

# Dynamic Factor GARCH: Multivariate Volatility Forecast for a Large Number of Series

Lucia Alessi\*  
Matteo Barigozzi†  
Marco Capasso‡

Financial Market Analysis Research Centre (CAFiM)  
Sant'Anna School of Advanced Studies, Pisa

January 2007

## Abstract

We propose a new method for multivariate forecasting which combines the Generalized Dynamic Factor Model (GDFM) and the multivariate Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model. We assume that the dynamic common factors are conditionally heteroskedastic. The GDFM, applied to a large number of series, captures the multivariate information and disentangles the common and the idiosyncratic part of each series; it also provides a first identification and estimation of the dynamic factors governing the data set. A time-varying correlation GARCH model applied on the estimated dynamic factors finds the parameters governing their covariances' evolution. Then a modified version of the Kalman filter gets a more precise estimation of the static and dynamic factors' in-sample levels and covariances. A method is suggested for predicting conditional out-of-sample variances and covariances of the original data series. Finally, we carry out an empirical application aiming at comparing volatility forecasting results of our Dynamic Factor GARCH model against the univariate GARCH.

**Keywords:** Dynamic Factors, Multivariate GARCH, Covolatility Forecasting

**JEL-classification:** C32, C52, C53

---

\**E-mail:* lucia.alessi@sssup.it

†*E-mail:* matteo.barigozzi@gmail.com

‡*E-mail:* marco.capasso@gmail.com

# 1 Introduction

Exploiting all the information contained in a large dataset may be troublesome, if an increase in the cross-dimension corresponds to a much higher increase in the number of parameters. This phenomenon cannot be avoided whenever many different links exist among the variables. Indeed, it may become overwhelming when such links are operating not only at the level of the first moment, but also at the covariance level. For example, a forecast of large covariance matrices is needed for several financial tasks, including the construction of an optimal portfolio and the price determination of options based on many underlying returns. Many procedures have already been suggested in order to make estimation simpler for the case of many series linked one another by a relation in their conditional heteroskedasticities, and all of them have to face the trade-off between the reduction of complexity and the strength of the simplifying assumptions. The estimation and forecast method proposed here does not avoid this trade-off, but suggests a way to model and predict conditional covariances for a large number of series by simultaneously exploiting the information contained in the entire dataset. It represents a possible way for forecasting multivariate volatility by means of a factor model in which the dynamic factors are conditionally heteroskedastic and have a multivariate GARCH evolution. The main pitfall of multivariate GARCH models in most specifications is the very large number of parameters, which rapidly makes the estimation unfeasible as the number of series increases. Those specifications which bypass this problem, on the other hand, pay the price in terms of a severe loss of generality<sup>1</sup>. Neither multivariate SV models, although relatively more parsimonious, are able to handle more than a few number of series because of their complexity of estimation<sup>2</sup>. For both streams of literature, the key for dimensionality reduction stands in the idea of the existence of a few latent variables, the so called *factors*, as driving forces for the whole dataset. Back to finance, models as CAPM explain theoretically why we may speak of factors in the market. Indeed, the use of factor models allows to disentangle within each stock the component which is directly linked to these common forces and the component which is peculiar to the stock itself. Doing this way, the factor analysis makes use of co-movements across stocks in order to improve forecasts.

Here we focus on the GARCH side of the story<sup>3</sup>. After the original ARCH and GARCH univariate specifications, respectively by Engle [1982] and Bollerslev [1986], many multivariate versions have been proposed, notably the VECH model of Bollerslev et al. [1988], the constant correlation model of Bollerslev [1990], the BEKK model of Engle and Kroner [1995] and the dynamic conditional correlation model by Engle [2000]. The idea of a factor approach to conditional heteroskedasticity has first been suggested by Engle [1987]. Soon after, Diebold and Nerlove [1989] have developed a static one-factor model on return series where the covariance matrix of factors is conditionally heteroskedastic, while the conditional covariance of the idiosyncratic part is homoskedastic. The estimation of the model is pursued by using a Kalman filter whose errors are found by maximizing the likelihood function. The authors prefer this simultaneous method to a two-step one in which static factors are extracted from the unconditional covariance matrix before being modelled as univariate GARCH processes. This latter method, with some modifications, is instead used by Engle et al. [1990] for a more complex model in the asset pricing context; Sentana [1998] proves that this model is nested in the previous by Diebold and Nerlove. Harvey et al. [1992] build a modified version of the Kalman

---

<sup>1</sup>See Bauwens et al. [2006].

<sup>2</sup>See Harvey et al. [1994].

<sup>3</sup>For multivariate SV models within the factor approach, see Chib et al. [2006].

filter for models with unobservable heteroskedastic factors; among the possible applications of the filter, they consider also the case of a dynamic factor model, which is called structural ARCH (STARARCH), and the aforementioned latent factor model by Diebold and Nerlove [1989]. The modified version of the Kalman filter is used by King et al. [1994], who build a static factor model with both observed and unobserved components. Such model is allowed to have a diagonal time-varying conditional covariance matrix of the idiosyncratic components, and is estimated by a two-step procedure in which factor loadings are first found by means of principal component analysis, while the static factors and their conditional covariances are next estimated by the modified Kalman filter with maximization of the likelihood function. More recently, the Orthogonal GARCH model by Alexander [2001], typically used for Value-at-Risk modelling, provides a method for obtaining large positive semi-definite conditional covariance matrices by modelling the principal components of the financial returns' unconditional covariance matrix as univariate GARCH processes. The GO-GARCH model by van der Weide [2002] generalizes the Orthogonal GARCH approach by allowing for the linear map that links components and observed data to be non orthogonal; a three-step estimation procedure has been proposed by van der Weide [2004].

We suggest a Dynamic Factor GARCH (DF-GARCH) model that might be able to explain and forecast the conditional covariances of a large number of series by means of a relatively small number of parameters. We assume that each series is composed of a common part, which depends on some dynamic factors, and an idiosyncratic part. We also assume that dynamic factors and idiosyncratic parts are conditionally heteroskedastic, where both heteroskedasticities evolve according to a GARCH rule. Our aim is predicting the conditional covariance matrix of all the series at time  $T + 1$ , that is for the first out-of-sample period. Whenever a factor structure lies behind observable data, the conditional variances and covariances of the observable series do not depend only on idiosyncratic elements, but partially derive from the conditional heteroskedasticity of the common factors. As a consequence, we can model the evolution of the conditional covariances of observable series by simply modelling the evolution of the conditional covariances of few factors, thus using a small number of parameters. Moreover, if the factors are dynamic, that is their influence on observable data is not only contemporaneous, the number of factors can be further reduced, and the parameters required for the estimation and forecast of conditional variance-covariance becomes even smaller. This feature of our model acquires more and more importance as the cross-dimension of the dataset becomes larger, because the number of parameters required by a multivariate GARCH to model and forecast conditional covariances would be overwhelming. At the same time, the existence of a factor structure that drives the movements of data cannot be ignored when predicting the conditional volatility of each observable series, as the common and the idiosyncratic parts of each series should be modelled separately.

For the estimation of the Dynamic Factor GARCH model, we propose a the two-step methodology that takes the same approach to factor analysis as in Giannone et al. [2005], in that we mix a first step based on non-parametric procedures, that provides an estimate of the parameters governing the factor dynamics, with a second step in which a Kalman filter corrects the estimate of dynamic and static factors. The framework has already been extended by Doz et al. [2005], who test the estimation consistency of a similar two-step procedure, and by Giannone et al. [2006], who apply the procedure to macroeconomic data for nowcasting GDP and inflation. The difference of our methodology lies in the explicit consideration we take of the dynamic factor conditional heteroskedasticity. Such conditional heteroskedasticity, which is

not unlikely in finance, can be modelled and used in order to forecast conditional variances and covariances of a large number of series, when the dynamics of our dataset depends on a small number of dynamic factors. We first operate an initial division of each series into common and idiosyncratic part by applying the Generalized Dynamic Factor Model (GDFM) by Forni et al. [2000]. The GDFM generalizes on the one hand the dynamic factor model proposed by Sargent and Sims [1977] and Geweke [1977] by allowing for mildly correlated idiosyncratic components; on the other hand the approximate factor model by Chamberlain [1983] and Chamberlain and Rothschild [1983] which is static. In the same stream of literature, Stock and Watson [2002] deal with forecasting issues, although in a macroeconomic context, by means of an approximate dynamic factor model which is estimated in a static way. Here the GDFM is initially applied through the one sided estimation method by Forni et al. [2006b], that provides us with an initial estimate of the static factors and of the common components of the observed series. We then apply a procedure that draws on Giannone et al. [2005] to obtain an estimate of the parameters that govern the links among dynamic factors, static factors and observed data; we also get a first estimate of the dynamic factors. By univariately modelling the idiosyncratic parts, we find their conditional means and variances. By multivariately modelling the estimated dynamic factors, we find the GARCH parameters linking them. We then run a modified Kalman filter on data series; this filter will use the parameters obtained up to this point, and the conditional variances of the idiosyncratic parts obtained by the previous univariate models. We thus get a new estimate of the dynamic factors. Once we retrieve the dynamic factors and their conditional covariances for the last period  $T$  of the sample, it is easy to build a prediction for the conditional covariance matrix of original series at time  $T+1$ .

