An EM Algorithm for Conditionally Heteroskedastic Factor Models [1]

Antonis Demos

(Athens University of Economics and Business)

Enrique Sentana

(CEMFI)

December 1991

Current Version: May 1997

[1] This is a condensed version of Demos and Sentana (1996), which in turn is an extensively revised version of Demos and Sentana (1992). We would like to thank Gabriele Fiorentini, Clive Granger, Neil Shephard, George Tauchen, and seminar participants at CEMFI, LSE, the ZEW Financial Markets Econometrics Conference (Manheim, 1992), the Royal Economic Society Conference (London, 1992) and the European Meeting of the Econometric Society (Brussels, 1992) for very useful comments and suggestions. An associate editor and two anonymous referees have also helped us greatly improve the paper. This research was initiated when both authors were based at the LSE Financial Markets Group, whose financial support as part of the ESRC project "The Efficiency and Regulation of Financial Markets" is gratefully acknowledged. Correspondence should be addressed to E. Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (e-mail: sentana@cemfi.es).

Abstract

This paper discusses the application of the EM algorithm to factor models with dynamic heteroskedasticity in the common factors. It demonstrates that the EM algorithm reduces the computational burden so much that researchers can estimate such models with a large number of series. Two empirical applications with 11 and 266 stock returns are presented, which confirm that the EM algorithm yields significant speed gains, and that it makes unnecessary the computation of good initial values. However, near the optimum it slows down significantly. Then, the best practical strategy is to switch to a first derivative-based method.

Keywords: Maximum Likelihood, Kalman Filter, Volatility, Asset Pricing, Stock Returns.

1 Introduction

One of the most popular approaches to multivariate dynamic heteroskedasticity employs the same idea as traditional factor analysis to obtain a parsimonious representation of conditional second moments. The factor GARCH model of Engle (1987) and the latent factor GARCH model introduced by Diebold and Nerlove (1989) and extended by King, Sentana and Wadhwani (1994) are the best known examples. Such models are particularly appealing in finance, as the concept of factors plays a fundamental role in asset pricing (e.g. the Arbitrage Pricing Theory of Ross 1976). However, whereas traditional applications have often used a very large collection of assets, those that take into account variation in conditional moments have considered a much smaller number (see e.g. Engle, Ng and Rothschild 1990, King et al. 1994, Ng, Engle and Rothschild 1992, or Sentana 1995). This is due to the fact that the estimation of such models involves a very time consuming procedure, which is disproportionately more so as the number of series considered increases. For standard factor models, however, a fast and reliable algorithm is available based on the EM procedure of Dempster, Laird and Rubin (1977) (see Rubin and Thayer 1982), which has been successfully employed to handle a very large dataset by Lehmann and Modest (1988).

The purpose of this paper is to study the application of the EM algorithm to factor models in which the factors are subject to dynamic heteroskedasticity-type effects. We also include risk premium components and weakly exogenous explanatory variables in the specification of the conditional mean. The paper is organized as follows. In section 2 we introduce the model, and include a general discussion on the use of the EM algorithm in this context. Then, in section 3 we adapt the algorithm to the two most widely used conditional variance parameterizations. Illustrative applications to two different data sets are presented in section 4. Our conclusions can be found in section 5.

2 Conditionally Heteroskedastic Factor Models

Consider the following multivariate model:

$$\mathbf{x}_t = \mathbf{B}\mathbf{z}_t + \mathbf{C}\mathbf{\Lambda}_t\boldsymbol{\tau} + \mathbf{C}\mathbf{f}_t + \mathbf{w}_t \tag{1}$$

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{w}_t \end{pmatrix} |\mathbf{X}_{t-1} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{pmatrix}\right] \tag{2}$$

