Strathprints Institutional Repository

Pan, Jiazhu and Polonik, Wolfgang and Yao, Qiwei (2010) *Estimating factor models for multivariate volatilities: an innovation expansion approach.* [Proceedings Paper]

Estimating Factor Models for Multivariate Volatilities: An Innovation Expansion Method

Jiazhu Pan[1], Wolfgang Polonik[2], and Qiwei Yao[3]

[1] Department of Mathematics and Statistics, University of Strathclyde
26 Richmond Street, Glasgow, G1 1XH, UK, *jiazhu.pan@strath.ac.uk*
[2] Division of Statistics, University of California at Davis
Davis, CA 95616, USA, *wpolonik@ucdavis.edu*
[3] Department of Statistics, London School of Economics
London WC2A 2AE, UK, *q.yao@lse.ac.uk*

Abstract. We introduce an innovation expansion method for estimation of factor models for conditional variance (volatility) of a multivariate time series. We estimate the factor loading space and the number of factors by a stepwise optimization algorithm on expanding the "white noise space". Simulation and a real data example are given for illustration.

Keywords: dimension reduction, factor models, multivariate volatility.

1 Introduction

Factor modelling plays an important role in the analysis of high-dimensional multivariate time series(see Sargent and Sims, 1977; Geweke, 1977) because it is both flexible and parsimonious. Most of factor analysis in the literature is for the mean and conditional mean of a multivariate time series and panel data, see Pan and Yao (2008) and a series of papers of article by Forni, Hallin, Lippi and Reichlin (2000,2004), and Hallin and Liška (2007).

For the conditional variance, which is so-called volatility, the multivariate generalized autoregressive conditional heteroskedastic (GARCH) models are commonly used, see Engle and Kroner (1995), Engle (2002), Engle & Sheppard (2001). But a multivariate GARCH model often has too many parameters so that it is difficult to estimate the model, which is a high-dimensional optimization problem. Factor models for volatility are useful tools to overcome the overparametrisation problem, e.g. Factor-ARCH (Engle, Ng and Rothschild 1990).

In this paper, we consider a frame work of factor analysis for the multivariate volatility, including factor ARCH as a special case. We introduce a innovation expansion method for the estimation of the factor loading space and the number of factors. Our method can change a high-dimensional optimization problem to a stepwise optimization algorithm by expanding the "white noise space" (innovation space) one step each time.

2 Models and methodology

Let $\{Y_t\}$ be a $d\times 1$ time series, and $E(Y_t|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, \cdots)$. Assume that $E(Y_t Y_t^\tau)$ exists, and we use the notation $\Sigma_y(t) = var(Y_t|\mathcal{F}_{t-1})$. Pan et al. (2009) consider a common factor model

$$Y_t = AX_t + \varepsilon_t, \tag{1}$$

where X_t is a $r\times 1$ time series, $r < d$ is unknown, A is a $d\times r$ unknown constant matrix, $\{\varepsilon_t\}$ is a sequence of i.i.d. innovations with mean 0 and covariance matrix Σ_ε, and ε_t is independent of X_t and $\mathcal{F}_{t-1}$. This assumes that the volatility dynamics of Y is determined effectively by a lower dimensional volatility dynamics of X_t plus the static variation of ε_t, as

$$\Sigma_y(t) = A\Sigma_x(t)A^\tau + \Sigma_\varepsilon, \tag{2}$$

where $\Sigma_x(t) = var(X_t|\mathcal{F}_{t-1})$. The component variables of X_t are called the factors. There is no loss of generality in assuming $\mathrm{rk}(A) = r$ and requiring the column vectors of $A = (a_1, \cdots, a_r)$ to be orthonormal, i.e. $A^\tau A = I_r$, where I_r denotes the $r \times r$ identity matrix.

We are concerned with the estimation for the factor loading space $\mathcal{M}(A)$, which is uniquely defined by the model, rather than the matrix A itself. This is equivalent to the estimation for orthogonal complement $\mathcal{M}(B)$, where B is a $d \times (d - r)$ matrix for which (A, B) forms a $d \times d$ orthogonal matrix, i.e. $B^\tau A = 0$ and $B^\tau B = I_{d-r}$. Now it follows from (1) that

$$B^\tau Y_t = B^\tau \varepsilon_t. \tag{3}$$

Hence $B^\tau Y_t$ are homoscedastic components since

$$E\{B^\tau Y_t Y_t^\tau B|\mathcal{F}_{t-1}\} = E\{B^\tau \varepsilon_t \varepsilon_t^\tau B\} = E\{B^\tau Y_t Y_t^\tau B\} = B^\tau var(Y_t)B.$$

This implies that

$$B^\tau E[\{Y_t Y_t^\tau - var(Y_t)\}I(Y_{t-k} \in C)]B = 0, \tag{4}$$

for any $t, k \geq 1$ and any measurable $C \subset R^d$.

