Omega\right)-b\lambda\lambda' y_t y_t'\right] D_n^{+\prime} \\
= D_n^+ & \left[\begin{array}{c} a\lambda\lambda'\otimes\lambda\lambda' - b\lambda\lambda'\otimes\Omega+(1-c)\Theta- \\ 4(1-c)\Omega\otimes\Omega-2(1-c)(vec\Omega)(vec\Omega)' \end{array}\right] D_n^{+\prime}
\end{aligned}
$$

where:

$$
\phi(y_t,\Omega)=(Vec\ y_t y_t')(Vec\ y_t y_t')'-4\Omega\otimes y_t y_t'-(Vec\ \Omega)(Vec\ y_t y_t')'-(Vec\ y_t y_t')(Vec\ \Omega)'
$$

In practice, one would not like to make inference about the huge number of unknown parameters involved in the matrix Θ through all the conditional moment restrictions associated to $\phi(y_{t+1},\Omega)$. In other words, only some components, for instance corresponding to diagonal terms as in proposition 3.7, may be considered. Moreover, note that, in the one factor case, the second order moment restrictions of proposition 3.9 can also be written:

$$
E_t\left[\left(\bar{\lambda}\bar{\bar{y}}_{t+1}-\bar{\bar{\lambda}}\bar{y}_{t+1}\right)y_{t+1}'\right]=vec\left[\bar{\lambda}\Omega_{2.}-\bar{\bar{\lambda}}\Omega_{1.}\right]
$$

These restrictions were precisely the ones tested in section 3.1 to determine the number of factors, in the line of common features restrictions *à la* Engle and Kozicki (1993). However, we keep the formulas in terms of $b=\bar{\bar{\lambda}}\bar{\lambda}^{-1}$, since they shed more light on the multifactor extensions with $B=\bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$. All the results of this subsection can actually be extended to a SV multifactor model at the cost of tedious notations. For sake of illustration, we provide in the appendix the generalization of propositions 3.7 and 3.9 to the two factors case.

4 Model with linear risk premiums

Following King, Sentana and Wadhwani (1994) (KSW hereafter), we consider specification of time-varying risk premiums that can be understood as a dynamic version of the Arbitrage Pricing Theory. As in KSW, the time variation in the conditional variances of factors allows to identify the factor risk premiums. Even more importantly, when the prices of factor risks are non-zero, identification of corresponding risk premiums precludes any transfer of a part of factor variance into the residual variance, as put forward in section 2. Therefore identification and IV estimation of all the parameters of interest are made possible from expectation and variance of returns without resorting to higher order moments. Moreover, the statistical sequential procedure that we have settled in section 3.1 to determine the number of factors is easily generalized to the case of APT-like time varying risk premium.

Finally, we show that, if we want to relax the APT specification, one can also identify and estimate a fully unconstrained set of idiosyncratic risk premiums. Following KSW, these estimators can be used to assess the APT specification.

4.1 The general framework

Introducing risk premiums is akin to revisit model (3.1) in a more general form:

$$y_{t+1} = E_t\left(y_{t+1}\right) + \Lambda f_{t+1} + u_{t+1} \tag{4.1}$$

which allows to consider a vector of time varying expected returns $E_t\left(y_{t+1}\right)$. In this section, we always consider returns mesured in excess of the riskless asset and thus, expected returns $E_t\left(y_{it+1}\right)$ are unambiguously interpreted as risk premiums.

Following the APT literature or more generally the linear factor pricing principle, we assume that risk premiums are linear combinations of return betas:

$$E_t\left(y_{t+1}\right) = \Lambda V_t\left(f_{t+1}\right)\tau_t \tag{4.2}$$

where τ_t is interpreted as the vector of prices of risk for each of the factors. Of course, this economic interpretation implies that τ_t belongs to the agent's information set at time t. (4.2) is actually the risk premium specification choosen by KSW. Notice that, while KSW maintain the assumption of an exact conditional K-factors structure, which means a diagonal residual matrix, this is no longer the case in our model. Therefore, there may be less theoretical underpinnings for the APT-like assumption of zero risk premium for idiosyncratic risks. Some arguments will be made explicit in subsection 4.3 below to warrant specification (4.2) as well as to define a statistical testing procedure of it.

Before studying IV estimation of the parameters of interest that takes into account the extra risk premium terms and corresponding additional unknown parameters, it is important to address the model choice issue, that is the determination of the number K of factors. We basically want to extend the approach proposed in section 3.1 to the more general factor model (4.1)/(4.2). The crucial trick of section 3.1 was a sequential testing procedure based on conditional moment restrictions:

$$E_t\left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)y'_{t+1} - D\right] = 0 \tag{4.3}$$

for unknown matrices B and D. We considered that, when the overidentification test fails to reject (4.3) for a given K-dimensional subvector $\bar{y}_{t+1}$ of y_{t+1}, it means that a K-factor model:

$$\begin{bmatrix} \bar{y}_{t+1} \\ \bar{\bar{y}}_{t+1} \end{bmatrix} = \begin{bmatrix} \bar{\Lambda} \\ \bar{\bar{\Lambda}} \end{bmatrix} f_{t+1} + u_{t+1} \tag{4.4}$$

is valid with $B = \bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$.

Let us now consider the generalization of (4.4) according to the risk premium specification (4.2):

$$\begin{bmatrix} \bar{y}_{t+1} \\ \bar{\bar{y}}_{t+1} \end{bmatrix} = \begin{bmatrix} \bar{\Lambda} \\ \bar{\bar{\Lambda}} \end{bmatrix} V_t\left(f_{t+1}\right)\tau_t + \begin{bmatrix} \bar{\Lambda} \\ \bar{\bar{\Lambda}} \end{bmatrix} f_{t+1} + \begin{bmatrix} \bar{u}_{t+1} \\ \bar{\bar{u}}_{t+1} \end{bmatrix} \tag{4.5}$$

Then, if $B = \bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$, $\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1} = \bar{\bar{u}}_{t+1} - B\bar{u}_{t+1}$ still has constant conditional covariances with each of the returns y_{it+1}. Therefore proposition 3.1 remains valid in the more general factor model (4.1) with risk premiums and inference about H_{0k} can still be based on the conditional moment restrictions:

$$E_t\left[\left(\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right)y'_{t+1} - D\right] = 0$$

for unknown matrices B and D. In other words, the identification strategy of the number of factors will be exactly the same as in section 3.1.

4.2 Identification and IV estimation with APT-like risk premiums

For sake of notational simplicity, let us consider a one-factor version of the model (4.1)/(4.2):

$$y_{t+1} = \lambda\sigma_t^2\tau_t + \lambda f_{t+1} + u_{t+1}.$$

As usual, the respective roles of σ_t^2 and τ_t within the risk premium cannot be disentangled without specifying more precisely the dynamics of the risk premium process τ_t. Following KSW, we will maintain here the simplifying assumption that the price of risk is positive and constant over time:

$$\tau_t = \tau > 0 \text{ for all } t.$$

As stressed by KSW, this does not though imply that the overall price of risk for each asset, that is the Sharpe ratio $\frac{E_t y_{it+1}}{(V_t y_{it+1})^{1/2}}$, is constant.

Whatever, we focus here on the following specification:

$$y_{t+1} = \lambda\tau\sigma_t^2 + \lambda f_{t+1} + u_{t+1} \tag{4.6}$$

Of course, the occurence of σ_t^2 in $E_t\left(y_{t+1}\right)$ implies that $E_t\left(y_{t+1}y'_{t+1}\right)$ will involve σ_t^4. Therefore, observable moment restrictions about returns volatility cannot be obtained without specifying a forecasting model for σ_{t+1}^4. We will maintain here the following assumption :

Assumption 4.1 *There exists a^*, b^*, c^* such that*

$$E_t\left(\sigma_{t+1}^4\right) = a^* + b^*\sigma_t^2 + c^*\sigma_t^4$$

with $0 < c^* < 1$ *and* $a^* + b^* + c^* > 1$.

One way to get some intuition on this assumption is to compute:

$$\begin{aligned} V_t\left(\sigma_{t+1}^2\right) &= E_t\left(\sigma_{t+1}^4\right) - \left[(1-\gamma) + \gamma\sigma_t^2\right]^2 \\ &= \left[a^* - (1-\gamma)^2\right] + \left[b^* - 2\gamma\left(1-\gamma\right)\right]\sigma_t^2 + \left[c^* - \gamma^2\right]\sigma_t^4 \end{aligned} \tag{4.7}$$

By comparison of (4.7) with proposition 3.5, one can realize that assumptions 3.4 and 4.1 are actually equivalent when:

$$a^* = \frac{a}{\kappa},\ b^* = \frac{b}{\kappa},\ \text{and } c^* = c. \tag{4.8}$$

In this respect, assumption 4.1 is tightly related to previous assumption 3.4 and could be justified by the same examples of strong GARCH or affine process of conditional variance. However, by contrast with section 3.3 we are looking here for conditional moment restrictions in the spirit of proposition 3.3, that is involving only conditional expectations and variances of returns. This has two important consequences in terms of identification.

