The Stochastic Volatility Factor Model.

M. Escobar P. Olivares.

January 29, 2009

Abstract

In this paper we study the properties of a new multidimensional process, the Stochastic Volatility Factor model. In particular, conditions for stationarity, ergodicity and mixing properties are studied following Escobar et. al. 2008. The conditional characteristic function as well as the properties of the instantaneous dependence structure are described. The calibration of the model is performed along the lines of the method of continuum of moments conditions by Carrasco 2007. The impact of this process on Risk Measures and Portfolio Theory are described.

Today

1 Introduction

Continuous time Stochastic Volatility models have been used in the pricing of financial derivatives since the beginning of the 90s; one of the best known papers in this area is the seminal work of Heston in 1993, [?]. Although these models are more realistic and explain some behaviors than classical Black-Scholes model can not explain, they become quite intractable in dealing with derivatives involving a large number of underlines. These difficulties have created an interest in continuous time multidimensional modeling during the last decade. Most work has been dedicated to discrete time models as generalizations of GARCH (see Engle 2001).

The best known representative of this effort in continuous time is the Stochastic Wishart model first studied by Bru 1991 and most recently brought
to finance by Gourieroux 2005. Even though this family of models capture several important stylized features they fail to remain simple enough for estimation and simulation purposes.

In this paper, we introduce the multivariate Stochastic Volatility Factor process (SVF) as a simplified-practical stochastic volatility model for financial stock prices. This model has several theoretical and practical strengths. First, it allows for reducing dimensionality which is a way to control the ”dimensionality curse”. This is critical when dealing with multidimensional derivatives as for example CDO, CFO and Mountain Range. As a matter of fact, other authors have used axioms of principal component analysis (as well as factor analysis) before in the context of static modelling and they are in support of the fact that few eigenvalues are sufficient to describe most of the variation in standard portfolios (see Alexander (2000), (2001)). Secondly, for these eigenvalues and principal components we suggest a system of stochastic processes for the eigenvalues allowing for several ”levels” of stochasticity: stochastic volatility for the underlying (as proved in financial data since Engle 82), stochastic correlation among stocks (see Engle 2001), variances (Heston 2007) and stocks-correlations (Da Fonseca et. al. 2007). This stochasticity has been argued to capture several stylized features of the historical data as well as of derivatives, for instance smile and skew volatilities and correlations (Da Fonseca et. al. 2007, Heston 2007).

Thirdly, a closed form expression for the joint characteristic function (useful for derivative and risk management purposes) is available. In general, at any time the infinitesimal variance of the actual process is decomposed into a reduced number of (unobservable) ”principal components”, whose eigenvalues follow orthogonal stochastic differential equations. In spite of the similarities to the Wishart family of process, we show that the SVPC is not a particular case of Wishart and therefore it creates a whole new group of processes.

We study the problem of parameter estimation for these class of processes from observations at discrete times. In particular we use empirical moments based on their ergodic properties following the lines of [?] whose results can be easily extended to several dimension processes in our framework.

The organization of the paper is the following: In the next section the continuous SVF model is introduced together with some properties of relevance for practitioners. In section 3 a discrete time variant is defined together with the definition of hidden Markov process provided under a multivariate setting. Then it is shown that our discretely observed process satisfy the
conditions of a hidden Markov chain. Thanks to the ergodicity and mixing conditions of the unidimensional volatilities, we can imply the same for the process of the increments. Some numerical results from simulated and real data are provided in section 4.

2 Stochastic Volatility Factor Process.

Definition 2.1. A m-dimensional stochastic process \( Y(t) \) follows a SPCM if it satisfies the following diffusion:

\[
\begin{align*}
    dY_i(t) &= \left[ \mu_i - \frac{1}{2} \sum_{j=1}^{p} a_{i,j}^2 V_j(t) - \frac{1}{2} V^{(i)}(t) \right] dt + \sum_{j=1}^{p} a_{i,j} \sigma_j(t) dW_j(t) + \sigma^{(i)}(t) dW^{(i)}(t) \\
    dV_j(t) &= \frac{c_j}{2} (\beta_j - V_j(t)) dt + c_j \sqrt{V_j(t)} dB_j(t) \\
    dV^{(i)}(t) &= \frac{c^{(i)}}{2} (\beta^{(i)} - V^{(i)}(t)) dt + c^{(i)} \sqrt{V^{(i)}(t)} dB^{(i)}(t)
\end{align*}
\]

for \( i = 1, 2, \ldots, m \). Here \( W_i(t) \) and \( B_i(t) \) are Brownian motions for every \( i \) with quadratic variations: \( < W_i(t), B_j(t) >= \delta_{ij} \rho_i \), \( < W_j(t), W^{(i)}(t) >= < B_j(t), B^{(i)}(t) >= 0 \). \( V_i(t) = \sigma_i^2(t) \) with initial conditions \( V_i(0) = \eta_i \), \( Y_0 = 0 \).

Here \( \sigma_i(t) \) and \( \alpha_i = (\alpha_{i,1}, \ldots, \alpha_{i,p}) \) are the eigenvalues and the eigenvector of the instantaneous covariance matrix \( \Sigma(t, dt) \) of the underline process after removing the intrinsic factor influence: \( \sigma^{(i)}(t) dW^{(i)}(t) \).

We denote by \( A = (\alpha)_{ij} \) a \( m \times p \) dimensional matrix. Without lost in generality we assume \( 0 < \beta_j < \beta_{j-1} \), for \( i = 1, 2, \ldots, p \).

Note that for any fixed \( t > 0 \) the process decomposes into \( p \) orthogonal directions given by eigenvectors \( \alpha_i \) non depending on \( t \). Moreover, the number of parameters is \( p(2 + m) + m \), which could be fairly small compared with number of parameters in the complete process of returns. For example if we consider \( m = 100 \) and \( p = 5 \) it will give 610 parameters instead of the original 10300.

The main ingredient of this multivariate process are a family of unidimensional stochastic processes for the eigenvalues. We assume for simplicity
Heston-type processes but the approach works for other kind of processes. The condition $\beta > 0$ ensures stationarity, ergodicity and mixing conditions for the unidimensional processes $V_i$ and later on it is shown that this implies the same features on the increments.

**Remark 2.2.** We will see that if correlation is assumed $dB_i(t)dW_j(t)$ ($i \neq j$) or $dB_i(t)dB_j(t)$ then there is no closed form expression for the characteristic function. On the other hand, the model still allows for correlation on $dW_i(t)dW_j(t)$ and $dB_i(t)dW_i(t)$. Allowing correlation on $dB_i(t)dW_i(t)$ then several type of correlations are achieved, i.e. between the driver of $V$ and the driver of the volatility of $V$, the driver of $V$ and the driver of the correlation of $V_i, V_j$.

**Remark 2.3.** This model is not a particular case of Gourieroux (2005) or Da Fonseca et. al. 2007. It is a generalization of Escobar and Olivares 2008.

We give some properties of the SVF changing slightly our notations: We begin by assuming the following diffusions for our underlines (spot prices, volatilities, correlations) under a risk neutral measure:

$$S_j(t) = \exp\{V_j(t)\}$$

for $j = 1, ..., n$.