For matrix $H = (h_{ij})$, let $||H|| = \{tr(H^\tau H)\}^{1/2}$ denote its norm. Then (4) implies that

$$\sum_{k=1}^{k_0} \sum_{C\in\mathcal{B}} w(C) \Big|\Big| \sum_{t=k_0+1}^{n} E[B^\tau \{Y_t Y_t^\tau - var(Y_t)\}BI(Y_{t-k} \in C)]\Big|\Big|^2 = 0 \tag{5}$$

where $k_0 \geq 1$ is a prescribed integer, $\mathcal{B}$ is a finite or countable collection of measurable sets, and the weight function $w(\cdot)$ ensures the sum on the right-hand side finite. In fact we may assume that $\sum_{C\in\mathcal{B}} w(C) = 1$. Even without the stationarity on Y_t, $var(Y_t)$ in (5) may be replaced by $\hat{\Sigma}_y \equiv$

$(n-k_0)^{-1}\sum_{k_0<t\le n} Y_t Y_t^\tau$. This is due to the fact $B^\tau var(Y_t)B = B^\tau \Sigma_\varepsilon B$, and

$$(n-k_0)^{-1}\sum_{t=k_0+1}^{n} B^\tau Y_t Y_t^\tau B \;=\; (n-k_0)^{-1}\sum_{t=k_0+1}^{n} B^\tau \varepsilon_t \varepsilon_t^\tau B \;\stackrel{a.s.}{\to}\; B^\tau \Sigma_\varepsilon B,$$

see (3). Therefore $B^\tau \hat{\Sigma}_y B$ is a consistent estimator for $B^\tau var(Y_t)B$ for all t. Denote

$$D_k(C) = (n-k_0)^{-1}\sum_{t=k_0+1}^{n} (Y_t Y_t^\tau - \hat{\Sigma}_y) I(Y_{t-k}\in C).$$

Now (5) suggests to estimate $B \equiv (b_1, \cdots, b_{d-r})$ by minimizing

$$\begin{aligned}\Phi_n(B) &= \sum_{k=1}^{k_0}\sum_{C\in\mathcal{B}} w(C)\big|\big|B^\tau D_k(C)B\big|\big|^2 \qquad (6)\\ &= \sum_{k=1}^{k_0}\sum_{1\le i,j\le d-r}\sum_{C\in\mathcal{B}} w(C)\big\{b_i^\tau D_k(C)b_j\big\}^2\end{aligned}$$

subject to the condition $B^\tau B = I_{d-r}$. This is a high-dimensional optimization problem. Further it does not explicitly address the issue how to determine the number of factors r. We present an algorithm which expands the innovation space step by step and which also takes care of these two concerns. Note for any $b^\tau A = 0$, $Z_t \equiv b^\tau Y_t (= b^\tau \varepsilon_t)$ is a sequence of independent random variables, and therefore, exhibits no conditional heterosedasticity. The determination of the r is based on the likelihood ratio test for the null hypothesis that the conditional variance of Z_t given its lagged valued is a constant against the alternative that it follows a GARCH(1,1) model with normal innovations. See also Remark 1(vii) below.

Put

$$\Psi(b) = \sum_{k=1}^{k_0}\sum_{C\in\mathcal{B}} w(C)[b^\tau D_k(C) b]^2,$$

$$\Psi_m(b) = \sum_{k=1}^{k_0}\Big\{2\sum_{i=1}^{m-1}\sum_{C\in\mathcal{B}} w(C)[\hat{b}_i^\tau D_k(C)\, b]^2 + \sum_{C\in\mathcal{B}} w(C)[b^\tau D_k(C) b]^2\Big\}.$$

An Innovation Expansion Algorithm for estimating B and r: let p be an integer between 1 and k_0 and $\alpha\in(0,1)$ specify the level of significance test.