In the next three sections we will describe respectively a heteroskedastic GDFM, its representation and its estimation. In sections 5 and 6, we will concentrate on the Kalman filter correction which takes into account the GARCH evolution of the dynamic factors' conditional covariance matrix and allows to provide a multivariate prediction. In section 7, our method is applied to a financial dataset. Final remarks are reported in section 8. Sometimes we will use the word "returns" when referring to the original data series and "volatilities" when referring to their covariances, as finance is the first field of application for our method.

## 2 The model

We denote as  $x_t = (x_{1t} \dots x_{Nt})'$  an  $N$ -dimensional vector process. Each of the series is stationary and second order moments  $\gamma_{ik} = E[x_{it}x'_{it-k}]$  exist finite for all  $i$  and  $k$ . In the Generalized Dynamic Factor Model (GDFM), as proposed by Forni et al. [2000], it is assumed that each series  $x_{it}$  can be written as the sum of two mutually orthogonal unobservable components, the common component  $\chi_{it}$  and the idiosyncratic component  $\xi_{it}$ . The common component is driven by a small number  $q$  of dynamic common factors  $u_{jt}$  with  $j = 1, \dots, q$ , which are loaded with possibly different coefficients and lags. Formally:

$$x_{it} = \chi_{it} + \xi_{it} = d_{i1}(L)u_{1t} + d_{i2}(L)u_{2t} + \dots + d_{iq}(L)u_{qt} + \xi_{it} \quad i = 1, \dots, N \quad (1)$$

The  $q$ -dimensional vector process  $u_t = (u_{1t} \dots u_{qt})'$  is an orthonormal white noise. The  $N$ -dimensional vector process  $\xi_t = (\xi_{1t} \dots \xi_{Nt})'$  has zero mean and is stationary. Moreover,  $\xi_{it}$  is orthogonal to  $u_{jt-k}$  for all  $k, i$  and  $j$ . The polynomials in the lag operator  $d_{i1}(L) \dots d_{iq}(L)$  are square-summable, one-sided filters in principle of infinite order.

We assume that the elements of the spectral density matrix of the sample are bounded, and we denote the spectral density matrices of the common part and the idiosyncratic part respectively as  $\Sigma^x(\theta)$  and  $\Sigma^\xi(\theta)$ , with  $\theta \in [-\pi, \pi]$ . Finally, we assume that the  $q$  eigenvalues of  $\Sigma^x(\theta)$  diverge almost everywhere as the number of series goes to infinity, while the largest eigenvalue of  $\Sigma^\xi(\theta)$  is uniformly bounded for  $\theta \in [-\pi, \pi]$ . This last condition, in other words, relaxes the assumption of mutual orthogonality of idiosyncratic components by allowing for a limited amount of cross-sectional correlation.

Throughout the paper, we will focus on the particular case in which dynamic factors are heteroskedastic and distributed as:

$$u_{t|t-1} \sim N(0, Q_t) \quad (2)$$

$Q_t$  being a non-diagonal matrix changing over time. The conditional variance of the dynamic factors is governed by the following equation:

$$Q_t = C_0' C_0 + C_1' u_{t-1} u_{t-1}' C_1 + C_2' Q_{t-1} C_2 \quad (3)$$

Here we adopt the full BEKK representation of multivariate GARCH models as in Engle and Kroner [1995]; we also suppose that in empirical cases the number of dynamic factors should be very small. We also assume that idiosyncratic parts evolve according to a univariate ARMA-GARCH model:

$$\xi_{t|t-1} \sim N(\mu_t, R_t) \quad (4)$$

where, for each time, the diagonal matrix  $R_t$  contains the conditional covariances of each idiosyncratic series:

$$R_{it} = \rho_{0i} + \rho_{1i} \xi_{it-1} + \rho_{2i} R_{it-1} \quad (5)$$

We do not define a particular structure of the ARMA process governing the conditional mean of  $\xi$ . Indeed, as our first empirical application deals with financial returns, whose idiosyncratic parts are unlikely to have strong dynamics, the empirical section of this paper will not consider any ARMA structure of the idiosyncratic components. However, the general description of our procedure will allow for this possibility.

### 3 VAR representation of factor dynamics

In a factor model the  $N$ -dimensional variable considered  $x_t$  can be written as the sum of a common part  $\chi_t$  and an idiosyncratic part  $\xi_t$ . In a static framework the common part of the factor model is made of  $r \ll N$  common factors, thus the factor model is written as:

$$x_t = \chi_t + \xi_t = \Lambda F_t + \xi_t \quad (6)$$

where  $F_t = (F_{1t} \dots F_{rt})'$  is the  $r$ -vector of common factors and  $\Lambda$  is a  $N \times r$  matrix of loadings. We assume that  $F_t$  is driven by  $q < r$  common shocks and it has the following VAR(1) representation:

$$F_t = A F_{t-1} + B u_t \quad (7)$$

where  $u_t = (u_{1t} \dots u_{qt})'$  is the  $q$ -vector of common shocks,  $A$  is  $r \times r$  and  $B$  is  $r \times q$ . The common shocks are precisely the dynamic factors that we want to estimate. Before doing this, we show that the static model of equation (6) contains a special case of the usual dynamic

factor model (see Forni et al. [2006a] and Bai and Ng [2005]). Indeed, by inverting equation (7) and using the lag operator  $L$ , we have:

$$F_t = (\mathbf{I} - AL)^{-1}Bu_t$$

Hence we have immediately the dynamic factor model:

$$x_t = \Lambda(\mathbf{I} - AL)^{-1}Bu_t + \xi_t = D(L)u_t + \xi_t$$

where

$$D(L) = \Lambda \left( \sum_{k=0}^{\infty} H_k L^k \right) B = \sum_{k=0}^{\infty} D_k L^k$$

where  $D_k$  are  $N \times q$ . It's clear that in this model the common part has an MA( $\infty$ ) representation that allows for both MA and AR loading of the dynamic factors, which is an important generalization with respect to the static model by Stock and Watson [2002]. We now assume that each dynamic factor is loaded together with its own lags and we keep a maximum of  $s$  lags. Then if we define  $r = q(s + 1)$  and

$$\begin{aligned} F_t &= (u'_t, u'_{t-1} \dots u'_{t-s})' \\ \Lambda &= (D_0 \dots D_s) \end{aligned}$$

we have that the static factors are just the dynamic factors with all their lags and we have recovered the usual Generalized Dynamic Factor Model. This proves that the VAR(1) representation is general enough for us.

## 4 Estimating the linear model

For the estimation of the GDFM, we follow the two-step procedure proposed in Forni et al. [2006b]. In the first step the spectral density matrix of  $x_t$ ,  $\Sigma^x(\theta)$ , is estimated by applying the Fourier transform to the sample auto-covariance matrices  $\hat{\Gamma}_k^4$ . Then the dynamic principal component decomposition is applied, thereby selecting the first  $q$  largest eigenvalues of  $\hat{\Sigma}^x(\theta)$  and the corresponding eigenvectors. After calling  $\Phi(\theta)$  the matrix having on its diagonal the  $q$  largest eigenvalues of  $\hat{\Sigma}^x(\theta)$ , and  $\Psi(\theta)$  the matrix containing the corresponding eigenvectors as columns, we compute the estimated spectral density matrix of  $\chi_t$  as:  $\hat{\Sigma}^\chi(\theta) = \Psi(\theta)\Phi(\theta)\Psi(\theta)'$ . It's worth noticing at this point the key difference between this dynamic approach and the static principal component method used by Stock and Watson [2002]: while the first exploits the information contained in lagged covariance matrices, the latter makes use of contemporaneous covariances only. By applying to  $\hat{\Sigma}^\chi(\theta)$  the inverse Fourier transform we retrieve estimates of the covariance matrices of the common component,  $\hat{\Gamma}_k^\chi$ ; the estimate of the covariance matrices of  $\xi_t$ ,  $\hat{\Gamma}_k^\xi$ , is obtained by difference. In the second step of the procedure we move to a static representation of the model in which we estimate the first  $r$  generalized eigenvectors of  $\hat{\Gamma}_k^\chi$  with respect to  $\hat{\Gamma}_k^\xi$ . The first generalized eigenvector solves the following maximization problem:

$$\begin{cases} z^{(1)} &= \operatorname{argmax}_{a \in \mathbb{R}^n} a' \hat{\Gamma}_0^\chi a \\ \text{s.t.} & a' \hat{\Gamma}_0^\xi a = 1 \end{cases} \quad (8)$$

---

<sup>4</sup>Later on, hatted symbols will indicate estimates.

