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

Lucia Alessi∗
Matteo Barigozzi†
Marco Capasso‡

Laboratory of Economics and Management
Sant’Anna School of Advanced Studies, Pisa

October 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

We thank Domenico Giannone, Marc Giannoni, Marc Hallin and Marco Lippi for helpful comments and suggestions. Of course, the responsibility for any error is entirely our own.
As members of the Financial Market Analysis Research Centre (CAFiM) at Sant’Anna School of Advanced Studies in Pisa, we had access to the Bloomberg dataset used in the paper.
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.\(^1\) 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.\(^2\) 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.\(^3\) After the original ARCH and GARCH univariate specifications, respectively by Engle [1982] and Bollerslev [1986], many multivariate versions have been proposed, notably the VECCH 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 [2002b]. 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

---

\(^1\)See Bauwens et al. [2006].

\(^2\)See Harvey et al. [1994].

\(^3\)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 (STARCH), 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. [2004], 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. [2005], 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. [2004] 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 describe respectively a heteroskedastic GDFM, its representation and its estimation. In sections 5 and 6, we 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 and the assumptions of the DF-GARCH are tested. Section 8 and 9 compare volatility and covolatility predictions for different specifications of the DF-GARCH and, in the case of volatility, for a univariate GARCH. A discussion of the results and some final remarks are reported in section 10. Sometimes we 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 $x_t = (x_{1t} \ldots x_{Nt})'$ an $N$-dimensional vector process. Each of the series is covariance stationary, standardized and the second order moments $\Gamma_k = E[x_t x_{t-k}']$ exist finite for all $k$. 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 $j = 1, \ldots, q$, which are loaded with possibly different coefficients and lags. Formally

$$x_{it} = \chi_{it} + \xi_{it} = d_{i1}(L)u_{1t} + d_{i2}(L)u_{2t} + \ldots + d_{iq}(L)u_{qt} + \xi_{it} \quad i = 1, \ldots, N. \quad (1)$$
The q-dimensional vector process \( u_t = (u_{1t}, \ldots, u_{qt})' \) is an orthonormal white noise. The N-dimensional vector process \( \xi_t = (\xi_{1t}, \ldots, \xi_{Nt})' \) 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 \( d_{i1}(L) \ldots d_{iq}(L) \) 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 focus on the particular case in which dynamic factors are heteroskedastic and distributed as

\[
\xi_{it \mid t-1} \sim N(0, Q_t),
\]

\( Q_t \) being a non-diagonal matrix changing over time, and where the subscript \( t \mid t-1 \) means that the values at time \( t \) are conditioned upon the information available at time \( t-1 \). The conditional variance of the dynamic factors is governed by

\[
Q_t = C_0' C_0 + C_1' u_{t-1} u_{t-1}' C_1 + C_2' Q_{t-1} C_2.
\]

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 is very small. Alternatively \( Q_t \) could be modelled according to the Dynamic Conditional Correlation (DCC) specification as in Engle [2002b]. The main advantage of the DCC formulation is that it always requires the estimation of only two parameters when considering a multivariate GARCH of order one. This is particularly useful and it seems that, when dealing with many series, we could even avoid the factor decomposition. However, if the number of series is large, two parameters are too few to describe the whole dynamics of conditional correlations. This is the practical reason for using the factor decomposition before estimating a multivariate GARCH even with a DCC representation. Given that with few series it is feasible to estimate also a full BEKK we consider both possibilities.

We assume that idiosyncratic parts evolve according to a univariate ARMA-GARCH model

\[
\xi_{it \mid t-1} \sim N(\mu_t, R_t),
\]

where, for each time, the diagonal matrix \( R_t \) contains the conditional covariances of each idiosyncratic series

\[
R_{it} = \rho_{0i} + \rho_{1i} \xi_{it-1}^2 + \rho_{2i} R_{it-1}.
\]