Step 1. Compute $\hat{b}_1$ which minimises $\Psi(b)$ subject to the constraint $b^\tau b = 1$. Let $Z_t = \hat{b}_1^\tau Y_t$. Compute the 2log-likelihood ratio test statistic

$$T = (n-k_0)\big\{1+\log\big(\frac{1}{n-k_0}\sum_{t=k_0+1}^{n} Z_t^2\big)\big\} - \min\sum_{t=k_0+1}^{n}\big\{\frac{Z_t^2}{\sigma_t^2}+\log(\sigma_t^2)\big\}, \qquad (7)$$

where $\sigma_t^2 = \alpha + \beta Z_{t-1}^2 + \gamma \sigma_{t-1}^2$, and the minimisation is taken over $\alpha > 0$, $\beta, \gamma \geq 0$ and $\beta + \gamma < 1$. Terminate the algorithm with $\hat{r} = d$ and $\hat{B} = 0$ if T is greater than the top α-point of the χ_2^2-distribution. Otherwise proceed to Step 2.

Step 2. For $m = 2, \cdots, d$, compute $\hat{b}_m$ which minimizes $\Psi_m(b)$ subject to the constraint

$$b^\tau b = 1, \quad b^\tau \hat{b}_i = 0 \quad \text{for } i = 1, \cdots, m-1. \tag{8}$$

Terminate the algorithm with $\hat{r} = d-m+1$ and $\hat{B} = (\hat{b}_1, \cdots, \hat{b}_{m-1})$ if T, calculated as in (7) but with $Z_t = |\hat{b}_m^\tau Y_t|$ now, is greater than the top α-point of the χ_2^2-distribution.

Step 3. In the event that T_p never exceeds the critical value for all $1 \leq m \leq d$, let $r = 0$ and $\hat{B} = I_d$.

Remark 1. (i) The algorithm grows the dimension of $\mathcal{M}(B)$ by 1 each time until a newly selected direction $\hat{b}_m$ being relevant to the volatility dynamics of Y_t. This effectively reduces the number of the factors in model (1) as much as possible without losing significant information.

(ii) The minimization problem in Step 2 is a d-dimensional subject to constraint (8). It has only $(d-m+1)$ free variables. In fact, the vector b satisfying (8) is of the form

$$b = A_m u, \tag{9}$$

where u is any $(d-m+1) \times 1$ unit vector, A_m is a $d \times (d-m+1)$ matrix with the columns being the $(d-m+1)$ unit eigenvectors, corresponding to the $(d-m+1)$-fold eigenvalue 1, of matrix $I_d - B_m B_m^\tau$, and $B_m = (\hat{b}_1, \cdots, \hat{b}_{m-1})$. Note that the other $(m-1)$ eigenvalues of $I_d - B_m B_m^\tau$ are all 0.

(iii) We may let $\hat{A}$ consist of the $\hat{r}$ (orthogonal) unit eigenvectors, corresponding to the common eigenvalue 1, of matrix $I_d - \hat{B}\hat{B}^\tau$ (i.e. $\hat{A} = A_{d-\hat{r}+1}$). Note that $\hat{A}^\tau \hat{A} = I_{\hat{r}}$.

(iv) A general formal $d \times 1$ unit vector is of the form $b^\tau = (b_1, \cdots, b_d)$, where

$$b_1 = \prod_{j=1}^{d-1} \cos\theta_j, \quad b_i = \sin\theta_{i-1} \prod_{j=i}^{d-1} \cos\theta_j \;\; (i = 2, \cdots, d-1), \quad b_d = \sin\theta_{d-1},$$

where $\theta_1, \cdots, \theta_{d-1}$ are $(d-1)$ free parameters.

(v) We may choose $\mathcal{B}$ consisting of the balls centered at the origin in R^d. Note that $EY_{t-k} = 0$. When the underlying distribution of Y_{t-k} is symmetric and unimodal, such a $\mathcal{B}$ is the collection of the minimum volume sets of the distribution of Y_{t-k}, and this $\mathcal{B}$ determines the distribution of Y_{t-k} (Polonik 1997). In numerical implementation we simply use $w(C) = 1/K$, where K is the number the balls in $\mathcal{B}$.