**Proposition 2.4.** The log-assets exhibit stochastic volatility and correlation:

$$\sigma_{Y_j}(dt) = d\langle Y_j, Y_j \rangle = \sum_{i=1}^{p} a_{j,i}^2 V_i(t) + V^{(j)}(t) \ dt$$  \hspace{1cm} (4)

$$\rho_{(Y_j,Y_k)}(t) = \frac{d\langle Y_j,Y_k \rangle}{\sqrt{d\langle Y_j,Y_j \rangle \ d\langle Y_k,Y_k \rangle}}$$  \hspace{1cm} \hspace{1cm} (5)

$$\sqrt{\sum_{i=1}^{p} a_{j,i}^2 V_i(t)} \sqrt{\sum_{i=1}^{p} a_{k,i}^2 V_i(t)}$$

$\sqrt{\sum_{i=1}^{p} a_{j,i}^2 V_i(t) + V^{(j)}(t)} \sqrt{\sum_{i=1}^{p} a_{k,i}^2 V_i(t) + V^{(k)}(t)}$
Correlations between stock-volatility, volatility-volatility and correlation-stock respectively are:

$$\rho_{Y_i,\sigma_i}(t, dt) = \frac{d\langle Y_i, \sigma^2_i \rangle}{\sqrt{d\langle Y_i \rangle} \sqrt{d\langle \sigma^2_i \rangle}} = \frac{\sum_{j=1}^{p} a_{i,j}^3 c_j V_j(t) \rho_j}{\sqrt{\sum_{j=1}^{p} a_{i,j}^2 V_j(t)} + V^{(i)}(t) \sqrt{\sum_{j=1}^{p} a_{i,j}^4 c_j^2 V_j(t) + (c^{(i)})^2 V^{(i)}(t)}}$$

$$\rho_{\sigma_i,\sigma_j}(t, dt) = \frac{d\langle \sigma^2_i, \sigma^2_j \rangle}{\sqrt{d\langle \sigma^2_i \rangle} \sqrt{d\langle \sigma^2_j \rangle}} = \frac{\sum_{k=1}^{p} a_{i,k}^2 a_{j,k}^2 c_k^2 V_k(t)}{\sqrt{\sum_{k=1}^{p} a_{i,k}^4 c_k^2 V_k(t) + (c^{(i)})^2 V^{(i)}(t)} \sqrt{\sum_{k=1}^{p} a_{j,k}^4 c_k^2 V_k(t) + (c^{(j)})^2 V^{(j)}(t)}}$$

$$\rho_{Y_i,\rho_{i,j}}(t, dt) = \frac{d\langle Y_i, \rho_{Y_i,Y_j} \rangle}{\sqrt{d\langle Y_i \rangle} \sqrt{d\langle \rho_{Y_i,Y_j} \rangle}} = \frac{\sum_{k=1}^{p} a_{i,k} c_k \frac{\partial \rho_{Y_i,Y_j}}{\partial V_k} V_k(t) \rho_k}{\sqrt{\sum_{k=1}^{p} a_{i,k}^2 V_k(t) + (c^{(i)})^2 V^{(i)}(t)} \sqrt{\sum_{k=1}^{p} c_k V_k^{1/2} + c^{(i)} \frac{\partial \rho_{Y_i,Y_j}}{\partial V^{(i)}} V^{(i)} + c^{(j)} \frac{\partial \rho_{Y_i,Y_j}}{\partial V^{(j)}} V^{(j)}}}$$

**Proposition 2.5.** For any $t > 0$ the joint characteristic function of the vector of the log-assets $Y$ given by the model (??-??) satisfies the following
PDE:
\[
0 = \frac{\partial f}{\partial t} + \sum_{j=1}^{m} \frac{\partial f}{\partial y_j} \left( \mu_j - \sum_{i=1}^{p} a_i^2 V_i(t) - \frac{V^{(i)}(t)}{2} \right)
+ \frac{1}{2} \sum_{j=1}^{m} \left( \sum_{i=1}^{p} a_{j,i}^2 V_i(t) \right) + \frac{m}{j>k} \left( \sum_{i=1}^{p} a_j V_i(t) \right) + \frac{1}{2} \sum_{i=1}^{p} a_i^2 c_i^2 V_i(t)
+ \sum_{j=1}^{m} \left[ \frac{1}{2} \sum_{i=1}^{p} \frac{\partial^2 f}{\partial y_i \partial y_j} \rho_i a_{j,i} c_i^2 V_i \right] + \sum_{j=1}^{p} \left( \sum_{i=1}^{p} \frac{\partial^2 f}{\partial y_i^2} \left( \beta_j - V^{(j)} \right) + \frac{1}{2} \frac{\partial^2 f}{\partial v^{(j)}^2} \right) \left( V^{(j)} \right)^2
f(\phi;Y,V,T) = \exp \{ i\phi Y \}
\]
the last expression being the boundary condition.
It has an affine solution given by:
\[
f(\phi; y, v, t) = \exp \{ C(T - t) + D(T - t)v + i\phi y \}
\]
with coefficients \( C, D \) given in (??).

Another important element of this model that will be used later on is the infinitesimal generator of \( (Y, V) \) and that of \( V \) which are respectively (see [?]):
\[
\Xi_{Y,V} f(y, v) = \sum_{i=1}^{m} \mu^Y_i(v_i) \frac{\partial f}{\partial y_i} + \sum_{i=1}^{p} \mu^Y_i(v_i) \frac{\partial f}{\partial v_i} + \sum_{i=1}^{m} \mu^{(i)}_i(v^{(i)}) \frac{\partial f}{\partial v^{(i)}}
+ \sum_{1 \leq i, j \leq n} \left( \sum_{k=1}^{p} a_{i,k} a_{j,k} v_k \right) \frac{\partial^2 f}{\partial y_i \partial y_j} + \sum_{1 \leq i \leq m; 1 \leq j \leq p} a_{i,j} v_j \beta_j \frac{\partial^2 f}{\partial y_i \partial v_j}
+ \sum_{1 \leq j \leq p} v_j c_j \frac{\partial^2 f}{\partial v^2_j} + \sum_{1 \leq j \leq p} v^{(j)} c^{(j)} \frac{\partial^2 f}{\partial v^{(j)}^2}
\]
\[
\Xi_{V} g(v) = \sum_{i=1}^{p} \mu^V_i(v_i) \frac{\partial g}{\partial v_i} + \sum_{1 \leq i \leq p} c_i^2 v_i \frac{\partial^2 g}{\partial v^2_i} + \sum_{i=1}^{n} \mu^{(i)}_i(v^{(i)}) \frac{\partial g}{\partial v^{(i)}} + \sum_{1 \leq i \leq p} c^{(i)}_i v^{(i)} \frac{\partial^2 g}{\partial v^{(i)}^2}
\]
where \( \mu^Y_i \) and \( \mu^V_i \) are the drifts of equations (??) and (??) respectively.
Remark 2.6. It is easy to see that the vector of eigenvalues $V$ is strictly stationary. This comes from the independence and the strict stationarity of each eigenvalue process, therefore the stationary distribution for the vector is the product of their corresponding marginals distribution. Another feature of this model that will be used later on is the infinitesimal generator of $(Y, V)$ and that of $V$ which are respectively:

Theorem 2.7. For the process in (??), we have that $V = (V_1, \ldots, V_p, V^{(1)}, \ldots, V^{(m)})$ is $\rho$-mixing and therefore $\alpha$-mixing and ergodic.

The next result uses proposition ?? to show how a multidimensional financial derivative with maturity time $T$ and payoff:

$$\prod_{i=1}^{m} \left( \exp \{ Y_i(T) \} - \exp \{ k_i \} \right)^+$$

(9)

can be priced, in closed form, under a PCSV model. In general the idea can be extended to other multidimensional derivatives like, for example, the “best of” $G(T) = \left( \max_i \{ S_i(T) \} - K_i \right)^+$.

Proposition 2.8. Let $f$ be the characteristic function defined in (??) and $f^*$ the characteristic function of the derivative price with payoff given by (??), then the following holds for any positive vector $\alpha$, any real vector $k$ and a fixed interest rate $r$:

$$f^*(s) = \frac{e^{-i\alpha^T s} f(s_1 - \alpha_1 i, \ldots, s_m - (\alpha_m + 1)i)}{\prod_{j=1}^{n} (\alpha_j + i s_j)(\alpha_j + 1 + i s_j)}$$

Moreover its price is:

$$\frac{e^{-i\alpha^T k}}{(2\pi)^m} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} e^{-i\alpha^T k} f^*(s) ds$$
3 Stationarity, ergodicity and mixing properties of a SVF process

The definition of a Hidden Markov Chain (HMC) is provided in a multivariate setting. Then it is shown that a discretely defined SVPC process satisfies the conditions of a HMC, this discrete process is related to the observable process. Thanks to the ergodicity and mixing conditions of the unidimensional volatilities, it is possible to imply the same properties for the process of the increments.