We do not define a particular structure of the ARMA process governing the conditional mean of \( \xi \). Indeed, as our empirical application deals with financial returns, whose idiosyncratic parts are unlikely to have strong dynamics, we do not consider any ARMA structure of the idiosyncratic components. However, the general description of our procedure allows 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
A factor model is made of \( r << N \) common factors, thus the factor model is written as

\[
x_t = \chi_t + \xi_t = \Lambda F_t + \xi_t ,
\]

where \( F_t = (F_1 \ldots F_r)' \) 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 VAR(1) representation

\[
F_t = AF_{t-1} + Bu_t ,
\]

where \( u_t = (u_{1t} \ldots u_{qt})' \) is the \( q \)-vector of common shocks, \( A \) is \( r \times r \) and \( B \) is \( r \times q \). According to the static representation of the dynamic factor model the common shocks are precisely the dynamic factors that we want to estimate. Indeed the static model of equation (6) contains a special case of the usual dynamic factor model (see Forni et al. [2006] and Bai and Ng [2005]). By inverting equation (7) and using the lag operator \( L \), we have

\[
F_t = (I - AL)^{-1} Bu_t .
\]

Hence we have immediately the dynamic factor model

\[
x_t = \Lambda (I - AL)^{-1} Bu_t + \xi_t = D(L)u_t + \xi_t ,
\]

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

\[
F_t = (u'_t, u'_{t-1} \ldots u'_{t-s})' ,
\]

\[
\Lambda = (D_0 \ldots D_s) ,
\]

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. [2005]. 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 With hatted symbols we indicate estimates.

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 \( \hat{\Sigma}_x(\theta) = \Psi(\theta)\Phi(\theta)\Psi(\theta)' \). It is 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 \( \hat{\Sigma}(\theta) \) the inverse Fourier transform we retrieve estimates of the covariance matrices of the common component, \( \hat{\Gamma}_k \); 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 \( \hat{\Gamma}_k \) with respect to \( \hat{\Gamma}_k^\xi \). The first generalized eigenvector solves the maximization problem

\[
\begin{align*}
z^{(1)} &= \arg\max_{a \in \mathbb{R}^n} a' \hat{\Gamma}_0^\chi a \\
\text{s.t.} \quad a' \hat{\Gamma}_0^\xi a &= 1.
\end{align*}
\]

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

\[
\hat{\chi}_t = \hat{\Gamma}_0^\chi Z (Z' \hat{\Gamma}_0^\chi Z)^{-1} Z' x_t \quad \forall \ t = 1, \ldots, T.
\]

We obtain the idiosyncratic component simply as difference between the original series \( x_t \) and \( \hat{\chi}_t \).

Given the results of the previous section, we can write the GDFM is state-space form

\[
\begin{align*}
x_t &= \Lambda F_t + \xi_t = D(L) u_t + \xi_t \\
F_t &= A F_{t-1} + B u_t.
\end{align*}
\]

Using the one-sided estimator proposed by Forni et al. [2005] 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 \( \Omega \Omega' = I \). Hereafter we suppose that \( \Omega = I \), 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:

\[
\begin{align*}
\hat{\chi}_t &= \text{the estimated common part}, \\
\hat{\xi}_t &= \text{the estimated idiosyncratic part}, \\
\hat{\Gamma}_0^\chi &= E[\hat{\chi}_t \hat{\chi}_t'], \\
\hat{\Gamma}_0^\xi &= E[\hat{\xi}_t \hat{\xi}_t'] = \hat{\Gamma}_0^\chi - \hat{\Gamma}_0^\xi, \\
\hat{F}_t &= \text{the estimated generalized principal components}.
\end{align*}
\]

Following the procedure used by Giannone et al. [2004], from (11) we have an estimation for \( \Lambda \) given a sample length \( T \)

\[
\hat{\Lambda} = \frac{1}{T} \sum_{t=1}^{T} x_t \hat{F}_t' \left( \sum_{t=1}^{T} \hat{F}_t \hat{F}_t' \right)^{-1}.
\]

Analogously, we have an estimation for \( A \)

\[
\hat{A} = \frac{1}{T-1} \sum_{t=2}^{T} \hat{F}_t \hat{F}_{t-1}' \left( \sum_{t=2}^{T} \hat{F}_{t-1} \hat{F}_{t-1}' \right)^{-1},
\]

\[6\]
and we can also estimate the covariance of $Bu_t$ 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 $Bu_t$ which is a linear 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