(vi) Under the additional condition that

$$c^\tau A\{E(X_t X_t^\tau | \mathcal{F}_{t-1}) - E(X_t X_t^\tau)\} A^\tau c = 0 \tag{10}$$

if and only if $A^\tau c = 0$, (4) is equivalent to

$$E\{(b_i^\tau Y_t Y_t^\tau b_i - 1) I(Y_{t-k} \in C)\} = 0, \quad 1 \le i \le d - r, \;\; k \ge 1 \text{ and } C \in \mathcal{B}.$$

See model (1). In this case, we may simply use $\Psi(\cdot)$ instead of $\Psi_m(\cdot)$ in Step 2 above. Note that for b satisfying constraint (8), (9) implies

$$\Psi(b) = \sum_{k=1}^{k_0} \sum_{C \in \mathcal{B}} w(C) \big(u^\tau A_m^\tau D_k(C) A_m u\big)^2. \tag{11}$$

Condition (10) means that all the linear combinations of AX_t are genuinely (conditionally) heteroscadastic.

(vii) When the number of factors r is given, we may skip all the test steps, and stop the algorithm after obtaining $\hat{b}_1, \cdots, \hat{b}_r$ from solving the r optimization problems.

Remark 2. The estimation of A leads to a dynamic model for $\Sigma_y(t)$ as follow:

$$\hat{\Sigma}_y(t) = \hat{A}\hat{\Sigma}_z(t)\hat{A}^\tau + \hat{A}\hat{A}^\tau \hat{\Sigma}_y \hat{B}\hat{B}^\tau + \hat{B}\hat{B}^\tau \hat{\Sigma}_y,$$

where $\hat{\Sigma}_y = n^{-1} \sum_{1 \le t \le n} Y_t Y_t^\tau$, and $\hat{\Sigma}_z(t)$ is obtained by fitting the data $\{\hat{A}^\tau Y_t, \; 1 \le t \le n\}$ with, for example, the dynamic correlation model of Engle (2002).

3 Consistency of the estimator

For $r < d$, let $\mathcal{H}$ be the set consisting of all $d \times (d - r)$ matrices H satisfying the condition $H^\tau H = I_{d-r}$. For $H_1, H_2 \in \mathcal{H}$, define

$$D(H_1, H_2) = ||(I_d - H_1 H_1^\tau) H_2|| = \{d - r - tr(H_1 H_1^\tau H_2 H_2^\tau)\}^{1/2}. \tag{12}$$

Denote our estimator by $\hat{B} = argmin_{B \in \mathcal{H}_\mathcal{D}} \Phi_n(B)$.

Theorem 1. Let $\mathcal{C}$ denote the class of closed convex sets in $\mathcal{R}^d$. Under some mild assumptions (see Pan et al. (2009)), if the collection $\mathcal{B}$ is a countable subclass of $\mathcal{C}$, then $D(\hat{B}, B_0) \xrightarrow{P} 0$.

4 Numerical properties

We always set $k_0 = 30$, $\alpha = 5\%$, and the weight function $C(\cdot) \equiv 1$. Let $\mathcal{B}$ consist of all the balls centered at the origin.

4.1 Simulated examples

Consider model (1) with $r = 3$ factors, and $d \times 3$ matrix A with $(1, 0, 0)$, $(0, 0.5, 0.866)$ $(0, -0.866, 0.5)$ as its first 3 rows, and $(0, 0, 0)$ as all the other $(d-3)$ rows. We consider 3 different settings for $X_t = (X_{t1}, X_{t2}, X_{t3})^\tau$, namely, two sets of GARCH(1,1) factors $X_{ti} = \sigma_{ti} e_{ti}$ and $\sigma_{ti}^2 = \alpha_i + \beta_i X_{t-1,i}^2 + \gamma_i \sigma_{t-1,i}^2$, where $(\alpha_i, \beta_i, \gamma_i)$, for $i = 1, 2, 3$, are

$$(1,\ 0.45,\ 0.45), \qquad (0.9,\ 0.425,\ 0.425), \qquad (1.1,\ 0.4,\ 0.4), \tag{13}$$

or

$$(1,\ 0.1,\ 0.8), \qquad (0.9,\ 0.15,\ 0.7), \qquad (1.1,\ 0.2,\ 0.6), \tag{14}$$

and one mixing setting with two ARCH(2) factors and one stochastic volatility factor:

$$\begin{aligned} X_{t1} &= \sigma_{t1} e_{t1}, & \sigma_{t1}^2 &= 1 + 0.6 X_{t-1,1}^2 + 0.3 X_{t-2,1}^2, \\ X_{t2} &= \sigma_{t2} e_{t2}, & \sigma_{t2}^2 &= 0.9 + 0.5 X_{t-1,2}^2 + 0.35 X_{t-2,2}^2, \\ X_{t3} &= \exp(h_t/2) e_{t3}, & h_t &= 0.22 + 0.7 h_{t-1} + u_t. \end{aligned} \tag{15}$$