We collect the first  $r$  generalized eigenvectors in the matrix  $Z = (z^{(1)} \dots z^{(r)})$  and by means of such matrix and of the contemporaneous covariance matrix, estimated in the first step, we are able to estimate the common component as:

$$\hat{\chi}_t = \hat{\Gamma}_0^x Z (Z' \hat{\Gamma}_0 Z)^{-1} Z' x_t \quad \forall t = 1, \dots, T \quad (9)$$

We obtain the idiosyncratic component simply as difference between the original series  $x_t$  and  $\hat{\chi}_t$ . Indeed, the one-sided estimator allows to forecast the common component at  $T + h$  by substituting the estimated lagged covariances  $\hat{\Gamma}_h^x$  to the contemporaneous covariance  $\hat{\Gamma}_0^x$  in (9). Following the results of the previous section, we restate the equations from which we started<sup>5</sup>:

$$x_t = \Lambda F_t + \xi_t = D(L)u_t + \xi_t \quad (10)$$

$$F_t = A F_{t-1} + B u_t \quad (11)$$

Using the one-sided estimator proposed by Forni et al. [2006b] we have an estimation of the common part and also an estimation of the  $r$  static common factors, at least up to a rotation, i.e. we actually get the generalized principal components  $G_t = \Omega F_t$  with  $\Omega \Omega' = \mathbf{I}$ . Hereafter we suppose that  $\Omega = \mathbf{I}$ , since  $F_t$  and  $G_t$  span the same space and we are not interested in recovering the “true” static factors; investigating the influence that a factor rotation would have on our estimation procedure is left to further research. Summarizing from the one-sided estimator, we have:

$$\begin{aligned} \hat{\chi}_t & \quad \text{the estimated common part} \\ \hat{\xi}_t & \quad \text{the estimated idiosyncratic part} \\ \hat{\Gamma}_0^x & = E[\hat{\chi}_t \hat{\chi}_t'] \\ \hat{\Gamma}_0^\xi & = E[\hat{\xi}_t \hat{\xi}_t'] = \hat{\Gamma}_0^x - \hat{\Gamma}_0^x \\ \hat{F}_t & \quad \text{the estimated generalized principal components} \end{aligned}$$

The following part of the section draws on the procedure shown by Giannone et al. [2005]. From (11), we have an estimation for  $\Lambda$  given a sample length  $T$ :

$$\hat{\Lambda} = \sum_{t=1}^T \hat{\chi}_t \hat{F}_t' \left( \sum_{t=1}^T \hat{F}_t \hat{F}_t' \right)^{-1} \quad (12)$$

In the previous regression, we use  $\chi$  instead of  $x$  to get rid of the heteroskedasticity coming from the idiosyncratic part. Analogously, we have an estimation for  $A$ :

$$\hat{A} = \sum_{t=2}^T \hat{F}_t \hat{F}_{t-1}' \left( \sum_{t=2}^T \hat{F}_{t-1} \hat{F}_{t-1}' \right)^{-1} \quad (13)$$

and we can also estimate the covariance of  $B u_t$  which is given by:

$$\hat{\Gamma}_0^{Bu} = \frac{1}{T-1} \sum_{t=2}^T \hat{F}_t \hat{F}_t' - \hat{A} \left( \frac{1}{T-1} \sum_{t=2}^T \hat{F}_{t-1} \hat{F}_{t-1}' \right) \hat{A}'$$

The estimation of  $A$ , as given by (13), is not efficient. The equation (11) not only represents an autoregression of the dependent variable, but also shows an error term  $B u_t$  which is a linear

---

<sup>5</sup>The whole procedure has to be applied on standardized and covariance stationary data.

combination of GARCH processes; as a consequence, the autoregression of the static factors  $F_t$  involves a weak GARCH evolution of the error (see Nijman and Sentana [1996]). However, the unbiasedness of this estimation and the limit we must impose to the number of parameters drive us towards the decision of using equation (13).

Consider now the matrix  $S$ , that has the  $q$  largest eigenvalues of  $\hat{\Gamma}_0^{Bu}$  on its diagonal, while the corresponding eigenvectors are in the matrix  $M$ . Then we have:

$$\hat{\Gamma}_0^{Bu} = E[Bu_t u_t' B'] = BB' = MSM' = MS^{1/2} S^{1/2} M'$$

from which we get an estimation of  $B$ :

$$\hat{B} = MS^{1/2} \tag{14}$$

By inverting (11) and using the estimations of  $F_t$ ,  $A$ ,  $B$  and  $\Lambda$  we have an estimate for  $u_t$ :

$$\hat{u}_t = S^{-1/2} M' (\mathbf{I} - \hat{A}L) \hat{F}_t \tag{15}$$

## 5 Correcting for heteroskedasticity

### Modelling the covariances of dynamic factors and idiosyncratic parts

Now we apply a modified version of the Kalman filter that will provide us with an estimation of the static factor conditional covariances, which depend on the GARCH specification of the dynamic factors' dynamics. While the factors, together with their covariances, are re-estimated by the filter, the parameters of the linear part of the model remain fixed; therefore, the Kalman filter operates without a final likelihood maximization. The filter can only be quasi-optimal because, at each step, past disturbances are not observable, and therefore we are not sure that the distribution of current disturbances is conditionally Gaussian (see the comments by Harvey et al. [1992] about a different modification of the Kalman filter).

Before starting the filter, we apply a univariate ARMA+GARCH model to each time series of  $\xi$  as in (5), and so we obtain conditional mean  $\hat{\mu}_t$  and variance  $\hat{R}_t$  for each idiosyncratic series and for each time. On the other side, we apply a multivariate GARCH model to the estimate of the dynamic factors  $\hat{u}$  obtained previously using the GDFM, that is we apply on  $\hat{u}$ 's the model described by (3) and we get estimates of the parameters  $\hat{C}_0$ ,  $\hat{C}_1$ ,  $\hat{C}_2$  as well as of the dynamic factors' conditional covariances  $\hat{Q}_t \forall t$ .

### New State-Space Model

In order to concentrate on the heteroskedastic process governing the common part of each series of returns, we do not use the return  $x_t$  as such as the observable variable of the model, but we use a new variable  $\tilde{x}_t = x_t - \mu_t$  that is obtained as the return minus the conditional mean of the old idiosyncratic part. As a consequence, the error term  $\xi$  of the model (11) will be replaced by a new idiosyncratic part  $\varepsilon_t$ , whose process has conditional mean equal to zero. Our new state-space model is:

$$\tilde{x}_t = \hat{\Lambda} F_t + \varepsilon_t \quad \text{measurement equation} \tag{16}$$

$$F_t = \hat{A} F_{t-1} + \hat{B} u_t \quad \text{transition equation} \tag{17}$$

where

$$\begin{aligned} \varepsilon_{t|t-1} &\sim N(0, \hat{R}_t) \quad \hat{R}_t \text{ diagonal} \\ u_{t|t-1} &\sim N(0, Q_t) \\ Q_t &= \hat{C}'_0 \hat{C}_0 + \hat{C}'_1 u_{t-1} u'_{t-1} \hat{C}_1 + \hat{C}'_2 Q_{t-1} \hat{C}_2 \end{aligned} \tag{18}$$

The conditional heteroskedasticity of dynamic factors comes into the model through (18), in which the parameter matrices  $\hat{C}_0$  and  $\hat{C}_1$  (that have already been estimated on  $\hat{u}$ 's) govern a multivariate GARCH process of the  $u$ 's.  $Q_t$  is the corresponding conditional covariance matrix of the dynamic factors  $u_t$ , while  $\hat{R}_t$ 's have been obtained by univariate GARCH's estimated on  $\hat{\xi}_t$ , as said previously. All the other parameters ( $\hat{\Lambda}$ ,  $\hat{A}$  and  $\hat{B}$ ) have been obtained by the GDFM one-sided estimation and the subsequent operations we have already described. These parameters will be kept fixed over time, while running the Kalman filter. On the other hand, we need a safer estimation for the static and dynamic factors (and their covariances), which are varying over time, because our prediction strongly depends on their estimation for the last period of our sample. That is why we run a modified version of the Kalman filter and obtain new estimates of static and dynamic factors for each period of the sample. At this point, one could implement an EM algorithm as in Doz et al. [2006], by using the new estimate of static factors as a starting point for re-applying the procedure described in sections 4 and 5. The empirical test of our methodology avoids such implementation in order to focus the attention on just the main intuitions of our procedure. In the remaining part of this section, it is explained in detail how our modification of the Kalman filter works.

### Kalman Filter Initialization

Initial values are built as:

$$\begin{cases} F_{1|1} & = \hat{F}_1 \\ P_{1|1} & \text{sufficiently large} \\ E_1(u_1) & = \hat{u}_1 \\ E_1(Q_1) & = \hat{Q}_1 \\ E_1(u_1 u_1') & = (E_1(u_1))(E_1(u_1))' + E_1(Q_1) , \end{cases} \quad (19)$$

where the variables with the hat have been obtained during the previous GDFM estimation,  $\hat{Q}_1$  has been obtained by the multivariate GARCH model, and the state initial covariance matrix  $P_{1|1}$  must represent the high uncertainty about the initial value of the state vector.

