Laboratory of Economics and Management

Sant'Anna School of Advanced Studies
Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)
Tel. +39-050-883-343 Fax +39-050-883-344
Email: lem@sssup.it Web Page: http://www.lem.sssup.it/

LEM

Working Paper Series

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

Lucia Alessi*, Matteo Barigozzi*, Marco Capasso*

*Sant'Anna School of Advanced Studies, Pisa, Italy.

2006/25 **December 2007**

ISSN (online) 2284-0400

Dynamic Factor GARCH

Multivariate Volatility Forecast for a Large Number of Series

Lucia Alessi * Matteo Barigozzi * Marco Capasso *

December 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. A method is suggested for estimating and predicting conditional variances and covariances of the original data series. We suggest also a modified version of the Kalman filter as a way to get a more precise estimation of the static and dynamic factors' in-sample levels and covariances in order to achieve better forecasts. Simulation results on different panels with large time and cross sections are presented. Finally, we carry out an empirical application aiming at comparing estimates and predictions of the volatility of financial asset returns. The Dynamic Factor GARCH model outperforms the univariate GARCH.

Keywords: Dynamic Factors, Multivariate GARCH, Covolatility Forecasting
JEL-classification: C32, C52, C53

*Sant'Anna School of Advanced Studies - Laboratory of Economics and Management
Piazza Martiri della Libertà 33, I-56127 Pisa, Italy.

E-mail: lucia.alessi@sssup.it
E-mail: matteo.barigozzi@gmail.com
E-mail: marco.capasso@gmail.com

We thank for helpful comments and suggestions Marc Hallin and Marco Lippi. We thank also all the participants to the Second Italian Congress of Econometrics and Empirical Economics (ICEEE) held in Rimini (Italy) in January 2007 and to the conference "Multivariate Volatility Models" held in Faro (Portugal) in October 2007. Of course, the responsibility for any error is entirely our own.

As members of the Center for the Analysis of Financial Markets (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 second moments, 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 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 [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

[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].

previous by Diebold and Nerlove. Harvey et al. [1992] build a modified version of the Kalman 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 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 number of 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 two-step methodology that takes the same approach to factor analysis as in Giannone et al. [2004]. Indeed, 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 section we describe the DF-GARCH model. Section 3 is devoted to the estimation of the model. Section 4 presents a modified Kalman filter useful for forecasting, here 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 (details are in the Appendix). In section 5 we illustrate some results obtained when applying the estimation to simulated panels. In section 6 our method is applied to a financial dataset and the assumptions of the DF-GARCH are tested. Section 7 and 8 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 9. Sometimes we use the word "returns" when referring to the original data series and "(co)-volatilities" when referring to their (co)-variances, 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 covariance stationary, standardized and the second order moments $\Gamma_k = \mathrm{E}[x_t x'_{t-k}]$ exist finite for all $k \in \mathbb{N}$. In the Generalized Dynamic Factor Model (GDFM), as proposed by Forni et al. [2005], it is assumed that each series x_{it} can be written as the sum of two mutually orthogonal unobservable components, the common component χ_{it} and the idiosyncratic component ξ_{it}. The common component is driven by a q-dimensional vector of dynamic common factors (or shocks) $u_t = (u_{1t} \dots u_{qt})'$, and usually, in empirical applications, $q << N$. Each factor is

loaded with possibly different coefficients and lags. Formally in vector notation

$$x_t = \chi_t + \xi_t = D(L)u_t + \xi_t \,. \tag{1}$$

By assumption u_t is an orthonormal white noise and ξ_t has zero mean and is stationary. Moreover, ξ_{it} is orthogonal to u_{jt-k} for all integers k, i and j. The entries of $D(L)$ are square-summable polynomials in the lag operator $d_{ij}(L)$. They are one-sided filters, and, in principle, of infinite order. We assume that the q largest eigenvalues of the spectral density matrix of the common component, $\Sigma^{\chi}(\theta)$, diverge almost everywhere, for $\theta \in [-\pi, \pi]$, as the number of series goes to infinity, while the largest eigenvalue of the spectral density matrix of the idiosyncratic component is uniformly bounded. 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. Both these assumptions on the eigenvalues of the spectral density matrices are necessary to guarantee the identification of the common component.

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

$$u_t | \, \mathcal{I}_{t-1} \sim \mathrm{N}\left(0, Q_t\right) \,,$$

Q_t being a non-diagonal matrix changing over time, and where $\mathcal{I}_t$ contains all the information available at time t. In the rest of the paper we assume that the conditional variance of the dynamic factors is governed by

$$Q_t = C_0' C_0 + C_1' u_{t-1} u_{t-1}' C_1 + C_2' Q_{t-1} C_2 \,. \tag{2}$$

This is the full BEKK representation of multivariate GARCH models as in Engle and Kroner [1995]. As we said above, 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 might lead to think that many series could be handled by doing without a factor decomposition. However, this is not true, two parameters are too few to describe the whole dynamics of conditional correlations when the number of series is large. Moreover, the estimation of a DCC model would require the inversion of a $N \times N$ matrix which may be computationally very slow. This is the practical reason for applying a 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 in the empirical applicaton, but focus on the BEKK representation in the next section. Concerning the BEKK representation (2), notice that as in the case of univariate GARCH we have $\mathrm{E}[Q_t] = \mathrm{E}[u_t u_t'] = \mathrm{I}_q$. Hence the additional condition $C_0' C_0 = \mathrm{I}_q - C_1' C_1 - C_2' C_2$ on the coefficients applies.
We assume that idiosyncratic parts evolve according to a univariate ARMA-GARCH model

$$\xi_t | \, \mathcal{I}_{t-1} \sim \mathrm{N}\left(\mu_t, R_t\right) \,,$$

where, for each point in time, R_t is a diagonal matrix containing the conditional variances of each idiosyncratic series

$$R_{iit} = \rho_{0i} + \rho_{1i} \xi_{it-1}^2 + \rho_{2i} R_{iit-1} \qquad i = 1, \ldots, N \,. \tag{3}$$

Since in a factor model the bulk of the dynamics of the series is usually concentrated in the common part, we do not look for a way to model conditional covariances between idiosyncratic parts, and we limit ourselves to modeling only the conditional variances. Moreover, we do not define a particular structure of the ARMA process governing the conditional mean of ξ_t. 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 in what follows (i.e. $\mu_t = 0$). However, notice that the general description of our procedure allows also for this possibility.

In the static representation the common part of the factor model is made of $r < N$ common static factors F_t, thus the factor model is written as

$$x_t = \Lambda F_t + \xi_t \,, \tag{4}$$

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

$$F_t = AF_{t-1} + Bu_t \,, \tag{5}$$

where u_t is the 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 (4) contains a special case of the usual dynamic factor model (see Forni et al. [2006] and Bai and Ng [2005]). By inverting equation (5) and using the lag operator L, the static factors are an MA(∞) of the dynamic factors, namely

$$F_t = (\mathrm{I} - AL)^{-1} B u_t \,.$$

We are then back to the dynamic representation of the factor model

$$x_t = \Lambda(\mathrm{I} - AL)^{-1} B u_t + \xi_t \,,$$

where the loadings of (1) are now

$$D(L) = \sum_{k=0}^{\infty} D_k L^k = \sum_{k=0}^{\infty} \Lambda A^k B L^k \,,$$

It is clear that a VAR(1) representation of the static factors is general enough for the common part to be an MA(∞) representation. This 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]. Usually we truncate the infinite summation at a maximum lag s such that we have a number of static factors $r = q(s+1)$. This is equivalent to assume that the static factors are just the dynamic factors with all their lags: $F_t = (u_t', u_{t-1}' \dots u_{t-s}')'$.