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

from which we get an estimation of $B$

$$\hat{B} = MS^{1/2} . \quad (14)$$

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

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

5 Correcting for heteroskedasticity

Modelling the covariances of dynamic factors and idiosyncratic parts

Now we apply a modified version of the Kalman filter that provides 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 need to apply a univariate GARCH model to each time series of $\xi$ as in (5), and so we obtain the conditional variance $\hat{R}_t$ for each idiosyncratic series and for each time. If we model also the conditional mean of the idiosyncratic part then also an ARMA model should be estimated for each time series of $\xi$ in order to obtain an estimated of the conditional mean $\hat{\mu}_t$. 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, if the ARMA model for the idiosyncratic parts is specified, 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) is replaced by a new idiosyncratic part \( \varepsilon_t \), whose process has conditional mean equal to zero. Our new state-space model is

\[
\begin{align*}
\tilde{x}_t &= \hat{A}F_t + \varepsilon_t \quad \text{measurement equation}, \\
F_t &= \hat{A}F_{t-1} + \hat{B}u_t \quad \text{transition equation},
\end{align*}
\]

where

\[
\begin{align*}
\varepsilon_{t|t-1} &\sim N(0, \hat{R}_t) \quad \hat{R}_t \text{ diagonal}, \\
u_{t|t-1} &\sim N(0, \hat{Q}_t), \\
Q_t &= \hat{C}_0(\hat{C}_1u_{t-1}u_{t-1}'\hat{C}_1) + \hat{C}_2\hat{Q}_{t-1}\hat{C}_2.
\end{align*}
\]

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 are 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 Appendix we explain in detail how our modification of the Kalman filter works.

6 Forecasting

Once the Kalman filter estimation is completed we have a new estimate of the static factors \( F_0T \) and of their conditional covariance matrix \( P_0T \). We consider \( \chi^*_t = \hat{A}F_0T \) as the new estimated common part of each return observation \( x_t \); by difference we obtain a new idiosyncratic part \( \xi^*_t = x - \chi^* \). We assume that \( \xi^*_t \) are the realizations of a process having the conditional variances \( \hat{R}_t \) that we found before applying the filter by univariate GARCH modelling of \( \hat{\xi} \). If we specified an ARMA model for the idiosyncratic parts then \( \xi^*_t \) have as conditional mean the previously estimated \( \hat{\mu}_t \). The predicted diagonal covariance matrix \( \Gamma^\chi_{T+1|T} \) for \( \xi^*_t \) 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 \( \xi \) before the Kalman filter. The predicted covariance matrix for the common part is

\[
\Gamma^\chi_{T+1|T} = \hat{A}P_{T+1|T}\hat{A}' ,
\]

where \( P_{T+1|T} \) has been obtained by applying the prediction steps to the smoothed estimations of last period \( T \) (see the Appendix for details). 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

\[
\Gamma^\xi_{T+1|T} = \Gamma^\chi_{T+1|T} + \Gamma^\xi_{T+1|T} .
\]
A forecast for the terms of the conditional correlation matrix can now be obtained as

$$
\rho_{ij,T+1|T} = \frac{\gamma_{ij,T+1|T}}{\sqrt{\gamma_{ii,T+1|T} \gamma_{jj,T+1|T}}} \tag{21}
$$

Notice that in (20) the off-diagonal terms of $\Gamma_{T+1|T}$ do not consider the mild correlations (and conditional correlations) among the idiosyncratic terms. Correlation estimations given by (21) might thus be biased; for instance, conditional correlations are 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 (20).

7 Preliminary analysis of the dataset

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 1st October 2001 to 31st July 2003 (457 working days). Transaction prices have been cleaned from outliers by using the procedure described in Brownlees and Gallo [2006]; 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 $i$th factor as

$$
EV_i = \frac{\int_{-\pi}^{\pi} \lambda_i(\theta) d\theta}{\sum_{j=1}^{N} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta} \tag{22}
$$

We require that, as $N \to \infty$,

$$
\begin{cases}
EV_i \to \infty & \text{for} & i = 1, \ldots, q \\
\exists M \in \mathbb{R}^+ & \text{s.t.} & EV_{q+1} \leq M \tag{23}
\end{cases}
$$

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.

