The Likehood Function of Conditionally Heteroskedastic Factor Models

Enrique SENTANA *

ABSTRACT – We derive the likelihood function and score of factor models with dynamic heteroskedasticity, and the KUHN-TUCKER conditions defining the inequality restricted maximum likelihood estimators that guarantee a positive definite covariance matrix. We present three methods to compute the likelihood function, its gradient and factor scores, which are numerically efficient and reliable, and statistically sound. We show that the incidence of zero idiosyncratic variance estimates (HEYWOOD cases) depends on the correlation of a variable with the rest.

La fonction de vraisemblance dans les modèles de facteurs conditionnellement hétéroscédastiques

RÉSUMÉ. – Nous calculons la fonction de vraisemblance pour les modèles factoriels avec hétéroscédasticité dynamique dans les facteurs, et les conditions de KUHN-TUCKER que définissent les estimateurs sous contraintes à l'inégalité qui garantissent une matrice de covariance définie positive. Nous présentons trois méthodes pour calculer la fonction de vraisemblance, son gradient, et les estimateurs des facteurs qui sont numériquement efficaces et fiables, et statistiquement justifiées. Nous démontrons que l'occurrence d'estimations avec variance spécifique nulle (cas de HEYWOOD) dépend de la corrélation d'une variable avec le reste.

---

* E. SENTANA: CEMFI

This is an extensively revised version of SENTANA [1994], which concentrated on the evaluation of the likelihood function. Financial support for that version from the Spanish Ministry of Education and Science and the LSE Financial Markets Group as part of the ESRC project *"The Efficiency and Regulation of Financial Markets"* is gratefully acknowledged. The author is also grateful to M. ARELLANO, A. DEMOS, J.J. DOLADO, G. FIORENTINI and N. SHEPHARD, as well as seminar participants at CEMFI for their comments. The suggestions of an anonymous referee have also substantially improved the paper. Of course, the usual caveat applies.

1 Introduction

Time variation in the volatility of financial markets is nowadays well documented and widely recognized as one of the main empirical characteristics of asset prices. As a result, in recent years many statistical models for time-varying variances have been developed. By and large, though, most theoretical and applied work in this area has been on univariate series, as the application of these models in a multivariate context has been hampered by the large number of parameters involved.

Hence, it is perhaps not surprising that one of the most popular approaches to multivariate dynamic heteroskedasticity uses the same idea as traditional factor analysis to obtain a parsimonious representation of conditional second moments. That is, it is assumed that each of several observed variables is a linear combination of a smaller number of conditionally orthogonal factors plus an idiosyncratic noise term, but allowing for dynamic heteroskedasticity-type effects in the underlying factors. The factor GARCH model discussed by BOLLERSLEV and ENGLE [1993] and the latent factor GARCH model of DIEBOLD and NERLOVE [1989] are the best known examples (see, SENTANA, [1998] for details). Importantly, if both common and idiosyncratic factors are covariance stationary, these models often imply an unconditional factor structure, which makes them compatible with traditional factor analysis (see *e.g.* LAWLEY and MAXWELL [1971]). They are also in line with the long tradition of factor, or multi-index models in finance. In particular, they can be closely integrated with dynamic versions of ROSS' [1976] APT theory (see, KING, SENTANA and WADHWANI [1994]).

A non-trivial advantage of these models is that they automatically guarantee a positive semi-definite covariance matrix for the observed series, once we ensure that the variances of common and specific factors are non-negative. The required parameter restrictions should be taken into account during estimation, since the solution of the likelihood equations often involves some negative estimates of the idiosyncratic variances. In those instances, known as HEYWOOD cases in the static factor analysis literature (see *e.g.* BARTHOLOMEW [1987]), the inequality constrained maximum likelihood estimates will be zero, unlike the corresponding gradient.

In this paper, we derive the likelihood function and score of a class of conditionally heteroskedastic factor models analysed in DEMOS and SENTANA [1998], as well as the KUHN-TUCKER first-order conditions which define the inequality restricted maximum likelihood estimators. We also present three methods to compute the likelihood function, its gradient, and the best filtered estimates of the factors, which are not only numerically efficient and reliable, but also intuitive from a statistical point of view.