Definition 3.1. For $h > 0$ we define, for $1 \leq i \leq n$ and $k = 1, 2, \ldots, p$ the following processes:

$$Z_{i,k} = \mu_k \sqrt{h} - \frac{1}{2 \sqrt{h}} \sum_{j=1}^{p} a_{k,j}^2 \int_{(i-1)h}^{ih} V_j(t) dt - \frac{1}{2 \sqrt{h}} \int_{(i-1)h}^{ih} V^{(j)}(t) dt$$

$$+ \frac{1}{\sqrt{h}} \sum_{j=1}^{p} a_{kj} \int_{(i-1)h}^{ih} \sigma_j(t) dW_j(t) + \frac{1}{\sqrt{h}} \int_{(i-1)h}^{ih} \sigma^{(j)}(t) dW^{(j)}(t)$$

$$\Delta V_i = V(ih) - V((i-1)h)$$

$$U_i = (\bar{V}_i, \Delta V_i)$$

$$\bar{V}_i = \left( \frac{1}{h} \int_{(i-1)h}^{ih} V_1(s) ds, \ldots, \frac{1}{h} \int_{(i-1)h}^{ih} V_p(s) ds, \ldots, \frac{1}{h} \int_{(i-1)h}^{ih} V^{(n)}(s) ds \right)$$

This definition is useful to achieve a non zero random variable for all $h$. $Z$ is related to $Y$ but the expression $Z_i = \frac{Y_{ih} - Y_{(i-1)h}}{\sqrt{h}}$.

Lemma 3.2. Conditionally on $(V_s, s \geq 0)$, for any $1 \leq i \leq n$ the random vector $(Z_i)$ with components $Z_{ik}$ distributes normally with mean:

$$\mu \cdot \sqrt{h} + \frac{1}{\sqrt{h}} \left( -\frac{1}{2} (A \circ A) \bar{V}_i + A \left[ (\frac{\Delta V_i}{c} - \frac{1}{2} (\beta - \bar{V}_i) \circ \rho) \right] \right)$$

and variance

$$A \left[ dg \left( \bar{V}_i, (1 - \rho_t^2) \right) \right] A' + dg \left( \bar{V}_i \right).$$
Here $\mu$ is a vector of components $\mu_k$'s, $A \circ B$ indicates the Hadamard product and $\text{diag}(A)$ is the diagonal matrix associate with A.

The concept of a Hidden Markov Chain is extended in straightforward to the multivariate case:

**Definition 3.3.** A stochastic vector process $(X_n, n \geq 1)$, with state space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, is a hidden Markov model if the following conditions hold:

(i) (Hidden chain) We are given a strictly stationary Markov chain $T_1, T_2, ..., T_n, ...$ (here $T_i = (T_{i,1}, ..., T_{i,2p})$ is a vector of $2p$ independent components) with state space $(\mathbb{R}^{2m}, \mathcal{B}(\mathbb{R}^{2m}))$.

(ii) For all $n$, given $(T_1, T_2, ..., T_n)$, the vectors $X_i, i = 1, ..., n$, are conditionally independent, and the conditional distribution of $X_i$ depends only on $T_i$.

(iii) The conditional distribution of $X_i$ given $T_i = u$ does not depend on $i$.

**Theorem 3.4.** For the process in ??, we have:

(i) $(U_i, i \geq 1)$ is a strictly stationary, ergodic and $\alpha$–mixing, Markov vector chain with state-space $(0, +\infty)^{2p}$;

(ii) $(Z_i, i \geq 1)$ is a hidden Markov model with hidden chain $(U_i, i \geq 1)$.

(iii) The process $(Z_i, i \geq 1)$ is strictly stationary, ergodic and $\alpha$-mixing, with $\alpha_Z(k) \leq \alpha_U(k)$, $\alpha_U(k) \leq \alpha_V((k-1)h)$, where $\alpha_X$ is the coefficient corresponding to process $X$.

### 4 Estimation of the SVF model.

First, some comments regarding estmation: Factor models are not easy to estimate because Available methods for handling communality are based on a recursive estimation of both factor’s loading ($A$) and communality, where iterations are performed till the communality values converged (we describe in the appendix the most standard ones, Principal Factor Method, Centroid Method and Maximum Likelihood). It is known that for large covariance matrixes (more than 50 variables), . . . ve iterations bring the communalities closed enough to converged values.
It is important to realize that "there is no solution to the communality problem" (from Comrey 1996, page 76), "there is no way of obtaining the correct communalities, even converged communality values are not necessarily correct", this means that there may be serious errors in the estimation of communality values.

In this section, we detail the estimation strategy that we adopt for the SVPC model. First, we briefly review the classical GMM procedure based on the characteristic function, before extending it the case of a continuum of moment conditions. Finally, we discuss the estimation of the covariance matrix of the C-GMM estimators. In this section, we thoroughly follow Carrasco et al. (2007).

4.1 Integrating out volatility

In the SVF case, we need to deal with an additional difficulty when using conditional characteristic function-based estimation methods, since the covariances are not directly observed. Thus, the usual moment condition cannot be conditioned upon $V(t)$. Even if some authors propose to use the integrated volatility as a proxy for the volatility process (see e.g. Bollerslev and Zhou (2002) and Garcia et al. (2007)), Chacko and Viceira (2003) and Rockinger and Semenova (2005) proposed to integrate the volatility process out from the expectation expression in equation (10). In the SVF case, such calculations may be complex and may increase the numerical complexity of the problem we are faced with. Therefore, we propose to condition the characteristic function on $V(0)$ instead of $V(t)$. $V(0)$ will have to be estimated separately. Now, the moment conditions should involve the following "forward characteristic function":

$$
Ψ_{V(0),Y(t)}(τ, w, t) = E[\exp\{i\langle w, Y(t + τ) - Y(t)\rangle\} \mid V(0), Y(t)]
$$

where $Y(t) = \log S(t)$, and $S(t)$ is the vector containing the price of the financial assets at time $t$ and $V(0)$ the initial co-volatility matrix that is assumed to be observable. It is important to remark that $τ$ measures the sampling frequency of the dataset. The log-returns are thus computed as $r(t + τ) = Y(t + τ) - Y(t)$. 

10
More precisely, it is now possible to write previous equation as:

\[ \Psi_{V(0),Y(t)}(\tau, w, t) = E[\exp \{i \langle w, Y(t + \tau) - Y(t) \rangle \} \mid V(0), r(t)] \]

**Proposition 4.1.** The forward characteristic function \( \Psi_{V(0),Y(t)}(\tau, w, t) \) satisfies the following equation:

\[ \Psi_{V(0),Y(t)}(\tau, w, t) = e^{C(\tau, w)} \prod_{i=1}^{p} f(D_i(\tau, w), V_i(0), t) \prod_{i=1}^{n} f(D^{(i)}(\tau, w), V^{(i)}(0), t) \]

where

\[
\begin{align*}
    f(\phi, V(0), t) &= E[\exp \{i\phi V(t)\} \mid V(0)] \\
    &= \exp \{b(t, i\phi) + a(t, i\phi)V(0)\} \\
    b &= \frac{2\gamma\beta}{c^2} \log \left\{1 + i\phi \frac{c^2}{2\gamma} \left(1 - \exp \{-\gamma t\}\right)\right\} \\
    a &= \frac{i\phi \exp \{-\gamma t\}}{1 + \frac{c^2\phi}{2t} \left[1 - \exp \{-\gamma t\}\right]} \\
\end{align*}
\]

### 4.2 The Generalized Method of Moments

Let

\[ \Psi(w_j) = E[\exp \{i w_j X_{t+\tau}\} \mid F_t] \]

be the characteristic function associated to the stochastic process \((X_t)_{t \in \mathbb{Z}}\) conditionally upon the information available at time \(t\), with \(w_j \in \mathbb{R}^d\). Let us now define

\[ g(w_j, X_{t+\tau}) = \exp \{i w_j X_{t+\tau}\} - \Psi(w_j) \]