First, there is no hope to take advantage of an assumption of fixed conditional kurtosis κ for the factor process. This assumption is not maintained here. Second, one cannot identify the unconditional variance of σ_t^2, or equivalently, the unconditional variance of the risk premium vector. Note that, according to assumption 4.1:

$$Var\sigma_t^2 = \frac{a^* + b^*}{1 - c^*} - 1 \tag{4.9}$$

Therefore, the necessary degree of freedom in $Var\sigma_t^2$ can be taken into account by considering a free parameter a^* for given b^* and c^*. Up to this degree of freedom, we will get IV estimation and identification of all the parameters of interest as stated in proposition 4.2:

Proposition 4.2 *In the one SV factor model with risk premium (4.1)/(4.2):*

$$y_{t+1} = \lambda\sigma_t^2\tau + \lambda f_{t+1} + u_{t+1}$$

for any given value of a^*, the parameters $b^*,\ c^*,\ \lambda,\ \tau,\ \gamma$ and $\Omega(a^*)$ are characterized by:

$$E_t\left[y_{t+1}\lambda' - \lambda y'_{t+1}\right] = 0 \tag{4.10}$$

$$E_t\left[\left(\bar{\lambda}\bar{\bar{y}}_{t+1} - \bar{\bar{\lambda}}\bar{y}_{t+1}\right)y'_{t+1}\right] = \bar{\lambda}\Omega_{2.} - \bar{\bar{\lambda}}\Omega_{1.} \tag{4.11}$$

$$E_{t-1}\left[(1-\gamma L)\,y_{t+1} - \lambda\tau(1-\gamma)\right] = 0 \tag{4.12}$$

$$VecE_{t-1}\left[\begin{array}{c} y_{t+1}y'_{t+1} - c^* y_t y'_t - y_{t+1}\frac{\lambda'}{\tau} - \\ y_t\lambda'\left(\frac{c^*}{\tau} - \tau b^*\right) - (1-c^*)\,\Omega(a^*) - \lambda\lambda'\tau^2 a^* \end{array}\right] = 0 \tag{4.13}$$

The notation $\Omega(a^*)$ means that only the identification of the residual covariance matrix Ω is contaminated by the non-identification of a^*, which is actually akin to non-identification of $Var\sigma_t^2$. Indeed, $\Omega = \Omega(a^*)$ is identified only through $(1-c^*)\Omega(a^*) + \lambda\lambda'\tau^2 a^*$, when the free parameter a^* is fixed.

On the contrary, we claim that parameters b^*, c^*, λ, τ and γ are fully identified from the conditional moment restrictions (4.10), (4.11), (4.12) and (4.13). The intuition behind this is the following. The APT-like risk premium specification first adds a set of common features restrictions:

$$E_t\left[\lambda_i y_{jt+1} - \lambda_j y_{it+1}\right] = 0 \tag{4.14}$$

to the common features restrictions already provided by the one factor model of conditional covariances:

$$E_t\left[\left(\lambda_i y_{jt+1} - \lambda_j y_{it+1}\right)y'_{t+1}\right] = \lambda_i\Omega_{j.} - \lambda_j\Omega_{i.} \tag{4.15}$$

Conditions (4.14) were actually already ensured in section 3 through the maintained assumption $E_t(y_{t+1}) = 0$. As in section 3, the common features set of restrictions, even augmented by (4.14), provides identification of the factor loadings λ_i's only up to a scale factor. According to proposition 2.2, this scale factor $\left(1-\underline{\sigma}^2\right)$ may be associated to a variance transfer:

$$\begin{cases} \sigma_t^2 \text{ replaced by } \frac{\sigma_t^2 - \underline{\sigma}^2}{1-\underline{\sigma}^2} \\ \Omega \text{ replaced by } \Omega + \underline{\sigma}^2\lambda\lambda' \end{cases}$$

Further, these common features restrictions do not bring any information about the free parameter a^* in $\Omega(a^*)$. The added value, in terms of identification, of proposition 4.2 is to allow full identification of the factor loadings through additional moment restrictions (4.13) resulting from the risk premium model:

$$E_t(y_{t+1}) = \lambda\tau\sigma_t^2$$

Under the maintained assumption of a non-zero price of risk τ, this model brings additional identifying information about the latent process σ_t^2 which precludes the aforementionned transfer of variance. To see this, it is worth noticing that the (i, j) coefficient of (4.13) provides, through the coefficients of y_{jt+1} and y_{it}, separate identification of:

$$\frac{\lambda_j}{\tau} \text{ and } \frac{\lambda_i}{\tau}\left(c^* - b^*\tau^2\right)$$

Moreover, identification of $\frac{\lambda_j}{\tau}$ leads to a separate identification of λ_j and $\tau > 0$, thanks to the additional information (implied by (4.12)):

$$Ey_{jt+1} = \lambda_j \tau$$

Then b^* is identified from $\frac{\lambda_i}{\tau}\left(c^* - b^*\tau^2\right)$ since c^* is identified as the coefficient of $y_t y_t'$. The volatility persistence parameter γ is identified from (4.12).

4.3 Testing for the zero-price of idiosyncratic risk

Following KSW, we can test the APT-like specification of risk premiums by allowing the idiosyncratic volatility ω_{ii} of each asset i to affect the corresponding risk premium through an additive term μ_i:

$$y_{t+1} = \mu + \lambda\sigma_t^2\tau + \lambda f_{t+1} + u_{t+1} \tag{4.16}$$

Note that, since we do not assume that the idiosyncratic covariance matrix Ω is diagonal, μ_i may also involve risk premium terms related to the covariation with idiosyncratic risks of other assets $j \neq i$. This is only an issue for interpretation and does not play any role in the following testing procedure.

The crucial point is that the μ_i's are also identified, jointly with the other parameters of interest, from conditional moments restrictions like (4.12) and (4.13). To see this, let us just rewrite (4.12) and (4.13) with y_{t+1} replaced by $(y_{t+1} - \mu)$. We then get conditional moment restrictions consistent with the extended model (4.16):

$$E_{t-1}\left[(1-\gamma L)\left(y_{t+1} - \mu\right) - \lambda\tau\left(1-\gamma\right)\right] = 0 \tag{4.17}$$

and

$$E_{t-1}\begin{bmatrix} y_{t+1}y_{t+1}' - c^* y_t y_t' - (1-c^*L)\left(\mu y_{t+1}' + y_{t+1}\mu' + y_{t+1}\frac{\lambda'}{\tau}\right) \\ -\tau b^* y_t \lambda' - (1-c^*)\left(\Omega(a^*) - \mu\mu' - \mu\frac{\lambda'}{\tau}\right) - \mu\lambda'\tau b^* - \lambda\lambda'\tau^2 a^* \end{bmatrix} = 0 \tag{4.18}$$

Then, for $i \neq j$, the (i,j) coefficient of (4.18) provides, through the coefficients of y_{it+1}, y_{jt+1} and y_{it}, separate identification of :

$$\mu_j + \frac{\lambda_j}{\tau} \text{ , } \mu_i \text{ and } c^* \left(\mu_j + \frac{\lambda_i}{\tau} \right) - b^* \tau \lambda_j$$

Then, we get identification of μ and $\frac{\lambda}{\tau}$ while separate identification of λ_i and τ is obtained from the additional information (implied by (4.17)):

$$E\left(y_{jt+1} - \mu_j\right) = \lambda_j \tau$$

Then, b^* is identified from $c^* \left(\mu_j + \frac{\lambda_j}{\tau} \right) - b^* \tau \lambda_j$ since c^* is identified as the coefficient of $y_t y_t'$. The volatility persistence parameter γ is identified from (4.17).

To summarize, we still get IV estimation and identification of all the parameters of interest, up to the free parameter a^*. It is then possible to test the APT-like specification of risk premiums, either equation by equation (testing the null $H_{0i} : \mu_i = 0$ for any given i) or jointly (testing the null $H_0 : \mu = 0$).

5 Conclusion

The main contribution of this paper is to characterize to what extent SV factor structures are identified by conditional moment restrictions. Insofar as the announced goal of such structures is to afford a parsimonious representation of joint volatility dynamics, fully parametric models of conditional probability distributions should not be needed for their identification. We actually show that, when factor volatilities also show up in conditional means through well specified risk premium terms, identification of the SV factor structure is ensured from the first two conditional moments. On the contrary, without such time-varying risk premiums, higher order moments are needed for full identification of the SV factor structure. We focus here on conditional kurtosis under a maintained assumption of zero conditional skewness and no leverage effect. The way to accomodate in our framework any kind of multivariate asymetry effect is discussed in a companion paper (Dovonon, Doz and Renault (2004)).

Of course, identifying conditional moment restrictions naturally paves the way for GMM estimation and inference through a convenient choice of instruments. Practical implementation of such GMM interence open several kinds of issues. First, as it would also be the case with likelihood inference, a preliminary step of determination of the number of factors is needed. We have shown here how the Engle and Kozicki (1993) test procedure may be completed to fully take into account the information content of the factor structure. Second, the empirical goodness of fit of competing SV factor models is still an open question. While the empirical performance of similar SV structures has been documented in maximum likelihood settings (see Fiorentini, Sentana and Shephard (2003) and references therein), the semi-parametric structure considered in this paper may improve the statistical fit. Finally, depending upon the category of financial asset returns considered, additional asymetry effects along the line of Dovonon, Doz and Renault (2004) could be statistically and economically significant. An extensive horse race between the various possible SV factor specifications is still work in progress.

Appendix

Proof of Proposition 2.1:

We have $1 = E\sigma_{kt}^2 \geq \underline{\sigma}_k^2 > 0$. Let $\alpha_k \leq \underline{\sigma}_k^2$, $0 < \alpha_k < 1$.