where $\mathbf{x}_t$ is a $N \times 1$ vector of observable variables, $\mathbf{z}_t$ is a $m \times 1$ vector of weakly exogenous or lagged dependent explanatory variables, $\mathbf{B}$ is the $N \times m$ matrix of regression coefficients, $\mathbf{f}_t$ is a $k \times 1$ vector of unobserved common factors, $\mathbf{C}$ is the $N \times k$ matrix of factor loadings, with $N \geq k$ and rank $(\mathbf{B}, \mathbf{C}) = m + k$, $\mathbf{w}_t$ is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to $\mathbf{f}_t$, $\mathbf{\Gamma}$ is a $N \times N$ positive semidefinite matrix of constant idiosyncratic variances, $\boldsymbol{\tau}$ is a $k \times 1$ vector of price of risk coefficients, and $\mathbf{\Lambda}_t$ is a $k \times k$ diagonal positive definite matrix of time-varying factor variances, which generally involve some extra parameters, $\boldsymbol{\psi}$. Note that in line with the standard solution in the applied literature, we assume that the $\lambda'_{jt}s$ are measurable functions of the econometrician's information set, $\mathbf{X}_{t-1} = \{\mathbf{z}_t, \mathbf{x}_{t-1}, \mathbf{z}_{t-1}, \mathbf{x}_{t-2}, \ldots\}$ (cf. Harvey, Ruiz and Sentana 1992).

Our assumptions imply that the distribution of $\mathbf{x}_t$ conditional on $\mathbf{X}_{t-1}$ is $N(\mathbf{B}\mathbf{z}_t + \mathbf{C}\mathbf{\Lambda}_t\boldsymbol{\tau}, \mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma})$. For this reason, we refer to (1-2) as a conditionally heteroskedastic factor model. The parameters of interest $\boldsymbol{\phi} = \{vec'(\mathbf{B}),$ $vec'(\mathbf{C}),$ $vech'(\mathbf{\Gamma})$ or $vecd'(\mathbf{\Gamma}),$ $\boldsymbol{\tau}',$ $\boldsymbol{\psi}'\}'$ are usually estimated jointly from the log-likelihood function of the observed variables, $g(\mathbf{x}_t)$. Since this involves a very time consuming procedure, applications have been limited to N relatively small. But, if

a) the conditional variances of the common factors do not depend on $\mathbf{B}, \mathbf{C}, \mathbf{\Gamma}$.

b) the parameters $(\boldsymbol{\tau}, \boldsymbol{\psi})$ and $(\mathbf{B}, \mathbf{C}, \mathbf{\Gamma})$ are variation free.

c) the conditional variance parameters are specific to each factor and

d) $(\tau_1, \boldsymbol{\psi}_1), \ldots, (\tau_k, \boldsymbol{\psi}_k)$ are variation free

fully efficient estimates of $\boldsymbol{\phi}$ could be easily computed if the factors were observed. Define the factor representing portfolios, $\mathbf{r}_{ft}$, as $\mathbf{\Lambda}_t \boldsymbol{\tau} + \mathbf{f}_t$, so that $\mathbf{x}_t = \mathbf{B}\mathbf{z}_t + \mathbf{C}\mathbf{r}_{ft} + \mathbf{w}_t$. The joint log-likelihood function of $\mathbf{x}_t, \mathbf{r}_{ft}$ (conditional on $\mathbf{X}_{t-1}$ and $\mathbf{F}_{t-1} = \{\mathbf{f}_{t-1}, \mathbf{f}_{t-2}, \ldots\}$), $g(\mathbf{x}_t, \mathbf{r}_{ft})$, can be factorized as $g(\mathbf{r}_{ft}) + g(\mathbf{x}_t|\mathbf{r}_{ft})$. The marginal component has mean $\mathbf{\Lambda}_t \boldsymbol{\tau}$ and diagonal covariance matrix $\mathbf{\Lambda}_t$, whereas the conditional has mean $\mathbf{B}\mathbf{z}_t + \mathbf{C}\mathbf{r}_{ft}$, and covariance matrix $\mathbf{\Gamma}$. Given a) and b), we would have performed a sequential cut on $g(\mathbf{x}_t, \mathbf{r}_{ft})$ which would make $\mathbf{r}_{ft}$ weakly exogenous for $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{\Gamma}$ (see Engle, Hendry and Richard 1983). Hence, unrestricted MLE's of these parameters could be obtained from N univariate OLS regressions of each x_{it} on $\mathbf{z}_t$ and $\mathbf{r}_{ft}$. Similarly, MLE's of $(\tau_j, \boldsymbol{\psi}_j)$ would be obtained from k univariate dynamic heteroskedasticity in mean models for r_{fjt}.