Thus, the expectation of this equation is 0. Provided that \((X_t)_{t \in \mathbb{Z}}\) is a strictly stationary process, it is then possible to transform the moment condition defined in (28) into an estimation tool: the expectation can be computed using the strong Law of Large Numbers.
When the characteristic function $\Psi(w) \theta$ is parameter dependent – denoted $\Psi(w_j, \theta)$, where $\theta$ is a vector of parameters defining the distribution of $(X_t)_{t \in \mathbb{Z}}$ – and when the model is identified, the relation

$$E[g(w_j, X_{t+\tau})] = 0$$

should only be verified when $\theta = \theta_0$, with $\theta_0$ is the true parameters involved in the data generating process. When the system is over identified – i.e. when $w \in \mathbb{R}^k \times \mathbb{R}^d$ and $\theta \in \mathbb{R}^p$ with $kd >> p$– the parameters can be estimated by minimizing a quadratic criterion such as

$$Q(\theta) = \hat{G}(w, \theta)W^{-1}\hat{G}(w, \theta)$$

where $\hat{G}(w, \theta) = (\hat{G}(w_1, \theta), ..., \hat{G}(w_k, \theta))$ and W a symmetric positive semi definite matrix. The latter method is known as Spectral Generalized Method of Moments, see e.g. Chacko and Viceira (2003). When the number of moment conditions becomes infinite, the method is named GMM with a Continuum of Moment Conditions (C-GMM hereafter), after Carrasco and Florens (2000) and Carrasco et al. (2007). For a documented review on the GMM approaches we refer to Hall (2005).

**Remark 4.2.** We could try this approach taking:

$$\Psi(w) = \Psi(\omega, \gamma(t))(\tau, w_j)$$

we could even assume $\gamma(0)$ as part of the parametric space or simply take a proxy based on the history?.

### 4.3 Using a continuum of moment conditions.

There are two problems with the classical GMM approach: the first one is that the estimates obtained are not as efficient as the ones obtained in a standard Maximum Likelihood (ML) approach. A comparison of the estimators can be found in Zhou (2000). The second one is the choice of the value $w$ in equation (30): for example Chacko and Viceira (2003) chose an integer valued $w$, but any choice of this value could be model dependent.
One possibility consists in using an infinite number of moment conditions, as detailed in Carrasco and Florens (2000) and Singleton (2001). Nevertheless, two new difficulties arise in this case: on one hand the variance covariance matrix of the estimates may be singular. On the other hand, the optimal instrument to be used to reach the ML efficiency depends on the unknown probability distribution function (see Singleton (2001)).

Carrasco et al. (2007) proposed an elegant solution that circumvents all these problems, using a continuum of moment conditions. They showed that the CGMM is asymptotically efficient, provided that we use a particular double indexed instrument as illustrated in the following subsection.

The (conditional) moment condition required in the CGMM method for a given \( w_s \in \mathbb{R}^n \) involves the conditioning with respect to the information up to time \( t \). In order to span this information it is customary to introduce a so called instrument, which could be any function of \( F_t \)-measurable random variables. Given an instrument \( m(w_r, r_t) \) with \( w_r \in \mathbb{R}^n \), the moment function used in the estimation can be written as follows:

\[
h_t(w_s, w_r, \theta) = (e^{i(w_s, r_{t+\tau})} - \Psi_{V(o)}(\tau, w_s, t)) \ m(w_r, r_t) \tag{14}
\]

with the scalar product and \( \tau \) the vector-sized log-returns at time \( t \) over a period of a length equal to \( \tau \). Carrasco et al. (2007) showed that the subclass of instruments

\[
m(w_r, r_t) = e^{i(w_r, r_t)}
\]

leads to efficient estimates. Unlike the GMM case, the optimal instruments do not depend anymore on the score function, which is unknown in our case. See Carrasco et al. (2007) for a complete discussion of this point.

**Remark 4.3.** In the Wishart case, Fonseca restrict to the case where \( w_r = w_s = w \) – that is the simple index case – given that in the two assets case, \( w \) already contains two elements. Were \( w_r \) to be different from \( w_s \), we would face a 4-dimensional integration problem that is too time consuming for estimation purposes. He claims that following Carrasco et al. (2007), the efficiency of the double index case comes from the certainty that the function basis used to create the instruments spans the unknown score function. With the simple index procedure, this certainty vanishes but its possibility remains: it is still possible but unsure that the estimation method is still efficient. He claim the simple index estimation method stands a good chance to be the best estimation...
method for the WASC model, given that the estimators are consistent and as efficient as possible.

Next we use the dimension reduction feature of our factor to work with the double index case and therefore ensure efficiency.

Let now \( \hat{h}_t(.) \) be the sample mean of the moment condition specified in equation ??, that is a function from \( R^{2n} \) to \( \mathbb{C} \). In an infinite conditions framework, Carrasco et al. (2007) showed that the objective function to minimize is:

\[
\hat{\theta} = \arg \min_{\theta} \left\| K^{-1/2} \hat{h}(\theta) \right\|
\]

where \( K \) is the covariance operator, that is the counterpart of the covariance matrix in finite dimension – as in standard GMM approach and \( \|\cdot\| \) is the weighted norm.

\[
\|f\| = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(w)f(w')\pi(w)dw
\]

where \( \pi \) denotes any probability measure (we will take it as the independent \( 2n \) Gaussian distribution). Carrasco et al. (2007) showed that the covariance operator \( K \) can be written as follows:

\[
Kf(w) = \int_{\mathbb{R}^n} \int k(w,z)f(z)\pi(z)dz
\]

where the function \( k \) is the so called kernel of the integral operator \( K \) and is defined by:

\[
k(w,z) = \sum_{j=-\infty}^{\infty} E_{\theta_0} \left( h_t(w,\theta_0)\overline{h_{t-j}(z,\theta_0)} \right)
\]

In order to construct an estimator of the covariance operator, Carrasco et al. (2007) proposed a two-step procedure. The first step consists in finding:

\[
\theta_{2c6_1} = \arg \min_{\theta} \left\| \hat{h}(\theta) \right\|
\]

The next proposition shows that the large dimensional integration of this approach could be reduced by using the Factor structure.
Proposition 4.4. The weighted norm defined in ... becomes a lower dimensional integral for the factor model:

The second step consists in estimating the kernel. In the WASC case, the \((h_t)_{t \in R}\) consist in a martingale difference sequence, which make the computation of the covariance operator a lot easier:

\[ k(w, z) = E^{\theta_0} (h_t(w, \theta_0) h_t(z, \theta_0)) \]

The latter expectation can be estimated by the usual moment estimator. It is necessary to remark that when neglecting the previous simplification, the estimation is numerically heavier, but similar.

Once the covariance operator is estimated, the minimization in (39) requires the computation of the inverse of \(K\). Unfortunately, \(K\) has typically a countable infinity of eigenvalues decreasing to zero, so that its inverse is not bounded. We need then to regularize the inverse of \(K\), which can be done by replacing \(K\) by a nearby operator that has a bounded inverse, due to the presence of a penalizing term. Carrasco et al. (2007) used the Tikhonov approximation of the generalized inverse of \(K\): let \(\alpha\) be a strictly positive parameter, then \(K^{-1}\) is replaced by:

\[ (K^{\alpha})^{-1} = (K^2 + \alpha I)^{-1} K \]

As outlined in Carrasco et al. (2007), the choice of \(\alpha\) is important but does not jeopardize the consistency of the estimates. Carrasco and Florens (2000) investigated an empirical method to select its value, and the optimal value for it should represent a tradeoff between the instability of the generalized inverse (for small values of \(\alpha\)) and the distance from the true inverse as \(\alpha\) increases. Furthermore we found much more convenient to compute \((K^{\alpha})^{-1}\) using the Cholesky’s decomposition than the spectral decomposition: it is sufficient for the evaluation of (39) and avoids the numerically difficult problem of eigenvectors computation.