We let $\{\varepsilon_{ti}\}$, $\{e_{ti}\}$ and $\{u_t\}$ be sequences of independent $N(0,1)$ random variables. Note that the (unconditional) variance of X_{ti}, for each i, remains unchanged under the above three different settings. We set the sample size $n = 300, 600$ or 1000. For each setting we repeat simulation 500 times.

Table 1. Relative frequency estimates of r with $d = 5$ and normal innovations

		$\hat{r}$					
Factors	n	0	1	2	**3**	4	5
GARCH(1,1) with	300	.000	.046	.266	**.666**	.014	.008
coefficients (13)	600	.000	.002	.022	**.926**	.032	.018
	1000	.000	.000	.000	**.950**	.004	.001
GARCH(1,1) with	300	.272	.236	.270	**.200**	.022	.004
coefficients (14)	600	.004	.118	.312	**.500**	.018	.012
	1000	.006	.022	.174	**.778**	.014	.006
Mixture (15)	300	.002	.030	.166	**.772**	.026	.004
	600	.000	.001	.022	**.928**	.034	.014
	1000	.000	.000	.000	**.942**	.046	.012

We conducted the simulation with $d = 5, 10, 20$. To measure the difference between $\mathcal{M}(\mathcal{A})$ and $\mathcal{M}(\hat{\mathcal{A}})$, we define

$$D(A, \hat{A}) = \{|(I_d - AA^\tau)\hat{A}|_1 + |AA^\tau \hat{B}|_1\}/d^2, \tag{16}$$

where $|A|_1$ is the sum of the absolute values of all the elements in matrix A.

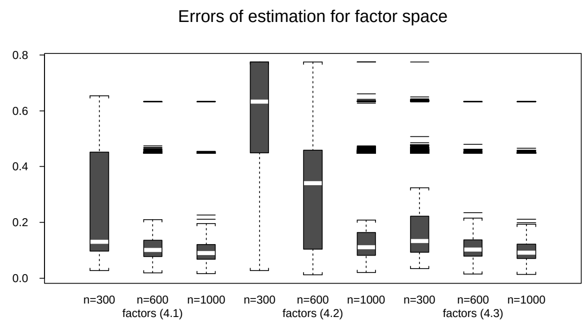

Fig. 1. Boxplots of $D(A, \hat{A})$ with two sets of GARCH(1,1) factors specified, respectively, by (13) and (14), and mixing factors (15). Innovations are Gaussian and $d = 5$.

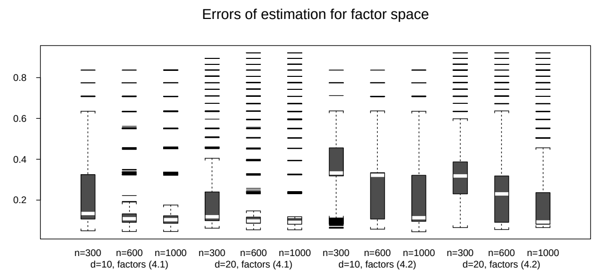

Fig. 2. Boxplots of $D(A, \hat{A})$ with two sets of GARCH(1,1) factors specified in (13) and (14), normal innovations and $d = 10$ or 20.

We report the results with $d = 5$ first. Table 1 lists for the relative frequency estimates for r in the 500 replications. When sample size n increases, the relative frequency for $\hat{r} = 3$ (i.e. the true value) also increases. Even for $n = 600$, the estimation is already very accurate for GARCH(1,1) factors (13) and mixing factors (14), less so for the persistent GARCH(1,1) factors (14). For $n = 300$, the relative frequencies for $\hat{r} = 2$ were non-negligible, indicating the tendency of underestimating of r, although this tendency disappears when n increases to 600 or 1000. Figure 1 displays the boxplots of $D(A, \hat{A})$. The estimation was pretty accurate with GARCH factors (13) and mixing factors (15), especially with correctly estimated r. Note with $n = 600$ or 1000, those outliers (lying above the range connected by dashed lines) typically correspond to the estimates $\hat{r} \neq 3$.