Unfortunately, the $\mathbf{f}_t's$ are generally unobserved. Nevertheless, the EM algorithm can be used to obtain values for $\boldsymbol{\phi}$ as close to the optimum as desired (see Ruud 1991). At each iteration, the EM algorithm maximizes the expected value of $g(\mathbf{x}_t|\mathbf{r}_{ft}) + g(\mathbf{r}_{ft})$ conditional on $\mathbf{X}_T$ and the current parameter estimates, $\boldsymbol{\phi}^{(n)}$. The rationale stems from the fact that $g(\mathbf{x}_t, \mathbf{r}_{ft})$ can also be factorized as $g(\mathbf{x}_t) + g(\mathbf{r}_{ft}|\mathbf{x}_t)$. Since the expected value of the latter, conditional on $\mathbf{X}_T$ and $\boldsymbol{\phi}^{(n)}$, reaches a maximum at $\boldsymbol{\phi} = \boldsymbol{\phi}^{(n)}$, *any increase* in the expected value of $g(\mathbf{x}_t, \mathbf{r}_{ft})$ must represent an increase in

$g(\mathbf{x}_t)$. This is the generalized EM principle.

Let $\mathbf{f}_{t|T} = E(\mathbf{f}_t|\mathbf{X}_T)$, $\mathbf{\Lambda}_{t|T} = V(\mathbf{f}_t|\mathbf{X}_T)$ and $\mathbf{r}_{ft|T} = \mathbf{C}\mathbf{\Lambda}_t\boldsymbol{\tau} + \mathbf{f}_{t|T}$, which can be evaluated via the Kalman filter. The EM objective function is

$$T \log|\mathbf{\Gamma}| + \sum_{t=1}^{T} tr\left\{\mathbf{\Gamma}^{-1}\left[(\mathbf{x}_t - \mathbf{B}\mathbf{z}_t - \mathbf{C}\mathbf{r}^{(n)}_{ft|T})(\mathbf{x}_t - \mathbf{B}\mathbf{z}_t - \mathbf{C}\mathbf{r}^{(n)}_{ft|T})' + \mathbf{C}\mathbf{\Lambda}^{(n)}_{t|T}\mathbf{C}'\right]\right\} \tag{3}$$

$$+\sum_{j=1}^{k}\left\{\sum_{t=1}^{T}\left[\log|\lambda_{jt}| + [(r^{(n)}_{fjt|T} - \tau_j\lambda_{jt})^2 + \lambda^{(n)}_{jt|T}]/\lambda_{jt}\right]\right\} \tag{4}$$

where $^{(n)}$ means evaluated at $\boldsymbol{\phi}^{(n)}$. If a) and b) hold, we get that:

$$\begin{pmatrix} \mathbf{B}^{(n+1)\prime} \\ \mathbf{C}^{(n+1)\prime} \end{pmatrix} = \left[\sum_{t=1}^{T}\begin{pmatrix} \mathbf{z}_t\mathbf{z}_t' & \mathbf{z}_t\mathbf{r}^{(n)\prime}_{ft|T} \\ \mathbf{r}^{(n)}_{ft|T}\mathbf{z}_t' & \mathbf{r}^{(n)}_{ft|T}\mathbf{r}^{(n)\prime}_{ft|T} + \mathbf{\Lambda}^{(n)}_{t|T} \end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix} \mathbf{z}_t \\ \mathbf{r}^{(n)}_{ft|T} \end{pmatrix}\mathbf{x}_t'\right] \tag{5}$$

$$\mathbf{\Gamma}^{(n+1)} = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_t\mathbf{x}_t' - \begin{pmatrix} \mathbf{B}^{(n+1)} & \mathbf{C}^{(n+1)} \end{pmatrix}\begin{pmatrix} \mathbf{z}_t \\ \mathbf{r}^{(n)}_{ft|T} \end{pmatrix}\mathbf{x}_t'\right] \tag{6}$$

Similarly, if c) and d) hold, $\tau_j^{(n+1)}, \boldsymbol{\psi}_j^{(n+1)}$can be obtained by numerically minimizing the expression in curly brackets in (4). Independently of the number of series considered, this is the only price paid for modelling the conditional variance of the factors. Moreover, the price is small because this is a rather fast procedure, and the Kalman filter is used only once per EM iteration. Furthermore, in practice we can do a few iterations over (5) and (6) alone, before minimizing (4).