Under mild regularity conditions (conditions A.1. to A.5. in Carrasco et al. (2007)), Carrasco et al. (2007) proved that the optimal C-GMM estimator of 3b8 is obtained by:

\[ \hat{\theta} = \arg \min_{\theta} \left\| (K_T^{\alpha})^{-1/2} \tilde{h}_T(\theta) \right\| \]
(putting all the pieces together, that leads to a very unfriendly objective function) and is asymptotically Normal with:

\[ \sqrt{T}(\hat{\theta}_T - \theta_0) \to N(0, \left( \left\langle E^{\theta_0}(\nabla_{\theta} h) , E^{\theta_0}(\nabla_{\theta} h) \right\rangle_K \right)^{-1}) \]

as T and \( T^{\alpha}T^{5/4} \) go to infinity and \( \alpha \) goes to zero. (\( \nabla_{\theta} h \) denotes the Jacobian matrix of \( h(.) \)).

Finally, it is important to mention that Carrasco et al. (2007) present a matrix-based version of their estimation method that may be more appealing than the one presented here for a WASC model based on more than two assets or for other models.

There is an alternative approach in section 3.3 of Carrasco which requires a kernel \( \omega \) with certain properties (page 537). The good thing is that here is more clear that we basically need to get \( \int \int \hat{h}_t(w)\hat{h}_t(w)\pi(w)dw \) using fewer integrals.

Assumption A.7. The stochastic process \( X_t \) is a px1-vector of random variables. It is stationary and \( \alpha \)-mixing with coefficients \( \alpha_j \) that satisfy \( \sum_{j=1}^{\infty} j^2 \alpha_j < \infty \). The conditional pdf of \( X_{t+1} \) given \( X_t, f_{\theta}(x_{t+1} | x_t; \theta) \) is indexed by a finite dimensional parameter \( \theta \in \Theta \) and \( \Theta \) is compact. \( f_{\theta}(x_{t+1} | x_t; \theta) \) is continuously differentiable with respect to \( \theta \).

Assumption A.2. \( \pi \) is the pdf of a distribution that is absolutely continuous with respect to Lebesgue measure on \( \mathbb{R}^d \) and admits all its moments. \( \pi(t) > 0 \) for all \( t \in \mathbb{R}^d \).

Assumption A.3. The equation \( E^{\theta_0}(h_t(t, \theta)) = 0 \) for all \( t \in \mathbb{R}^d ; \pi \)-almost everywhere has a unique solution \( \theta_0 \) which is an interior point of \( \Theta \). \( E^{\theta_0} \) denotes the expectation with respect to the distribution of \( Yt \) for \( \theta = \theta_0 \).

Assumption A.8 (i) For all the following inequality holds:

\[ E^{\theta_0} \left[ \left( 1 - \frac{f_{\theta}(x_{t+1} | x_t; \theta)}{f_{\theta_0}(x_{t+1} | x_t; \theta_0)} \right)^2 \right] < \infty \]

and there exists a neighborhood \( N \) about \( \theta_0 \) such that:

\[ E^{\theta_0} \left[ \frac{\nabla_{\theta} f_{\theta}(x_{t+1} | x_t; \theta) \nabla_{\theta} f_{\theta}(x_{t+1} | x_t; \theta)'}{f_{\theta_0}(x_{t+1} | x_t; \theta_0)^2} \right] < \infty \]
for all $\theta$ in $\mathbb{N}$.

(ii) 

(iii).

See Carrasco page 551-554 for more detail.

Note that for A7, $\sum_{j=1}^{\infty} j^2 \alpha_j (\sigma(O_0), \sigma(O_{j\Delta}))$, but we know that $\beta_O(t) = O(e^{-at})$, so the fact that $O$ is beta-mixing is enough. For A2, note that we know the stationary of the wishart process (next remark) but we need that of the increments (still working on that). As for A# and A8 they are difficult to verify (see page 553, 570 where these two are verified in the context of a CIR).

Remark 4.5. Note that Carrasco paper is for discrete observations and she offers no comments on the property of the estimators when $\Delta t \to 0$.

Using the laplace transform of the covariance matrix given in Gourieroux 2005, page 15, we can show that the stationary distribution of the wishart process (as long as $A$ is definite positive) is multidimensional gamma as described by Krishnamoorthy and Parthasarathy in the book by Kotz et al. 2000.

$$\lim_{h \to \infty} \exp \left\{ Tr \left[ M(h) \Gamma (I - 2\Sigma(h)\Gamma)^{-1} M(h)Y_t \right] \right\} = \frac{1}{[\det (I - \Gamma + A^{-1}\Gamma)]^{K/2}}$$

In order to compute the stationary distribution of the stocks' increments, we need their Laplace transform. Da Fonseca 2008 (page 14, file: Da Fonseca - Estimating Wishart) gives the joint characteristic function of increments and covariance matrix (Wishart). I think we could find the stationary conditional distribution by means of computing the limit on the joint laplace.

By the way, we could also prove that the wishart process is mixing, besides going straight to the increments as we do in this paper.

5 Numerical Estimation Examples.

5.1 Multidimensional Examples

We first start this section by showing empirical evidence of few nonzero principal components. Then the recovery of parameters via simulated processes
is performed in order to show the reasonable levels of errors implied by the method of moments.

The following graph shows the explanatory power of each principal component as well as the importance of the respective intrinsic factor on the returns from the companies in SP100, for the period 1983-2006. Less than 7 components explain 97% of the total variance and the factor is almost unimportant.

For this graph, the x-axis is the principal component and the y-axis is the value of the total explain beta parameters (expected value of the total variance). This is evidence of the benefits of working with few principal components.

Next we show some multidimensional runs with zero drift. Model:

\[ dY_i(t) = \sum_{j=1}^{p} a_{i,j} \sigma_j(t) dW_j(t) \]  

\[ dV_j(t) = \frac{c_j}{2} (\beta_j - V_j(t)) \, dt + c_j \sqrt{V_j(t)} dB_j(t) \]  

\[ E[dW_i dB_j] = 0 \]

Here, \( i = 1, ..., m, m = 40, p = 3 \), the parameters \( (A_{nxp}, \beta_{px1}, c_{px1}) \) where taken randomly, while the measures used for assessing the quality of the estimation are given percentage-wise:

\[ E_A = \frac{\sum_{i,j} |a_{ij} - \hat{a}_{ij}|}{\sum_{i,j} |a_{ij}|} \]

\[ E_{\beta,c} = \sum_{j=1}^{p} \left[ \left| \frac{\beta_j - \hat{\beta}_j}{\beta_j} \right| + \left| \frac{c_j - \hat{c}_j}{c_j} \right| \right] \]

In the next figure, one can find the error for 1000 runs of the previous setting. One curve represents \( E_A \) while \( E_{\beta,c} \) is in the second curve.
6 Applications to Finance.

6.1 Risk Measures

Here the sensitivity of several risk measures to the parameters and features of the model are presented.

6.2 Portfolio Theory

Relying on simulations, the optimizations of risk measures in previous subsection will be performed for various levels of returns.

7 Conclusions

The following objectives were accomplished in this paper.

- The SVF model was defined and empirical evidence of its usefulness shown.
- The volatility-correlations properties as well as its characteristic function were described.
- The stationarity, ergodicity and rho-mixing behaviour of the model were proved under certain standard conditions.
- The Continuum of Moments method was used to estimate its parameters, several examples were shown.
- The impact of the various parameters in risk measures and Portfolio Theory was described.
References


A Proofs

Proof of proposition ??