### Kalman Filter Prediction

The steps described in this and the following section must be repeated together for time  $t = 2 \dots T$ . First we predict the unobserved state vector

$$F_{t|t-1} = \hat{A}F_{t-1|t-1} \quad (20)$$

and its conditional covariance matrix

$$P_{t|t-1} = \hat{A}P_{t-1|t-1}\hat{A}' + \hat{B}E_{t-1}(u_t u_t')\hat{B}' , \quad (21)$$

where

$$\begin{cases} E_{t-1}(u_t u_t') & = E_{t-1}(Q_t) \\ E_{t-1}(Q_t) & = \hat{C}_0' \hat{C}_0 + \hat{C}_1' E_{t-1}(u_{t-1} u_{t-1}') \hat{C}_1 + \hat{C}_2' E_{t-1}(Q_{t-1}) \hat{C}_2 . \end{cases} \quad (22)$$

The conditional covariance matrix for the state vector is obtained by using the GARCH estimated parameters  $\hat{C}_0$ ,  $\hat{C}_1$  and  $\hat{C}_2$ ; they are applied on the updated conditional covariance of the transition error  $(u_{t-1} u_{t-1}')$ , which in turn has been obtained by the Kalman update, as we will see in the next step.

The prediction error is given by

$$\eta_{t|t-1} = \tilde{x}_t - \tilde{x}_{t|t-1} = \tilde{x}_t - \hat{\Lambda}F_{t|t-1}, \quad (23)$$

whose conditional covariance is built by using the predicted conditional covariance of the static factors and the known conditional covariance of the measurement errors, as obtained previously by univariate modelling of the idiosyncratic parts:

$$W_{t|t-1} = \hat{\Lambda}P_{t|t-1}\hat{\Lambda}' + \hat{R}_t. \quad (24)$$

### Kalman Filter Update

Finally, we compute the Kalman gain

$$K_t = P_{t|t-1}\Lambda'W_{t|t-1}^{-1} \quad (25)$$

and we build more accurate inferences, exploiting information up to time  $t$ , as in the following:

$$F_{t|t} = F_{t|t-1} + K_t\eta_{t|t-1} \quad (26)$$

$$P_{t|t} = P_{t|t-1} - K_t\hat{\Lambda}P_{t|t-1}. \quad (27)$$

By inverting the transition equation and recalling (14), we get

$$\mathbf{E}_t(u_t) = S^{-1/2}M'(\mathbf{I} - \hat{A}L)F_{t|t} \quad (28)$$

and then

$$\mathbf{E}_t(u_t u_t') = (\mathbf{E}_t(u_t))(\mathbf{E}_t(u_t))'. \quad (29)$$

Equation (29), when put in the context of the following prediction step (22), is not precise. As noted by Harvey et al. [1992], a correction term should be added on the right hand side in order to take into account the conditional variance of the dynamic factor. However, the same authors show that, when applied to the factor model by Diebold and Nerlove [1989], the effect of this correction may be empirically negligible. The differences between their estimation procedure and ours, including the update passage described in (28), let us prefer avoiding the estimation of the correction term.

### Smoothing

Smoothing would be especially useful when extending our procedure to a higher number of lags in the GARCH structure of dynamic factors' conditional covariances. In any case, the smoothing procedure is recommended for getting a more precise estimate of the common and idiosyncratic components of the dataset, as it will be described in the next section. Following de Jong [1989] and Durbin and Koopman [2001], the following fixed interval smoother can be applied for  $t = T, T - 1, \dots, 2$  in order to find more precise in-sample values of the static factors and of dynamic factors' conditional covariances. First we compute:

$$r_{t-1} = L_t' r_t + \Lambda' W_{t|t-1}^{-1} \eta_{t|t-1} \quad (30)$$

$$F_{t|T} = F_{t|t-1} + P_{t|t-1} r_{t-1} \quad (31)$$

where  $L_t = A(\mathbf{I} - K_t\Lambda)$ ,  $r_T = 0$ . At each step, we also find the smoothed state variance matrix:

$$P_{t|T} = P_{t|t-1} - P_{t|t-1}\Theta_{t-1}P_{t|t-1}, \quad (32)$$

where  $\Theta_t$  has been obtained by

$$\Theta_{t-1} = \Lambda'W_t^{-1}\Lambda + L_t'\Theta_tL_t \quad (33)$$

with initial value  $\Theta_T = 0$ . At the end of each step, we get smoothed values for the dynamic factors and their conditional covariances  $Q_t$  as in the following:

$$E_T(u_t) = E_{t-1}(Q_t)B'r_t \quad (34)$$

$$E_T(Q_t) = E_{t-1}(Q_t) - E_{t-1}(Q_t)B'\Theta_tBE_{t-1}(Q_t). \quad (35)$$

## 6 Forecasting

We consider  $\chi_t^* = \hat{\Lambda}F_{t|T}$  as the new estimated common part of each return observation  $x_t$ ; by difference we obtain a new idiosyncratic part  $\xi^* = x - \chi^*$ . We assume that  $\xi_t^*$  are the realizations of a process having the conditional means  $\hat{\mu}_t$  and the conditional variances  $\hat{R}_t$  that we found before applying the filter by univariate GARCH modelling of  $\hat{\xi}$ . The predicted diagonal covariance matrix  $\Gamma_{T+1|T}^{\xi^*}$  for  $\xi^*$  at time  $T+1$  is obtained by applying to  $\xi_T^*$  (excluding the first period) the parameters of the univariate GARCH models that we have run onto  $\hat{\xi}$  before the Kalman filter. The predicted covariance matrix for the common part will be equal to

$$\Gamma_{T+1|T}^{\chi^*} = \hat{\Lambda}P_{T+1|T}\hat{\Lambda}', \quad (36)$$

where  $P_{T+1|T}$  has been obtained by applying the prediction steps (22) and (21) to the smoothed estimations of last period  $T$ . Given the assumed orthogonality of common and idiosyncratic part, we get the one-step-ahead out-of-sample conditional covariance of the observable data series as:

$$\Gamma_{T+1|T}^x = \Gamma_{T+1|T}^{\chi^*} + \Gamma_{T+1|T}^{\xi^*}. \quad (37)$$

A forecast for the terms of the conditional correlation matrix can now be obtained as:

$$\rho_{ij,T+1|T}^x = \frac{\gamma_{ij,T+1|T}^x}{\sqrt{\gamma_{ii,T+1|T}^x\gamma_{jj,T+1|T}^x}} \quad (38)$$

Notice that in (37) the off-diagonal terms of  $\Gamma_{T+1|T}^x$  do not consider the mild correlations (and conditional correlations) among the idiosyncratic terms. Correlation estimations given by (38) might thus be biased; for instance, conditional correlations will be often underestimated in a dataset for which the conditional covariance among idiosyncratic parts is often positive. We do not face this problem if we are just interested in the conditional correlations arising from the common factor dynamics; in this case, conditional correlations are computed by using just the conditional covariances of the common part, without the step described in (37).

## 7 Empirics

### 7.1 The data

The dataset we use for the empirical investigation includes all the transaction prices of the 89 stocks traded on the London Stock Exchange and participating in the construction of the FTSE100 index for the whole considered time span, that is from 1<sup>st</sup> October 2001 to 31<sup>st</sup> July 2003 (457 working days). Transaction prices have been cleaned from outliers by using the procedure described in Brownlees and Gallo [2006]<sup>6</sup>; we choose respectively 60 and 0.02 as neighborhood and granularity parameters. Returns have been computed by using the last transaction recorded each day before the closing time of the LSE. Daily realized volatilities and covolatilities for out-of-sample evaluation are computed on a 5-minute frequency after removing the first 15 minutes of each day, as Barndorff-Nielsen and Shephard [2005] have done on LSE data in order to avoid open effects. When computing realized covolatilities, we do not use leads and lags of intra-daily returns, as the 5-minute frequency should be low enough to avoid the non-synchronicity bias (see Martens [2004]).

Firstly, we verify that our dataset does fulfill GDFM assumptions on the eigenvalues  $\lambda_i(\theta)$  of the spectral density matrix of  $x_t$ . According to Brillinger [1981], we define the variance explained by the  $i^{\text{th}}$  factor as:

$$EV_i = \frac{\int_{-\pi}^{\pi} \lambda_i(\theta) d\theta}{\sum_{j=1}^N \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta} \quad (39)$$

We require that, as  $N \rightarrow \infty$ :

$$\left\{ \begin{array}{ll} EV_i \rightarrow \infty & \text{for } i = 1, \dots, q \\ \exists M \in \mathbb{R}^+ & \text{s.t. } EV_i \leq M \quad \text{for } i = q + 1, \dots, N \end{array} \right. \quad (40)$$

Indeed, as shown for example in figure 1.1 for the subsample including only the first 350 observations, this is the case. Figure 1.2 shows the cumulated explained variance relative to the first two eigenvalues for the same sample.