3 Conditional Variance Parameterizations

3.1 Factor GARCH

Engle's (1987) k factor GARCH(p,q) model can be written as a particular case of (1-2), with the conditional variances of the factors given by:

$$\lambda_{jt} = \sum_{s=1}^{q} \alpha_{js} \dot{\varepsilon}_{jt-s}^2 + \sum_{r=1}^{p} \beta_{jr} \lambda_{jt-r}$$

where $\dot{\boldsymbol{\varepsilon}}_t = \mathbf{D}'(\mathbf{x}_t - \mathbf{C}\boldsymbol{\Lambda}_\mathbf{t}\boldsymbol{\tau} - \mathbf{B}\mathbf{z}_t) = \dot{\mathbf{x}}_t - \boldsymbol{\Lambda}_\mathbf{t}\boldsymbol{\tau} - \dot{\mathbf{B}}\mathbf{z}_t$ and $\mathbf{D}$ is a $N \times k$ matrix of full column rank satisfying $\mathbf{D}'\mathbf{C} = \mathbf{I}_k$ (see Sentana 1997).

If the conditional mean contains linear regression terms, $g(\mathbf{r}_{ft})$ will depend on $\dot{\mathbf{B}}$, so that condition b) no longer holds. But even if there are no regressors, the restriction $\mathbf{D}'\mathbf{C} = \mathbf{I}_k$ implies two things:

1) $\mathbf{r}_{ft}$ would not be weakly exogenous for $\mathbf{C}$ and $\boldsymbol{\Gamma}$ unless $\mathbf{D}$ is known.

2) even if $\mathbf{D}$ is known, there are linear restrictions on $\mathbf{C}$, so that (5) and (6) have to be replaced by SURE-type estimators.

In practice, both problems are less serious than it may seem. Since most empirical applications of this model have been carried out under the assumption that the matrix $\mathbf{D}$ is known, we shall maintain such an assumption and only consider what Lin (1992) calls restricted maximum likelihood estimators. Similarly, SURE-type estimators that impose the restriction $\mathbf{D}'\mathbf{C} = \mathbf{I}_k$ are particularly easy to obtain. Let $\bar{\mathbf{D}} = (\mathbf{D}, \ddot{\mathbf{D}})$ be a known $N \times N$ matrix of full rank, with $\ddot{\mathbf{D}}$ arbitrary, and let $\bar{\mathbf{x}}_t' = \mathbf{x}_t'\bar{\mathbf{D}} = \left(\mathbf{x}_t'\mathbf{D}, \mathbf{x}_t'\ddot{\mathbf{D}}\right) = (\dot{\mathbf{x}}_t', \ddot{\mathbf{x}}_t')$ denote the transformed observations. It is easy to see that the factor structure for $\mathbf{x}_t$ is preserved in $\bar{\mathbf{x}}_t$, and that $\mathbf{f}_{t|T}$ and $\boldsymbol{\Lambda}_{t|T}$ are not affected by the change of variables.

As before, $g(\bar{\mathbf{x}}_t, \mathbf{r}_{ft})$ can be factorized as $g(\mathbf{r}_{ft}) + g(\bar{\mathbf{x}}_t|\mathbf{r}_{ft})$. But the latter can be factorized in turn as $g(\ddot{\mathbf{x}}_t|\dot{\mathbf{x}}_t, \mathbf{r}_{ft}) + g(\dot{\mathbf{x}}_t|\mathbf{r}_{ft})$, so that the restriction $\mathbf{D}'\mathbf{C} = \mathbf{I}_k$ only affects the last term. Let $\bar{\mathbf{\Gamma}}^*_{22} = \bar{\mathbf{\Gamma}}_{22} - \bar{\mathbf{\Gamma}}_{21}\bar{\mathbf{\Gamma}}^{-1}_{11}\bar{\mathbf{\Gamma}}'_{21}$, $\ddot{\mathbf{E}}^* = \bar{\mathbf{\Gamma}}_{21}\bar{\mathbf{\Gamma}}^{-1}_{11}$, $\ddot{\mathbf{B}}^* = \ddot{\mathbf{B}} - \ddot{\mathbf{E}}^*\dot{\mathbf{B}}$ and $\ddot{\mathbf{C}}^* = \ddot{\mathbf{C}} - \ddot{\mathbf{E}}^*$ be the implicit one-to-one reparameterization, where $\bar{\mathbf{B}}'= \mathbf{B}'\bar{\mathbf{D}}$, $\bar{\mathbf{C}}'= \mathbf{C}'\bar{\mathbf{D}}$ and $\bar{\mathbf{\Gamma}} = \bar{\mathbf{D}}'\mathbf{\Gamma}\bar{\mathbf{D}}$. The arguments made in section 2 imply that:

$$\begin{pmatrix} \ddot{\mathbf{B}}^{*(n+1)\prime} \\ \ddot{\mathbf{E}}^{*(n+1)\prime} \\ \ddot{\mathbf{C}}^{*(n+1)\prime} \end{pmatrix} = \left[\sum_{t=1}^{T} \begin{pmatrix} \mathbf{z}_t\mathbf{z}'_t & \mathbf{z}_t\dot{\mathbf{x}}'_t & \mathbf{z}_t\mathbf{r}^{(n)\prime}_{ft|T} \\ \dot{\mathbf{x}}_t\mathbf{z}'_t & \dot{\mathbf{x}}_t\dot{\mathbf{x}}'_t & \dot{\mathbf{x}}_t\mathbf{r}^{(n)\prime}_{ft|T} \\ \mathbf{r}^{(n)}_{ft|T}\mathbf{z}'_t & \mathbf{r}^{(n)}_{ft|T}\dot{\mathbf{x}}'_t & \mathbf{r}^{(n)}_{ft|T}\mathbf{r}^{(n)\prime}_{ft|T} + \mathbf{\Lambda}^{(n)}_{t|T} \end{pmatrix}\right]^{-1} \left[\sum_{t=1}^{T} \begin{pmatrix} \mathbf{z}_t \\ \dot{\mathbf{x}}_t \\ \mathbf{r}^{(n)}_{ft|T} \end{pmatrix} \ddot{\mathbf{x}}'_t\right]$$

$$\bar{\mathbf{\Gamma}}^{*(n+1)}_{22} = \frac{1}{T}\sum_{t=1}^{T}\left[\ddot{\mathbf{x}}_t\ddot{\mathbf{x}}'_t - \begin{pmatrix} \ddot{\mathbf{B}}^{*(n+1)} & \ddot{\mathbf{E}}^{*(n+1)} & \ddot{\mathbf{C}}^{*(n+1)} \end{pmatrix} \begin{pmatrix} \mathbf{z}_t \\ \dot{\mathbf{x}}_t \\ \mathbf{r}^{(n)}_{ft|T} \end{pmatrix} \ddot{\mathbf{x}}'_t\right]$$

Similarly, since we can concentrate $\bar{\mathbf{\Gamma}}_{11}$ as

$$\bar{\mathbf{\Gamma}}_{11}(\dot{\mathbf{B}}) = \frac{1}{T}\sum_{t=1}^{T}\left[(\dot{\mathbf{x}}_t - \mathbf{r}^{(n)}_{ft|T} - \dot{\mathbf{B}}\mathbf{z}_t)(\dot{\mathbf{x}}_t - \mathbf{r}^{(n)}_{ft|T} - \dot{\mathbf{B}}\mathbf{z}_t)' + \mathbf{\Lambda}^{(n)}_{t|T}\right]$$

the k-variate function that we have to minimize numerically with respect to $\boldsymbol{\psi}, \boldsymbol{\tau}$ and $\dot{\mathbf{B}}$ at each EM iteration is

$$\sum_{j=1}^{k}\left\{\sum_{t=1}^{T}\left[\log|\lambda_{jt}| + [(r^{(n)}_{jt|T} - \lambda_{jt}\tau_j)^2 + \lambda^{(n)}_{jt|T}]/\lambda_{jt}\right]\right\} + T\log\left|\bar{\mathbf{\Gamma}}_{11}(\dot{\mathbf{B}})\right|$$

Finally, ML estimates of $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{\Gamma}$ can be obtained by simply inverting the transformations. Consistent initial values for $\boldsymbol{\psi}, \boldsymbol{\tau}, \dot{\mathbf{B}}$ and $\bar{\mathbf{\Gamma}}_{11}$ can be obtained by numerically maximizing $g(\dot{\mathbf{x}}_t)$.

3.2 Latent Factor Models

These are special cases of (1-2) in which the conditional variances are parameterized as univariate ARCH models, but taking into account that the factors are unobserved. In particular, for the GQARCH(1,1) formulation of

Sentana (1995),

$$\lambda_{jt} = \psi_{j0} + \psi_{j1} f_{jt-1|t-1} + \alpha_{j1}(f^2_{jt-1|t-1} + \lambda_{jt-1|t-1}) + \beta_{j1}\lambda_{jt-1} \tag{7}$$

with ψ_{j0} restricted to avoid scale indeterminacy.