We denote by D a $K \times K$ matrix the coefficients of which are all zero, except the k^{th} diagonal coefficient, equal to α_k. We have:

$$\Sigma_t = \Lambda \left(D_t - D\right) \Lambda' + \Lambda D \Lambda' + \Omega$$

with

$$\begin{cases} \tilde{\Omega} = \Lambda D \Lambda' + \Omega \neq \Omega \\ \tilde{\Omega} - \Omega = \Lambda D \Lambda' \gg 0 \end{cases}$$

Let us define: $\Delta = Id_K - D$.By construction, Δ is a diagonal matrix with positive diagonal coefficients. Therefore $\Delta^{1/2}$ and $\Delta^{-1/2}$ are defined without any ambiguity and we can consider:

$$\tilde{\Lambda} = \Lambda \ \Delta^{1/2} \tag{A.1}$$

and

$$\tilde{D}_t = \Delta^{-1/2} \left(D_t - D\right) \Delta^{-1/2} \tag{A.2}$$

Then, we have:

$$\Sigma_t = \Lambda \left(D_t - D\right) \Lambda' + \tilde{\Omega} = \tilde{\Lambda} \tilde{D}_t \tilde{\Lambda}' + \tilde{\Omega}$$

with:

$$\tilde{\Lambda} = \Lambda \quad \Delta^{1/2} \quad \text{and} \quad E\tilde{D}_t = \Delta^{-1/2} \left(Id_K - D\right) \Delta^{-1/2} = Id_K.$$

Proof of Proposition 2.2:

While (2.10) means that we have a one-factor model of conditional variance:

$$\Sigma_t = \lambda\lambda'\sigma_t^2 + \Omega$$

where $\sigma_t^2 = Var\left(f_{t+1} \left| J_t\right.\right)$, we know, from Proposition 2.1, that we can write:

$$\Sigma_t = \tilde{\lambda}\tilde{\lambda}'\tilde{\sigma}_t^2 + \tilde{\Omega}$$

where by (A.1) and (A.2): $\tilde{\sigma}_t^2 = \dfrac{\sigma_t^2 - \underline{\sigma}^2}{1 - \underline{\sigma}^2}$, $\tilde{\lambda} = \lambda\sqrt{1 - \underline{\sigma}^2}$, and $\tilde{\Omega} = \Omega + \underline{\sigma}^2 \lambda\lambda'$. Therefore, we will prove the announced result by characterizing a factor $\tilde{f}_{t+1}$ such that:

$$\begin{cases} y_{t+1} = \tilde{\lambda}\ \tilde{f}_{t+1} + \tilde{u}_{t+1} \\ Cov\left(\tilde{u}_{t+1}, \tilde{f}_{t+1} \,|J_t\right) = 0 \\ Var\left(\tilde{f}_{t+1} \,|J_t\right) = \tilde{\sigma}_t^2 \end{cases} \tag{A.3}$$

We write: $\tilde{f}_{t+1} = \dfrac{f_{t+1}}{\sqrt{1-\underline{\sigma}^2}} + \xi_{t+1}$ which then means that:

$$\begin{aligned} y_{t+1} &= \lambda\ f_{t+1} + u_{t+1} = \tilde{\lambda}\ \frac{f_{t+1}}{\sqrt{1-\underline{\sigma}^2}} + u_{t+1} \\ &= \tilde{\lambda}\left(\tilde{f}_{t+1} - \xi_{t+1}\right) + u_{t+1} \\ &= \tilde{\lambda}\ \tilde{f}_{t+1} + \tilde{u}_{t+1} \end{aligned}$$

if and only if: $\tilde{u}_{t+1} = u_{t+1} - \tilde{\lambda}\xi_{t+1}$.

Therefore, the second equation of (A.3) is tantamount to:

$$Cov\left(u_{t+1} - \tilde{\lambda}\ \xi_{t+1}, \frac{f_{t+1}}{\sqrt{1-\underline{\sigma}^2}} + \xi_{t+1} \,|J_t\right) = 0$$

that is: $Cov\left(u_{t+1}, \xi_{t+1} \,|J_t\right) = \lambda\ Cov\left(\xi_{t+1}, f_{t+1} \,|J_t\right) + \lambda\sqrt{1-\underline{\sigma}^2} Var\left(\xi_{t+1} \,|J_t\right)$.

In other words:

$$\exists \rho_t \in J_t \qquad Cov\left(u_{t+1}, \xi_{t+1} \,|J_t\right) = \ \rho_t\ \lambda \tag{A.4}$$

with:

$$Cov\left(f_{t+1}, \xi_{t+1} \,|J_t\right) = \rho_t - \sqrt{1-\underline{\sigma}^2}\ Var\left(\xi_{t+1} \,|J_t\right) \tag{A.5}$$

On the other hand, the last equation of (A.3) means that:

$$\frac{1}{1-\underline{\sigma}^2}\ \ \sigma_t^2 + Var\left(\xi_{t+1} \,|J_t\right) + \frac{2}{\sqrt{1-\underline{\sigma}^2}}\ Cov\left(f_{t+1}, \quad \xi_{t+1} \,|J_t\right) = \frac{\sigma_t^2 - \underline{\sigma}^2}{1-\underline{\sigma}^2}$$

that is:

$$\frac{\underline{\sigma}^2}{1-\underline{\sigma}^2} + Var\left(\xi_{t+1} \,|J_t\right) + \frac{2}{\sqrt{1-\underline{\sigma}^2}} Cov\left(f_{t+1}, \xi_{t+1} \,|J_t\right) = 0 \tag{A.6}$$

We will rewrite the set of conditions (A.4), (A.5) and (A.6) on the following equivalent form:

There exists a J_t-measurable random variable ρ_t such that:

$$\begin{cases} Cov\left(u_{t+1}\ , \xi_{t+1} \,|J_t\right) = \rho_t\ \lambda \\ Var\left(\xi_{t+1} \,|J_t\right) = \frac{\underline{\sigma}^2}{1-\underline{\sigma}^2} + \frac{2\rho_t}{\sqrt{1-\underline{\sigma}^2}} \\ Cov\left(f_{t+1}, \quad \xi_{t+1} \,|J_t\right) = -\rho_t - \frac{\underline{\sigma}^2}{\sqrt{1-\underline{\sigma}^2}} \end{cases} \tag{A.7}$$

In other words, the proof will be completed if we succeed to build a random variable ξ_{t+1} such that the existence of $\rho_t \in J_t$ conformable to (A.7) is guaranteed. In order to do this, we define ξ_{t+1} from its conditional linear regression on (f_{t+1}, u_{t+1}) given J_t:

$$\begin{cases} \xi_{t+1} = \alpha_t \ f_{t+1} + \beta_t' \ u_{t+1} + z_{t+1} \\ Cov\left(z_{t+1}, u_{t+1} \left| J_t \right.\right) = 0 \\ Cov\left(z_{t+1}, f_{t+1} \left| J_t \right.\right) = 0 \end{cases}$$

We first notice that the value of α_t and β_t are imposed by the first and the last equations of (A.7) and are respectively given by:

$$Var\left(u_{t+1} \left| J_t\right.\right) \beta_t = \rho_t \lambda \Longleftrightarrow \beta_t = \rho_t \Omega^{-1} \lambda$$

and:

$$\alpha_t \sigma_t^2 = -\rho_t - \frac{\underline{\sigma}^2}{\sqrt{1-\underline{\sigma}^2}} \Longleftrightarrow \alpha_t = -\frac{1}{\sigma_t^2}\left(\rho_t + \frac{\underline{\sigma}^2}{\sqrt{1-\underline{\sigma}^2}}\right).$$

By computing $Var\left(\xi_{t+1} \left| J_t \right.\right)$ with the above values of α_t and β_t, we conclude that the conjunction of the three conditions of (A.7) will be fulfilled if and only if:

$$\frac{1}{\sigma_t^2}\left(\rho_t + \frac{\underline{\sigma}^2}{\sqrt{1-\underline{\sigma}^2}}\right)^2 + \rho_t^2 \ \lambda' \ \Omega^{-1} \ \lambda + Var\left(z_{t+1} \left| J_t \right.\right) = \frac{\underline{\sigma}^2}{1-\underline{\sigma}^2} + \frac{2\rho_t}{\sqrt{1-\underline{\sigma}^2}} \tag{A.8}$$

Note that if we find ρ_t conformable to equation (A.8), $Var\left(\xi_{t+1} \left| J_t\right.\right)$ as defined by the second equation of (A.7) will be positive by construction. In other words, the only thing to prove is that we are able to define a random variable z_{t+1} such that the equation (A.8) admits at least one solution ρ_t. But (A.8) can be rewritten as:

$$\rho_t^2\left(\frac{1}{\sigma_t^2} + \lambda'\Omega^{-1}\lambda\right) + \frac{2\rho_t}{\sqrt{1-\underline{\sigma}^2}}\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right) + \frac{\underline{\sigma}^2}{1-\underline{\sigma}^2}\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right) + Var\left(z_{t+1} \left| J_t \right.\right) = 0$$

Therefore, we have to find a random variable $Var\left(z_{t+1} \left| J_t\right.\right)$ such that the discriminant of this equation is positive a.s.:

$$\frac{1}{1-\underline{\sigma}^2}\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right)^2 - \left(\frac{1}{\sigma_t^2} + \lambda'\Omega^{-1}\lambda\right)\left[\frac{\underline{\sigma}^2}{1-\underline{\sigma}^2}\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right) + Var\left(z_{t+1} \left| J_t\right.\right)\right] \geq 0$$

Equivalently, we have to check that:

$$\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right)^2 - \left(\frac{1}{\sigma_t^2} + \lambda'\Omega^{-1}\lambda\right)\underline{\sigma}^2\left(\frac{\underline{\sigma}^2}{\sigma_t^2} - 1\right) \geq 0.$$

But, by Assumption (2.12): $\frac{\underline{\sigma}^2}{\sigma_t^2} - 1 \leq 0 \quad$ a.s.

Then, we have to show that:

$$\frac{\underline{\sigma}^2}{\sigma_t^2} - 1 - \underline{\sigma}^2 \left(\frac{1}{\sigma_t^2} + \lambda' \Omega^{-1} \lambda \right) \leq 0 \quad \text{a.s.}$$

that is: $1 + \underline{\sigma}^2 \lambda' \Omega^{-1} \lambda \geq 0$, which is obviously true and completes the proof.