---

5 All computations for this section have been performed by using the standard Matlab software package (v.7.0) plus the freely available toolboxes MATNEM by Christian T. Brownlees and ucsd_garch by Kevin K. Sheppard; the Matlab code used in the next subsection has been kindly provided by Roman Liška.
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 Liška [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 of the covariogram-smoothing version of the criterion. For given $N$ and $T$, it consists in choosing the number of factors $\hat{q}$ that minimizes the variance explained by the remaining $N - \hat{q}$ factors. In principle, the maximum number of factors allowed $q_{\text{max}}$ is the number of series in the dataset. Therefore, we introduce a penalty function to avoid overestimation of $\hat{q}$, but at the same time it should not overpenalize. Multiplying the penalty function by a positive constant $c$ is a way to tune the penalizing power. Hallin and Liška propose an automatic procedure for selecting $\hat{q}$ which basically explores the behavior of the variance of the selected $q$ 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 $\hat{q} < N$. In figure 2.1, relative to the subsample of the first 350 daily returns of the 89 stocks, the solid lines indicate the value of $q$ while the dashed lines represent its variance when varying the sample size and the time length of the dataset. We are looking for detecting stability intervals. In other words, $\hat{q}$ corresponds to the plateau of the solid line associated with the second flat zero-level dashed line. This procedure, suggested by Hallin and Liška, 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 could 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

---

6 Hallin and Liška suggest that this form has better finite sample performance.
7 Notice that the first stability interval always corresponds to $q_{\text{max}}$.
2.1: Hallin-Liška criterion.  
2.2: IC1 criterion. 

Figure 2: Determining the number of factors. 

3.1: The five largest dynamic eigenvalues.  
3.2: Spectral density of the common component. 

Figure 3: Evidence of dynamics in asset returns. 

squared distance is computed for all the possible numbers of static factors between 0 and $r_{\text{max}}$, 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 $r_{\text{max}}$ as an input, only $PC$ criteria explicitly take into account its resulting minimum squared distance. However it is recognized by the literature (e.g. see Forni et al. [2006]) that these criteria highly depend on $r_{\text{max}}$. Indeed, 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 apply our modification of the original criteria by Bai and Ng [2002] that simply introduces a positive constant $c$ in the penalty function. The procedure to determine the number of static factors $\hat{r}$ for given $N$ and $T$ is the same as the one used in determining the number of the dynamic factors (for details on this new criterion see Alessi et al. [2007]). In figure 2.2 we show the plots for the modified Bai-Ng’s IC1 criterion and once again we are looking for the second stable region (i.e. dashed line equal to zero) corresponding to a plateau in the number of factors. With this procedure we find 5 static factors. The same result holds also for IC2 and IC3, and all PC criteria. An heuristic argument is based on the explained variance by the
first $q$ factors. We need a number of static factors that should explain the same amount of variance which in our case is 44% of the total sample variance. If we look at the eigenvalues of the sample variance-covariance matrix, two static factors would be able to explain only 38% of the total sample variance, while we would need between 5 and 6 static factors to explain the same percentage of variance that is explained by the two dynamic factors. This result is consistent with the one suggested by our criterion and therefore we choose $r = 6$.

7.4 A comment on the dynamics of asset returns

In figure 3.1 we plot the five largest dynamic eigenvalues at different frequencies. It is clear that the first dynamic eigenvalue explains the bulk of the variance (precisely 36%) but at least at low frequencies (i.e. in the long run) a second factor gives an appreciable contribute. In what follows we therefore assume that $q = 2$ for a total explained variance of 44% corresponding to an average variance of the common part of 44% (with a maximum of 75%). Moreover, given that $r = q(s + 1)$ and given the result of previous section, we choose $s = 2$. Although it is usually assumed that asset returns have no appreciable dynamics (they are often modelled as white noise), we decide to adopt a dynamic factor model for at least two good reasons. First the pattern of peaks and troughs followed by the first eigenvalue at different frequencies indicates the high influence that the factors exert on returns at different specific frequencies, thus supporting the idea that the factors are really dynamic. Second figure 3.2 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 $\Sigma \chi(\theta) = D(e^{i\theta}) \Sigma u D(e^{-i\theta})^\prime$. 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 $\Sigma \chi(\theta)$ 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 $D \Sigma u D$ 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 suggests common factors’ dynamics.