The paper is organized as follows. In section 2, we formally introduce the model and discuss some of its properties. We then cast it in time-series and cross-sectional state space form, and explain how to obtain factor estimates by means of the KALMAN filter. The likelihood function, score and first-order conditions are presented in section 3, where we discuss in detail the incidence of HEYWOOD cases, *i.e.* zero estimates of some idiosyncratic variances. Finally,

in section 4, we derive three alternative computational methods based on the decomposition of joint densities as the product of marginal and conditional components. An alternative algebraic derivation based on the WOODBURY formula is developed in the appendix.

2 Conditionally Heteroskedastic Factor Models

2.1 Definition and Basic Properties

Consider the following multivariate model:

$$\mathbf{x}_t = \mathbf{C}\mathbf{f}_t + \mathbf{w}_t \tag{1}$$

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

where $\mathbf{x}_t$ is a $N \times 1$ vector of observable random variables, $\mathbf{f}_t$ is a $k \times 1$ vector of unobserved common factors, $\mathbf{C}$ is the associated $N \times k$ matrix of constant factor loadings, with $N \geqslant k$ and rank $(\mathbf{C}) = k$, $\mathbf{w}_t$ is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to $\mathbf{f}_t$, Γ is a $N \times N$ diagonal positive semidefinite (p.s.d.) matrix of constant idiosyncratic variances, Λ_t is a $k \times k$ diagonal positive definite (p.d.) matrix of time-varying factor variances, which generally involve some extra parameters ψ, and I_{t-1} is an information set that contains the values of $\mathbf{x}_t$ and $\mathbf{f}_t$ up to, and including time $t-1$. These assumptions imply that the distribution of $\mathbf{x}_t$ conditional on I_{t-1} is $N(\mathbf{0}, \Sigma_t)$, where the conditional covariance matrix $\Sigma_t = \mathbf{C}\Lambda_t\mathbf{C}' + \Gamma$ has the usual factor structure. For this reason, we shall refer to the data-generation process specified by (1) and (2) as a conditionally heteroskedastic exact factor model.

Such a formulation nests several models widely used in the empirical literature, which typically assume that the unobserved factors follow dynamic heteroskedastic processes, but differ in the particular functional form of the conditional variances of the factors. In this respect, it is important to distinguish between $I_{t-1} = \{\mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \mathbf{x}_{t-2}, \mathbf{f}_{t-2}, \ldots\}$, and the econometrician's information set, $X_{t-1} = \{\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots\}$, which only includes lagged values of $\mathbf{x}_t$, where $I_{t-1} \equiv X_{t-1} \cup F_{t-1}$, with $F_{t-1} = \{\mathbf{f}_{t-1}, \mathbf{f}_{t-2}, \ldots\}$. If the conditional variances of the factors depend on their past values, as in the original latent factor model with ARCH effects introduced by DIEBOLD and NERLOVE [1989], the form of the distribution of $\mathbf{x}_t$ conditional on X_{t-1} alone is unknown, and one has to resort to approximate analytical solutions (see, HARVEY, RUIZ and SENTANA [1992]), Markov Chain Monte Carlo simulation methods (see, FIORENTINI, SENTANA and SHEPHARD [1998]) or indirect inference procedures (see, GOURIEROUX, MONFORT and RENAULT [1993] and CALZOLARI, FIORENTINI

and SENTANA [1999]). In line with the standard solution in the applied literature, though, we follow DEMOS and SENTANA [1998] in assuming that the diagonal elements of Λ_t are effectively measurable functions of the observable information, and thus, that the distribution of $\mathbf{x}_t$ conditional on X_{t-1} is also $N(\mathbf{0}, \Sigma_t)$. For clarity of exposition, we also assume that the conditional variances of the common factors do not depend on $\mathbf{C}$ or Γ, and that the parameters ψ and $(\mathbf{C}, \Gamma)$ are variation free.

A non-trivial advantage of these models is that they automatically guarantee a p.s.d. covariance matrix for $\mathbf{x}_t$ once we ensure that both $\Gamma$and the covariance matrix of the common factors are p.s.d. In this respect, it is clear that we need $\Gamma \geqslant 0$. Moreover, given that Γ is diagonal, it is straightforward to see that a necessary and sufficient condition for Σ_t to be p.d. when Γ is p.s.d., but not p.d., is that the matrix formed with the rows of $\mathbf{C}$ corresponding to the zero elements of Γ must be of full *row* rank.