When $d = 10$ and 20, comparing with Table 1, the estimation of r is only marginally worse than that with $d = 5$. Indeed the difference with $d = 10$ and 20 is not big either. Note the D-measures for different d are not comparable; see (16). Nevertheless, Figure 2 shows that the estimation for A becomes more

Table 2. Relative frequency estimates of r with GARCH(1,1) factors, normal innovations and d=10 or 20

			$\hat{r}$							
Coefficients	d	n	0	1	2	**3**	4	5	6	≥ 7
(13)	10	300	.002	.048	.226	**.674**	.014	.001	.004	.022
	10	600	.000	.000	.022	**.876**	.016	.012	.022	.052
	10	1000	.000	.000	.004	**.876**	.024	.022	.022	.052
	20	300	.000	.040	.196	**.626**	.012	.008	.010	.138
	20	600	.000	.000	.012	**.808**	.012	.001	.018	.149
	20	1000	.000	.000	.000	**.776**	.024	.012	.008	.180
(14)	10	300	.198	.212	.280	**.248**	.016	.008	.014	.015
	10	600	.032	.110	.292	**.464**	.018	.026	.012	.046
	10	1000	.006	.032	.128	**.726**	.032	.020	.016	.040
	20	300	.166	.266	.222	**.244**	.012	.004	.001	.107
	20	600	.022	.092	.220	**.472**	.001	.001	.012	.180
	20	1000	.006	.016	.092	**.666**	.018	.016	.014	.172

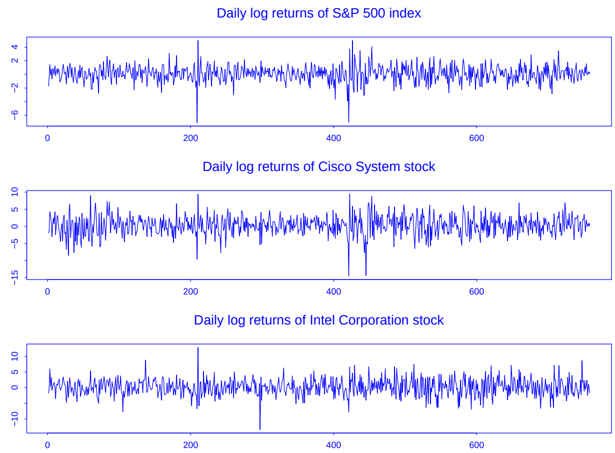

Fig. 3. Time plots of the daily log-returns of S&P 500 index, Cisco System and Intel Coprporation stock prices.

accurate when n increases, and the estimation with the persistent factors (14) is less accurate than that with (13).

4.2 A real data example

Figure 3 displays the daily log-returns of the S&P 500 index, the stock prices of Cisco System and Intel Corporation in 2 January 1997 – 31 December 1999. For this data set, $n = 758$ and $d = 3$. The estimated number of factors

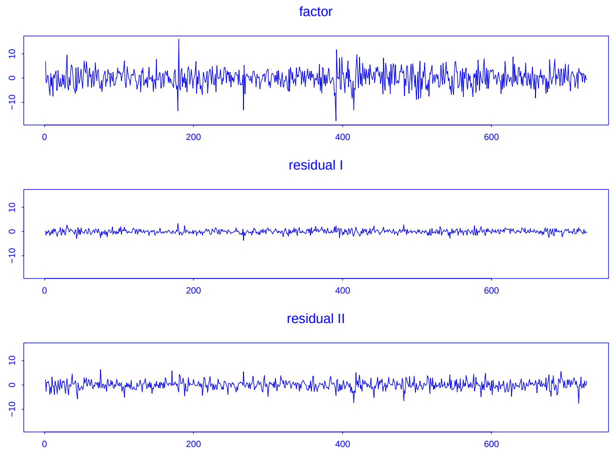

Fig. 4. Time plots of the estimated factor and two homoscedastic components for the S&P 500, Cisco and Intel data.