Proof of Proposition 2.3:

Let us assume, without loss of generality, that the first K rows of Λ define a nonsingular matrix $\bar{\Lambda}$ of size K. Then, by denoting $\bar{\bar{\Lambda}}$ the last $(n-K)$ rows of Λ, the $n-K$ rows of the matrix $A = (-\bar{\bar{\Lambda}}\bar{\Lambda}^{-1} \quad Id_{n-K})$ define a basis of the orthogonal space $\Lambda^{\perp}$ of the range of Λ.
Thus, the equality :

$$\Lambda D_t \Lambda' + \Omega = L \Delta_t L' + W$$

implies that $A\left(L\Delta_t L' + W\right)$ is a constant matrix, equal to its unconditional expectation: $A\left(LL' + W\right)$.
By difference, we get:

$$AL\left(\Delta_t - Id_J\right) L' = 0$$

From the linear independence of the J columns of L, we conclude that:

$$AL\left(\Delta_t - Id_J\right) = 0$$

and thus, the J columns of AL are zero since none of the random diagonal coefficients of the diagonal matrix $(\Delta_t - Id_J)$ is identically zero. Therefore, the rows of A belong to the orthogonal space $L^{\perp}$ of the range of L, that is : $\Lambda^{\perp} \subset L^{\perp}$.
Hence: $\text{Span}(L) \subset \text{Span}(\Lambda)$.
Finally, as L and Λ play symmetric roles: $\text{Span}(L) = \text{Span}(\Lambda)$ and $K = J$.

Proof of Proposition 2.4:

From the two factor decompositions:

$$\Lambda \ D_t \ \Lambda' + \Omega = L \ \Delta_t \ L' + W$$

we get, by considering unconditional expectations :

$$\Lambda \ \Lambda' + \Omega = L \ L' + W$$

and then, by difference of these two equations:

$$\Lambda(D_t - Id_K)\Lambda' = L(\Delta_t - Id_K)L'$$

From proposition 2.3, we know that the ranges of L and Λ concide. As L and Λ have full rank, it exists a non singular matrix M such that : $L = \Lambda M$. Thus:

$$\Lambda(D_t - Id_K)\Lambda' = \Lambda M(\Delta_t - Id_K)M'\Lambda'$$

As Λ is of full column rank, this implies (using a left multiplication by $(\Lambda'\Lambda)^{-1}\Lambda'$ and a right multiplication by $\Lambda(\Lambda'\Lambda)^{-1}$):

$$D_t - Id_K = M(\Delta_t - Id_K)M'$$

If we denote $M = (m_{ij})_{1\leq i,j\leq K}$ and if $D_t = \text{diag}(\sigma^2_{kt})$ and $\Delta_t = \text{diag}(\tilde{\sigma}^2_{kt})$ we thus obtain :

$$\sum_{k=1}^{K}(\tilde{\sigma}^2_{kt} - 1)m_{ik}m_{jk} = \begin{cases} \sigma^2_{it} - 1 & \text{if } \ i = j \\ 0 & \text{otherwise.} \end{cases}$$

As $\delta_t = (\tilde{\sigma}^2_{kt})_{1\leq k\leq K}$ is supposed to have a non singular covariance matrix, we then obtain : $m_{ik}m_{jk} = 0$ if $i \neq j$. This proves that in each column m_k of M there is at most one element m_{ik} which is different from 0. But, as M is non singular, there is in fact exactly one element m_{ik} which is different from 0 in each column m_k. For each k, let us denote by $m_{\tau(k)k}$ this element. As no row of M can be equal to 0, τ is a permutation on $\{1, \cdots, K\}$.
Then, the relation $L = \Lambda M$ can be written :

$$\forall (i,j) \quad l_{ij} = \sum_{k=1}^{K}\lambda_{ik}m_{kj} = \lambda_{i\tau(j)}m_{\tau(j)j}$$

Let us then define a permutation matrix Q by :

$$q_{ij} = \begin{cases} 1 & \text{if } \ i = \tau(j) \\ 0 & \text{otherwise} \end{cases}$$

and let us denote $\Delta = \text{diag}(m_{1\tau^{-1}(1)} \cdots m_{K\tau^{-1}(K)})$.
Straightforward calculations show that : $L = \Lambda\Delta Q$.

Proof of Proposition 2.5:

We know by proposition 2.4 that $L = \Lambda\Delta Q$ where Q is the permutation matrix defined in the proof of proposition 2.4 and $\Delta = \mathrm{diag}(\delta_1 \cdots \delta_K)$ with $\delta_k = m_{k,\tau^{-1}(k)} \neq 0$ for $k = 1 \cdots K$. Then the relation $\Lambda D_t \Lambda' + \Omega = L\Delta_t L' + W$ can be written :

$$\Lambda(D_t - \Delta Q \Delta_t Q' \Delta)\Lambda' = W - \Omega \tag{A.9}$$

so that :

$$D_t - \Delta Q \Delta_t Q' \Delta = (\Lambda'\Lambda)^{-1}\Lambda'(W - \Omega)\Lambda(\Lambda'\Lambda)^{-1}$$

Let $A = (\Lambda'\Lambda)^{-1}\Lambda'(W - \Omega)\Lambda(\Lambda'\Lambda)^{-1}$, we then have :

$$\begin{aligned} a_{kk} &= \sigma_{kt}^2 - \delta_k^2 \tilde{\sigma}_{\tau^{-1}(k)t}^2 \quad \forall k = 1 \cdots K \\ a_{ij} &= 0 \qquad \text{if } i \neq j. \end{aligned}$$

Then, since $\delta_k^2 \neq 0$, these equalities can be consistent with the zero lower bound for both σ_{kt}^2 and $\tilde{\sigma}_{\tau^{-1}(k)t}^2$ (identification condition) if and only if $a_{kk} = 0$ for any $k = 1 \cdots K$, that is $A = 0$.

But on the one hand, $A = 0$ means $\Lambda'(W - \Omega)\Lambda = 0$ and on the other hand, taking the expectation of (A.9) implies that $\Lambda(Id_K - \Delta^2)\Lambda' = W - \Omega$. These two relations imply that $\Lambda'\Lambda(Id_K - \Delta^2)\Lambda'\Lambda = 0$. We then obtain that $\Delta^2 = Id_K$ which in turns implies that $W = \Omega$ and completes the proof.

Proof of Proposition 2.7:

We know from proposition 2.4 that the factor loadings λ and $\tilde{\lambda}$ must be proportional:

$$\tilde{\lambda} = k\lambda \text{ for some } k \in \mathbb{R}.$$

Then the decomposition of the conditional variance of y_{t+1} gives:

$$\lambda\lambda'\sigma_t^2 + \Omega = k^2\lambda\lambda'\tilde{\sigma}_t^2 + \tilde{\Omega}$$

and, by difference with unconditional expectations:

$$\lambda\lambda'\left(\sigma_t^2 - 1\right) = k^2\lambda\lambda'\left(\tilde{\sigma}_t^2 - 1\right).$$

Thus:

$$\sigma_t^2 - 1 = k^2(\tilde{\sigma}_t^2 - 1) \tag{A.10}$$

As we assume that both f_t^2 and $\tilde{f}_t^2$ have a GARCH (1,1) structure:

$$\begin{cases} \sigma_{t+1}^2 - 1 = -\gamma + \alpha f_{t+1}^2 + (\gamma - \alpha)\sigma_t^2 \\ \tilde{\sigma}_{t+1}^2 - 1 = -\tilde{\gamma} + \tilde{\alpha} \tilde{f}_{t+1}^2 + (\tilde{\gamma} - \tilde{\alpha})\tilde{\sigma}_t^2 \end{cases}$$

we then obtain, by applying (A.10): $0 = -\gamma + \tilde{\gamma} k^2 + \alpha f_{t+1}^2 - \tilde{\alpha} k^2 \tilde{f}_{t+1}^2 + (\gamma - \alpha)\sigma_t^2 - (\tilde{\gamma} - \tilde{\alpha})\, \tilde{k}^2 \tilde{\sigma}_t^2$. By computing conditional variances given J_t, we get:

$$\alpha^2 \kappa \sigma_t^2 = \tilde{\alpha}^2 k^4 \tilde{\kappa} \tilde{\sigma}_t^2 \tag{A.11}$$

where $(\kappa + 1)$ and $(\tilde{\kappa} + 1)$ are respectively the kurtosis coefficients of the conditional probability distribution of (f_{t+1}/σ_t) and $(\tilde{f}_{t+1}/\tilde{\sigma}_t)$ given J_t.

By plugging (A.11) into (A.10) to eliminate $\tilde{\sigma}_t^2$, we get:

$$\sigma_t^2 - 1 = k^2 \left[\frac{\alpha^2 \kappa}{\tilde{\alpha}^2 k^4 \tilde{\kappa}} \sigma_t^2 - 1 \right].$$

Since σ_t^2 is by definition a non degenerate random variable, this imply: $k^2 = 1$ and in turn by (A.10) and (2.21): $\sigma_t^2 = \tilde{\sigma}_t^2$ and $\gamma = \tilde{\gamma}$.

Then, by identification of the two GARCH equations:

$$\alpha(f_{t+1}^2 - \sigma_t^2) = \tilde{\alpha}(\tilde{f}_{t+1}^2 - \sigma_t^2).$$

But, using $k^2 = 1$, (A.11) gives $\alpha = \pm\tilde{\alpha}$, under the maintained assumtion: $\kappa = \tilde{\kappa}$. Thus, as α and $\tilde{\alpha}$ are nonnegative, this assumption gives: $\alpha = \tilde{\alpha}$ and $f_{t+1}^2 = \tilde{f}_{t+1}^2$.

Proof of Proposition 3.1

We first show that (i) $\Rightarrow$ (ii). With obvious notations, since:

$$\bar{y}_{t+1} = \bar{\Lambda} f_{t+1} + \bar{u}_{t+1},$$

we get:

$$\bar{\bar{y}}_{t+1} = \bar{\bar{\Lambda}} \bar{\Lambda}^{-1} \left[\bar{y}_{t+1} - \bar{u}_{t+1} \right] + \bar{\bar{u}}_{t+1}$$

that is, with $B = \bar{\bar{\Lambda}} \bar{\Lambda}^{-1}$:

$$\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1} = \bar{\bar{u}}_{t+1} - B\bar{u}_{t+1}$$

is conditionally homoskedastic, as a (constant) linear function of the homoskedastic vector u_{t+1}.