Proof. Expressions ??-?? can be obtained by using the bilinearity of the quadratic variation and Ito’s formula. For example the instantaneous correlation between assets and correlations is defined as:

\[ \rho(Y_i, \rho_{ij})(t) = \frac{\langle Y_i, \rho(Y_i, Y_j) \rangle}{\sqrt{\langle Y_i \rangle} \sqrt{\langle \rho(Y_i, Y_j) \rangle}}. \]

Using Ito’s formula on \( \rho(Y_i, Y_j) \) together with the independence among \( W_k(t) \) and the dependence between \( W_k(t), V_k(t) \) we get

\[
\langle Y_i, \rho(Y_i, Y_j) \rangle = \left( \sum_{k=1}^{p} a_{i,k} \sqrt{V_k(t)} W_k(t), \sum_{k=1}^{p} \frac{\partial \rho(Y_i, Y_j)}{\partial v_k} V_k(t) \right) \\
= \sum_{k=1}^{p} a_{i,k} \sqrt{V_k(t)} \frac{\partial \rho(Y_i, Y_j)}{\partial v_k} \rho_{kt} \\
= \sum_{k=1}^{p} a_{i,k} \sqrt{V_k(t)} \frac{\partial \rho(Y_i, Y_j)}{\partial v_k} \rho_{kt}
\]

The expressions \( \sigma_{Y_j}(t), \rho_{(\sigma_i, \sigma_k)}(t), \rho(Y_j, Y_k)(t) \) and \( \rho(Y_i, \sigma_i)(t) \) follow similarly. \( \Box \)
Proof of proposition (17)

Proof. Consider \( g(\phi, y, v, t) \) a bounded twice differentiable function with continuous second derivatives for \( t > 0 \) and \((y, v) \in \mathbb{R}^{m+p+m}\). Apply Ito’s formula to it together with the processes \((Y(t), V(t), t)\). Then for \( T > t > 0 \)

\[
g(\phi, Y(T), V(T), T) - \int_0^T \Xi(Y, V)g(\phi, Y(s), V(s), s)ds
\]

is a zero- mean martingale. Taking conditional expectations on both sides we see that

\[
f(\phi; y, v, t) = E[g(\phi, Y(T), V(T)) | Y(t) = y, V(t) = v]
\]  

(17)
satisfies (??) when \( g(\phi, y, v, t) = \exp\{i\phi'y\} \).

Next, we consider a solution in the form:

\[
f(\phi; y, v, t) = \exp\left\{C(T - t) + \sum_{j=1}^p D_j(T - t)v_j + \sum_{j=1}^m D^{(j)}(T - t)v^{(j)} + i \sum_{j=1}^m \phi_jy_j\right\}
\]

for some functions \( C, D_j \) and \( D^{(j)} \).

By substituting into (??) the PDE is transformed into the ODE:

\[
0 = \left[-C' - \sum_{i=1}^p D'_iV_i - \sum_{i=1}^m D^{(i)'}V^{(i)}\right]
\]

\[
+ \sum_{j=1}^m \left[i\phi_j \left(\mu_j - \sum_{i=1}^p a_j^2V_i - \frac{1}{2}V^{(j)}\right) - \frac{1}{2}\phi_j^2 \left(\sum_{i=1}^p a_j^2V_i + V^{(j)}\right)\right]
\]

\[
- \sum_{j>k=1}^m \left[\phi_j\phi_k \left(\sum_{i=1}^p a_j a_k V_i\right)\right]
\]

\[
+ \sum_{j=1}^m \left[\frac{1}{2} \sum_{i=1}^p i\phi_jD_i \rho_i a_j c_j^2 V_i\right] + \sum_{j=1}^p \left[D_j \frac{\epsilon_j}{2} (\beta_j - V_j) + \frac{1}{2}D_j^2 c_j^2 V_j\right]
\]

\[
+ \sum_{j=1}^m \left[D^{(j)} \frac{c^{(j)}}{2} (\beta^{(j)} - V^{(j)}) + \frac{1}{2} \left(D^{(j)}\right)^2 (c^{(j)})^2 V^{(j)}\right]
\]

where the prime indicates the derivative with respect to time \( t \). This
equation must hold for all \( v_j(t), v^{(i)}(t) \) then it leads to the Riccati system

\[
D_j' = -i \sum_{i=1}^{m} \frac{a^2_{i,j}}{2} \phi_i - \frac{1}{2} \sum_{i,k=1}^{m} a_{i,j} a_{k,j} \phi_i \phi_k + D_j \frac{c_j}{2} + \frac{1}{2} D_j^2 c_j^2 + i \frac{1}{2} \sum_{i=1}^{m} \phi_i D_j \rho_j a_{i,j} c_j^2
\]

\[
D_j(0) = 0
\]

\[
D^{(j)\nu} = -\frac{1}{2} \phi_j - \frac{1}{2} \phi_j^2 + D^{(j)} \frac{c_j}{2} + \frac{1}{2} \left( D^{(j)} \right)^2 \left( c^{(j)} \right)^2
\]

\[
D^{(j)}(0) = 0
\]

\[
C' = \sum_{j=1}^{m} i \phi_j \mu_j + \sum_{j=1}^{p} D_j \beta_j \frac{c_j}{2} + \sum_{j=1}^{m} D^{(j)} \beta^{(j)} \frac{c^{(j)} j}{2}
\]

\[
C(0) = 0
\]

which is a system of Riccati equations with closed form solution given by:

\[
D_j(\tau) = \frac{2 \arctan \left( \frac{b^* + 2 \tau c^*}{\sqrt{4a^* c^* - b^2}} \right)}{\sqrt{4a^* c^* - b^2}}
\]

\[
D^{(j)}(\tau) = \frac{2 \arctan \left( \frac{b^{**} + 2 \tau c^{**}}{\sqrt{4a^{**} c^{**} - b^{**}^2}} \right)}{\sqrt{4a^{**} c^{**} - b^{**2}}}
\]

\[
C(\tau) = \sum_{j=1}^{m} i \phi_j \mu_j \tau + \sum_{j=1}^{p} \beta_j \frac{c_j}{2} \int_{0}^{\tau} D_j(s) ds + \sum_{j=1}^{m} \beta^{(j)} \frac{c^{(j)} j}{2} \int_{0}^{\tau} D^{(j)}(s) ds
\]

where

\[
a^* = -i \sum_{i=1}^{m} \frac{a^2_{i,j}}{2} \phi_i - \frac{1}{2} \sum_{i,k=1}^{m} a_{i,j} a_{k,j} \phi_i \phi_k
\]

\[
b^* = c_j \frac{c_j}{2} + i \frac{1}{2} \sum_{i=1}^{m} \phi_i \rho_j a_{i,j} c_j^2
\]

\[
c^* = \frac{1}{2} c_j^2
\]

\[
a^{**} = -\frac{1}{2} \phi_j - \frac{1}{2} \phi_j^2
\]

\[
b^{**} = \frac{c^{(j)}}{2}
\]

\[
c^{**} = \frac{1}{2} \left( c^{(j)} \right)^2
\]

24
See [?] for details.

Proof of proposition ??.

Proof. The price of this option according to Black-Scholes is:

\[ \Pi_1(k) = e^{-rT} \int_{k_1}^{+\infty} \ldots \int_{k_n}^{+\infty} \prod_{i=1}^{n} (e^{y_i} - e^{k_i}) \phi(y_1, \ldots, y_n) dy \]

where \( \phi(y_1, \ldots, y_n) \) is the density of \( y(T) \) conditionally to \( y(0) = y_0 \).