3 Estimation

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, $\hat{\Sigma}^x(\theta)$, is estimated by applying

the Fourier transform to the contemporaneous and lagged sample covariance matrices $\hat{\Gamma}_k^x$.[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, with which we compute $\hat{\Sigma}^\chi(\theta)$ and $\hat{\Gamma}_k^\chi$, using the inverse Fourier transform. The estimate of the covariance matrix of the idiosyncratic part, $\hat{\Gamma}_k^\xi$, is then obtained by difference.
In the second step of the procedure we move to a static representation of the model in which we estimate the r largest generalized eigenvectors $Z = (z^{(1)} \dots z^{(r)})$ of $\hat{\Gamma}_k^\chi$ with respect to $\hat{\Gamma}_k^\xi$. The common component is finally estimated as

$$\hat{\chi}_t = \hat{\Gamma}_0^\chi Z (Z' \hat{\Gamma}_0^\chi Z)^{-1} Z' x_t \,.$$

We obtain the idiosyncratic component simply as difference between the original series x_t and the estimated common component. It is worth noticing the key difference between this dynamic approach and the static principal component method used by Stock and Watson [2002]. Indeed, while the first exploits the information contained in lagged covariance matrices, the latter makes use of contemporaneous covariances only.

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 $\hat{G}_t = Z' x_t$. These are identified only up to an orthogonal transformation, i.e. we actually estimate the generalized principal components $\hat{G}_t = \Omega \hat{F}_t$ with $\Omega\Omega' = \mathrm{I}$. Hereafter we suppose that $\Omega = \mathrm{I}$, since $\hat{F}_t$ and $\hat{G}_t$ 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. Notice however that, as shown by Forni et al. [2005], the space spanned by the static factors and the common component are always identified. As a consequence, also the conditional covariance matrix of χ_t is identified.
From (4), given a sample length T, we also have

$$\hat{\Lambda} = \hat{\Gamma}_0^x Z \left(Z' \hat{\Gamma}_0^x Z \right)^{-1} .$$

We thus have an estimate of equation (4).

In order to estimate also equation (5) we need estimates of A and B. Following Giannone et al. [2004] and Forni et al. [2006], we have

$$\hat{A} = Z' \hat{\Gamma}_1^x Z \left(Z' \hat{\Gamma}_0^x Z \right)^{-1} , \tag{6}$$

and we can also estimate the covariance of Bu_t which is given by

$$\hat{\Gamma}_0^{Bu} = \frac{1}{T-1} \left(Z' \hat{\Gamma}_0^x Z - \hat{A} Z' \hat{\Gamma}_0^x Z \hat{A}' \right) .$$

The estimation of A, as given by (6), is not efficient. Indeed, equation (5) not only represents an autoregression of the dependent variable, but also shows an error term Bu_t which is a linear

[4] Hereafter we consider the estimates

$$\hat{\Gamma}_k^x = \frac{1}{T-1} \sum_{t=1}^{T} x_t x_{t-k}' \,.$$

With hatted symbols we always denote estimates.

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 (6). 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

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

from which we get

$$\hat{B} = MS^{1/2} \,. \tag{7}$$

Finally, by inverting (5) and using the estimates for F_t, A, B and Λ we have an estimate of the dynamic factors

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

We now take into account the hetroskedasticity of the idiosyncratic part and of the dynamic factors. We apply a univariate GARCH model to each series ξ_{it} as in (3), and so we obtain the conditional variance matrix of the idiosyncratic component $\hat{R}_t$. As in the estimation of the GDFM we do not consider out-of-diagonal elements of the covariance matrix of the idiosyncratic part (see Forni et al. [2005] for a justification in the unconditional case). If we model also the conditional mean of the idiosyncratic part then also an ARMA model should be estimated for each series, in order to obtain an estimate of the conditional mean $\hat{\mu}_t$. On the other side, we apply a multivariate GARCH model to the estimate of the dynamic factors $\hat{u}_t$ obtained previously, that is we apply the model described by (2) and we get maximum likelihood estimates of the parameters $\hat{C}_0$, $\hat{C}_1$, $\hat{C}_2$ as well as of the dynamic factors' conditional covariances $\hat{Q}_t$.

Summing up, from the estimates of equations (4), (5), (2), and (3), we have the estimated DF-GARCH model written in state-space form as

$$\begin{aligned} x_t &= \hat{\Lambda}\hat{F}_t + \hat{\xi}_t \qquad \text{measurement equation}\,, \\ \hat{F}_t &= \hat{A}\hat{F}_{t-1} + \hat{B}\hat{u}_t \quad \text{transition equation}\,, \end{aligned} \tag{8}$$

where

$$\begin{aligned} \hat{\xi}_t|\,\mathcal{I}_{t-1} &\sim \mathrm{N}(\mu_t, \hat{R}_t) \qquad \hat{R}_{iit} = \hat{\rho}_{0i} + \hat{\rho}_{1i}\hat{\xi}_{it-1}^2 + \hat{\rho}_{2i}\hat{R}_{iit-1} \\ \hat{u}_t|\,\mathcal{I}_{t-1} &\sim \mathrm{N}(0, \hat{Q}_t) \qquad \hat{Q}_t = \hat{C}_0'\hat{C}_0 + \hat{C}_1'\hat{u}_{t-1}\hat{u}_{t-1}'\hat{C}_1 + \hat{C}_2'\hat{Q}_{t-1}\hat{C}_2 \,. \end{aligned} \tag{9}$$

The dynamics of x_t is specified through an unobserved component model as in (8), while the conditional heteroskedasticity of x_t comes into the model through (9). $\hat{Q}_t$ is the conditional covariance matrix estimated from a multivariate GARCH on the dynamic factors, while $\hat{R}_t$ has been obtained by from a univariate GARCH estimated on the idiosyncratic component. All other parameters ($\hat{\Lambda}$, $\hat{A}$ and $\hat{B}$) have been obtained by the one-sided estimation and the subsequent operations we have already described.

From (8), we have a complete specification of the dynamics of the static factors, whose first two conditional moments are

$$\mathrm{E}[\hat{F}_t|\,\mathcal{I}_{t-1}] = \hat{A}\hat{F}_{t-1} \,, \tag{10}$$

and

$$\mathrm{E}[(\hat{F}_t - E[\hat{F}_t|\,\mathcal{I}_{t-1}])^2|\,\mathcal{I}_{t-1}] = \hat{B}\hat{Q}_t\hat{B}' \,. \tag{11}$$

Therefore the estimated conditional variance of the common component is

$$\hat{\Gamma}_t^{\chi} = \hat{\Lambda}\hat{B}\hat{Q}_t\hat{B}'\hat{\Lambda}' \,. \tag{12}$$

In the empirical application that follows we are interested only in equation (11), because when dealing with financial data we usually do not specify any conditional mean model. Nevertheless the DF-GARCH is able to provide joint estimates of both the first two conditional moments of large datasets that may be useful in other applications different from finance (e.g. with macroeconomic data). Although an estimate of χ_t is already available from the one-sided estimator by Forni et al. [2005], the conditional mean model specified in (10) turns out to be very useful when forecasting the level of a series (see Giannone et al. [2004] and Barigozzi and Capasso [2007]). The conditional covariance matrix of x_t is obtained by adding to (12) the conditional covariance matrix of the idiosyncratic part. Once again, notice also that usually we do not specify a conditional mean model for the idiosyncratic part, as it is almost always the case in financial applications. If we would like to impose an ARMA model on the idiosyncratic part, then the conditional mean of x_t should take it into account.

Finally, notice that (11) holds if F_t where observed, otherwise an additional term expressing the uncertainty in the estimate $\hat{F}_t$ should be added. This term will be considered in the Kalman filter step presented in the next section, when a new estimate $\tilde{F}_t$ will be compared with $\hat{F}_t$. This term however is very small given that the Kalman filter operates in order to minimize it.

4 Forecasting with Kalman filter

The estimation method presented in the previous section can be straightforwardly applied also when forecasting. However, in this case we propose an additional estimation step based on a modified Kalman filter that provides us with new, and theoretically better, estimates of the most recent realizations of static and dynamic factors. Indeed, we need a safer estimation for the static and dynamic factors (and consequently of their conditional covariances), 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. While the factors, together with their conditional 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). 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 the previous section. 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.
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 $x_t^* = x_t - \mu_t$ that is obtained as the original series minus the conditional mean of the idiosyncratic part. As a consequence, the error term ξ_t of the model (8) is replaced by a new term ξ_t^*, whose process has zero conditional mean.