### 7.2 The number of dynamic factors

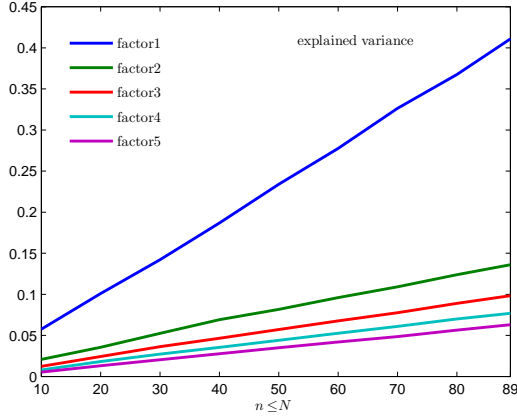
We do not rely only on the intuition coming from the graphs for determining the number of dynamic factors to include in the model, but we apply the Hallin and Liska [2007] information criterion for determining the minimum number of dynamic factors that satisfy the hypotheses of GDFM. The criterion by Hallin and Liška exploits the relation in the GDFM between the number of dynamic factors and the number of diverging eigenvalues of the spectral density matrix of the observations; we choose the logarithmic form<sup>7</sup> of the covariogram-smoothing version of the criterion. For given  $N$  and  $T$ , it consists in choosing the number of factors  $\hat{q}_N^T$  so to minimize the following:

$$IC(q_N^T) = \log \left[ \frac{1}{N} \sum_{i=q+1}^N \frac{1}{2M_T + 1} \sum_{h=-M_T}^{M_T} \lambda_{Ni}^T(\theta_h) \right] + q \, c p(N, T), \quad 0 \leq q_N^T \leq q_{max} < N \quad (41)$$

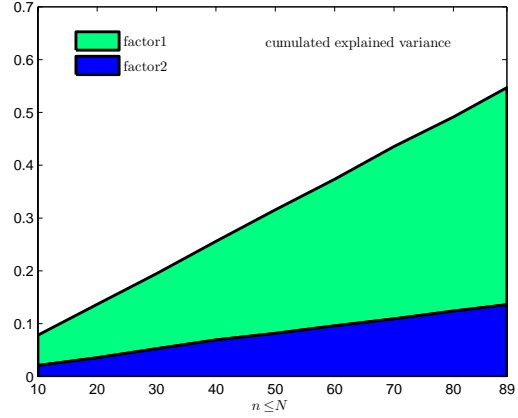
---

<sup>6</sup>All computations for this section have been performed by using the standard Matlab software package plus the freely available toolboxes “Matnem” by Christian T. Brownlees and “UCSD Garch” by Kevin K. Shephard; the Matlab code used in the next subsection has been kindly provided by Roman Liška.

<sup>7</sup>Hallin and Liška suggest that this form has better finite sample performance.

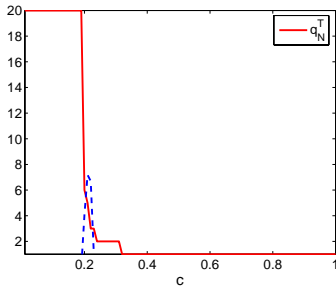


1.1: Explained variance.

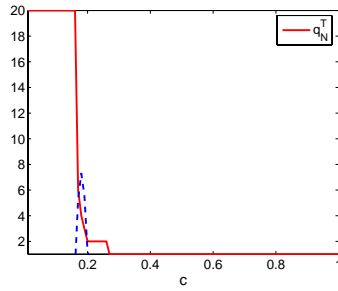


1.2: Cumulated explained variance.

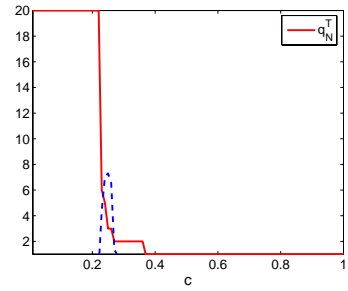
Figure 1: Plots of diverging variances as  $N \rightarrow \infty$



2.1: First penalty function.



2.2: Second penalty function.



2.3: Third penalty function.

Figure 2: Hallin-Liška criterion plots for different penalty functions.

where  $M_T$  is a truncation lag parameter satisfying  $M_T \rightarrow \infty$  and  $M_T/T \rightarrow 0$  as  $T \rightarrow \infty$  (for instance, we choose  $M_T \sim \lceil \sqrt{T}/2 \rceil$ ),  $\theta_h = \pi h / (2M_T + 1)$  are the  $2M_T + 1$  points for which the spectral density is estimated, and  $p(N, T)$  is a penalty function satisfying

$$\lim_{N \rightarrow \infty} p(N, T) = 0 \quad \text{and} \quad \lim_{N \rightarrow \infty} Np(N, T) = \infty. \quad (42)$$

In principle, the maximum number of factors allowed  $q_{max}$  is the number of series in the dataset. Therefore, the penalty function should be large enough to avoid overestimation of  $\hat{q}_N^T$ , but at the same time it should not overpenalize. Multiplying the penalty function by a constant  $c$  is a way to tune the penalizing power of  $p(N, T)$ .

Hallin and Liška propose an automatic procedure for selecting  $\hat{q}_N^T$  which basically explores the behavior of the variance of the selected  $q_N^T$  for the whole region of values of the constant  $c$  for  $N$  and  $T$  going to infinity. What we seek is the first stability region compatible with  $\hat{q}_N^T < N$ . In figure 2, relative to the subsample of the first 350 daily returns of the 89 stocks, the solid lines indicate the value of  $q_N^T$  while the dashed lines represent the variance of  $q_N^T$ , thus detecting stability intervals. In other words,  $\hat{q}_N^T$  corresponds to the second plateau of the solid line associated with a flat zero-level dashed line. For all the three penalty functions suggested by Hallin and Liška, the criterion thus indicates the existence of two common dynamic factors.

### 7.3 The number of static factors

In order to find the number  $r$  of static factors, we compute the six consistent criteria suggested by Bai and Ng [2002] and the criterion by Onatski [2006]. All Bai-Ng's criteria look for the number of static factors that minimizes the mean squared distance between observed data and their common part as estimated by a static principal component analysis. The mean squared distance is computed for all the possible numbers of static factors between 0 and  $r_{max}$ , and is counterbalanced by a penalty function in order to avoid the criterion to point a number of static factors higher than the true one. Although both  $PC$  and  $IC$  criteria need a maximum number of factors  $r_{max}$  as an input, only  $PC$  criteria explicitly take into account its resulting minimum squared distance.

When applied to the first 350 multivariate observations of our sample, some criteria show difficulties in finding convergence or are sensitive to variations in the maximum number of factors fed to the algorithm. Therefore, we record the suggested number of factors for each criterion and for each  $r_{max}$ , showing results for  $PC$  and  $IC$  criteria respectively in figures 3 and 4. For high  $r_{max}$ , the suggested number of factors corresponds to the maximum possible and the graph thus follows a 45 degrees line. When  $r_{max}$  becomes smaller, the line becomes convex in the cases of  $PC$  criteria, and flat for  $IC$  criteria.  $IC$  criteria indicate the presence of two static factors, which would imply no dynamics in our sample as  $r = q(s + 1)$  where  $s$  is the number of lags of dynamic factors. The same suggestion would come from  $PC$  criteria, if we decide to use the rule by Schwert [1989], modified as in Bai and Ng [2002], for which  $r_{max} = 8[(\min\{N, T\}/100)^{1/4}]$ . However, Bai and Ng [2002] use a larger  $r_{max}$  (equal to 15) for an empirical application to financial data, and Forni et al. [2006a] choose  $r_{max} = 30$ , when using a sample with the same number of series of (and a smaller time-dimension than) our sample. If we choose  $r_{max} = 30$ ,  $PC$  criteria suggest a number of static factors between 7 and 14. In the same range, we can find the first plateaux of  $PC$  curves when lowering the  $r_{max}$  from 89 to 1, that is we can find the first intervals in which  $PC$  results tend to stabilize.

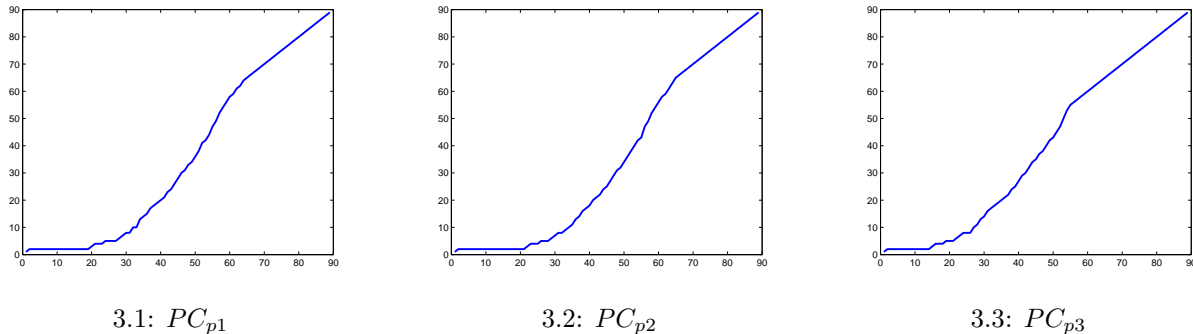
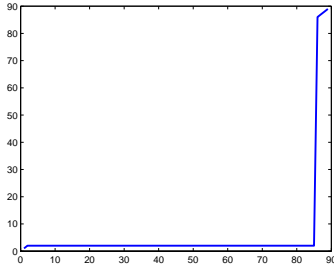
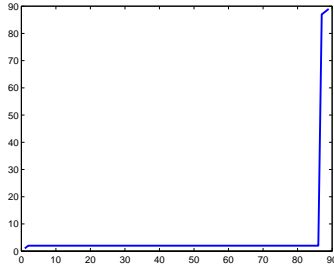


Figure 3: Number of static factors (vertical axis) suggested by PC criteria for all possible  $r_{max}$ 's (horizontal axis).