There are two complications associated with this conditional variance parameterization from the point of view of applying the EM algorithm:

1) λ_{jt} depends on $\mathbf{B}, \mathbf{C}$ and $\mathbf{\Gamma}$ via $f_{jt-1|t-1}$ and $\lambda_{jt-1|t-1}$, so that $g(\mathbf{r}_{ft})$ is an indirect function of these parameters.

2) even if $\mathbf{B}, \mathbf{C}$ and $\mathbf{\Gamma}$ were known, the numerical minimization of (4) with respect to $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$, would involve the use of the Kalman filter to produce estimates of $f_{jt-1|t-1}$ and $\lambda_{jt-1|t-1}$ once per parameter per quasi-Newton iteration.

As a result, the expressions derived in section 2 have to be interpreted with care. For instance, it is still true that (5) and (6) minimize (3), but they may increase or decrease (4). Fortunately, the generalized EM principle works provided that at each iteration we increase the expected value of $g(\mathbf{x}_t, \mathbf{r}_{ft})$, even though we do not maximize it. On this basis, our proposal is to minimize, at each EM iteration, an approximation to (3)+(4) such that the new optimum is easy to compute. In particular, we propose to replace the conditional variance (7) by:

$$\lambda^{(n)}_{jt} = \psi_{j0} + \psi_{j1} f^{(n)}_{jt-1|t-1} + \alpha_{j1}[(f^{(n)}_{jt-1|t-1})^2 + \lambda^{(n)}_{jt-1|t-1}] + \beta_{j1}\lambda^{(n)}_{jt-1} \tag{8}$$

so that (5) and (6) are still extrema of the approximating function, while $\boldsymbol{\tau}^{(n+1)}$ and $\boldsymbol{\psi}^{(n+1)}$ can be obtained via k univariate optimizations. Such an approximation is exact at $\boldsymbol{\phi} = \boldsymbol{\phi}^{(n)}$, and improves as we approach the optimum, so we would expect that the minimum of the approximating function does not increase the sum of (3) and (4). Nevertheless, we recommend monitoring $g(\mathbf{x}_t)$ after each iteration, and switching to its direct maximization if the variation were negative.

4 Empirical Applications

We investigate the performance of the procedures discussed in the previous sections in two illustrative applications. For comparison purposes, we also employ the NAG library version of the BFGS quasi-Newton (QN) algorithm. Table 1 contains the relevant timing information.

4.1 US Size-Ranked Portfolios

As part of their analysis, Ng et al. (1992) estimate a 1-factor GARCH(1,1) model for excess returns on the stock market index (VW) and 10 decile portfolios under the assumption that the matrix $\mathbf{D}$ is $(1, 0, \ldots, 0)'$. Given the large number of parameters involved, they use a consistent two-step estimation procedure. First, a univariate model is fitted to the VW returns. Then, the estimated conditional variance of the VW return is taken as data in the estimation of 10 univariate models for each of the portfolios. As the authors acknowledge, and the simulation results in Lin (1992) confirm, this two-step procedure ignores cross-asset correlations and parameter restrictions, and thus it sacrifices efficiency.

Here, we *jointly* estimate by restricted maximum likelihood a model like theirs over the same sample period (1964:8 to 1985:11). The number of parameters (including mean constants) is $11 + 10 + 66 + 2 = 89$. In order to get sensible initial values, we estimated a univariate GARCH(1,1) model for the VW return series, and 10 univariate market model regressions. This procedure is biased in favour of QN methods, as the parameter values obtained in this way are efficient estimators of a special case of the factor GARCH model in which the distribution of $\ddot{\mathbf{x}}_t$ given $\dot{x}_t$ has a linear mean and a constant covariance matrix. Given the structure of $\mathbf{D}$, it is not necessary to transform the observations.

The QN method makes steady progress for a while, but then the log-likelihood becomes very flat, and fails to converge. This is probably due to the fact that $\mathbf{D}$ only involves the VW index, which is almost a linear combination of the size-ranked portfolios ($\mathrm{R}^2 = 0.976$). The EM algorithm is initially much faster. Each iteration takes 0.18 seconds, a third being spent computing the Kalman filter, and the remaining maximizing the expected log-likelihood function. Not surprisingly, though, the algorithm slows down near the optimum.