Notice that our model includes also the static case that is obtained just by imposing $s = 0$. In what follows, results obtained by performing the DF-GARCH estimation in a static way, thus with only two static factors, is used as a benchmark for evaluating the prediction accuracy of the DF-GARCH with two dynamic factors and two lags. The other benchmark is represented by the traditional univariate GARCH model. Moreover the DF-GARCH is estimated using either a full BEKK and a DCC model for the dynamic factors.

7.5 Testing for heteroschedasticity

The main hypothesis of the DF-GARCH is that the dynamic factors are heteroschedastic. In order to test this assumption we employ the usual ARCH test up to 10 lags. As shown in
Table 1: ARCH-test on $u_t$ for heteroskedasticity (* significant at 99%). Observations from $t = 1$ to $t = 350$ i.e. first in-sample.

table 1, for the first in-sample used we find that one of the two dynamic factors is highly heteroschedastic. In particular this is always the largest of the two factors, as it is shown in the same table when estimating the model with only one dynamic factor. The same result holds for different in-samples considered. It is the largest factor that determines the heteroschedasticity of asset returns. Good news for our model since having at least one heteroschedastic dynamic factor justifies our assumptions. Notice that he model itself is flexible enough to accomodate cases with some heteroschedastic and some homoschedastic dynamic factors.

8 Volatility prediction

8.1 Performance evaluation

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 is used both in the proper way (two dynamic factors and two lags) and in a static way (two dynamic factors corresponding to just two static factors with no lags) using either the full BEKK and the DCC specification. Whenever we name just “DF-GARCH”, we refer to the proper dynamic estimation of the model independently of the chosen multivariate GARCH specification. 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. As for the volatility proxy we always use realized volatilities (see section 7.1), but results are robust when we compare our predictions with other proxies as the naïve squared returns or the more sophisticated squared adjusted range.\(^8\)

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

\(^8\)We report only average results and only compared to realized volatility. Detailed results for all proxies and all series are available upon request.
all the predictions we have computed at each iteration of our rolling scheme and we run a regression based upon real and predicted conditional standard deviations:

\[ V_{T+k+1} = b_0 + b_1 \hat{\sigma}^2_{T+k+1} + e_{T+k+1} \quad k = 0, \ldots, 99, \]

where \( T \) is the last period of the first in-sample considered (i.e. \( T = 350 \)) and the volatility proxy \( V_{T+k+1} \) is the realized volatility at time \( T + k + 1 \). On the other side, \( \hat{\sigma}^2_{T+k+1} \) represents the one-step-ahead volatility forecast at time \( T + k + 1 \), as predicted at time \( T + k \). 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>
<thead>
<tr>
<th>Model</th>
<th>( \hat{b}_0 )</th>
<th>( \hat{b}_1 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>0.0033</td>
<td>0.0025</td>
<td>0.0791</td>
</tr>
<tr>
<td>DF-GARCH (BEKK)(no lags)</td>
<td>0.0002</td>
<td>1.6063</td>
<td>0.1084</td>
</tr>
<tr>
<td>DF-GARCH (BEKK)(2 lags)</td>
<td>0.0002</td>
<td>1.6714</td>
<td>0.1100</td>
</tr>
<tr>
<td>DF-GARCH (DCC)(no lags)</td>
<td>0.0000</td>
<td>1.9384</td>
<td>0.0981</td>
</tr>
<tr>
<td>DF-GARCH (DCC)(2 lags)</td>
<td>0.0000</td>
<td>2.0188</td>
<td>0.1000</td>
</tr>
</tbody>
</table>

Table 2: Volatilities. Average results for MZ regressions.

Table 2 presents a summary of the results, obtained by averaging the results of all \( N \) regressions. The model forecasts are obtained without mean predicting. For 71 series (i.e. 80% of the total number of series) the DF-GARCH with full BEKK specification obtains a higher coefficient \( R^2 \) than the traditional GARCH model, whose performance is consistent with the empirical results found in the literature. If we use the DCC specification the performance is slightly worse and the DF-GARCH has a higher \( R^2 \) than univariate GARCH for 64 series (i.e. 72% of the total). The static application of DF-GARCH also performs better than univariate GARCH with results similar to the dynamic case.