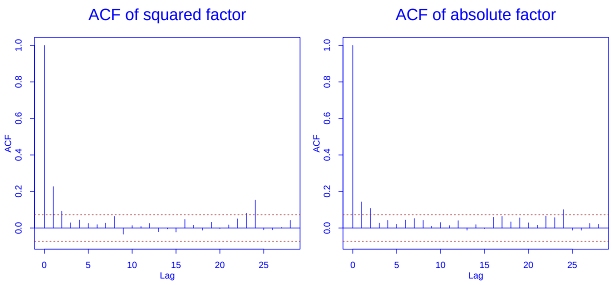

Fig. 5. The correlograms of squared and absulote factor for the the S&P 500, Cisco and Intel data

is $\hat{r} = 1$ with $\hat{A}^\tau = (0.310,\ 0.687,\ 0.658)$. The time plots of the estimated factor $Z_t \equiv \hat{A}^\tau Y_t$ and the two homoscedastic components $\hat{B}^\tau Y_t$ are displayed in Figure 4. The P-value of the Gaussian-GARCH(1,1) based likelihood ratio test for the null hypothesis of the constant conditional variance for Z_t is 0.000. The correlograms of the squared and the absolute factor are depicted in Figure 5 which indicates the existence of heteroscedasticity in Z_t. The fitted GARCH(1,1) model for Z_t is $\hat{\sigma}_t^2 = 2.5874 + 0.1416 Z_{t-1}^2 + 0.6509 \hat{\sigma}_{t-1}^2$. In contrast, Figure 6 shows that there is little autocorrelation in squared or absolute components of $\hat{B}^\tau Y_t$. The estimated constant covariance matrix is

$$\hat{\Sigma}_0 = \begin{pmatrix} 1.594 & & \\ 0.070 & 4.142 & \\ -1.008 & -0.561 & 4.885 \end{pmatrix}.$$

The overall fitted conditional variance process is given with $\hat{\Sigma}_z(t) = \hat{\sigma}_t^2$.

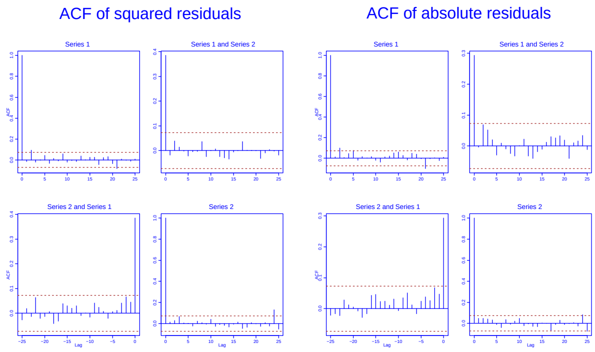

Fig. 6. The correlograms of squared and absulote homoscedastic compoments for the the S&P 500, Cisco and Intel data

References

ENGLE, R. F. (2002): Dynamic conditional correlation a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20, 339350.

ENGLE, R. F. and KRONER, K.F. (1995): Multivariate simultaneous generalised ARCH. *Econometric Theory*, 11, 122-150.

ENGLE, R. F., NG, V. K. and ROTHSCHILD, M. (1990): Asset pricing with a factor ARCH covariance structure: empirical estimates for Treasury bills. *Journal of Econometrics*, 45, 213-238.

FORNI, M., HALLIN, M., LIPPI, M. and REICHIN, L. (2000): The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*,82, 540-554.

FORNI, M., HALLIN, M. LIPPI, M. and REICHIN, L. (2004): The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics*,119, 231-255.

GEWEKE, J. (1977): The dynamic factor analysis of economic time series. In: D.J. Aigner and A.S. Goldberger (eds.): *Latent Variables in Socio-Economic Models*, Amsterdam: North-Holland, 365383.

HALLIN, M. and LIŠKA, R. (2007): Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102, 603-617.

LIN, W.-L. (1992): Alternative estimators for factor GARCH models – a Monte Carlo comparison. *Journal of Applied Econometrics*, 7, 259-279.

PAN, J., POLONIK,W., YAO,Q., and ZIEGELMANN,F. (2009): Modelling multivariate volatilities by common factors.*Research Report*, Department of Statistics, London School of Economics.

PAN, J. and YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95, 365-379.

SARGENT,T. J. and SIMS, C.A. (1977): Business cycle modelling without pretending to have too much a priori economic theory. In: C. A. Sims (ed.): *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis, 45109.