Once the Kalman filter estimation is completed, we have a new estimate of the static factors $\tilde{F}_t$, of the dynamic factors $\tilde{u}_t$, and of their conditional covariance matrix $\tilde{Q}_t$. We therefore have new estimates of the common and of the idiosyncratic parts: $\tilde{\chi}_t$ and $\tilde{\xi}_t$.[5] The predicted covariance matrix for the dynamic factors, $\tilde{Q}_{T+1|T}$, is obtained by applying the prediction steps to the smoothed estimations of last period T (see equation (21) in the Appendix for details). The predicted diagonal covariance matrix $\tilde{R}_{T+1|T}$ of the idiosyncratic part is obtained by applying to $\tilde{\xi}_T$ the parameters of the univariate GARCH models that we have estimated before the Kalman filter. Given the assumed orthogonality of common and idiosyncratic part, and given (12), we get the one-step-ahead out-of-sample forecast of the conditional covariance of the common part and of the observable data series as

$$\tilde{\Gamma}^{\chi}_{T+1|T} = \hat{\Lambda}\hat{B}\tilde{Q}_{T+1|T}\hat{B}'\hat{\Lambda}' \ ,$$

$$\tilde{\Gamma}^{x}_{T+1|T} = \tilde{\Gamma}^{\chi}_{T+1|T} + \tilde{R}_{T+1|T} \ . \tag{13}$$

Notice that this conditional covariance matrix is positive definite by construction. Indeed $\tilde{R}_t$ and and $\tilde{Q}_t$ come respectively from N univariate GARCH and a BEKK multivariate GARCH, and the first term of (13) is clearly a quadratic form.

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

$$\tilde{\rho}^{x}_{ij,T+1|T} = \frac{\tilde{\Gamma}^{x}_{ij,T+1|T}}{\sqrt{\tilde{\Gamma}^{x}_{ii,T+1|T}\tilde{\Gamma}^{x}_{jj,T+1|T}}} \ . \tag{14}$$

The off-diagonal terms of $\tilde{\Gamma}^{x}_{T+1|T}$ as computed in (13) do not consider the mild correlations (and conditional correlations) among the idiosyncratic terms. Correlation estimations given by (14) 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 second term on the right-hand-side of (13).

5 Simulation results

In order to assess the validity of our model, we apply our estimation method to simulated panels that differ in the cross and time dimension, in the number of dynamic and static factors and in the amount of variance explained by the common part with respect to the total. As possible

[5] If we consider also the conditional mean of the idiosyncratic part, the new estimate of the common part is $\tilde{\chi}_t^*$, and by difference we obtain a new idiosyncratic part $\tilde{\xi}_t^* = x_t - \tilde{\chi}_t^*$ different from $\tilde{\xi}_t$. We assume that $\tilde{\xi}_t^*$ are the realizations of a process having the conditional variances $\tilde{R}_t$ and as conditional mean the previously estimated $\hat{\mu}_t$.

values we choose $N = 75, 150$ and $T = 250, 500, 750$ and for every possible combination we take 2 or 3 dynamic factors, loaded with 2 or 4 lags, and an average variance ratio (VR) between idiosyncratic and common parts of 0.3 or 0.5. We simulate u_t as a multivariate GARCH follwing the full BEKK as in (2). Idiosyncratic parts are simply simulated as univariate GARCH(1,1) as in (3).[6] Parameters of GARCH and BEKK are extracted from uniform distributions with range determined, according to usual empirical estimates, as follows:

1. C_1 has diagonal elements in $[0.1, 0.5]$ and off-diagonal elements in $[-0.2, 0.2]$;
2. C_2 has diagonal elements in $[0.8, 0.95]$ and off-diagonal elements in $[-0.15, 0.15]$;
3. ρ_1 has values in $[0, 0.1]$ and ρ_2 has values in $[0.8, 0.95]$.

At each extraction of the parameters, positive definiteness of the simulated conditional variances has been checked before proceeding.
We prefer to simulate the GDFM as in (1) instead of simulating it in its static form (4). Indeed, (1) is the real way in which dynamic factors are loaded, while (4) and (5) are just a possible way to represent the data, which is necessary for estimation. Such a choice avoids also the need of simulating the static factors. Therefore, for every model considered, we do not extract values for Λ, A, and B, but we instead simulate the loadings $D(L)$ by extracting them from a standard normal distribution. These loadings are then renormalized in such a way that on average x_t has unit variance and zero mean, and the chosen VR is on average respected. Although in principle each series in the factor model should have exactly unit variance, we prefer not to standardize data before the estimation to avoid an additional step.

A first visual proof of the goodness of our estimation method is given in figure 5, when considering the confidence interval at 90% level. Figures 2.1, 2.2, 3.1 and 3.2 show four examples of estimated and simulated conditional variances and covariances, with remarkably good performance of our estimation method.

For each simulated dataset we repeat the procedure 250 times. At every replication, in order to compare our estimation of χ_t with the results obtained in the same way by Forni et al. [2005], we compute

$$H_1 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\chi}_{it} - \chi_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\chi_{it}^2} \,. \tag{15}$$

We also compute an analogous measure for the elements of the conditional covariance matrix of the common part

$$H_2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\Gamma}_{iit}^{\chi} - \Gamma_{iit}^{\chi})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(\Gamma_{iit}^{\chi})^2} \,. \tag{16}$$

Concerning H_2, we consider only the diagonal elements of the conditional covariance matrix, thus obtaining a measure of the error made when estimating volatilities. We then consider only the out-of-diagonal elements in the upper-triangular part of the conditional covariance matrix, in order to measure the error made when estimating covolatilities. We thus compute the analogue of H_2 when summing over all the $N(N-1)/2$ elements of the upper-triangular part of $\hat{\Gamma}_t^{\chi}$. In tables 1 and 2 we report the mean values and the standard errors of H_1 and

[6] All computations and simulations in this paper have been performed by using the standard Matlab software packages (v.7.0) plus the freely available toolboxes `MATNEM` by Christian T. Brownlees and `ucsd_garch` by Kevin K. Sheppard; the code used in the next section has been kindly provided by Roman Liška.

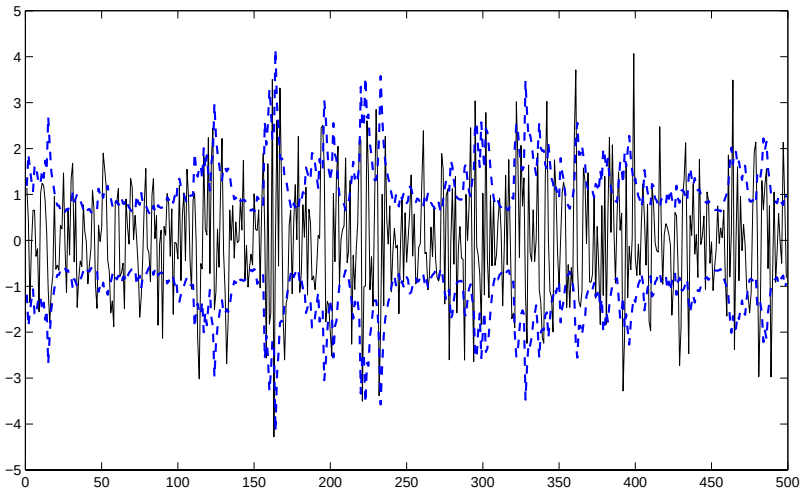


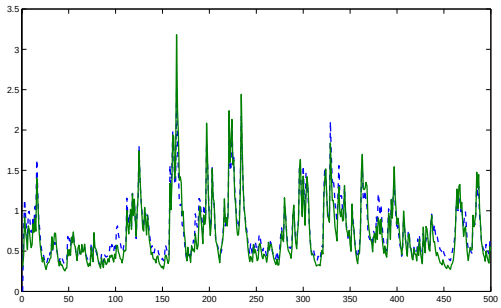
Figure 1: Estimated confidence interval for $N = 150$, $T = 500$, $q = 2$, $s = 4$, and VR= 0.3. Simulated χ_t: solid line. Estimated 5^{th} and 95^{th} conditional percentiles: dashed line.

the two H_2's across the 250 replications. Results are shown for the case in which we do not perform the Kalman filter step and for the last quarter of the sample when performing also the Kalman filter step. Concerning the common component, we find results that are similar to the ones obtained by Forni et al. [2005]. It is evident the improvement on the last observations when using the Kalman filter.