To have a comparative performance evaluation for each series, we take the prediction of the two models and compute one-step-ahead Root Mean Squared Errors (RMSE) against the realized volatility \( V_{it} \). For each series \( i \), we compute

\[
RMSE_i = \sqrt{\frac{1}{K} \sum_{k=0}^{K-1} \left( \hat{\sigma}^2_{iT+k+1} - V_{iT+k+1} \right)^2} \quad K = 100 \quad i = 1, \ldots, N ,
\]

where \( \hat{\sigma}^2_{iT+k+1} \) 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 DF-GARCH and the RMSE obtained for the same series using the traditional GARCH model. Average results are listed in table 3, together with two statistics that help to compare the two models:
1. $P$ corresponds to the percentage of series for which the DF-GARCH outperforms the univariate GARCH, i.e. the percentage of the cases for which

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

For the majority of the series in the sample, the DF-GARCH using full BEKK turns out to be a better predictor than the traditional GARCH. For our sample, $P$ is equal to 82.02%; when DF-GARCH is estimated in a static way, $P$ is slightly lower (80.90%).

2. $Q$ is the geometric mean of the RMSE ratios:

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

$Q$ 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 and with full BEKK specification, $Q$ is equal to 0.9798; when DF-GARCH is estimated in a static way, $Q$ becomes 0.9823.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>$P$</th>
<th>$Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>3.4938 $10^{-3}$</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>DF-GARCH (BEKK)(no lags)</td>
<td>3.3988 $10^{-3}$</td>
<td>80.899</td>
<td>0.9823</td>
</tr>
<tr>
<td>DF-GARCH (BEKK)(2 lags)</td>
<td><strong>3.3935 $10^{-3}$</strong></td>
<td><strong>82.022</strong></td>
<td><strong>0.9798</strong></td>
</tr>
<tr>
<td>DF-GARCH (DCC)(no lags)</td>
<td>3.4088 $10^{-3}$</td>
<td>77.528</td>
<td>0.9862</td>
</tr>
<tr>
<td>DF-GARCH (DCC)(2 lags)</td>
<td>3.4021 $10^{-3}$</td>
<td>79.755</td>
<td>0.9839</td>
</tr>
</tbody>
</table>

Table 3: Average RMSEs, $P$ and $Q$ with respect to univariate GARCH.

8.2 Testing for equal predictive accuracy

We evaluate the performance of the DF-GARCH by means of two different tests of predictive accuracy. In table 4 we show the results of the usual Diebold and Mariano [1995] test of equal predictive accuracy. Given the predictions of two competing models (say $a$ and $b$) we compute the difference $d_t = E[(x^a_{t+h|t} - x_{t+h})^2 - (x^b_{t+h|t} - x_{t+h})^2]$ between the squared errors obtained with the two models and we test for $d_t = 0$. If the computed statistic is significantly larger than zero, model $b$ has a better forecast performance than model $a$. Viceversa, if the statistic is significantly smaller than zero then model $a$ is better than model $b$. Notice that, although some of the models we are comparing may be considered as nested, this test is already useful to make a first distinction between them. When the null hypothesis of equal predictive accuracy is rejected with high significance levels, then, no matter if the models are nested, we already have an indication of which one is better.

In order to test the significance of the difference between the RMSEs of two models when
one of the models nests the other, Clark and West [2007] show that a correction is needed on the RMSE of the DF-GARCH to account for the possible errors made in estimating more parameters. The univariate GARCH could be considered as a model nested in the DF-GARCH when \( N = 1 \) and \( q = 1 \). In particular, the following difference must be computed for each time \( t \) and each series \( i \)

\[
\hat{f}_{it} = (V_{it} - \hat{\sigma}^2_{itG})^2 - \left( (V_{it} - \hat{\sigma}^2_{itDF})^2 - (\hat{\sigma}^2_{itG} - \hat{\sigma}^2_{itDF})^2 \right),
\]

(24)

where \( V_{it} \) represents the realized volatility at time \( t \) and \( \hat{\sigma}^2_{it} \) represents the volatility forecast at time \( t \), as predicted at time \( t - 1 \) by the simple GARCH (subscript \( G \)) or the DF-GARCH (subscript \( DF \)). We then test for equal mean square prediction error by regressing \( \hat{f}_i \) on a constant and using the resulting t-statistic for a zero coefficient. In table 4 we show a summary of our results of the tests by Diebold and Mariano [1995] and Clark and West [2007] both for the case of DF-GARCH estimated with two lags and for the case of DF-GARCH estimated in a static way, against the traditional univariate GARCH model. At 10% significance, DF-GARCH with BEKK specification and estimated with two lags performs better than traditional GARCH for 70 series (79% of total).