But the most distinctive feature of factor models is that they provide a parsimonious specification of the dynamic and cross-sectional dependence of a vector of observable random variables. In this case, in particular, the factor structure, together with the constancy of Γ, implies that the time-varying component of Σ_t is of rank k rather than N (see, ENGLE, NG and ROTHSCHILD [1990]). This results in a significant reduction in the number of parameters, and allows the estimation of these models with a large number of series (see, DEMOS and SENTANA [1998]).

Finally, notice that if $\mathbf{f}_t$ is conditionally homoskedastic, the above model reduces to the static one used in statistical factor analysis (see *e.g.* LAWLEY and MAXWELL [1971]). At the same time, it is a special case of the general approximate conditional factor structure $\Sigma_t = \mathbf{C}_t \mathbf{C}_t' + \Gamma_t$ (*cf.* CHAMBERLAIN and ROTHSCHILD [1983]), where $\mathbf{C}_t$ and Γ_t are matrices of functions of the information set, and Γ_t is such that its eigenvalues remain bounded with probability one as N increases (as in band-diagonal matrices). In this framework, we can rewrite (1) – (2) as $\mathbf{x}_t = \mathbf{C}_t \mathbf{f}_t^* + \mathbf{w}_t$, with Γ_t diagonal and constant, the conditional covariance matrix of $\mathbf{f}_t^*$ equal to the identity matrix, and $\mathbf{C}_t = \mathbf{C}\Lambda_t^{1/2}$, so that the loadings of different variables on each conditionally homoskedastic factor change proportionately over time (see, ENGLE, NG. and ROTHSCHILD [1990]). Such a parsimonious specification of the time variation in $\mathbf{C}_t$ has been the only one adopted so far in applications. It also empirically plausible, in the sense that it is compatible with traditional factor analysis even if the common factors are conditionally heteroskedastic, because provided that they are covariance stationary, the unconditional covariance matrix of $\mathbf{x}_t$, Σ, can be written as:

$$\Sigma = \mathbf{C}\Lambda\mathbf{C}' + \Gamma$$

where $V(\mathbf{f}_t) = E(\Lambda_t) = \Lambda$. Notice that even if Γ_t is diagonal, the unconditional covariance of a process characterized by a general conditional factor representation may very well lack an unconditional factor structure for any $k < N$ (see, HANSEN and RICHARD [1987]). In fact, it is possible to show that an exact conditional factor structure is compatible with an unconditional one with the same number of factors if and only if $\mathbf{C}_t = \mathbf{C}\Lambda_t^{1/2}$ with Λ_t p.d. but not necessarily diagonal.

2.2 State-Space Representations and Factor Scores

The conditionally heteroskedastic factor model in (1) – (2) can be regarded as a bidimensional stochastic process (or random field) with indices $i = 1,\dots N$ and $t = 1,\dots,T$. Therefore, it is not surprising that it has both a time-series state-space representation, with $\mathbf{f}_t$ as the state, (1) as the measurement equation, and $\mathbf{f}_t = \mathbf{0}\cdot\mathbf{f}_{t-1} + \mathbf{f}_t$ as the transition equation, and a cross-sectional state-space representation. For fix t, the measurement equation is given by:

$$x_{it} = \mathbf{c}_i'\mathbf{f}_{it} + w_{it} \tag{3}$$

where w_{it} is distributed as $N(0,\gamma_i)$ independently across i, $\mathbf{c}_i' = (c_{i1},\dots,c_{ik})$ is the i^{th} row of $\mathbf{C}$, with $\mathbf{c} = (\mathbf{c}_1',\dots,\mathbf{c}_N')' = \text{vec}(\mathbf{C}')$, and γ_i is the i^{th} diagonal element of Γ, so that $\gamma = (\gamma_1,\dots,\gamma_N)' = \text{vec}d(\Gamma)$. Since the factors are the same for all x_{it}, the transition equation is simply $\mathbf{f}_{it} = \mathbf{f}_{i-1t}$, with initial condition $\mathbf{f}_{0t}\sim N(0,\Lambda_t)$. These equations correspond to a cross-sectional common trends model without innovations in the transition equation (see, HARVEY [1989]), and as such, provide a precise characterization of the cross-sectional dependence in $\mathbf{x}_t$. [1]