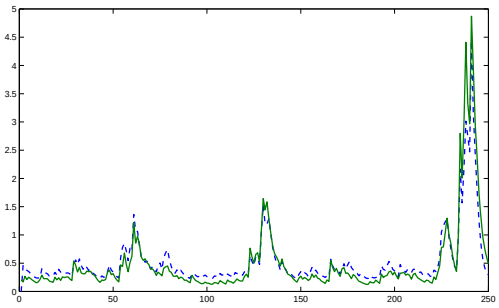
Finally, we run Mincer and Zarnowitz [1969] in-sample regressions. At each replication and for each series i, we consider all the estimates we have computed we run a regression based upon simulated and estimated conditional variances

$$\Gamma^{\chi}_{iit} = b_0 + b_1 \hat{\Gamma}^{\chi}_{iit} + e_{it} \qquad i = 1, \ldots, N\,.$$

We run similar regressions for the $N(N-1)/2$ conditional covolatilities series. 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 estimated conditional covariance that can be explained by the model, thus giving a general idea of its potentialities. Tables 4 and 3 report the average R^2 over all series and over all 250 replications, Once again notice the improvement achieved on the estimate of the last in-sample observations, when using the Kalman filter. This supports our suggestion of running a modified Kalman filter when we are interested in making accurate out-of-sample forecasts. Finally notice from both tables that, as T increases, the estimation greatly improves.

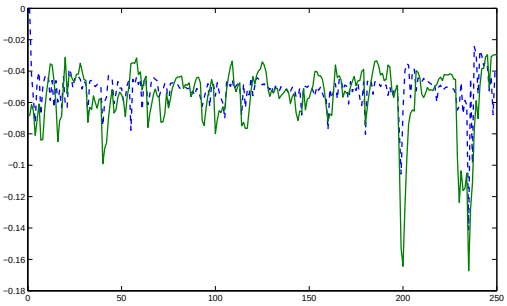


2.1: $N = 150$, $T = 500$, $q = 2$, $s = 4$, and VR= 0.3.

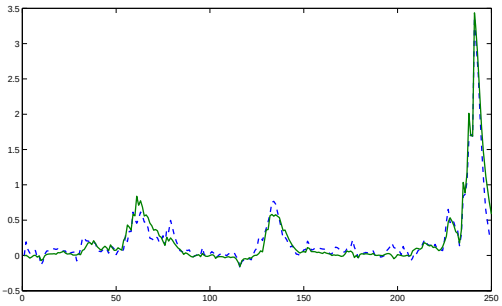


2.2: $N = 150$, $T = 250$, $q = 3$, $s = 2$, and VR= 0.3.

Figure 2: Conditional variances. Simulated: solid line. Estimated: dashed line.



3.1: $N = 150$, $T = 500$, $q = 2$, $s = 4$, and VR= 0.3.



3.2: $N = 150$, $T = 250$, $q = 3$, $s = 2$, and VR= 0.3.

Figure 3: Conditional covariances. Simulated: solid line. Estimated: dashed line.

N	T	q	r	VR	No Kalman			Kalman		
					H_1	H_2		H_1	H_2	
						vol	covol		vol	covol
75	250	2	2	0.3	0.0306	0.0567	0.0812	0.0222	0.0513	0.0747
					0.0114	0.0294	0.0366	0.0087	0.0225	0.0287
75	250	2	2	0.5	0.0542	0.0699	0.1028	0.0398	0.0652	0.0970
					0.0168	0.0337	0.0426	0.0130	0.0309	0.0400
75	250	2	4	0.3	0.0481	0.0729	0.1066	0.0324	0.0684	0.1006
					0.0128	0.0331	0.0433	0.0122	0.0330	0.0425
75	250	2	4	0.5	0.0917	0.0880	0.1354	0.0663	0.0806	0.1256
					0.0316	0.0400	0.0500	0.0297	0.0380	0.0468
75	250	3	2	0.3	0.0335	0.0746	0.1330	0.0250	0.0682	0.1258
					0.0039	0.0309	0.0463	0.0034	0.0289	0.0460
75	250	3	2	0.5	0.0592	0.0883	0.1591	0.0445	0.0807	0.1500
					0.0093	0.0336	0.0495	0.0081	0.0306	0.0469
75	250	3	4	0.3	0.0562	0.0922	0.1668	0.0393	0.0866	0.1592
					0.0081	0.0331	0.0474	0.0096	0.0340	0.0490
75	250	3	4	0.5	0.1069	0.1203	0.2260	0.0831	0.1107	0.2138
					0.0218	0.0426	0.0597	0.0226	0.0390	0.0575
150	250	2	2	0.3	0.0193	0.0576	0.0835	0.0144	0.0519	0.0773
					0.0048	0.0246	0.0316	0.0039	0.0242	0.0317
150	250	2	2	0.5	0.0323	0.0659	0.0979	0.0239	0.0591	0.0904
					0.0062	0.0295	0.0370	0.0047	0.0261	0.0345
150	250	2	4	0.3	0.0325	0.0700	0.1021	0.0216	0.0638	0.0951
					0.0076	0.0285	0.0373	0.0053	0.0281	0.0364
150	250	2	4	0.5	0.0534	0.0847	0.1252	0.0355	0.0752	0.1139
					0.0115	0.0487	0.0461	0.0081	0.0474	0.0426
150	250	3	2	0.3	0.0213	0.0672	0.1217	0.0167	0.0626	0.1161
					0.0026	0.0240	0.0371	0.0025	0.0231	0.0357
150	250	3	2	0.5	0.0387	0.0804	0.1452	0.0306	0.0749	0.1386
					0.0061	0.0348	0.0489	0.0050	0.0349	0.0485
150	250	3	4	0.3	0.0372	0.0809	0.1462	0.0269	0.0775	0.1420
					0.0061	0.0285	0.0413	0.0063	0.0298	0.0422
150	250	3	4	0.5	0.0740	0.1074	0.2008	0.0563	0.1030	0.1942
					0.0146	0.0362	0.0517	0.0136	0.0366	0.0518

Table 1: Mean and standard errors of H_1 and H_2.

N	T	q	r	VR	No Kalman			Kalman		
					H_1	H_2		H_1	H_2	
						vol	covol		vol	covol
75	500	2	2	0.3	0.0274	0.0382	0.0545	0.0174	0.0341	0.0496
					0.0052	0.0213	0.0266	0.0036	0.0202	0.0251
75	500	2	2	0.5	0.0461	0.0425	0.0618	0.0293	0.0366	0.0552
					0.0076	0.0265	0.0306	0.0054	0.0177	0.0235
75	500	2	4	0.3	0.0453	0.0480	0.0682	0.0255	0.0420	0.0613
					0.0084	0.0268	0.0301	0.0056	0.0209	0.0242
75	500	2	4	0.5	0.0774	0.0651	0.0932	0.0464	0.0558	0.0825
					0.0157	0.0368	0.0416	0.0119	0.0285	0.0335
75	500	3	2	0.3	0.0406	0.0638	0.1087	0.0258	0.0588	0.1026
					0.0061	0.0328	0.0441	0.0042	0.0309	0.0424
75	500	3	2	0.5	0.0680	0.0744	0.1274	0.0433	0.0676	0.1189
					0.0106	0.0375	0.0505	0.0073	0.0350	0.0480
75	500	3	4	0.3	0.0667	0.0775	0.1320	0.0404	0.0712	0.1238
					0.0105	0.0318	0.0414	0.0081	0.0315	0.0414
75	500	3	4	0.5	0.1115	0.1087	0.1830	0.0747	0.0971	0.1693
					0.0156	0.0481	0.0579	0.0147	0.0456	0.0561
75	750	2	2	0.3	0.0190	0.0325	0.0438	0.0120	0.0310	0.0424
					0.0034	0.0215	0.0251	0.0027	0.0221	0.0259
75	750	2	2	0.5	0.0325	0.0383	0.0527	0.0208	0.0355	0.0496
					0.0063	0.0208	0.0245	0.0049	0.0220	0.0254
75	750	2	4	0.3	0.0309	0.0441	0.0583	0.0172	0.0414	0.0554
					0.0053	0.0288	0.0325	0.0038	0.0297	0.0333
75	750	2	4	0.5	0.0548	0.0557	0.0773	0.0337	0.0508	0.0719
					0.0145	0.0298	0.0338	0.0131	0.0298	0.0340
75	750	3	2	0.3	0.0378	0.0468	0.0817	0.0253	0.0462	0.0802
					0.0045	0.0243	0.0349	0.0035	0.0256	0.0359
75	750	3	2	0.5	0.0707	0.0615	0.1088	0.0487	0.0621	0.1082
					0.0185	0.0286	0.0385	0.0192	0.0316	0.0418
75	750	3	4	0.3	0.0648	0.0613	0.1064	0.0427	0.0604	0.1041
					0.0082	0.0296	0.0394	0.0103	0.03096	0.0406
75	750	3	4	0.5	0.1231	0.0887	0.1608	0.0963	0.0857	0.1553
					0.0271	0.0493	0.0669	0.0382	0.0498	0.0671

Table 2: Mean and standard errors of H_1 and H_2.