We define:

\[ c(k) = e^{\alpha'k} \Pi_1(k) \]

where \( \alpha \) is a positive vector. The Fourier transform of \( c(k) \) (denoted as \( f^*(s) \)) is related to the Fourier transform of \( \phi(y) \), denoted as \( f(x) \) as follows:

\[
\begin{align*}
    f^*(s) &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{-is'k} c(k) dk \\
    &= e^{-rT} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{-is'k} \prod_{i=1}^{m} (e^{y_i} - e^{k_i}) \phi(y) dy dk \\
    &= e^{-rT} \int_{-\infty}^{\infty} \phi(y) \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{(\alpha- is')k} \prod_{i=1}^{m} (e^{y_i} - e^{k_i}) dk dy \\
    &= e^{-rT} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \phi(y) \frac{\exp \left\{ \sum_{j=1}^{m} (\alpha_j + 1 + is_j) y_j \right\}}{\prod_{j=1}^{m} (\alpha_j + is_j)(\alpha_j + 1 + is_j)} dy \\
    &= e^{-rT} f(s_1 - \alpha_1 i, \ldots, s_m - (\alpha_m + 1)i) \prod_{j=1}^{m} (\alpha_j + is_j)(\alpha_j + 1 + is_j)
\end{align*}
\]

Thus if the characteristic function \( f(x) \) is known in closed form, the Fourier transform \( f^*(s) \) of the option price will also be available analytically,
yielding the option price itself via an inverse transform:

\[
\Pi_1(k) = e^{-ia'k} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-is'k} f^*(s) ds
\]

Proof of theorem ??.

*Proof.* We prove that \( V \) is geometrically ergodic which implies it is \( \rho \)-mixing and therefore \( \alpha \)-mixing. This is done by showing that the extended infinitesimal generator for \( V \) satisfies the Foster-Lyapunov inequality (see condition CD3 in [?]).

We consider a solution of

\[
\Xi_V(f) = -a \cdot f
\]

in the form \( f(v) = \Pi_1^{i=1} f_i(v_i) \) for some positive functions \( f_i \) depending on \( e_i > 0, i = 1, 2, \ldots, p \) and \( a \geq \sum_{i=1}^{p} e_i \).

The PDE given by (??) is separable into the equations:

\[
e_1 f_1(v_1) = a_1(v_1 - b_1) f_1'(v_1) + c_1^2 v_1 f_1''(v_1)
\]

\[
\vdots
\]

\[
e_{p-1} f_{p-1}(v_{p-1}) = a_{p-1}(v_{p-1} - b_{p-1}) f_{p-1}'(v_{p-1}) + c_{p-1}^2 v_{p-1} f_{p-1}''(v_{p-1})
\]

\[
(a - \sum_{i=1}^{p} e_i) f_p(v_p) = a_p(v_p - b_p) f_p'(v_p) + c_p^2 v_p f_p''(v_p)
\]

which are Confluent Hypergeometric Differential Equations. They have an unique solution (see 13.6 in [?] (1985)). It proves the Foster-Lyapunov condition.

*Proof.* Take for any \( 1 \leq j \leq p \) the process \( G_j(t) \) independent with \( B_j(t) \) such
that $W_j(t) = \sqrt{1 - \rho_j^2}G_j(t) + \rho_jB_j(t)$. It follows:

$$Z_{i,k} = \mu_k\sqrt{h} - \frac{1}{2\sqrt{h}} \sum_{j=1}^{p} a_{k,j}^2 \int_{(i-1)h}^{ih} V_j(t)dt - \frac{1}{2\sqrt{h}} \int_{(i-1)h}^{ih} V^{(k)}(t)dt$$

$$+ \frac{1}{\sqrt{h}} \sum_{j=1}^{p} a_{kj} \left[ \sqrt{1 - \rho_j^2} \int_{(i-1)h}^{ih} \sigma_j(t)dG_j(t) + \rho_j \int_{(i-1)h}^{ih} \sigma_j(t)dB_j(t) \right]$$

$$+ \frac{1}{\sqrt{h}} \int_{(i-1)h}^{ih} \sigma^{(k)}(t)dW^{(k)}(t)$$

Notice, from the definition of $V$, we have:

$$\int_{(i-1)\Delta}^{i\Delta} \sigma_j(t)dB_j(t) = \frac{\Delta V_{i,j}}{c_j} - \left( \frac{\beta_j}{2} h - \int_{(i-1)h}^{ih} V_j(t)dt \right)$$

therefore:

$$Z_{i,k} = \mu_{Z_{i,k}} + \frac{1}{\sqrt{h}} \sum_{j=1}^{p} a_{kj} \left[ \sqrt{1 - \rho_j^2} \int_{(i-1)h}^{ih} \sigma_j(t)dG_j(t) \right] + \frac{1}{\sqrt{h}} \int_{(i-1)h}^{ih} \sigma^{(k)}(t)dW^{(k)}(t)$$

which follows a normal distribution with:

$$\mu_{i,k} = \mu_k\sqrt{h} - \frac{1}{2\sqrt{h}} \sum_{j=1}^{p} a_{k,j}^2 \int_{(i-1)h}^{ih} V_j(t)dt - \frac{1}{2\sqrt{h}} \int_{(i-1)h}^{ih} V^{(k)}(t)dt$$

$$+ \frac{1}{\sqrt{h}} \sum_{j=1}^{p} a_{kj} \rho_j \left[ \frac{\Delta V_{i,j}}{c_j} - \left( \frac{\beta_j}{2} h - \frac{1}{2} \int_{(i-1)h}^{ih} V_j(t)dt \right) \right].$$

$$+ \frac{1}{\sqrt{h}} \int_{(i-1)h}^{ih} \sigma^{(k)}(t)dW^{(k)}(t)$$

Hence $Z_i \sim N(\mu_i, \Sigma_{Z_i})$ where the component $(i, k)$ of the covariance matrix is given by:
\[
\sigma^2_{i,k} = \frac{1}{\sqrt{h}} \sum_{j=1}^{n} a_{kj}^2 (1 - \rho^2) \bar{V}_{i,j} + \frac{1}{\sqrt{h}} V_i^{(k)}
\]

Proof. The strict stationarity of \( U_i \) follows similarly to Theorem 3.1 in [?]. while ergodicity and \( \alpha \)-mixing follows from proposition 3.2 in the same reference.

Part (ii) follows easily from lemma ???. Indeed define the filtration \( G_t = \sigma (V_s, s \in [0, t]) \). Now conditionally on \( G_{nh} \), the random vectors \( Z_1, ..., Z_n \) are independent and \( Z_i \), has normal distribution. Thus, for real vectors \( \lambda_1, ..., \lambda_n \),

\[
E \left[ \exp \left\{ \sum_{j=1}^{n} i \lambda_j Z_j \right\} \mid G_{nh} \right] = E \left[ \exp \left\{ \sum_{j=1}^{n} i \lambda_j Z_j \right\} \mid U_1, ..., U_n \right] = \exp \left\{ \left( \sum_{j=1}^{n} \lambda_j \left[ \mu - \frac{1}{2} (A \circ A) \bar{V}_j + A \frac{\Delta V_j}{c} - \frac{1}{2} (\beta - \bar{V}_j) \right) \otimes \rho \right) \right\} \times \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n} \lambda_j A \left[ d g (V_j \circ (1 - \rho^2)) \right] A^t \lambda_j \right\}
\]

For part (iii), both inequalities follow from [?].

Proof. The conditioning only involves one of the lagged values of \( (Y)_{t \in \mathbb{R}} \), given that the process is Markov. Using the iterated expectation rule, we get:

\[
\Psi_{V(0), Y(t)}(\tau, w, t) = E \left[ E \left[ \exp \left\{ i \langle w, Y(t + \tau) - Y(t) \rangle \right\} \mid V(0), V(t), r(t) \right] \mid V(0), r(t) \right] = E \left[ f(w; 0, V, t) \mid V(0), r(t) \right] = E \left[ \exp \left\{ C(\tau, w) + D(\tau, w)V(t) \right\} \mid V(0), r(t) \right] = e^{C(\tau, w)} \prod_{i=1}^{p} f(D_i(\tau, w), V_i(0), t) \prod_{i=1}^{n} f(D^{(i)}(\tau, w), V^{(i)}(0), t)
\]

28
Where \( f(w; y, v, t) \) comes from ?? and \( f(\phi; v, t) \) is the well known laplace transform of the CIR process \( V \) (Heston process, see Gion 2005):

\[
\begin{align*}
f(\phi, V(0), t) &= E \left[ \exp \{ \phi V(t) \} \mid V(0) \right] \\
&= \exp \{ b(t, \phi) + a(t, \phi)V(0) \} \\
b &= \frac{2\gamma\beta}{c^2} \log \left\{ 1 + \phi \frac{c^2}{2\gamma} (1 - \exp \{ \gamma t \}) \right\} \\
a &= \frac{-\phi \exp \{ \gamma t \}}{1 + \frac{c^2\phi}{2t} [1 - \exp \{ \gamma t \}]}
\end{align*}
\]

We are also using the independence among the volatility factors. Clearly, the equation for \( \Psi_{V(0), Y(t)}(\tau, w, t) \) does not depend anymore on \( Y(t) \) but it does depend on \( t \) therefore we can rewrite it as \( \Psi_{V(0)}(\tau, w, t) \).