Conversely, let us show that (ii) $\Rightarrow$ (i) and $B = \bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$.

Let us assume, without loss of generality, that the asset return indices are such that:

$$y_{t+1} = \left(\bar{y}'_{t+1}, \bar{\bar{y}}_{t+1}\right)'.$$

Then, the $(k \times K)$ matrix $\bar{\Lambda}$ denotes the first k rows of Λ, and:

$$V_t\left[\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}\right] = \left(\bar{\bar{\Lambda}} - B\bar{\Lambda}\right) D_t \left(\bar{\bar{\Lambda}} - B\bar{\Lambda}\right)' + \left[-B, Id_{n-k}\right]\Omega\left[-B, Id_{n-k}\right]'.$$

Therefore, the assumption of conditional homoskedasticity of $\bar{\bar{y}}_{t+1} - B\bar{y}_{t+1}$ means that MD_tM' is a constant matrix, for $M = \bar{\bar{\Lambda}} - B\bar{\Lambda}$. But, the coefficients of the matrix MD_tM' are linear combinations of the conditional variances $\sigma^2_{jt}, j = 1, \cdots K$. By assumption, such linear combinations can be constant only if all their coefficients are zero. By considering the diagonal coefficients of MD_tM' we see in particular that: $\sum_{j=1}^{K} m^2_{ij}\sigma^2_{jt}$ is constant for all i, and thus: $m_{ij} = 0$ for all i and j.

Therefore $M = 0$, that is $\bar{\bar{\Lambda}} = B\bar{\Lambda}$. Then, if $\bar{\Lambda}^+$ denotes the Moore-Penrose inverse of $\bar{\Lambda}$, we get:

$$\bar{\bar{\Lambda}}\bar{\Lambda}^+\bar{\Lambda} = B\bar{\Lambda}\bar{\Lambda}^+\bar{\Lambda} = B\bar{\Lambda} = \bar{\bar{\Lambda}}$$

and thus:

$$\Lambda\bar{\Lambda}^+\bar{\Lambda} = \begin{bmatrix} \bar{\Lambda}\bar{\Lambda}^+\bar{\Lambda} \\ \bar{\bar{\Lambda}}\bar{\Lambda}^+\bar{\Lambda} \end{bmatrix} = \begin{bmatrix} \bar{\Lambda} \\ \bar{\bar{\Lambda}} \end{bmatrix} = \Lambda.$$

Since Λ is full column rank, we conclude that

$$\bar{\Lambda}^+\bar{\Lambda} = Id_K$$

Since the rank of $\bar{\Lambda}$ cannot be larger than k, we deduce from this that $k \geq K$ and then $k = K$. Thus, $\bar{\Lambda}$ is a square invertible matrix and $B = \bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$.

Proof of Proposition 3.2

As $vec\left(B\bar{y}_{t+1}y'_{t+1}\right) = \bar{y}_{t+1}y'_{t+1} \otimes I_{n-K} vecB$, we have :

$$\begin{aligned}
\frac{\partial\Phi(B,D)}{\partial\left[(vec\ B)',(vec\ D)'\right]} &= E\left[\begin{bmatrix} 1 \\ z_t \end{bmatrix} \otimes \begin{bmatrix} \bar{y}_{t+1}y'_{t+1} \otimes I_{n-K} & -I_{n(n-K)} \end{bmatrix}\right] \\
&= -E\left[\begin{bmatrix} \bar{y}_{t+1}y'_{t+1} \otimes I_{n-K} & I_{n(n-K)} \\ z_t\bar{y}_{t+1}y'_{t+1} \otimes I_{n-K} & z_t I_{n(n-K)} \end{bmatrix}\right]
\end{aligned}$$

As $\Lambda = \begin{bmatrix} I_K \\ B \end{bmatrix}$, if we denote $\Omega_{.1} = \begin{bmatrix} \Omega_{11} \\ \Omega_{21} \end{bmatrix}$, $Ez_t = \alpha$, and $E\left(z_t D_t\right) = \Delta$ we then obtain :

$$\frac{\partial\Phi(B,D)}{\partial\left[(vec\ B)',(vec\ D)'\right]} = \begin{bmatrix} (\Lambda + \Omega_{.1})' \otimes I_{n-K} & I_{n(n-K)} \\ (\Lambda\Delta + \alpha\Omega_{.1})' \otimes I_{n-K} & \alpha I_{n(n-K)} \end{bmatrix}$$

As $(\Delta - \alpha I_K) = Ez_tD_t - Ez_tED_t$ is assumed to be invertible (diagonal matrix with non-zero diagonal coefficients $cov(z_t, \sigma^2_{kt})$), it is then straightforward to show that $\frac{\partial\Phi(B,D)}{\partial[(vec\ B)',(vec\ D)']}$ is of full column rank. Actually, if it were not the case, it would be possible to find a non zero $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ such that :

$$\begin{bmatrix} (\Lambda + \Omega_{.1})' \otimes I_{n-K} & I_{n(n-K)} \\ (\Lambda\Delta + \alpha\Omega_{.1})' \otimes I_{n-K} & \alpha I_{n(n-K)} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = 0$$

By substracting α times the first equation to the second one, we would then get :

$$\left[\left(\Lambda\left(\Delta - \alpha I_K\right)\right)' \otimes I_{n-K}\right]\mu_1 = 0$$

As $(\Delta - \alpha I_K)$ is invertible, the rank of $\Lambda\left(\Delta - \alpha I_K\right)$ is equal to K so that the rank of $\left(\Lambda\left(\Delta - \alpha I_K\right)\right)' \otimes I_{n-K}$ is equal to $K\left(n-K\right)$. The above equality would then imply $\mu_1 = 0$ and in turn $\mu_2 = 0$. This completes the proof.

Proof of proposition 3.8

$$\begin{aligned}
vec\left(y_{t+1}y'_{t+1}\right) &= vec\left(\left(\lambda f_{t+1} + u_{t+1}\right)\left(\lambda' f_{t+1} + u'_{t+1}\right)\right) \\
&= vec\left(\lambda\lambda' f^2_{t+1} + f_{t+1}\left(u_{t+1}\lambda' + \lambda u'_{t+1}\right) + u_{t+1}u'_{t+1}\right)
\end{aligned}$$

with

$$
\begin{aligned}
vec\left(u_{t+1}\lambda'\right)=vec\left(\lambda u_{t+1}'\right)' &= K_{nn}vec\left(\lambda u_{t+1}'\right)\\
&= K_{nn}u_{t+1}\otimes\lambda.
\end{aligned}
$$

so that:

$$
vec\left(y_{t+1}y_{t+1}'\right)=\lambda\otimes\lambda f_{t+1}^2+\left(I+K_{nn}\right)u_{t+1}\otimes\lambda f_{t+1}+u_{t+1}\otimes u_{t+1}
$$

and (using $D_n^+K_{nn}=D_n^+$):

$$
\begin{aligned}
vech\left(y_{t+1}y_{t+1}'\right) &= D_n^+vec\left(y_{t+1}y_{t+1}'\right)\\
&= D_n^+\left(\lambda\otimes\lambda f_{t+1}^2+2u_{t+1}\otimes\lambda f_{t+1}+u_{t+1}\otimes u_{t+1}\right).
\end{aligned}
$$

Thus, using assumption 3.6, we get:

$$
\begin{aligned}
& E_t\left[\left(vech\left(y_{t+1}y_{t+1}'\right)\right)\left(vech\left(y_{t+1}y_{t+1}'\right)'\right]\\
&= D_n^+E_t\left[\begin{array}{c}\left(\lambda\otimes\lambda f_{t+1}^2+2u_{t+1}\otimes\lambda f_{t+1}+u_{t+1}\otimes u_{t+1}\right)\\ \left(\lambda'\otimes\lambda' f_{t+1}^2+2u_{t+1}'\otimes\lambda' f_{t+1}+u_{t+1}'\otimes u_{t+1}'\right)\end{array}\right]D_n^{+\prime}
\end{aligned}
$$

$$
= D_n^+\left[\begin{array}{c}\lambda\lambda'\otimes\lambda\lambda' E_tf_{t+1}^4+E_t\left(\lambda u_{t+1}'\otimes\lambda u_{t+1}'\right)\cdot E_tf_{t+1}^2+4\Omega\otimes\lambda\lambda' E_tf_{t+1}^2\\ +E_t\left(u_{t+1}\lambda'\otimes u_{t+1}\lambda'\right)\cdot E_t\left(f_{t+1}^2\right)+\Theta\end{array}\right]D_n^{+\prime}
$$

But:

$$
\begin{aligned}
E_t\left[\lambda u_{t+1}'\otimes\lambda u_{t+1}'\right] &= E_t\left[\lambda\otimes\lambda\cdot u_{t+1}'\otimes u_{t+1}'\right]\\
&= E_t\left[\left(vec\lambda\lambda'\right)\cdot\left(vec\ u_{t+1}u_{t+1}'\right)'\right]\\
&= \left(vec\lambda\lambda'\right)\left(vec\Omega\right)'
\end{aligned}
$$

and, in the same way:

$$
E_t\left[u_{t+1}\lambda'\otimes u_{t+1}\lambda'\right]=\left(vec\ \Omega\right)\left(vec\ \lambda\lambda'\right)'
$$

Finally we get:

$$
\begin{aligned}
& E_t\left[\left(vech\left(y_{t+1}y_{t+1}'\right)\right)\left(vech\left(y_{t+1}y_{t+1}'\right)\right)'\right]\\
&= D_n^+\left[\begin{array}{l}\lambda\lambda'\otimes\lambda\lambda' E_t\left(f_{t+1}^4\right)+\Theta\\ +E_t\left(f_{t+1}^2\right)\left[\left(vec\lambda\lambda'\right)\left(vec\ \Omega\right)'+\left(vec\ \Omega\right)\left(vec\lambda\lambda'\right)'+4\Omega\otimes\lambda\lambda'\right]\end{array}\right]D_n^{+\prime}
\end{aligned}
$$

This completes the proof of proposition 3.8.