N	T	q	lags	VR	No Kalman			Kalman		
					χ_t	Γ^{χ}_{iit}	Γ^{χ}_{ijt}	χ_t	Γ^{χ}_{iit}	Γ^{χ}_{ijt}
75	250	2	2	0.3	0.9673	0.5775	0.5616	0.9754	0.6344	0.5970
					0.0145	0.2169	0.2252	0.0113	0.2040	0.2194
75	250	2	2	0.5	0.9436	0.5733	0.5504	0.9562	0.6234	0.5823
					0.0187	0.2049	0.2084	0.0158	0.1949	0.2032
75	250	2	4	0.3	0.9509	0.5157	0.4930	0.9661	0.5687	0.5266
					0.0130	0.2292	0.2312	0.0136	0.2276	0.2330
75	250	2	4	0.5	0.9091	0.5067	0.4806	0.9299	0.5630	0.5178
					0.0301	0.2078	0.2100	0.0326	0.2011	0.2081
75	250	3	2	0.3	0.9653	0.4922	0.4689	0.9734	0.5316	0.4889
					0.0040	0.1763	0.1800	0.0038	0.1720	0.1815
75	250	3	2	0.5	0.9398	0.4695	0.4397	0.9524	0.5069	0.4587
					0.0094	0.1707	0.1729	0.0088	0.16846	0.1756
75	250	3	4	0.3	0.9437	0.4558	0.4283	0.9591	0.4991	0.4534
					0.0080	0.1681	0.1677	0.0102	0.1694	0.1713
75	250	3	4	0.5	0.8963	0.3865	0.3519	0.9134	0.4340	0.3788
					0.0201	0.1668	0.1608	0.0246	0.1624	0.1602
150	250	2	2	0.3	0.9792	0.5535	0.5300	0.9838	0.5561	0.5327
					0.0043	0.2443	0.2424	0.0035	0.2454	0.2435
150	250	2	2	0.5	0.9653	0.5182	0.4902	0.9728	0.5227	0.4944
					0.0067	0.2342	0.2331	0.0056	0.2353	0.2340
150	250	2	4	0.3	0.9666	0.5350	0.5110	0.9768	0.5401	0.5159
					0.0076	0.2384	0.2358	0.0057	0.2400	0.2373
150	250	2	4	0.5	0.9467	0.4973	0.4686	0.9623	0.5045	0.4755
					0.0106	0.2308	0.2272	0.0081	0.2338	0.2297
150	250	3	2	0.3	0.9774	0.5208	0.4942	0.9816	0.5640	0.5166
					0.0028	0.1811	0.1824	0.0027	0.1736	0.1803
150	250	3	2	0.5	0.9592	0.5228	0.4900	0.9663	0.5655	0.5116
					0.0067	0.1556	0.1584	0.0057	0.1542	0.1601
150	250	3	4	0.3	0.9622	0.4913	0.4618	0.9715	0.5347	0.4847
					0.0063	0.1678	0.1654	0.0069	0.1649	0.1651
150	250	3	4	0.5	0.9265	0.4418	0.4056	0.9405	0.4827	0.4291
					0.0143	0.1571	0.1527	0.0146	0.1536	0.1528

Table 3: Mean and standard errors for R^2 of the Mincer-Zarnowitz regressions.

N	T	q	lags	VR	No Kalman			Kalman		
					χ_t	Γ^{χ}_{iit}	Γ^{χ}_{ijt}	χ_t	Γ^{χ}_{iit}	Γ^{χ}_{ijt}
75	500	2	2	0.3	0.9717	0.6696	0.6484	0.9814	0.7117	0.6753
					0.0055	0.2152	0.2189	0.0039	0.2056	0.2151
75	500	2	2	0.5	0.9535	0.6156	0.5939	0.9690	0.6222	0.6006
					0.0074	0.2368	0.2383	0.0058	0.2390	0.2402
75	500	2	4	0.3	0.9551	0.6413	0.6176	0.9748	0.6898	0.6506
					0.0075	0.2131	0.2132	0.0049	0.2022	0.2101
75	500	2	4	0.5	0.9255	0.6058	0.5765	0.9550	0.6194	0.5900
					0.0137	0.2094	0.2123	0.0120	0.2128	0.2154
75	500	3	2	0.3	0.9594	0.5951	0.5558	0.9732	0.6265	0.5731
					0.0058	0.1651	0.1680	0.0043	0.1568	0.1653
75	500	3	2	0.5	0.9341	0.5655	0.5233	0.9554	0.6016	0.5445
					0.0093	0.1658	0.1656	0.0073	0.1565	0.1627
75	500	3	4	0.3	0.9355	0.5522	0.5135	0.9608	0.5932	0.5382
					0.0094	0.1760	0.1768	0.0076	0.1705	0.1773
75	500	3	4	0.5	0.8968	0.5096	0.4631	0.9295	0.5515	0.4892
					0.0129	0.1646	0.1630	0.0137	0.1612	0.1639
75	750	2	2	0.3	0.9805	0.8232	0.7963	0.9875	0.8500	0.8127
					0.0036	0.1453	0.1615	0.0028	0.1288	0.1525
75	750	2	2	0.5	0.9668	0.8109	0.7831	0.9782	0.8404	0.8022
					0.0062	0.1539	0.1627	0.0051	0.1368	0.1534
75	750	2	4	0.3	0.9689	0.7829	0.7566	0.9830	0.8172	0.7786
					0.0052	0.1729	0.1818	0.0037	0.1511	0.1706
75	750	2	4	0.5	0.9459	0.7756	0.7455	0.9671	0.8095	0.7689
					0.0142	0.1500	0.1595	0.0149	0.1362	0.1525
75	750	3	2	0.3	0.9623	0.6456	0.6011	0.9740	0.6765	0.6149
					0.0045	0.1577	0.1699	0.0037	0.1527	0.1705
75	750	3	2	0.5	0.9312	0.6326	0.5811	0.9499	0.6650	0.5968
					0.0166	0.1503	0.1550	0.0196	0.1425	0.1523
75	750	3	4	0.3	0.9372	0.6190	0.5747	0.9586	0.6528	0.5917
					0.0077	0.1565	0.1611	0.0107	0.1520	0.1602
75	750	3	4	0.5	0.8855	0.5423	0.4907	0.9065	0.5812	0.5115
					0.0228	0.1626	0.1630	0.0381	0.1580	0.1627

Table 4: Mean and standard errors for R^2 of the Mincer-Zarnowitz regressions.

6 Hypothesis testing on a real dataset

6.1 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-sinchronicity bias (see Martens [2004]).