<table>
<thead>
<tr>
<th></th>
<th>DF-GARCH 2 lags BEKK</th>
<th>DF-GARCH 0 lags BEKK</th>
<th>DF-GARCH 2 lags DCC</th>
<th>DF-GARCH 0 lags DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diebold-Mariano Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of series for which</td>
<td>53</td>
<td>58</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>DF-GARCH outperforms GARCH at 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of series for which</td>
<td>41</td>
<td>40</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>DF-GARCH outperforms GARCH at 5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Clark-West Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of series for which</td>
<td>70</td>
<td>70</td>
<td>68</td>
<td>67</td>
</tr>
<tr>
<td>DF-GARCH outperforms GARCH at 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of series for which</td>
<td>58</td>
<td>60</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>DF-GARCH outperforms GARCH at 5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Diebold-Mariano and Clark-West tests results against traditional univariate GARCH.

9 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 averaged over the whole sample are shown in table 5. The coefficient \( \hat{b}_1 \) of proportionality between
predictions and real proxies is low (about 0.2), witnessing a tendency of our predictor to overshoot the real conditional covariance as well as an amount of cases (about 28% 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 $R^2$ is always higher than in the previous case of conditional variance prediction, being slightly better for the static model.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{b}_0$</th>
<th>$\hat{b}_1$</th>
<th>$R^2$</th>
<th>Percentage of Right Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF-GARCH (BEKK)(no lags)</td>
<td>0.0000</td>
<td>0.1873</td>
<td>0.1569</td>
<td>72.014</td>
</tr>
<tr>
<td>DF-GARCH (BEKK)(2 lags)</td>
<td>0.0000</td>
<td>0.1951</td>
<td>0.1560</td>
<td>72.026</td>
</tr>
<tr>
<td>DF-GARCH (DCC)(no lags)</td>
<td>0.0000</td>
<td>0.2098</td>
<td>0.1323</td>
<td>72.215</td>
</tr>
<tr>
<td>DF-GARCH (DCC)(2 lags)</td>
<td>0.0000</td>
<td>0.2092</td>
<td>0.1274</td>
<td>72.218</td>
</tr>
</tbody>
</table>

Table 5: Covolatilities. Average results for MZ regressions and percentage of times in which the predicted sign is the same as the real sign.

10 Discussion 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 either a full BEKK or a DCC specification. 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: in the first step we disentangle common and idiosyncratic components of the dataset, we build the factor loading matrices and give a first estimate of static and dynamic factors; in the second step we provide 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 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.

Summing up the DF-GARCH performs always better than a univariate GARCH and almost identically in its specification with two lags (as suggested by criteria on factor numbers) or when using no lags (i.e. using a static specification). This is due to the well known fact that asset returns show poor dynamics in their levels. The model however is very flexible compared to purely static models as the O-GARCH or the GO-GARCH. Notice that a DF-GARCH with no lags presents almost no differences with respect to a model where the static factors (almost identical to the dynamic ones) are estimated as static principal components (see Stock and Watson [2002]). The comparison of the DF-GARCH with all this family of static factor models with heteroscedasticity is presently in progress by means of simulations, where different degrees of heteroscedasticity and dynamics in the returns can be modelled. Finally, concerning the multivariate GARCH model of dynamic factors, the full BEKK specification yields always better predictions than the DCC specification, independently of the number of lags used.

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. An ideal application of the DF-GARCH would be on series with rich dynamics in their levels and high heteroscedasticity. As a consequence, we are currently involved in integrating the method within a macroeconomic context, in which the multivariate inflation 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 (a first application is in Barigozzi and Capasso [2007]).
References

L. Alessi, M. Barigozzi, and M. Capasso. A robust criterion for determining the number of static factors in approximate factor models. LEM Working Papers 2007/19, LEM (Laboratory of Economics and Management), Sant’Anna School of Advanced Studies, Pisa, 2007.