Proof of proposition 3.9

The first three conditional restrictions are merely a re-statement of proposition 3.3 in the case where $K = 1$ and $\mu = 0$. Now, if we take the conditional expectation at time $t-1$ of the conditional kurtosis of y_{t+1} as obtained in proposition 3.8, we obtain :

$$E_{t-1}\left[\left(vech\left(y_{t+1}y_{t+1}'\right)\right)\left(vech\left(y_{t+1}y_{t+1}'\right)\right)'\right]$$

$$= D_n^+ \left[\begin{array}{l} \lambda\lambda' \otimes \lambda\lambda' E_{t-1}\left(f_{t+1}^4\right) + \Theta + \\ E_{t-1}\left(f_{t+1}^2\right)\left[\left(vec\lambda\lambda'\right)\left(vec\ \Omega\right)' + \left(vec\ \Omega\right)\left(vec\lambda\lambda'\right)' + 4\Omega\otimes\lambda\lambda'\right] \end{array}\right] D_n^{+\prime}$$

As $E_{t-1}\left(\lambda\lambda' f_{t+1}^2\right) = E_{t-1}\left(y_{t+1}y_{t+1}' - \Omega\right)$, this can also be written:

$$E_{t-1}\left[\left(vech\ y_{t+1}y_{t+1}'\right)\left(vech\ y_{t+1}y_{t+1}'\right)'\right]$$

$$= D_n^+ \left[\begin{array}{l} \lambda\lambda' \otimes \lambda\lambda' E_{t-1}\left(f_{t+1}^4\right) + \Theta + \\ E_{t-1}\left[\begin{array}{c} \left(vec\left(y_{t+1}y_{t+1}' - \Omega\right)\right)\left(vec\ \Omega\right)' + \left(vec\ \Omega\right)\left(vec\left(y_{t+1}y_{t+1}'\right) - \Omega\right)' \\ +4\Omega\otimes\left(y_{t+1}y_{t+1}' - \Omega\right) \end{array}\right] \end{array}\right] D_n^{+\prime}$$

$$= D_n^+ \left[\begin{array}{l} \lambda\lambda' \otimes \lambda\lambda' E_{t-1}\left(f_{t+1}^4\right) + \Theta - 2\left(vec\ \Omega\right)\left(vec\ \Omega\right)' - 4\Omega\otimes\Omega + \\ E_{t-1}\left[\left(vec\ y_{t+1}y_{t+1}'\right)\left(vec\ \Omega\right)' + \left(vec\ \Omega\right)\left(vec\ y_{t+1}y_{t+1}'\right)' + 4\Omega\otimes y_{t+1}y_{t+1}'\right] \end{array}\right] D_n^{+\prime}$$

We then obtain:

$$D_n^+ E_{t-1}\varphi(y_{t+1},\Omega) D_n^{+\prime}$$

$$\begin{aligned} = \; & E_{t-1}\left(vech\ y_{t+1}y_{t+1}'\right)\left(vech\ y_{t+1}y_{t+1}'\right) \\ & -D_n^+ E_{t-1}\left[4\Omega\otimes y_{t+1}y_{t+1}' - \left(vec\ \Omega\right)\left(vec\ y_{t+1}y_{t+1}'\right)' - \left(vec\ y_{t+1}y_{t+1}'\right)\left(vec\ \Omega\right)'\right] \end{aligned}$$

$$= D_n^+\left[\lambda\lambda'\otimes\lambda\lambda' E_{t-1} f_{t+1}^4 + \Theta - 2\left(vec\ \Omega\right)\left(vec\ \Omega\right)' - 4\Omega\otimes\Omega\right] D_n^{+\prime}$$

In the same way, we have:

$$D_n^+ E_{t-1}\varphi(y_t,\Omega) D_n^{+\prime} = D_n^+\left[\lambda\lambda'\otimes\lambda\lambda' E_{t-1} f_t^4 + \Theta - 2\left(vec\ \Omega\right)\left(vec\ \Omega\right)' - 4\Omega\otimes\Omega\right] D_n^{+\prime}$$

so that we get:

$$D_n^+ E_{t-1}\left(1-cL\right)\varphi\left(y_{t+1},\Omega\right) D_n^{+\prime} = D_n^+ \left[\begin{array}{c} \lambda\lambda'\otimes\lambda\lambda' E_{t-1}\left(1-cL\right) f_{t+1}^4 + \left(1-c\right)\Theta - \\ 2\left(1-c\right)\left(vec\ \Omega\right)\left(vec\ \Omega\right)' - 4\left(1-c\right)\Omega\otimes\Omega \end{array}\right] D_n^{+\prime}$$

Using assumption 3.4, we know that

$$E_{t-1}\left(1-cL\right)f_{t+1}^{4}=a+bE_{t-1}f_{t}^{2},$$

so, that:

$$\begin{aligned}\lambda\lambda' E_{t-1}\left(1-cL\right)f_{t+1}^{4} &= a\ \lambda\lambda'+b\ \lambda\lambda'\ E_{t-1}f_{t}^{2}\\ &= a\ \lambda\lambda'+b\ E_{t-1}\left(y_{t}y_{t}'-\Omega\right).\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}&D_{n}^{+}E_{t-1}\left(1-cL\right)\varphi\left(y_{t+1},\Omega\right)D_{n}^{+\prime}\\ =\ & D_{n}^{+}\left[\begin{array}{l} a\ \lambda\lambda'\otimes\lambda\lambda'+b\ \lambda\lambda'\otimes E_{t-1}\left(y_{t}y_{t}'\right)-b\ \lambda\lambda'\otimes\Omega\\ +\left(1-c\right)\Theta-2\left(1-c\right)\left(vec\ \Omega\right)\left(vec\ \Omega\right)'-4\left(1-c\right)\Omega\otimes\Omega\end{array}\right]D_{n}^{+\prime}\end{aligned}$$

which is the announced result.

Extension of proposition 3.9 to the two factors case

Proposition : In a two-factors SV model :

$$y_{t+1}=\lambda_{1}f_{1t+1}+\lambda_{2}f_{2t+1}+u_{t+1}$$

such that

$$E_{t}\left[u_{t+1}u_{t+1}'\otimes u_{t+1}u_{t+1}'\right]=\Theta$$

and

$$E_{t-1}\left[f_{kt+1}^{4}-a_{k}-b_{k}f_{kt}^{2}-c_{k}f_{kt}^{4}\right]=0\ \text{, for } k=1,2,$$

efficient instrumental variables estimation of Λ, Ω, γ_k, b_k, c_k, $k=1,2$, *can be obtained through the following set of conditional moment restrictions :*

$$\begin{aligned}&E_{t}\left(y_{t+1}\right)=0\\ &vecE_{t}\left[\left(\bar{\bar{y}}_{t+1}-\bar{\bar{\Lambda}}\bar{\Lambda}^{-1}\bar{y}_{t+1}\right)y_{t+1}'\right]=vec\left[\Omega_{2.}-\bar{\bar{\Lambda}}\bar{\Lambda}^{-1}\Omega_{1.}\right]\\ &vechE_{t-2}\left[\left(1-\gamma_{1}L\right)\left(1-\gamma_{2}L\right)\left(y_{t+1}y_{t+1}'-\Lambda\Lambda'-\Omega\right)\right]=0\end{aligned}$$

$$D_{n}^{+}E_{t-4}\left[\begin{array}{l}\left(1-\gamma_{1}L\right)\left(1-\gamma_{1}\gamma_{2}L\right)\left(1-c_{1}L\right)\left(1-c_{2}L\right)\phi\left(y_{t+1},\Omega\right)-\\ \left(1-c_{2}L\right)\left(1-\gamma_{1}\gamma_{2}L\right)\left(1-\gamma_{1}L\right)b_{2}\left(vecy_{t}y_{t}'\right)\left(vec\lambda_{2}\lambda_{2}'\right)'-\\ \left(1-c_{2}L\right)\left(1-c_{1}L\right)\gamma_{2}\left(1-\gamma_{1}\right)\left[\begin{array}{l}\left(vec\lambda_{1}\lambda_{1}'\right)\left(vecy_{t}y_{t}'\right)'+\left(vecy_{t}y_{t}'\right)\left(vec\lambda_{1}\lambda_{1}'\right)'\\ +4y_{t}y_{t}'\otimes\lambda_{1}\lambda_{1}'\end{array}\right]\end{array}\right]D_{n}^{+\prime}=\text{cst}$$

$$D_n^+ E_{t-4} \left[\begin{array}{ll} (1-\gamma_2 L)(1-\gamma_1\gamma_2 L)(1-c_1 L)(1-c_2 L)\,\phi(y_{t+1},\Omega) - & \\ (1-c_1 L)(1-\gamma_1\gamma_2 L)(1-\gamma_2 L)\,b_1 (vec y_t y_t')(vec \lambda_1\lambda_1')' - & \\ (1-c_2 L)(1-c_1 L)\,\gamma_1 (1-\gamma_2) & \left[\begin{array}{l} (vec\lambda_2\lambda_2')(vec y_t y_t')' + (vec y_t y_t')(vec\lambda_2\lambda_2')' \\ +4 y_t y_t' \otimes \lambda_2\lambda_2' \end{array} \right] \end{array} \right] D_n^{+\prime} = \text{cst}$$

where:

$$\phi(y_t,\Omega) = (Vec\ y_t y_t')(Vec\ y_t y_t')' - 4\Omega \otimes y_t y_t' - (Vec\ \Omega)(Vec\ y_t y_t')' - (Vec\ y_t y_t')(Vec\ \Omega)'$$

Of course, as it has been already said about proposition 3.9, the last two sets of conditional moments restrictions involve a huge number of parameters, but it is not necessary to use this whole set of restrictions to identify and estimate the parameters of interest. For instance, only the diagonal terms of the involved matrices can be used.