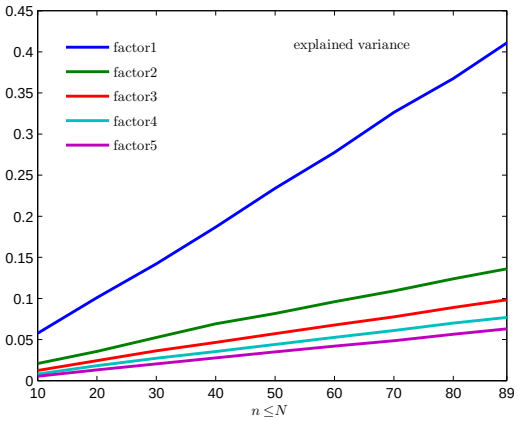
Firstly, we verify that our dataset does fulfill GDFM assumptions on the eigenvalues $\lambda_i(\theta)$ of the spectral density matrix of x_t. According to Brillinger [1981], we define the variance explained by the i^{th} factor as

$$EV_i = \frac{\int_{-\pi}^{\pi} \lambda_i(\theta) d\theta}{\sum_{j=1}^{N} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta} \,. \tag{17}$$

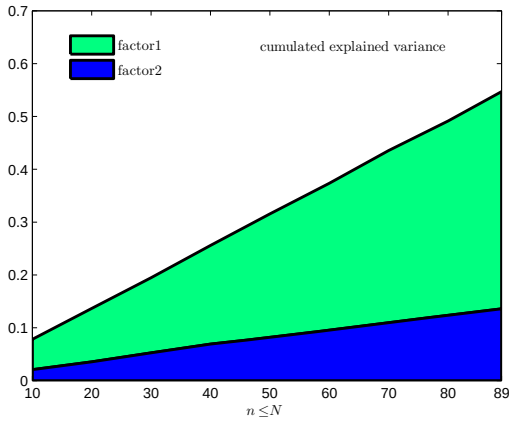
We require that, as $N \longrightarrow \infty$,

$$\begin{cases} EV_i \longrightarrow \infty & \text{for} \quad i = 1, \dots, q \\ \exists\, M \in \mathbb{R}^+ & \text{s.t.} \quad EV_{q+1} \leq M \,. \end{cases} \tag{18}$$

Indeed, as shown for example in figure 4.1 for the subsample including only the first 350 observations, this is the case. Figure 4.2 shows the cumulated explained variance relative to the first two eigenvalues for the same sample.

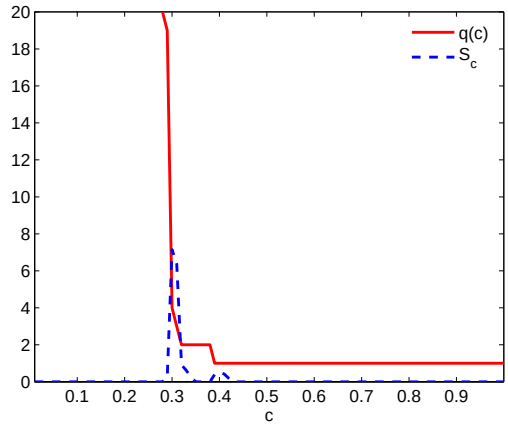


4.1: Explained variance.

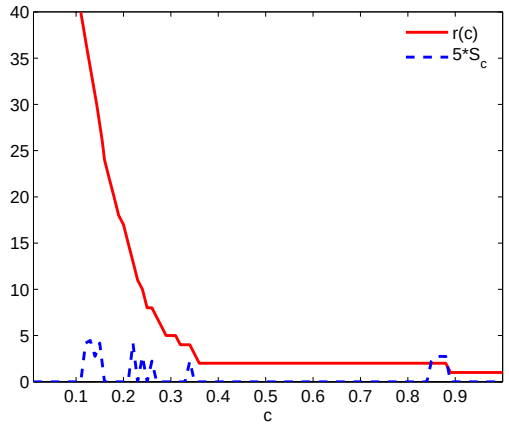


4.2: Cumulated explained variance.

Figure 4: Plots of diverging variances as $N \to \infty$.



5.1: Hallin-Liška criterion.



5.2: IC1 criterion.

Figure 5: Determining the number of factors.

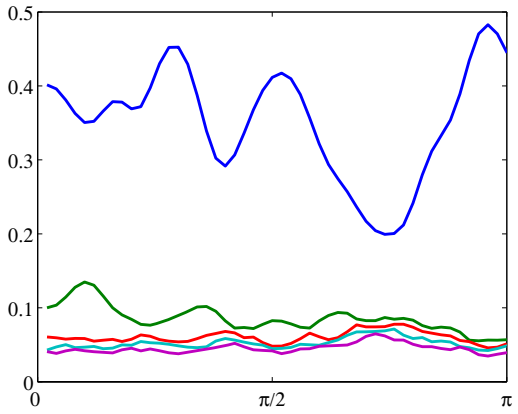
6.2 The number of dynamic factors

We do not rely only on the intuition coming from figures 4.1 and 4.2 for determining the number of dynamic factors to include in the model. We thus 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 of the covariogram-smoothing version of the criterion.[7] 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_{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 a 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 5.1, 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 stability intervals. In other words, $\hat{q}$ corresponds to the plateau of the solid line associated with the second flat zero-level dashed line.[8] This procedure, suggested by Hallin and Liška, indicates the existence of two common dynamic factors.

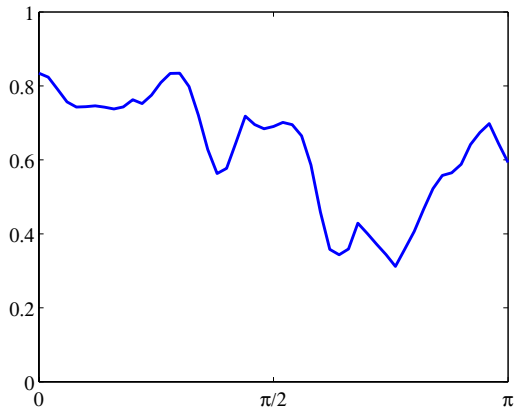
6.3 The number of static factors

In order to find the number r of static factors, some criteria are available (e.g. see Bai and Ng [2002] and Onatski [2006]). We first implement the six consistent criteria by Bai and Ng [2002] which 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.

[7] Hallin and Liška suggest that this form has better finite sample performance.
[8] Notice that the first stability interval always corresponds to q_{max}



6.1: The five largest dynamic eigenvalues.



6.2: Spectral density of the common component.

Figure 6: Evidence of dynamics in asset returns.

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. However it is recognized by the literature (e.g. see Forni et al. [2006]) that these criteria highly depend on r_{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 similar to the one used in determining the number of common dynamic factors (for details on this criterion see Alessi et al. [2007]). In figure 5.2 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.

A heuristic argument is based on the explained variance by the first q factors. We need as many static factors as it is necessary to explain the same amount of variance which is explained by the selected common dynamic factors (in our case 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 $r = 6$.

6.4 A comment on the dynamics of asset returns

In figure 6.1 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% cor-

responding 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 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 show 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 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 $\Sigma^{\chi}(\theta) = D\left(e^{i\theta}\right)\Sigma^{u}(\theta)D\left(e^{-i\theta}\right)'$. Notice that actually $\Sigma^{u}(\theta)$ is constant as u_t is assumed to be white noise, but $D\left(e^{i\theta}\right)$ 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 χ_t is equal to a matrix $D\frac{\sigma^2_{u_t}}{2\pi}D'$ 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 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.

6.5 Testing for conditional heteroskedasticity

The main hypothesis of the DF-GARCH is the conditional heteroskedasticity of the common dynamic factors. In order to test this assumption, we employ the usual ARCH test up to 10 lags. As shown in table 5, for the first in-sample used we find that one of the two dynamic factors is highly heteroskedastic. In particular, this happens for the most important 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. The most important factor determines the conditional heteroskedasticity of asset returns. Good news for our model since having at least one conditionally heteroskedastic dynamic factor justifies our assumptions. Notice that the model itself is flexible enough to accomodate cases with some conditionally heteroskedastic and some conditionally homoskedastic dynamic factors.

ARCH order	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Case $q = 2$										
u_{1t}	0.71	47.67*	72.44*	72.86*	81.43*	87.68*	88.04*	92.71*	93.18*	93.06*
u_{2t}	2.54	2.53	3.44	8.13	9.03	11.26	11.59	12.78	12.41	12.62
Case $q = 1$										
u_t	0.24	47.26*	75.35*	75.74*	85.02*	88.60*	88.77*	92.75*	96.22*	95.98*

Table 5: ARCH-test on u_t for heteroskedasticity (* significant at 99%). Observations from $t = 1$ to $t = 350$, i.e. first in-sample.