Appendix - A modified Kalman filter estimator

We explain here in detail the estimation of the state-space model
\[
\begin{align*}
\tilde{x}_t & = \hat{\Lambda}_t F_t + \varepsilon_t \quad \text{measurement equation}, \\
F_t & = \hat{\Lambda}_t F_{t-1} + \hat{B} u_t \quad \text{transition equation}, \\
\end{align*}
\]
where
\[
\begin{align*}
\varepsilon_{t|t-1} & \sim N(0, \hat{R}_t) \quad \text{diagonal}, \\
u_{t|t-1} & \sim N(0, Q_t), \\
Q_t & = \hat{C}'_0 \hat{C}_0 + \hat{C}'_1 u_{t-1} u_{t-1}' \hat{C}_1 + \hat{C}'_2 Q_{t-1} \hat{C}_2.
\end{align*}
\]

The multivariate GARCH representation considered here is a full BEKK, but the following procedure can be easily modified to allow for a DCC representation.

Initialization

Initial values are built as:
\[
\begin{align*}
F_{1|1} & = \hat{F}_1 \\
P_{1|1} & \quad \text{sufficiently large} \\
E_1 (u_1) & = \hat{u}_1 \\
E_1 (Q_1) & = \hat{Q}_1 \\
E_1 (u_1 u_1') & = (E_1 (u_1))(E_1 (u_1))' + E_1 (Q_1),
\end{align*}
\]
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.

Prediction

The steps described in this and the following section must be repeated together for time \( t = 2 \ldots T \). First we predict the unobserved state vector
\[
F_{t|t-1} = \hat{\Lambda}_t F_{t-1|t-1},
\]
and its conditional covariance matrix
\[
P_{t|t-1} = \hat{\Lambda} P_{t-1|t-1} \hat{\Lambda}' + \hat{B} E_{t-1} (u_t u_t') \hat{B}',
\]
where
\[
\begin{align*}
E_{t-1} (u_t u_t') & = E_{t-1} (Q_t) \\
E_{t-1} (Q_t) & = \hat{C}_0' \hat{C}_0 + \hat{C}_1' E_{t-1} (u_{t-1} u_{t-1}') \hat{C}_1 + \hat{C}_2' E_{t-1} (Q_{t-1}) \hat{C}_2.
\end{align*}
\]

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 see in the next step.

The prediction error is given by
\[
\eta_{t|t-1} = \tilde{x}_t - \bar{x}_{t|t-1} = \bar{x}_t - \hat{\Lambda} \bar{F}_{t|t-1}.
\]
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:

$$W_{t|t-1} = \hat{\Lambda}P_{t|t-1}\hat{\Lambda} + \hat{R}_t.$$  

**Update**

We compute the Kalman gain

$$K_t = P_{t|t-1}\Lambda W_{t|t-1}^{-1},$$

and we build more accurate inferences, exploiting information up to time $t$,

$$F_{t|t} = F_{t|t-1} + K_t\eta_{t|t-1},$$

$$P_{t|t} = P_{t|t-1} - K_t\hat{\Lambda}P_{t|t-1}.$$

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

$$E_t(u_t) = S^{-1/2}M'(1 - \hat{\Lambda}L)F_{t|t},$$  \hspace{1cm} (26)

and then

$$E_t(u_tu'_t) = (E_t(u_t))(E_t(u_t))'.$$  \hspace{1cm} (27)

Equation (27), when put in the context of the following prediction step (25), 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 the 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 (26), 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. Following de Jong [1989] and Durbin and Koopman [2001], the following fixed interval smoother can be applied for $t = T, T-1, \ldots, 2$ in order to find more precise in-sample values of the static factors and of dynamic factors’ conditional covariances. First we compute

$$r_{t-1} = L_t'r_t + \Lambda W_{t|t-1}\eta_{t|t-1},$$

$$F_{t|T} = F_{t|t-1} + P_{t|t-1}r_{t-1},$$

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

$$P_{t|T} = P_{t|t-1} - P_{t|t-1}\Theta_{t-1}P_{t|t-1},$$

where $\Theta_t$ has been obtained by

$$\Theta_{t-1} = \Lambda W_{t}^{-1}\Lambda + L'_t\Theta_t L_t,$$
with initial value $\Theta_T = 0$. At the end of each step, we get smoothed values for the dynamic factors and their conditional covariances $Q_t$

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

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