Proof:

The first three conditional moment restrictions are only a re-statement of proposition 3.3. We prove here the fourth moment condition, while the fifth one is just a corollary by commuting the roles of indexes 1 and 2. But before going into the detailed proof, which involves some tedious calculations, it can be useful to sketch the intuition.

Actually, the main trick in the proof of proposition 3.9 is to compute the conditional expectation at time t of the fourth order moments of y_{t+1}, and to apply assumption 3.4 to $\lambda\lambda' f_{t+1}^2$ instead of f_{t+1}^2, which allows to use observable variables through the fact that $E_{t-1}\left(\lambda\lambda' f_{t+1}^2\right) = E_{t-1}\left(y_{t+1}y_{t+1}' - \Omega\right)$.

In the case of two (or more) factors, things are a bit more complicated because assumption 3.4 is made for each factor f_{kt+1}, while only the sum $\lambda_1\lambda_1' f_{1t+1}^2 + \lambda_2\lambda_2' f_{2t+1}^2$ can be replaced by a function of the observable variables, through the relation:

$$E_{t-1}\left(\lambda_1\lambda_1' f_{1t+1}^2 + \lambda_2\lambda_2' f_{2t+1}^2\right) = E_{t-1}\left(y_{t+1}y_{t+1}' - \Omega\right).$$

Further, the calculations which are made in this case involve the unobservable term: $\sigma_{1t}^2\sigma_{2t}^2$ and, in order to get rid of this term, it will be necessary to use its own autoregressive structure, that is an AR(1) structure with an autoregressive parameter equal to $\gamma_1\gamma_2$.

In the two factors case, the detailed proof of the result is then the following one (it can be easily extended when there are more than two factors). As we have in this case:

$$y_{t+1}y_{t+1}' = \sum_{1\le j,k\le 2} \lambda_j\lambda_k'\ f_{jt+1}\ f_{kt+1} + \sum_{k=1}^{2}\left(\lambda_k u_{t+1}' + u_{t+1}\lambda_k'\right) f_{kt+1} + u_{t+1}u_{t+1}'$$

we get, using the relations: $vec\left(u_{t+1}\lambda_k'\right) = K_{nn} vec\left(\lambda_k u_{t+1}'\right), vec\left(\lambda_1\lambda_2'\right) = K_{nn} vec\left(\lambda_2\lambda_1'\right)$ and $D_n^+ K_{nn} = D_n^+$:

$$vech\left(y_{t+1}y_{t+1}'\right) = D_n^+ vec\left[\sum_{k=1}^{2}\lambda_k\lambda_k'\ f_{kt+1}^2 + 2\lambda_1\lambda_2'\ f_{1t+1}f_{2t+1} + 2\sum_{k=1}^{2}\lambda_k f_{kt+1}u_{t+1}' + u_{t+1}\ u_{t+1}'\right]$$

and

$$E_t\left[\left(vech\ y_{t+1}y_{t+1}'\right)\left(vech\ y_{t+1}\ y_{t+1}'\right)'\right]$$
$$= D_n^+ E_t\left[\begin{array}{l} \sum\limits_{k=1}^{2}\left(vec\lambda_k\lambda_k'\right)\left(vec\lambda_k\lambda_k'\right)' f_{kt+1}^4 + \Theta + \\ \sigma_{1t}^2\sigma_{2t}^2\left[\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_2\lambda_2'\right) + \left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_1\lambda_1'\right)' + 4\left(vec\lambda_1\lambda_2'\right)\left(vec\lambda_1\lambda_2'\right)'\right] + \\ \sum\limits_{k=1}^{2}\sigma_{kt}^2\left[\left(vec\lambda_k\lambda_k'\right)\left(vec\Omega\right)' + \left(vec\Omega\right)\left(vec\lambda_k\lambda_k'\right)' + 4\left(vec\lambda_k u_{t+1}'\right)\left(vec\lambda_k u_{t+1}'\right)'\right] \end{array}\right] D_n^{+\prime}$$

As: $E_t\left[\left(vec\lambda_k u_{t+1}'\right)\left(vec\lambda_k u_{t+1}'\right)'\right] = E_t\left[\left(u_{t+1}\otimes\lambda_k\right)\left(u_{t+1}'\otimes\lambda_k'\right)\right] = \Omega\otimes\lambda_k\lambda_k'$,
and as: $\sum\limits_{k=1}^{2}\lambda_k\lambda_k'\sigma_{kt}^2 = E_t\left(y_{t+1}y_{t+1}' - \Omega\right)$, we then get:

$$E_t\left[\left(vech\ y_{t+1}y_{t+1}'\right)\left(vech\ y_{t+1}\ y_{t+1}'\right)'\right]$$
$$= D_n^+ E_t\left[\begin{array}{l} \sum\limits_{k=1}^{2}\left(vec\lambda_k\lambda_k'\right)\left(vec\lambda_k\lambda_k'\right)' f_{kt+1}^4 + \\ \sigma_{1t}^2\sigma_{2t}^2\left[\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_2\lambda_2'\right)' + \left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_1\lambda_1'\right)' + 4\left(vec\lambda_1\lambda_2'\right)\left(vec\lambda_1\lambda_2'\right)'\right] + \\ \left(vec\left(y_{t+1}y_{t+1}' - \Omega\right)\right)\left(vec\Omega\right)' + \left(vec\Omega\right)\left(vec\left(y_{t+1}y_{t+1}' - \Omega\right)\right)' + \\ 4\Omega\otimes\left(y_{t+1}y_{t+1}' - \Omega\right) + \Theta \end{array}\right] D_n^{+\prime}$$

Thus:

$$D_n^+ E_{t-1}\left[\phi\left(y_{t+1},\Omega\right)\right] D_n^{+\prime} = D_n^+ E_{t-1}\left[\sum_{k=1}^{2}\left(vec\lambda_k\lambda_k'\right)\left(vec\lambda_k\lambda_k'\right)' f_{kt+1}^4 + \sigma_{1t}^2\sigma_{2t}^2 W_{12} - W\right] D_n^{+\prime} \tag{A.12}$$

where:

$$\begin{aligned} W &= 2\left(vec\Omega\right)\left(vec\Omega\right)' + 4\Omega\otimes\Omega - \Theta \\ W_{12} &= \left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_2\lambda_2'\right)' + \left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_1\lambda_1'\right)' + 4\left(vec\lambda_1\lambda_2'\right)\left(vec\lambda_1\lambda_2'\right)' \\ &= \left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_2\lambda_2'\right)' + \left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_1\lambda_1'\right)' + 4\lambda_2\lambda_2'\otimes\lambda_1\lambda_1'. \end{aligned}$$

For $k = 1, 2$, we have: $\sigma_{kt}^2 = 1 - \gamma_k + \gamma_k\sigma_{k+1}^2 + \nu_{kt}$, $E_{t-1}\nu_{kt} = 0$,
so that:

$$E_{t-1}\left(\sigma_{1t}^2\sigma_{2t}^2\right) = \gamma_1\gamma_2\sigma_{1t-1}^2\sigma_{2t-1}^2 + \left(1-\gamma_1\right)\left(1-\gamma_2\right) + \gamma_2\left(1-\gamma_1\right)\sigma_{2t-1}^2 + \gamma_1\left(1-\gamma_2\right)\sigma_{1t-1}^2$$

and:

$$E_{t-1}\left[\left(1-\gamma_1\gamma_2 L\right)\sigma_{1t}^2\sigma_{2t}^2\right]=\left(1-\gamma_1\right)\left(1-\gamma_2\right)+\gamma_2\left(1-\gamma_1\right)\sigma_{2t-1}^2+\gamma_1\left(1-\gamma_2\right)\sigma_{1t-1}^2 \qquad \text{(A.13)}$$

Besides, we have, for $k=1,2$:

$$E_{t-1}\left[\left(1-c_k L\right)f_{kt+1}^4\right]=E_{t-1}\left[a_k+b_k f_{kt}^2\right]=a_k+b_k\sigma_{kt-1}^2 \qquad \text{(A.14)}$$

Using (A.12), (A.13) and (A.14), we then deduce:

$$\begin{aligned}
&D_n^+E_{t-3}\left[\left(1-c_1L\right)\left(1-c_2L\right)\left(1-\gamma_1\gamma_2L\right)\phi\left(y_{t+1},\Omega\right)\right]D_n^{+\prime}\\
&=D_n^+E_{t-3}\left[\begin{array}{l}
\left(1-c_2L\right)\left(1-\gamma_1\gamma_2L\right)\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_1\lambda_1'\right)'\left(a_1+b_1\sigma_{1t-1}^2\right)+\\
\left(1-c_1L\right)\left(1-\gamma_1\gamma_2L\right)\left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_2\lambda_2'\right)'\left(a_2+b_2\sigma_{2t-1}^2\right)+\\
\left(1-c_1L\right)\left(1-c_2L\right)\left[\left(1-\gamma_1\right)\left(1-\gamma_2\right)+\gamma_2\left(1-\gamma_1\right)\sigma_{2t-1}^2+\gamma_1\left(1-\gamma_2\right)\sigma_{1t-1}^2\right]W_{12}+\\
\left(1-c_1\right)\left(1-c_2\right)\left(1-\gamma_1\gamma_2\right)W
\end{array}\right]D_n^{+\prime}
\end{aligned}$$

$$=D_n^+E_{t-3}\left[\begin{array}{l}
\left(1-c_1L\right)\left[\left(1-\gamma_1\gamma_2L\right)\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_1\lambda_1'\right)'b_1+\left(1-c_2L\right)\gamma_1\left(1-\gamma_2\right)W_{12}\right]\sigma_{1t-1}^2+\\
\left(1-c_2L\right)\left[\left(1-\gamma_1\gamma_2L\right)\left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_2\lambda_2'\right)'b_2+\left(1-c_1L\right)\gamma_2\left(1-\gamma_1\right)W_{12}\right]\sigma_{2t-1}^2+M
\end{array}\right]D_n^{+\prime}$$

with

$$\begin{aligned}
M &= \left(1-c_2\right)\left(1-\gamma_1\gamma_2\right)\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_1\lambda_1'\right)'a_1\\
&\quad+\left(1-c_1\right)\left(1-\gamma_1\gamma_2\right)\left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_2\lambda_2'\right)'a_2\\
&\quad+\left(1-c_1\right)\left(1-c_2\right)\left[\left(1-\gamma_1\right)\left(1-\gamma_2\right)W_{12}+\left(1-\gamma_1\gamma_2\right)W\right].
\end{aligned}$$