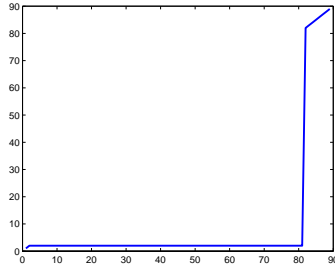
When we use the criterion by Onatski [2006], doubts persist about the right number of static factors for our sample. As for the previous criteria, we need a maximum number of factors  $r_{max}$  as an input, and the suggested number of factors corresponds to the plateau of a diagram in which the horizontal axis represents  $r_{max}$ . In figure 5, there are two plateaus, suggesting respectively the existence of 2 or 19 static factors.



4.1:  $IC_{p1}$



4.2:  $IC_{p2}$



4.3:  $IC_{p3}$

Figure 4: Number of static factors (vertical axis) suggested by IC criteria for all possible  $r_{max}$ 's (horizontal axis).

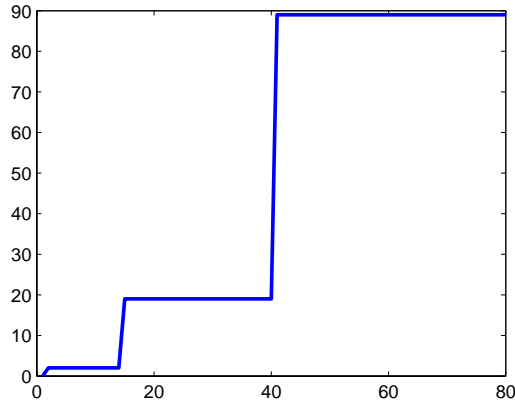
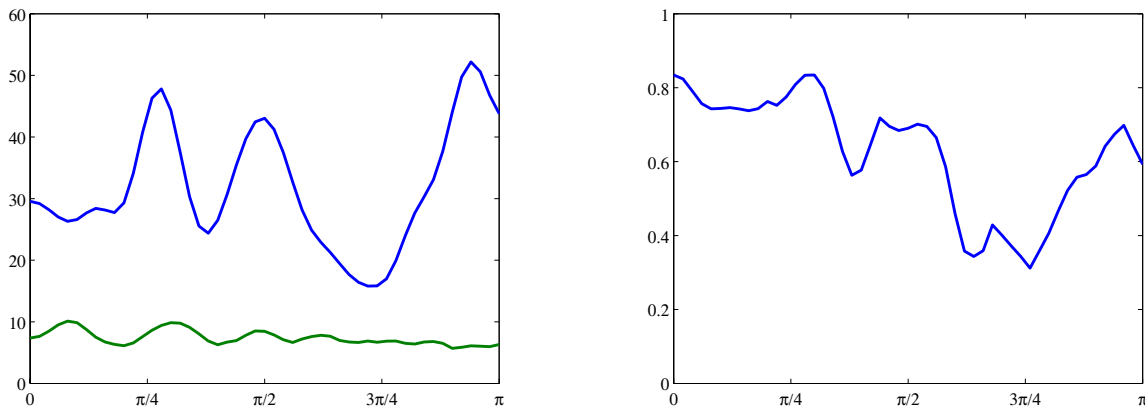


Figure 5: Number of static factors (vertical axis) suggested by Onatski's criterion for all possible  $r_{max}$ 's (horizontal axis).

An argument against the suggestion of two static factors can be found in the results of the previous subsection. The Hallin-Liška criterion previously described provide us with some additional information which could not be considered in the method by Bai and Ng [2005], who could look for the number of dynamic factors only after finding the number of static factors. By means of the Hallin-Liška criterion we have detected the number of dynamic factors as being equal to two, and the two largest eigenvalues of the spectral density matrix of our sample, computed at different frequencies, indicate that those two dynamic factors are able to explain 44.20% of the total sample variance. On the other side, if we look at the eigenvalues of the sample variance-covariance matrix, two static factors would be able to explain only 33.84% of the total sample variance, while we would need 9 static factors to explain the same percentage of variance that is explained by the two dynamic factors. In figure 6.1 we plot the two largest eigenvalues of the sample spectral density matrix, computed at different frequencies. The pattern of peaks and troughs followed by the first eigenvalue at different frequencies indicate the high influence that the factors exert on returns at different specific frequencies, thus supporting the idea that the factors are really dynamic. Figure 6.2 shows the spectral density of the common part of one return series from our dataset. Such common part is computed by using the two largest eigenvalues of the spectral density matrix of the whole sample. At each frequency, the spectral density of the common part has been divided by the total spectral density of the series, in order to plot relative values. We can easily prove that the spectral

density of the common part would be flat if the factors were not truly dynamic. Suppose that all the common components corresponding to the different return series of the dataset depend on just one factor. If there is dynamics in the returns then the factor is loaded with its lags, and we have  $\chi_t = D(L)u_t$ . In this case, the spectral density is  $\Sigma^x(\theta) = D(e^{i\theta})\Sigma^u D(e^{-i\theta})'$ . Notice that  $\Sigma^u$  is constant as  $u_t$  is assumed to be white noise, but  $D(e^{i\theta})$  cannot be constant over the frequency domain if the factor is loaded dynamically; therefore, the eigenvalues of  $\Sigma^x(\theta)$  cannot be constant either. On the other hand, if the factor is loaded with no lags, i.e. it is a static factor, the common part becomes  $\chi_t = Du_t$ , where  $D$  is a fixed parameter matrix that does not include any lag operator. In this case, the spectral density of  $\chi_t$  would be equal to a matrix  $D\frac{\sigma_{u_t}^2}{2\pi}D'$  that is constant over the frequency domain. Extending this result to a multi-factor framework, non-constancy over the frequency domain of the eigenvalues of the spectral density matrix represents a sign of common factors' dynamics. Actually, this is the case of our dataset, as shown in figure 6.



6.1: First two largest eigenvalues of the spectral density matrix. Frequencies on the horizontal axis.

6.2: Spectral density of the common part of the first series of the LSE dataset. Frequencies on the horizontal axis.

Figure 6: Evidence of common factors' dynamics.

Indeed, if we apply to our sample the one-sided estimator by Forni et al. [2006b], by using a number of static factors equal to 10, and we get a first estimation of the static factors, we can look for factors' dynamics by performing a Ljung-Box autocorrelation test. Results of the test are reported in table 1, where each number "1" indicates the presence of autocorrelation at significance 0.05 for a given static factor and a given number of lags. There is only one static factor for which the test would suggest no dynamics. As a consequence, we choose, consistently with Bai-Ng's *PC* criteria, a number of static factors  $r$  equal to 10, corresponding to four lags ( $s = 4$ ) of the dynamic factors, as  $r = q(s + 1)$ . However, results obtained by performing the DF-GARCH estimation in a static way, thus with only two static factors, will be used as a benchmark for evaluating the prediction accuracy of the DF-GARCH with two dynamic factors and four lags. The other benchmark will be represented by the traditional univariate GARCH model.

	1 lag	2 lags	3 lags	4 lags	5 lags
first static factor	0	0	1	1	1
second static factor	1	1	1	1	1
third static factor	0	0	0	0	0
fourth static factor	0	1	1	1	1
fifth static factor	0	1	1	1	1
sixth static factor	1	1	1	1	1
seventh static factor	1	1	1	1	1
eighth static factor	0	0	1	1	1
ninth static factor	1	1	1	1	1
tenth static factor	0	1	1	1	1

Table 1: Ljung-Box test on static factors (significance level = 0.05) for different numbers of lags of the sample autocorrelation function included in the Q-statistic; the number 1 indicates serial correlation

## 7.4 Volatility prediction

We want to compare the accuracy of the Dynamic Factor GARCH model and of the traditional GARCH (1,1) in predicting the conditional variance of all the series. The DF-GARCH will be used both in the proper way (two dynamic factors and four lags) and in a static way (two dynamic factors corresponding to just two static factors with no lags); whenever we name just “DF-GARCH”, we will refer to the proper dynamic estimation of the model. Returns are taken from 350 consecutive working days to infer a one-step-ahead volatility forecast; we follow a rolling scheme, for which our in-sample time span is fixed, while the forecast evaluation period includes the observations 351 up to 450 of the original dataset (100 one-step-ahead predictions). At each iteration, the DF-GARCH uses all the in-sample information to forecast the conditional variance-covariance matrix of the first out-of-sample day, while a traditional GARCH is applied univariately on each series (always reestimating the parameters) to generate  $N$  univariate volatility forecasts; indeed, at each iteration both models reestimate the parameters, but the number of dynamic and static factors of the DF-GARCH is kept fixed according to what we have already found for the first 350 working days of our sample.