Model	$\mathbf{Prob}\left\{\lvert\mathbf{x_{it}}\rvert > \mathbf{1.65}\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{\mathbf{ii,t}}\right\}$	$\mathbf{Prob}\left\{\mathbf{x_{it}} < -\mathbf{1.65}\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{\mathbf{ii,t}}\right\}$
GARCH	0.0881	0.0458
DF-GARCH (BEKK)(2 lags)	**0.0986**	0.0507
DF-GARCH (DCC)(2 lags)	0.0972	**0.0499**

Table 6: Interval predictions. Average results.

7 Empirical estimation and prediction of volatility

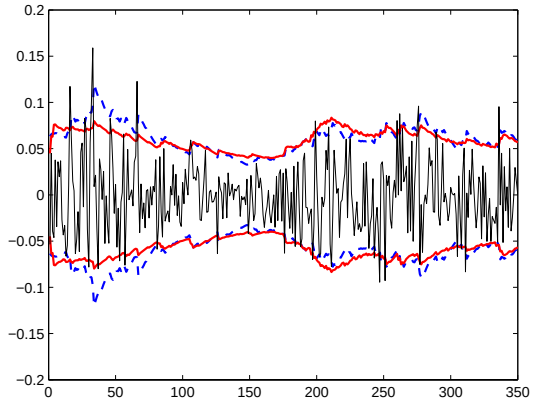
7.1 In-sample performance

We estimate the DF-GARCH with 2 dynamic and 6 static factors and a univariate GARCH using the first 350 observations (the in-sample window) and we look at the confidence intervals at 90% significance level under the assumption of normality.[9] Figure 7.1 shows the results for four asset returns. The two models have qualitatively a similar performance. If the estimation is good 10%, of the observations of series x_{it} should lie outside the interval $[-1.65\hat{\Gamma}^x_{ii,t}, +1.65\hat{\Gamma}^x_{ii,t}]$. Moreover, given that we are dealing with financial data we consider the 95% confidence level Value-at-Risk (VaR) prediction for each series, which is simply the 5-th conditional percentile of the distribution of returns. Therefore, under the assumption of normality of the retruns, we should check that $\text{Prob}\{x_{it} < -1.65\hat{\Gamma}^x_{ii,t}\} = 5\%$. VaR is considered as a measure of risk in financial applications, given that it concentrates only on the lower tail of the distribution. Table 6 reports the results averaged on the 89 series. The DF-GARCH has a comparable and sometimes even better performance than the univariate GARCH. These are encouraging results concerning the in-sample properties of our estimation method. We now move to a forecasting exercise which we believe to be the real test for the performance of a model.

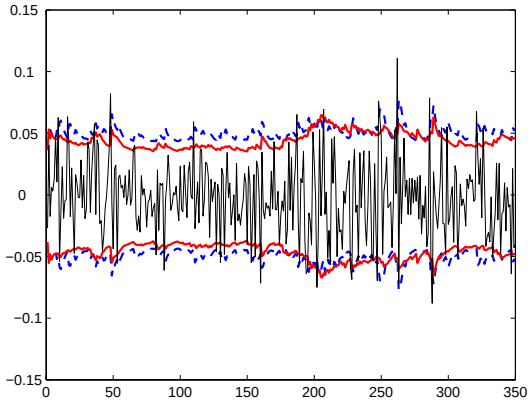
7.2 Out-of-sample performance

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 both the full BEKK and the DCC specification. Whenever we name just "DF-GARCH", we refer to the

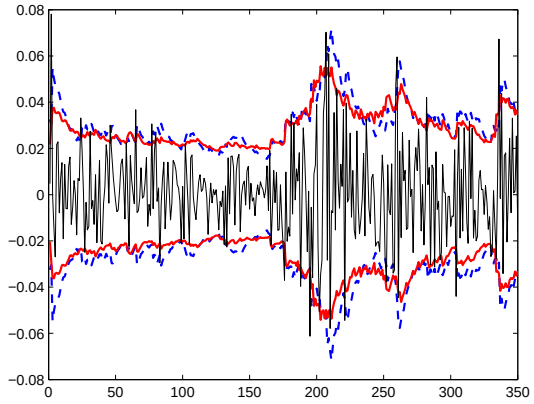
[9]Notice that we do not model the conditional mean of returns as it is always the case with financial data.



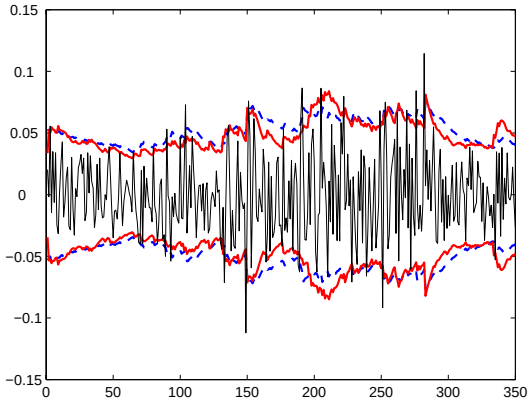
7.1: British Airways.



7.2: EMI.



7.3: Northern Rock.



7.4: Vodafone.

Figure 7: Confidence intervals. Univariate GARCH confidence interval: thick dashed line. DF-GARCH (BEKK) confidence interval: thick solid line. Return series: thin solid line.

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.[10]

Following Andersen et al. [2003], we evaluate the volatility forecasts of our model by run-

[10]We report only average results and only compared to realized volatility. Detailed results for all proxies and all series are available upon request.

Model	$\hat{\mathbf{b}}_\mathbf{0}$	$\hat{\mathbf{b}}_\mathbf{1}$	$\mathbf{R^2}$
GARCH	0.0026	0.3995	0.0780
DF-GARCH (BEKK)(no lags)	0.0002	1.6064	0.1084
DF-GARCH (BEKK)(2 lags)	0.0002	1.6715	**0.1100**
DF-GARCH (DCC)(no lags)	0.0000	1.9384	0.0981
DF-GARCH (DCC)(2 lags)	0.0000	2.0188	0.1000

Table 7: Volatilities. Average results for MZ regressions.

ning a Mincer-Zarnowitz regression. For each series i, we consider all the predictions we have computed at each iteration of our rolling scheme and we run a regression based upon real and predicted conditional variances:

$$V_{iT+k+1} = b_0 + b_1 \hat{\Gamma}^x_{iiT+k+1} + e_{iT+k+1} \qquad k = 0, \ldots, 99\,,$$

where T is the last period of the first in-sample considered (i.e. $T = 350$) and the volatility proxy V_{iT+k+1} is the realized volatility at time $T+k+1$ for series i. On the other side, $\hat{\Gamma}^x_{iiT+k+1}$ represents the one-step-ahead volatility forecast at time $T+k+1$, as predicted at time $T+k$, for series i. We 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 7 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 R^2 coefficient 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 the univariate GARCH for 64 series (i.e. 72% of the total). The static application of DF-GARCH also performs better than the 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

$$RMSE_i = \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}\left[\hat{\Gamma}^x_{iiT+k+1} - V_{iT+k+1}\right]^2} \qquad K = 100 \qquad i = 1, \ldots, N\,,$$

where $\hat{\Gamma}^x_{iiT+k+1}$ 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 8, together with three other statistics:

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

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

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 the DF-GARCH is estimated in a static way, Q becomes 0.9823.

3. In order to compare the out-of-sample performance with the results obtained from simulations we also compute a statistics that is the analogous of (15) for out-of-sample predictions. Namely, for each iteration we compute

$$H_3 = \frac{\sum_{i=1}^{N} (\hat{\Gamma}^x_{iiT+1} - V_{iT+1})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} V_{it}^2 / T} . \tag{19}$$

In table 8 we report the value of H_3 averaged on all the 100 iterations.