**Proof.** The weighted norm can be written as:

\[
\left\| \hat{n}(\theta) \right\|^2 = \int_{R^n} \int_{R^n} \left[ \frac{1}{T} \sum_{t=1}^{T} \left( e^{i(w, r_t) + i(w, r_t)} - \Psi_{V(0)}(\tau, w, t)e^{i(w, r_t)} \right) \right] \times \pi(w, w_s)dw
\]

The integrals involve in this expression are solvable in a lower dimensional space. To see this first note the following simplifications for the characteristic function:

\[
\int_{R^n} \Psi_{V(0)}(\tau, w, t)dw
\]

Let us assume the simple case of one common factor, denote:

\[
M^{(1)} = \int_{t}^{t+\tau} v_1(s)ds
\]

\[
M_1 = \int_{t}^{t+\tau} v_1(s)dW_1
\]
then the following relationships hold:

\[
\int_{\Omega_w} \Psi_{V(o)}(\tau, w, t)dw = \int_{\Omega_w} E \left[ \exp (i \langle w, y(t + \tau) - y(t) \rangle) \mid y(t), v_1(0), v_1^{(i)}(0) \right] dw
\]

\[
= \int_{\Omega_w} E \left[ \exp (i \langle w, y(t + \tau) - y(t) \rangle) \mid y(t), v_1^{(i)}(t), M_1, M^{(1)} \right] \mid y(t), v_1(0), v_1^{(i)}(0) \right] dw
\]

\[
= \int_{\Omega_w} E \left[ \prod_{i=1}^{m} f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)}) \mid r(t), v_1(0), v_1^{(i)}(0) \right] dw
\]

\[
= E \left[ \prod_{i=1}^{m} E \left[ \int_{\Omega_{w_i}} f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)})dw_i \mid r(t), v_1^{(i)}(0) \right] \right] \mid r(t), v_1(0), v_1^{(i)}(0) \right] \]

\[
= E \left[ \prod_{i=1}^{m} h_i(w_i, r(t), v_1^{(i)}(0), M_1, M^{(1)}) \mid r(t), v_1(t) \right] \mid r(t), v_1(0), v_1^{(i)}(0) \right] \]

\[
= \int_{R^+} \left( \prod_{i=1}^{m} \int_{R^+} \left( \int_{R^+} f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)}) dw_i \right) \right) g_{v_1(t)}(v_1(0))dv_1(t) \]

Note that what originally was a single \( m \)-dimensional integral becomes \( m \) 5 dimensional integrals.

Explanations:

In equation (20), \( f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)}) \) represents the one-dimensional Laplace transform of a log normal process with CIR volatility (see page 187 of Giou2005):

\[
f(\phi, y, v, t) = E[\exp \{ i\phi (Y(t + \tau) - Y(t)) \} \mid V(t) = v, Y(t) = y] \]

\[
= \exp \left\{ i\phi \left( \mu \tau - \frac{a_{11}}{2} M^{(1)} + a_{i1} M_1 \right) \right\} \left( i\phi - \frac{\phi^2}{2} \right) \int_{t}^{t+\tau} V(s)ds \]

where \( \Psi_{t,\tau}(u) \) is the Laplace transform of \( \int_{t}^{t+\tau} V(s)ds \) (see Giou2005).
Note that in general (Heston 1993):
\[ f(\phi, y, v, t) = E \{ \exp \{ i\phi (Y(t + \tau) - Y(t)) \} | V(t) = v, Y(t) = y \} \]
\[ = \exp \{ C(\tau, \phi) + D(\tau, \phi)v \} \]
\[ C = \mu \phi \tau + \frac{\gamma \beta}{\varsigma^2} \left[ (\gamma - \rho \phi \phi + d) \tau - 2 \ln \left( \frac{1 - g \exp(\rho \phi \tau)}{1 - g} \right) \right] \]
\[ D = \frac{\gamma - \rho \phi \phi + d}{\gamma - \rho \phi \phi - d} \left[ 1 - \exp(\rho \phi \tau) \right] \]
\[ g = \frac{\gamma - \rho \phi \phi + d}{\gamma - \rho \phi \phi - d} \left[ 1 - \exp(\rho \phi \tau) \right] \]
\[ d = \sqrt{(\rho \phi \phi - \gamma)^2 - \varsigma^2 (\phi \phi - \phi^2)} \]

For equation ???, the function \( h \) is defined as follows:
\[
h_i(w_i, r(t), v^{(i)}(0), M_1, M^{(1)}) \]
\[ = E \left[ \int_{\Omega_{w_i}} f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)})dw_i | r(t), v^{(i)}(0) \right] \]
\[ = \int_{\Omega_{w_i}} \left( \int_{\mathbb{R}^+} f_i(w_i, r(t), v^{(i)}(t), M_1, M^{(1)})g(v^{(i)}(t) | v^{(i)}(0))dv^{(i)} \right) dw_i \]

here \( g_{v^{(i)}(0)}(v^{(i)}(t)) \) is the density of a non centered gamma distribution (see Gio 2005, page 180) representing the conditional density of \( v^{(i)}(t) \mid v^{(i)}(0) \).

For equation ???, we first need to find
\[ E \left[ \prod_{i=1}^m h_i(w_i, r(t), v^{(i)}(0), M_1, M^{(1)}) | r(t), v_1(t) \right] \]

The basic information needed is the joint density of \((M_1, M^{(1)}) | v_1(t)\), denoted by \( g_{v_1(t)}(M_1, M^{(1)}) \), which can be found by using the joint Laplace transform, which is obtained next:

Denote
\[ \Psi_{t,\tau}(u_1, u_2) = E_t \exp \left\{ -u_1 \int_t^{t+\tau} V(s)ds - u_2 \int_t^{t+\tau} \sqrt{V(s)}dW(s) \right\} \]

31
Assume that there is an Affine solution:

$$\Psi_{t,\tau}(u_1, u_2) = \exp \left\{ -a(\tau, u_1, u_2)V(t) - b(\tau, u_1, u_2) \right\} \quad (25)$$

Then we will find conditions on $a$ and $b$:

$$\Psi_{t,\tau}(u_1, u_2) = E_t \left[ \exp \left\{ -u_1 \int_t^{t+dt} V(s)ds - u_2 \int_t^{t+dt} \sqrt{V(s)}dW(s) \right\} \Psi_{t+dt,\tau}(u_1, u_2) \right] \approx E_t \left[ \exp \left\{ -u_1 V(t)dt - u_2 \sqrt{V(t)}dW(t) \right\} \Psi_{t+dt,\tau-dt}(u_1, u_2) \right]$$

Equating (25) and (26) leads to a Riccati system of equations:

$$\frac{\partial a}{\partial \tau} = u_1 - u_2^2 - \gamma a - c^2 a^2$$

$$\frac{\partial b}{\partial \tau} = \beta \gamma a.$$

Then equation (26) follows from observing that the distribution of $v_1(t) \mid v_1(0)$ follows a non centered gamma.

Using the previous result we have, in particular:

$$\int \int_{R^n \times R^n} \left[ \Psi_{V(o)}(\tau, w_s, j) e^{-i(w_r, r_{r+\tau})} \right] \pi(w_r, w_s)dw$$

which leads to the wanted expression. \qed

mmm.