On the basis of this representation, we can apply the KALMAN filter cross-sectionally to obtain at each period t the so-called "*regression scores*", which are the best (in the conditional mean square error sense) estimates of the factors, $\mathbf{f}_{t|t} = E(\mathbf{f}_t|\mathbf{x}_t, X_{t-1})$, as well as the associated mean square errors, $\Omega_{t|t} = V(\mathbf{f}_t|\mathbf{x}_t, X_{t-1})$. Starting the recursions with $\mathbf{f}_{0t|0t} = 0$ and $\Omega_{0t|0t} = \Lambda_t$, the (cross-sectional) updating equations are

$$\begin{aligned} \mathbf{f}_{it|it} &= \mathbf{f}_{i-1t|i-1t} + \delta_{it}^{-1}\Omega_{i-1t|i-1t}\mathbf{c}_i\varepsilon_{it} \\ \Omega_{it|it} &= \Omega_{it-1|it-1} - \delta_{it}^{-1}\Omega_{i-1t|i-1t}\mathbf{c}_i\mathbf{c}_i'\Omega_{i-1t|i-1t} \end{aligned} \tag{4}$$

where

$$\begin{aligned} \varepsilon_{it} &= x_{it} - \mathbf{c}_i'\mathbf{f}_{i-1t|i-1t} \\ \delta_{it} &= \mathbf{c}_i'\Omega_{i-1t|i-1t}\mathbf{c}_i + \gamma_i \end{aligned} \tag{5}$$

are the cross-sectional prediction error and its variance.

After processing all N series in this way, it is possible to prove that we obtain the KALMAN filter (time-series) updating equations:

$$\begin{aligned} \mathbf{f}_{t|t} &= \Lambda_t\mathbf{C}'\Sigma_t^{-1}\mathbf{x}_t \\ \Omega_{t|t} &= \Lambda_t - \Lambda_t\mathbf{C}'\Sigma_t^{-1}\mathbf{C}\Lambda_t \end{aligned}$$

Importantly, given the degenerate nature of the (time-series) transition equation, smoothing is unnecessary in this case, so that $\mathbf{f}_{t|t} = E(\mathbf{f}_t|X_T;\phi)$ and $\Omega_{t|t} = V(\mathbf{f}_t|X_T;\phi)$ (see, DIEBOLD and NERLOVE [1989]). [2]

---

1. Alternatively, if we regard the $\mathbf{f}_t's$ as parameters and the $\mathbf{c}'s$ as regressors, we can also interpret them as the state-space representation of a standard (weighted) linear regression model, which uses $\mathbf{f}_{0t}\sim N(\mathbf{0},\Lambda_t)$ as an informative prior.

2. Smoothing is also unnecessary in the cross-section, as $\mathbf{f}_{it|Nt} = \mathbf{f}_{Nt|Nt} = \mathbf{f}_{t|t}$ and $\Omega_{it|Nt} = \Omega_{Nt|Nt} = \Omega_{t|t}$.

3 Maximum Likelihood Estimation

3.1 Derivative Based Optimization Methods

In this model, the parameters of interest $\phi'=(\mathbf{c}',\gamma',\psi')$ are usually estimated jointly from the log-likelihood function of the observed variables, $\mathbf{x}_t$. For those parameter configurations that imply rank $(\Sigma_t) = N$, the log-likelihood function of a sample of size T (ignoring initial conditions) takes the form $\sum_{t=1}^{T} l(\mathbf{x}_t|X_{t-1};\phi)$, where:

(6)

$$l(\mathbf{x}_t|X_{t-1};\phi) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\left|\mathbf{C}\Lambda_t\mathbf{C}'+\Gamma\right| - \frac{1}{2}\mathbf{x}_t'\left(\mathbf{C}\Lambda_t\mathbf{C}'+\Gamma\right)^{-1}\mathbf{x}_t$$

with $\Lambda_t = \text{diag}\,[\lambda_t(\psi)]$. [3] Given the nonlinear nature of the model, a numerical optimization approach is usually required to obtain maximum likelihood (ML) estimates of ϕ. This often involves first order derivatives, which can be approximated numerically by re-evaluating (6) with each parameter in turn shifted by a small amount. But in this case we can also obtain an analytical expression for the score. Following BOLLERSLEV and WOOLDRIDGE [1992], we can write the score function $\mathbf{s}(\mathbf{x}_t|X_{t-1};\phi) = \partial l(\mathbf{x}_t|X_{t-1};\phi)/\partial\phi$ for any conditionally Gaussian model with mean vector μ_t and covariance matrix Σ_t, as:

$$\mathbf{s}(\mathbf{x}_t|X_{t-1};\phi) = \frac{\partial\mu_t'}{\partial\phi}\Sigma_t^{-1}(\mathbf{x}_t-\mu_t) + \frac{1}{2}\frac{\partial\text{vec}'\,[\Sigma_t]}{\partial\phi}\left[\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right]\text{vec}\left[(\mathbf{x}_t-\mu_t)(\mathbf{x}_t-\mu_t)'-\Sigma_t\right]$$