Following Andersen et al. [2003], we evaluate the volatility forecasts of our model by running a Mincer-Zarnowitz regression (Mincer and Zarnowitz [1969]). For each series, we consider all the predictions we have computed at each iteration of our rolling scheme and we run the following regression based upon real and predicted conditional standard deviations:

$$(V_{iT+1})^{1/2} = b_0 + b_1 (\hat{\sigma}_{iT+1}^2)^{1/2} + e_{T+1} \quad i = 1, \dots, 100 \quad (43)$$

where  $T$  is the last period of the subsample we are using at iteration  $i$  and the volatility proxy  $V_{iT+1}$  is the realized volatility at time  $T + 1$ ; on the other side,  $\hat{\sigma}_{iT+1}^2$  represents the volatility forecast at time  $T + 1$ , as predicted at time  $T$ . We then run a number of regressions that is equal to the number of series, and we record the estimated regression parameters  $\hat{b}_0$  and  $\hat{b}_1$ , as well as the coefficient of multiple determination  $R^2$ . Should a model be correctly specified, we would obtain values of  $\hat{b}_0$  and  $\hat{b}_1$  that are close to 0 and 1, respectively. However, estimates are necessarily affected by estimation errors and downward biases in the estimation of  $b_1$  (see

e.g. Chow [1983]). Let us then focus upon the coefficient of multiple determination  $R^2$ , which roughly measures the amount of variability of the ex-post volatility that can be explained by the model, thus giving a general idea of its potentialities.

Table 2 lists the results we get from some of the  $N$  regressions, together with the synthetic values obtained by averaging the results of all  $N$  regressions. We used conditional standard deviations both in the real data, here approximated by the square roots of the realized volatilities, and in the model forecasts, obtained without mean predicting. For 66 series (i.e. 74.16% of the total number of series) the DF-GARCH obtains a higher coefficient  $R^2$  than the traditional GARCH model, whose performance is consistent with the empirical results found in the literature, and the static application of DF-GARCH, that performs even worse than univariate GARCH. For 47 series (i.e. 52.81% of the total number of series) the static estimation of DF-GARCH obtains a higher coefficient  $R^2$  than the traditional GARCH model.

We build another performance evaluation procedure as in the following: for each series, we take the prediction of the two models and compute one-step-ahead root mean square errors against the realized volatility. For each series  $i$ , we compute the RMSE as follows:

$$RMSE_i = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{\sigma}_{ikT+1}^2 - V_{ikT+1})^2} \quad K = 100 \quad i = 1, \dots, N \quad (44)$$

where  $\hat{\sigma}_{ikT+1}^2$  is the one-step-ahead volatility forecast of the considered model for series  $i$  at iteration  $k$ . We then compute the ratio between the RMSE obtained by the Dynamic Factor GARCH and the RMSE obtained for the same series by the traditional GARCH model. Results relative to the first series in the sample are listed in table 3, together with the the results obtained by averaging over all the series of the sample. Results can be summarized by means of two statistics:

- $P$  corresponds to the percentage of series for which the DF-GARCH outperforms the univariate GARCH, i.e. the percentage of the cases for which

$$\frac{RMSE_i(\text{DF-GARCH})}{RMSE_i(\text{GARCH})} < 1 \quad i = 1, \dots, N \quad (45)$$

For almost three quarters of the sample, the DF-GARCH turns out to be a better predictor than the traditional GARCH. For our sample,  $P$  is equal to 71.91%; when DF-GARCH is estimated in a static way,  $P$  lowers to 60.67%

- $Q$  is the geometric mean of the RMSE ratios:

$$Q = \left( \prod_{i=1}^N \frac{RMSE_i(\text{DF-GARCH})}{RMSE_i(\text{GARCH})} \right)^{\frac{1}{N}} \quad (46)$$

$Q$  is smaller than one, that is the average prediction accuracy of our method turns out to be better than the benchmark. In other words, the quantity  $(1 - Q)$  is a measure of the average gain obtained by using the DF-GARCH model with our estimation method. For our sample,  $Q$  is equal to 0.99156; when DF-GARCH is estimated in a static way,  $Q$  becomes 1.0018.

Series	Model	$\hat{b}_0$	$\hat{b}_1$	$R^2$
1	GARCH	0.0121	0.9700	0.0669
	DF-GARCH (no lags)	0.0062	1.1338	0.1117
	DF-GARCH (4 lags)	-0.0090	1.6440	<b>0.1945</b>
2	GARCH	0.0219	0.3922	0.0313
	DF-GARCH (no lags)	0.0187	0.5345	0.0108
	DF-GARCH (4 lags)	0.0004	1.1207	<b>0.0376</b>
3	GARCH	0.0112	0.8147	0.1433
	DF-GARCH (no lags)	0.0074	0.9400	0.1038
	DF-GARCH (4 lags)	0.0004	1.2903	<b>0.1667</b>
4	GARCH	0.0019	0.9730	0.1251
	DF-GARCH (no lags)	0.0021	1.0365	0.1141
	DF-GARCH (4 lags)	-0.0040	1.3226	<b>0.1471</b>
5	GARCH	0.0163	0.5898	0.0886
	DF-GARCH (no lags)	0.0075	0.9259	<b>0.1106</b>
	DF-GARCH (4 lags)	0.0076	0.9330	0.1030
6	GARCH	0.0118	0.7840	0.0987
	DF-GARCH (no lags)	0.0144	0.7205	0.0911
	DF-GARCH (4 lags)	-0.0066	1.2106	<b>0.2057</b>
7	GARCH	0.0057	0.7104	<b>0.1622</b>
	DF-GARCH (no lags)	0.0066	0.5572	0.0814
	DF-GARCH (4 lags)	0.0021	0.7487	0.1267
8	GARCH	-0.0747	1.8552	0.0072
	DF-GARCH (no lags)	0.0218	0.4698	0.0067
	DF-GARCH (4 lags)	0.0253	0.4546	<b>0.0106</b>
9	GARCH	-0.0076	2.0648	0.1437
	DF-GARCH (no lags)	-0.0012	1.5648	0.2096
	DF-GARCH (4 lags)	-0.0065	1.8886	<b>0.2528</b>
10	GARCH	0.0040	0.8418	0.1129
	DF-GARCH (no lags)	0.0026	0.8500	0.0775
	DF-GARCH (4 lags)	-0.0060	1.1973	<b>0.1199</b>
...	GARCH	...	...	...
	DF-GARCH (no lags)	...	...	...
	DF-GARCH (4 lags)	...	...	...
arit.	GARCH	0.0535	0.1178	0.1108
	DF-GARCH (no lags)	0.0052	0.9765	0.1074
mean	DF-GARCH (4 lags)	-0.0032	1.2865	<b>0.1435</b>

Table 2: MZ regressions - cond. volatility one-step-ahead predictions with rolling scheme.

Series	DF-GARCH RMSE	GARCH RMSE	RMSE ratio
1	0.00301	0.00307	0.97877
2	0.00670	0.00673	0.99524
3	0.00128	0.00131	0.98274
4	0.00109	0.00107	1.0197
...	...	...	...
arit. mean	0.00342	0.00349	0.99367

Table 3: Conditional volatility RMSEs

## 7.5 Covolatility prediction

We now build the same statistics for the off-diagonal elements of our prediction, i.e. for the predicted conditional covariances. A comparison with traditional univariate GARCH is not possible; we therefore do not compute root mean square errors, but just limit our attention to Mincer-Zarnowitz regressions. For each couple of different series, we regress the vector of realized covolatilities onto the vector of one-step-ahead predicted conditional covariances, following the same rolling scheme and time span of the previous subsections. Results for some couples of series and results averaged over the whole sample are shown in table 4. The coefficient  $\hat{b}_1$  of proportionality between predictions and real proxies is low (on average, 0.21), witnessing a tendency of our predictor to overshoot the real conditional covariance as well as an amount of cases (27.40% of total) in which the DF-GARCH has predicted the wrong sign of the conditional covolatility. Another possible explanation of this low value involves the mild correlation among the idiosyncratic parts: should this be the case, the estimated unconditional covariance between the idiosyncratic parts might be added to the conditional covariance we predict for the common parts. However, the coefficient of multiple determination  $R^2$  (on average, 0.11321) is only slightly lower than in the previous case of conditional variance prediction and is still higher than the coefficient obtained by the static estimation of DF-GARCH (on average 0.1040).