Model	**RMSE**	$\mathbf{H_3}$	**P**	**Q**
GARCH	$3.4938\ 10^{-3}$	0.6807	n.a.	n.a.
DF-GARCH (BEKK)(no lags)	$3.3988\ 10^{-3}$	**0.5994**	80.899	0.9823
DF-GARCH (BEKK)(2 lags)	$\mathbf{3.3935\ 10^{-3}}$	0.5998	**82.022**	**0.9798**
DF-GARCH (DCC)(no lags)	$3.4088\ 10^{-3}$	0.6007	77.528	0.9862
DF-GARCH (DCC)(2 lags)	$3.4021\ 10^{-3}$	0.6007	79.755	0.9839

Table 8: Average RMSEs, average H_3, P and Q with respect to the univariate GARCH.

7.3 Testing for equal predictive accuracy

We evaluate the performance of the DF-GARCH by means of two different tests of predictive accuracy. In table 9 we show the results of the usual Diebold and Mariano [1995] test of equal predictive accuracy. Given the predictions of a variable y from two competing models (say a and b) we compute the difference $d_t = \mathrm{E}[(y^a_{t+h|t} - y_{t+h})^2 - (y^b_{t+h|t} - y_{t+h})^2]$ 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

$$\hat{f}_{it} = \left(V_{it} - \hat{\Gamma}^x_{iitG}\right)^2 - \left[\left(V_{it} - \hat{\Gamma}^x_{iitDF}\right)^2 - \left(\hat{\sigma}^2_{itG} - \hat{\Gamma}^x_{itDF}\right)^2\right], \tag{20}$$

where V_{it} represents the realized volatility at time t and $\hat{\Gamma}^x_{iit}$ 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), for series i. 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 9 we show a summary of our results of the tests by Diebold and Mariano [1995] and Clark and West [2007] both for the 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, the DF-GARCH with BEKK specification and estimated with two lags performs better than the traditional GARCH for 70 series (79% of total).

	DF-GARCH 2 lags BEKK	**DF-GARCH 0 lags BEKK**	**DF-GARCH 2 lags DCC**	**DF-GARCH 0 lags DCC**
Diebold-Mariano Test				
number of series for which DF-GARCH outperforms GARCH at 10%	53	58	50	52
number of series for which DF-GARCH outperforms GARCH at 5%	41	40	41	41
Clark-West Test				
number of series for which DF-GARCH outperforms GARCH at 10%	70	70	68	67
number of series for which DF-GARCH outperforms GARCH at 5%	58	60	55	56

Table 9: Diebold-Mariano and Clark-West tests results against traditional univariate GARCH.

8 Empirical prediction of covolatility

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. These vectors are obtained by stacking all the $N(N-1)/2$ covariances. Results averaged over the whole sample are shown in table 10. 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 correlation be somehow not negligible in practice, it should 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. Moreover, the average RMSE is now an order of magnitude lower than the RMSE obtained when forecasting the conditional variances.

Model	$\hat{\mathbf{b}}_\mathbf{0}$	$\hat{\mathbf{b}}_\mathbf{1}$	$\mathbf{R^2}$	**Right Signs %**	**RMSE**	$\mathbf{H_3}$
DF-GARCH (BEKK)(no lags)	0.0000	0.1751	**0.1426**	72.018	$3.4360\ 10^{-4}$	0.0120
DF-GARCH (BEKK)(2 lags)	0.0000	0.1824	0.1417	72.030	$3.3067\ 10^{-4}$	0.0110
DF-GARCH (DCC)(no lags)	0.0000	0.1963	0.1204	72.217	$3.1075\ 10^{-4}$	**0.0097**
DF-GARCH (DCC)(2 lags)	0.0000	0.1957	0.1160	**72.218**	$\mathbf{3.1046\ 10^{-4}}$	**0.0097**

Table 10: Covolatilities. Average results for MZ regressions, percentage of times in which the predicted sign is the same as the real sign, average RMSEs, and average H_3.

9 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, which we call Dynamic Factor GARCH. 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 disentagle 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 behaviour. 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. Results of a Montecarlo experiment on different simulated panels show the goodness of our estimation method.
We evaluate the predictive performance of the DF-GARCH 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. However, the model 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 conditional heteroskedasticity is presently in progress by means of simulations, where different degrees of conditional heteroskedasticity 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 conditional heteroskedasticity. 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 use is 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.

C. Alexander. Orthogonal GARCH. In C. Alexander, editor, *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*, 25:52–60, 2005.

M. Barigozzi and M. Capasso. A multivariate perspective for modelling and forecasting inflation's conditional mean and variance. LEM Working Papers 2006/21, Scuola Superiore Sant'Anna, 2007.

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):79–109, 2006.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 127 (1):169–194, 2006.

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*, 127(2):341–371, 2006.

G. C. Chow. *Econometrics*. New York: Mc Graw-Hill, 1983.

T. E. Clark and K. D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 127(1):291–311, 2007.

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 R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63, 1995.

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. New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446, 2002a.

R. F. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–350, 2002b.

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, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840, 2005.

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 Economc Policy Research, 2006.

J. Geweke. The dynamic factor analysis of economic time series. In D. Aigner and A. Golberger, editors, *Latent Variables in Socio-Economic Models*. North-Holland, Amsterdam, 1977.

D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. In M. Gertler and K.Rogoff, editors, *NBER Macroeconomic Annual*. MIT Press, 2004.

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. Mills, editors, *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. Wadhwani. 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. S. et al., editor, *New Methods in Business Cycle Research*. Federal Reserve Bank of Minneapolis, Minneapolis, 1977.

E. Sentana. The relation between conditionally heteroskedastic factor models and factor GARCH models. *The Econometrics Journal*, 1:1–9, 1998.

J. H. Stock and M. W. Watson. Diffusion indexes. NBER Working Papers 6702, 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.

Appendix - A modified Kalman filter estimator

We explain here in detail the estimation of the state-space model

$$\begin{array}{ll} x_t & = \hat{\Lambda} F_t + \xi_t \qquad \text{measurement equation}\,, \\ F_t & = \hat{A} F_{t-1} + \hat{B} u_t \quad \text{transition equation}\,, \end{array}$$

where

$$\begin{array}{ll} \xi_{t|t-1} & \sim \mathrm{N}\,(0, \hat{R}_t) \quad \hat{R}_t \text{ diagonal}\,, \\ u_{t|t-1} & \sim \mathrm{N}\,(0, Q_t)\,, \\ Q_t & = \hat{C}_0' \hat{C}_0 + \hat{C}_1' u_{t-1} u_{t-1}' \hat{C}_1 + \hat{C}_2' Q_{t-1} \hat{C}_2\,. \end{array}$$

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:

$$\begin{cases} F_{1|1} & = \hat{F}_1 \\ P_{1|1} & \text{sufficiently large} \\ u_{1|1} & = \hat{u}_1 \\ Q_{1|1} & = \hat{Q}_1 \\ (u_1 u_1')_{|1} & = u_{1|1} u_{1|1}' + Q_{1|1}\,, \end{cases}$$

where the variables with the hat have been obtained during the estimation step presented in section 3, $\hat{Q}_1$ 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

$$F_{t|t-1} = \hat{A} F_{t-1|t-1}\,,$$

and its conditional covariance matrix

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

where

$$\begin{cases} (u_t u_t')_{|t-1} & = Q_{t|t-1} \\ \\ Q_{t|t-1} & = \hat{C}_0' \hat{C}_0 + \hat{C}_1' (u_{t-1} u_{t-1}')_{|t-1} \hat{C}_1 + \hat{C}_2' Q_{t-1|t-1} \hat{C}_2\,. \end{cases} \tag{21}$$

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 $\left(u_{t-1} u_{t-1}'\right)$, which in turn has been obtained by the Kalman update, as we see in the next step.
The prediction error is given by

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

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 (7), we get

$$u_{t|t} = S^{-1/2} M' \left(\mathrm{I} - \hat{A} L \right) F_{t|t} \,, \tag{22}$$

and then

$$(u_t u_t')_{|t} = u_{t|t} \, u_{t|t}' \,. \tag{23}$$

Equation (23), when put in the context of the following prediction step (21), 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 (22), 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

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

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

where $L_t = A \left(\mathrm{I} - K_t \Lambda \right)$, $r_T = 0$. 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 Θ_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

$$u_{t|T} = Q_{t|t-1} B' r_t \,,$$

$$Q_{t|T} = Q_{t|t-1} - Q_{t|t-1} B' \Theta_t B Q_{t|t-1} \,.$$