We also considered a combined procedure that switches to the QN method when the EM iterations make little progress ($<10^{-3}$) (see Ruud 1991). This time, the QN method converged to -4604.654 after 134 iterations. The time of the combined procedure is 11 minutes (1'12" EM + 10' 50" QN). The final parameter values are rather different from the initial ones. The main difference is that the ARCH parameter, α, is much lower than in the univariate model (0.0033 versus 0.0717), and correspondingly, that the factor loading coefficients, $\ddot{\mathbf{C}}$, are much larger. For instance, the coefficients for the first and second decile portfolios are 8.31 and 5.30 respectively, while the OLS regression estimates are 1.34 and 1.28.

In order to assess the influence of initial values, we re-started both procedures from plausible, but arbitrary values. In particular, we set the mean constants to 0, the factor loadings to 1, the idiosyncratic standard deviations to 3, the covariances to 0, and finally, the ARCH and GARCH parameters to 0.1 and 0.6 respectively. Not surprisingly, the starting log-likelihood is far away from the optimum (-7451.299). Nevertheless, the first two EM iterations take the function back to -4664.056 in 0.4 seconds, whereas the QN method needed 48 iterations (3'52.8") to reach a similar value. This demonstrates that it is not worth obtaining good starting values, especially if we take into account the researcher's own time.

4.2 UK Individual Stocks

We also estimate a latent factor model with constant means and a GQARCH(1,1) factor for monthly capital gains on 266 individual stocks included in the Financial Times Actuarial sectorial indices over the period 1965:2 - 1991:7. We also include a second common factor with constant variance in order to allow for a richer covariance structure. The results in Sentana (1992) guarantee that the factor loading matrix is uniquely identified. The model, therefore, contains $4 \times 266 + 3 = 1067$ parameters. Since we could not think of any easy way to compute initial estimates for this model, we set the mean parameters to 0, the factor loadings to 1, the idiosyncratic standard deviations to 0.8, the ARCH and GARCH parameters to 0.1 and 0.6, and the dynamic asymmetry parameter to 0.

The QN method stopped after five and a half days due to a power cut without achieving convergence. The intrinsic problem of the QN method is that it must use the Kalman filter at least 1067 times per iteration. On the other hand, the EM algorithm uses the filter only once per iteration. Not surprisingly, the EM algorithm is extremely faster. The log-likelihood function increases by almost 8984 points in 0.55 seconds by simply computing (5) and (6) once. This CPU time is dominated by the computation of the Kalman filter estimates (0.44 seconds, versus 0.11 for the M step). The first maximization of (4) with respect to the conditional variance parameters improves the function a further 37 points in 1.98 seconds. Given that most of the increments are due to changes in the static factor parameters, we would recommend doing some partial EM iterations over (5) and (6) alone. By the time the first QN iteration is completed, the EM algorithm has moved the log-likelihood function to -103021.478. But the function is increasing by less than 10^{-3} per iteration. If we then switch to the QN method, the log-likelihood function converges to -103021.254 in 47 iterations (16h 50' 26").

5 Conclusions

We discuss the application of the EM algorithm for maximum likelihood estimation of large factor models in which the common factors are subject to ARCH-type effects. We also include risk premium components and weakly exogenous explanatory variables in the specification of the conditional mean. We analyze the M step for those cases in which the common factors would be weakly exogenous for the static factor model parameters were they observed. We show that, irrespectively of the number of series under consideration, the only difference that modelling the conditional variances of the factors makes is that at each EM iteration we have to estimate k extra univariate dynamic heteroskedasticity in mean models. We also discuss modifications for the most widely used multivariate conditional variance parameterizations: the factor GARCH model of Engle (1987), and the latent factor model of Diebold and Nerlove (1989).

We consider two illustrative applications involving: (i) excess returns on the US value weighted stock market index and 10 size-ranked portfolios (89 parameters); and (ii) capital gains on 266 individual UK stocks (1067 parameters). Our proposed procedure yields significant speed gains in both applications. The problem with first derivative methods is that they have to compute the log-likelihood function at least once per parameter per iteration. In contrast, the EM algorithm only computes it once per iteration. Furthermore, the applications demonstrate that obtaining good starting parameter values is by no means worth pursuing in time efficiency terms. After just a few very fast iterations, the EM algorithm takes the parameters closer to their maximum likelihood estimates than a QN method after many very slow iterations. However, the EM algorithm slows down substantially when it gets very close to the optimum. For that reason, we recommend a combined procedure that switches to the QN method when the EM iterations make little progress.