Now: $E_{t-2}\left[\left(1-\gamma_1L\right)\sigma_{1t-1}^2\right]=1-\gamma_1$

and:

$$\begin{aligned}
E_{t-2}\left[\left(1-\gamma_1L\right)\lambda_2\lambda_2'\sigma_{2t-1}^2\right] &= E_{t-2}\left[\left(1-\gamma_1L\right)\left[y_ty_t'-\lambda_1\lambda_1'\sigma_{1t-1}^2-\Omega\right]\right]\\
&= E_{t-2}\left[\left(1-\gamma_1L\right)\left[y_ty_t'-\left(\lambda_1\lambda_1'+\Omega\right)\right]\right].
\end{aligned}$$

Using $W_{12}=\left(vec\lambda_1\lambda_1'\right)\left(vec\lambda_2\lambda_2'\right)'+\left(vec\lambda_2\lambda_2'\right)\left(vec\lambda_1\lambda_1'\right)'+4\lambda_2\lambda_2'\otimes\lambda_1\lambda_1'$, we get:

$$D_n^+E_{t-4}\left[\left(1-\gamma_1L\right)\left(1-c_1L\right)\left(1-c_1L\right)\left(1-\gamma_1\gamma_2L\right)\phi\left(y_{t+1},\Omega\right)\right]D_n^{+\prime}$$

$$= D_n^+ E_{t-4} \left[\begin{array}{l} \quad (1-\gamma_1)(1-c_1)\left[(1-\gamma_1\gamma_2)(vec\lambda_1\lambda_1')(vec\lambda_1\lambda_1')' b_1 + (1-c_2)\gamma_1(1-\gamma_2)W_{12}\right] \\ +(1-c_2L)\left[\begin{array}{c} (1-\gamma_1\gamma_2 L)(1-\gamma_1 L)\, vec\,(y_t y_t' - (\lambda_1\lambda_1' + \Omega))(vec\lambda_2\lambda_2')' b_2 \\ +(1-c_1L)\gamma_2(1-\gamma_1)\left[\begin{array}{c} (vec\lambda_1\lambda_1')(vec\,(y_t y_t' - (\lambda_1\lambda_1' + \Omega)))' \\ +(vec\,(y_t y_t' - (\lambda_1\lambda_1' + \Omega)))(vec\lambda_1\lambda_1')' \\ +4\,(y_t y_t' - (\lambda_1\lambda_1' + \Omega)) \otimes \lambda_1\lambda_1' \end{array} \right] \end{array} \right] \\ +(1-\gamma_1)M \end{array} \right] D_n^{+\prime}$$

$$= D_n^+ E_{t-4} \left[\begin{array}{l} (1-c_2L)(1-\gamma_1\gamma_2L)(1-\gamma_1L)\, b_2\,(vecy_t y_t')(vec\lambda_2\lambda_2')' + \\ (1-c_2L)(1-c_1L)\gamma_2(1-\gamma_1)\left[\begin{array}{l} (vec\lambda_1\lambda_1')(vecy_t y_t')' + (vecy_t y_t')(vec\lambda_1\lambda_1')' + \\ 4y_t y_t' \otimes \lambda_1\lambda_1' \end{array} \right] \\ +M_{12} \end{array} \right] D_n^{+\prime}$$

with M_{12} a constant matrix.

In the same way:

$$D_n^+ E_{t-4}\left[(1-\gamma_2L)(1-c_1L)(1-c_2L)(1-\gamma_1\gamma_2L)\,\phi\,(y_{t+1},\Omega)\right] D_n^{+\prime}$$

$$= D_n^+ E_{t-4} \left[\begin{array}{l} \quad (1-c_1L)(1-\gamma_1\gamma_2L)(1-\gamma_2L)\, b_1\,(vecy_t y_t')(vec\lambda_1\lambda_1')' + \\ (1-c_1L)(1-c_2L)\gamma_1(1-\gamma_2)\left[\begin{array}{l} (vec\lambda_2\lambda_2')(vecy_t y_t')' + (vecy_t y_t')(vec\lambda_2\lambda_2')' + \\ 4\lambda_2\lambda_2' \otimes y_t y_t' \end{array} \right] \\ +M_{21} \end{array} \right] D_n^{+\prime}$$

with M_{21} a constant matrix.

The first three conditional moment restrictions allow to identify the following parameters: γ_1, γ_2, $\Lambda\Lambda' + \Omega$ and $\bar{\bar{\Lambda}}\bar{\Lambda}^{-1}$.

The last two conditional moment restrictions allow to identify: $c_1, c_2, \Omega, \lambda_1\lambda_1', \lambda_2\lambda_2'$ and then b_1 and b_2.

Thus λ_1 and λ_2 are identified up to a sign so that Λ is identified up to a sign change in its columns. As previously, however: a_1, a_2 and Θ cannot be separately identified.

References

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

[2] Arrelano M., L.P. Hansen and E. Sentana (1999), "Underidentification ?", *Working Paper, CEMFI, Madrid.*

[3] Barndorff-Nielsen, O. E., and N. Shephard (2001), "Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics (with discussion)", *Journal of the Royal Statistical Society, Series B* 63, 167-241.

[4] Bollerslev, T. (1986), "Generalised Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 51, 307-327.

[5] Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988), "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy* 96, 116-131.

[6] Chamberlain, G., M. Rothschild (1983),"Arbitrage, factor structure and mean-variance analysis in large asset markets", *Econometrica* 51, 1305-1324.

[7] Cox, J.C., J. Ingersoll, and S. Ross (1984), "A Theory of the Term Structure of Interest Rates", *Econometrica* 53, 385-408.

[8] Dellaportas P., S. G. Giakoumatos, and D. M. Politis (1999), "Bayesian Analysis of the Unobserved ARCH Model", Department of Statistics, Athens University of Economics and Business.

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

[10] Dovonon P., C. Doz and E. Renault (2004), "Conditionally Heteroskedastic Factor Models with Skewness and Leverage Effects", *Manuscript, Université de Montréal* .

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

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

[13] Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica* 50, 987-1008.

[14] Engle, R. F. (2002), "New Frontiers for ARCH Models", *Journal of Applied Econometrics* 17 (5), 425-446.

[15] Engle, R. F., and S. Kozicki (1993), "Testing for Common Features", *Journal of Business and Economic Statistics,* Invited lecture with discussion 11(4), 369-395.

[16] Engle, R. F., V. Ng, and M. R. Rothschild (1990), "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimate for Treasury Bills", *Journal of Econometrics* 45, 213-238.

[17] Engle, R. F., V. Ng, and M. R. Rothschild (1992), "A Multi-Dynamic-Factor Model for Stock Returns", *Journal of Econometrics* 52, 245-266.

[18] Fiorentini G., and E. Sentana (2001), "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models", *Journal of Econometrics* 102, 143-164.

[19] Fiorentini, G., E. Sentana, and N. Shephard (2003), "Likelihood-Based Estimation of Latent Generalised ARCH Structures", *Econometrica*, forthcoming.

[20] Ghysels, E., A. C. Harvey, and E. Renault (1996), "Stochastic Volatility", in C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance,* 119-191, Amsterdam: North-Holland.

[21] Gouriéroux, C., A. Monfort, and E. Renault (1991), "Dynamic Factor Models", Working Paper GREMAQ, n^o 91.e.

[22] Gouriéroux, C., A. Monfort, and E. Renault (1995), "Inference in Factor Models" in *Advances in Econometrics and Quantitative Economics,* Maddala, Phillips and Srinivasan Eds, Blackwell.

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

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

[25] 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.

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

[27] 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.

[28] King, M.A, E. Sentana, and S. B. Wadhwani (1994), "Volatility and Links Between National Stock Markets", *Econometrica* 62, 901-933.

[29] Magnus, J.R., and H. Neudecker (1988), "*Matrix Differential Calculus with Applications in Statistics and Econometrics*", Wiley, Chichester.

[30] Meddahi, N., and E. Renault (2004), "Temporal Aggregation of Volatility Models", *Journal of Econometrics* 119, 355-379.

[31] Ross S.A. (1976), "The arbitrage theory of asset pricing ", *Journal of Economic Theory*, 13, 641-660.

[32] Sargan, J. D. (1983), "Identification and Lack of Identification", *Econometrica* 51, 1605-1634.

[33] Sentana, E. (1998), "The Relation between Conditionally Heteroskedastic Factor Models and Factor GARCH Models", *Econometrics Journal* 1, 1-9.

[34] Sentana, E. (2004), "Factor Representing Portfolios in Large Asset Markets", *Journal of Econometrics* 119, 257-290.

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