## 8 Conclusions and further research

In this paper we have proposed a new method for the estimation and forecast of conditional covariance when dealing with a large number of series. We operate within the framework of a Generalized Dynamic Factor Model in which the dynamic factors evolve according to a multivariate GARCH rule. The model can be considered as a special case of the structural ARCH described by Harvey et al. [1992]. The particular assumptions we make about the relationships among the observable series allow us to build a feasible predictor of a time-varying conditional covariance matrix when the cross-dimension of the dataset is large with respect to the time-dimension. Our estimation procedure consists of two parts: the first part disentangles common and idiosyncratic components of the dataset, builds the factor loading matrices and gives a first estimate of static and dynamic factors; the second part provides a more efficient estimate of static and dynamic factors while modelling their conditional covariances for all in-sample periods. The modification of the Kalman filter contained in the second part is propaedeutic to the multivariate volatility forecast, obtained by summing up the predicted conditional variances of the idiosyncratic terms and the conditional variance-covariance forecasts of the

Series $i$	Series $j$	Model	$\hat{b}_0$	$\hat{b}_1$	$R^2$
1	2	DF-GARCH (no lags)	0.0001	0.0698	0.0189
		DF-GARCH (4 lags)	0.0001	0.1072	<b>0.0331</b>
1	3	DF-GARCH (no lags)	0.0001	0.1959	<b>0.0873</b>
		DF-GARCH (4 lags)	0.0001	0.2210	0.0830
2	3	DF-GARCH (no lags)	0.0001	0.2127	0.0691
		DF-GARCH (4 lags)	0.0001	0.1728	<b>0.0771</b>
1	4	DF-GARCH (no lags)	0.0000	0.0643	<b>0.0311</b>
		DF-GARCH (4 lags)	0.0000	0.0635	0.0224
2	4	DF-GARCH (no lags)	0.0000	0.2293	0.2043
		DF-GARCH (4 lags)	0.0000	0.2756	<b>0.2162</b>
3	4	DF-GARCH (no lags)	0.0000	0.2145	0.1011
		DF-GARCH (4 lags)	0.0000	0.2581	<b>0.1080</b>
...	...	DF-GARCH (no lags)	...	...	...
		DF-GARCH (4 lags)	...	...	...
arit. mean		DF-GARCH (no lags)	0.0000	0.1727	0.1040
		DF-GARCH (4 lags)	0.0000	0.2094	<b>0.1132</b>

Table 4: MZ regressions - cond. covolatility one-step-ahead predictions with rolling scheme.

common part.

This method presents advantages with respect to the existing literature in that it allows for managing datasets in which the cross-dimension is high and the data present conditionally heteroskedastic behavior. Traditionally, these two features jointly cause estimation difficulties due to the number of parameters involved by the conditional covariance dynamics. The Generalized Dynamic Factor Model can reduce the complexity of the problem and give room for a volatility forecast that takes into account all the cross and time relationships within the entire information set.

We evaluate the predictive performance of our method by applying it onto a dataset of financial returns. We compare our method's conditional variance-covariance forecasts with realized volatilities and covolatilities built upon intra-daily data of the out-of-sample periods. For the diagonal elements of the conditional covariance matrix, our predictor performs better than the traditional univariate GARCH model, both in terms of root mean square errors and in terms of determination coefficients of Mincer-Zarnowitz regressions.

We have chosen to empirically test our model on financial data not only because of the growing interest on return multivariate modelling, but also for the availability of realized measures of conditional variance and covariance that can be used to evaluate the prediction accuracy of our procedure. However, the method presented here might be especially useful in predicting conditional covariances of large datasets when the only available measures of the data refer to levels and first moments. As a consequence, we are currently involved in integrating the method within a macroeconomic context, in which the multivariate volatility forecast can be used to evaluate the risk associated with different policies. Another appealing macroeconomic application would be the univariate variance prediction of an aggregate variable (e.g. inflation or GDP) by means of a multivariate analysis on disaggregated data.

# References

- L. Alessi, M. Barigozzi, and M. Capasso. Generalized dynamic factor model + GARCH: Exploiting multivariate information for univariate prediction. LEM Working Papers 2006/13, Scuola Superiore Sant'Anna, 2006.
- C. Alexander. Orthogonal GARCH. In C. Alexander, ed., *Mastering Risk volume 2*, FT Prentice Hall, 2001.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–626, 2003.
- J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.
- J. Bai and S. Ng. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 2005. Forthcoming.
- O. E. Barndorff-Nielsen and N. Shephard. Variation, jumps and high frequency data in financial econometrics. Economics Paper 2005-W16, Nuffield College, University of Oxford, 2005.
- L. Bauwens, S. Laurent, and J. V. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1), 2006. Forthcoming.
- J. Boivin and S. Ng. Are more data always better for factor analysis? NBER Working Papers 9829, National Bureau of Economic Research, Inc, July 2003.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31: 307–327, 1986.
- T. Bollerslev. Modeling the coherence in short-term nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, 72:498–505, 1990.
- T. Bollerslev, R. F. Engle, and J. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.
- D. Brillinger. *Time Series: Data Analysis and Theory*. Holden-Day, Inc., San Francisco, CA,, 1981.
- C. T. Brownlees and G. M. Gallo. Financial econometric analysis at ultrahigh frequency: Data handling concerns. Technical report, University of Florence, 2006.
- P. Burns. Multivariate GARCH with only univariate estimation, 2005. <http://www.burns-stat.com/>.
- G. Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51: 1281–1304, 1983.
- G. Chamberlain and M. Rothschild. Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, 51:1305–1324, 1983.
- S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 2006. Forthcoming.
- G. C. Chow. *Econometrics*. New York: Mc Graw-Hill, 1983.
- G. Connor, R. A. Korajczyk, and O. Linton. The common and specific components of dynamic volatility. *Journal of Econometrics*, 132:231–255, 2006.

- P. de Jong. Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association*, 84:1085–1088, 1989.
- F. X. Diebold and M. Nerlove. The dynamics of exchange rate volatility: A multivariate latent factor arch model. *Journal of Applied Econometrics*, 4(1):1–21, 1989.
- C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on kalman filtering. Unpublished, ECARES - Université Libre de Bruxelles, 2005.
- C. Doz, D. Giannone, and L. Reichlin. A quasi maximum likelihood approach for large approximate dynamic factor models. Working paper series no 674, European Central Bank, 2006.
- F. Drost and T. Nijman. Temporal aggregation of garch processes. *Econometrica*, 61/4:909–927, 1993.
- J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
- R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.
- R. F. Engle. Multivariate ARCH and factor structures - cointegration in variance. Discussion paper 87-2, UCSD, Department of Economics, 1987.
- R. F. Engle. Dynamic conditional correlation - a simple class of multivariate GARCH models. Working paper 2000-09, UCSD, Department of Economics, Aug. 2000.
- R. F. Engle. New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446, 2002.
- R. F. Engle and K. Kroner. Multivariate simultaneous GARCH. *Econometric Theory*, 11:122–150, 1995.
- R. F. Engle, V. K. Ng, and M. Rothschild. Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237, 1990.
- M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, 82:540–554, 2000.
- M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the black box: structural factor models with large cross-sections. CEPR discussion paper 4133, Center for Economic Policy Research, 2006a.
- M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 2006b. Forthcoming.
- J. Geweke. The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Golberger, eds., *Latent Variables in Socio-Economic Models*, 1977.
- D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. Technical report, C.E.P.R., 2005. CEPR Discussion Papers 4981.
- D. Giannone, L. Reichlin, and D. H. Small. Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases. Working paper series no 633, European Central Bank, 2006.
- M. Hallin and R. Liska. The generalized dynamic factor model: determining the number of factors. *Journal of the American Statistical Association*, 2007. Forthcoming.

- A. Harvey and G. De Rossi. Signal extraction. In K. Patterson and T.C. Mills, eds., *Palgrave Handbook of Econometrics: Volume 1 Econometric Theory*, Palgrave Macmillan, 2006.
- A. Harvey, E. Ruiz, and E. Sentana. Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, 52(1-2):341–349, 1992.
- A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–64, 1994.
- A. Harvey, E. Ruiz, and E. Sentana. Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics*, 102(2):143–164, 2001.
- C. J. Kim and C. R. Nelson. *State-Space Models with Regime Switching*. The MIT Press, 1999.
- M. King, E. Sentana, and S. Wadhvani. Volatility and links between national stock markets. *Econometrica*, 62:901–933, 1994.
- M. Martens. Estimating unbiased and precise realized covariances. Technical report, Econometric Institute, Erasmus University Rotterdam, June 2004.
- J. Mincer and V. Zarnowitz. *The Evaluation of Economic Forecasts*. Columbia University Press, 1969.
- T. Nijman and E. Sentana. Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics*, 71:71–87, 1996.
- A. Onatski. Determining the number of factors from empirical distribution of eigenvalues. Technical report, Economics Department, Columbia University, 2006.
- T. Sargent and C. Sims. Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims et al., eds., *New Methods in Business Cycle Research*, 1977.
- G. W. Schwert. Tests for unit roots: a monte carlo investigation. *Journal of Business and Economic Statistics*, 7:147–160, 1989. Forthcoming.
- E. Sentana. The relation between conditionally heteroskedastic factor models and factor GARCH models. *The Econometrics Journal*, 1:1–9, December 1998.
- J. H. Stock and M. W. Watson. Diffusion indexes. NBER working papers, National Bureau of Economic Research, Inc, Aug. 1998.
- J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–162, 2002.
- R. S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons Inc., 2002.
- R. van der Weide. GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564, 2002.
- R. van der Weide. Wake me up before you GO-GARCH. *Computing in economics and finance* 2004-316, Society for Computational Economics, Aug. 2004.