References

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Demos, A., and Sentana, E. (1992), "An EM-based Algorithm for Conditionally Heteroskedastic Factor Models", LSE Financial Markets Group Discussion Paper 140.

——— (1996), "An EM Algorithm for Conditionally Heteroskedastic Factor Models", CEMFI Working Paper 9615.

Diebold, F., and Nerlove, M. (1989), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model", *Journal of Applied Econometrics* 4, 1-21.

Engle, R. (1987), "Multivariate ARCH with Factor Structures - Cointegration in Variance", mimeo, University of California, San Diego.

Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983), "Exogeneity", *Econometrica* 51, 277-304.

Engle, R., Ng, V. K., and Rothschild, M. (1990), "Asset Pricing with a Factor-ARCH Structure: Empirical Estimates for Treasury Bills", *Journal of Econometrics* 45, 213-237.

Harvey, A., Ruiz, E., and Sentana, E. (1992), "Unobservable Component Time Series Models with ARCH Disturbances", *Journal of Econometrics* 52, 129-158.

King, M., Sentana, E., and Wadhwani, S. (1994), "Volatility and Links between National Stock Markets", *Econometrica* 62, 901-933.

Lehmann, B., and Modest, D. (1988), "The Empirical Foundations of the Arbitrage Pricing Theory", *Journal of Financial Economics* 21, 213-254.

Lin, W.L. (1992), "Alternative Estimators for Factor GARCH Models -

A Monte Carlo Comparison", *Journal of Applied Econometrics* 7, 259-279.

Ng, V. M., Engle, R. F., and Rothschild, M. (1992), "A Multi Dynamic Factor Model for Stock Returns", *Journal of Econometrics* 52, 245-266.

Ross, S.S. (1976), "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory* 13, 341-360.

Rubin, D., and Thayer, D. (1982), "EM Algorithms for ML Factor Analysis", *Psychometrika* 47, 69-76.

Ruud, P. (1991), "Extensions of Estimation Methods Using the EM Algorithm", *Journal of Econometrics* 49, 305-341.

Sentana, E. (1992), "Identification of Multivariate Conditionally Heteroskedastic Factor Models" LSE Financial Markets Group Discussion Paper 139.

——— (1995), "Quadratic ARCH Models", *Review of Economic Studies* 62, 639-661.

——— (1997), "The relation between Conditionally Heteroskedastic Factor Models and Factor GARCH Models", mimeo, CEMFI.

Table 1. Timing Information for EM and Quasi-Newton Algorithms

	EM		QN	
Iteration	Time	log-likelihood	Time	log-likelihood
US Size-Ranked Portfolios				
0	0	-4663.694	0	-4663.694
1	0.16”	-4654.370	4.83”	-4660.155
2	0.48”	-4651.025	9.67”	-4659.232
5	0.99”	-4640.709	24.06”	-4646.033
10	2.03”	-4630.191	47.84”	-4636.188
20	3.46”	-4620.541	1’35.62”	-4620.430
50	8.02”	-4612.068	3’59.75”	-4607.398
100	15.16”	-4608.967	8’0.16”	-4605.251
200	29.00”	-4607.312	16’14.87”	-4604.856
300	42.13”	-4606.700	25’17.92”	-4604.854
Last	27’15.99”	-4604.981	4h10’32.12”	-4604.846
UK Individual Stocks				
0	0	-114683.185	0	-114683.185
1	2.53”	-105662.441	1h25’02.93”	-114674.517
2	3.02”	-105296.583	1h49’26.09”	-114633.471
5	4.56”	-105451.319	2h01’06.10”	-113308.344
10	7.19”	-103154.490	3h58’58.19”	-110671.908
20	13.03”	-103109.745	7h56’48.17”	-116867.573
50	28.17”	-103084.101	19h52’49.92”	-103228.381
100	54.48”	-103072.705	39h40’37.93”	-103070.782
200	1’47.16”	-103061.298	79h29’10.35”	-103024.328
300	2’39.78”	-103055.266	119h7’4.82”	-103021.709
Last	25’43.90”	-103021.287	127h29’31.33”	-103021.675

Note: Computations carried out on a PC with a 155MHz Pentium Processor