Obviously, the first term disappears in our case because $\mu_t = 0$. Also, since the differential of Σ_t is

$$d(\mathbf{C}\Lambda_t\mathbf{C}'+\Gamma) = (d\mathbf{C})\Lambda_t\mathbf{C}' + \mathbf{C}(d\Lambda_t)\mathbf{C}' + \mathbf{C}\Lambda_t(d\mathbf{C}') + d\Gamma$$

3. Although expression (6) breaks down for Σ_t singular, the value of the log-likelihood function does not automatically become $+\infty$. On the contrary, a singular Σ_t generally imposes an unbounded penalty on the objective function, unless there are singularities in the data matched by the singularities in Σ_t. For instance, if $k = 0$, it is straightforward to prove that the log-likelihood function actually goes to $-\infty$ when γ_j tends to 0, unless $x_{jt}^2 = 0\ \forall t$.

(*cf.* MAGNUS and NEUDECKER [1988]), the three terms of the Jacobian corresponding to $\mathbf{c}, \gamma$ and ψ will be:

$$\frac{\partial \text{vec}\,[\Sigma_t]}{\partial \mathbf{c}'} = (\mathbf{I} + \mathbf{K}_{NN})(I \otimes \mathbf{C}\Lambda_t);$$

$$\frac{\partial \text{vec}\,[\Sigma_t]}{\partial \gamma'} = \mathbf{E}_N;$$

$$\frac{\partial \text{vec}\,[\Sigma_t]}{\partial \psi'} = (\mathbf{C} \otimes \mathbf{C})\mathbf{E}_k \frac{\partial \lambda_t(\phi)}{\partial \psi'}$$

where $\mathbf{E}_n$ is the unique $n^2 \times n$ "*diagonalization*" matrix which transforms $\text{vec}(\mathbf{A})$ into $vecd(\mathbf{A})$ as $\text{vec}d(\mathbf{A}) = \mathbf{E}_n'\text{vec}(\mathbf{A})$, and $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see, MAGNUS [1988]).

After some straightforward algebraic manipulations, we finally get

$$\mathbf{s}_{\mathbf{c}}(\mathbf{x}_t|X_{t-1};\phi) = \text{vec}\left[\Lambda_t\mathbf{C}'\Sigma_t^{-1}\mathbf{x}_t\mathbf{x}_t'\Sigma_t^{-1} - \Lambda_t\mathbf{C}'\Sigma_t^{-1}\right]$$

$$\mathbf{s}_{\gamma}(\mathbf{x}_t|X_{t-1};\phi) = \frac{1}{2}\text{vec}d\left[\Sigma_t^{-1}\mathbf{x}_t\mathbf{x}_t'\Sigma_t^{-1} - \Sigma_t^{-1}\right]$$

$$\mathbf{s}_{\psi}(\mathbf{x}_t|X_{t-1};\phi) = \frac{1}{2}\partial\lambda_t'(\psi)/\partial\psi \cdot \text{vec}d\left[\mathbf{C}'\Sigma_t^{-1}\mathbf{x}_t\mathbf{x}_t'\Sigma_t^{-1}\mathbf{C} - \mathbf{C}'\Sigma_t^{-1}\mathbf{C}\right]$$

3.2 HEYWOOD Cases

However, it is important to realize that the usual definition of ML estimates as a root of the unrestricted first-order conditions $\sum_{t=1}^{T}\mathbf{s}(\mathbf{x}_t|X_{t-1};\phi) = 0$ is not necessarily appropriate in this case, as it ignores that γ must be non-negative for the model to make sense. Without that restriction, in fact, the solution of the likelihood equations may involve some negative estimates of the idiosyncratic variances. But since a negative variance estimate is clearly an undesirable outcome, it is more appropriate to use constrained optimization algorithms which take into account the parameter restrictions which guarantee that Σ_t is p.d.

In this respect, the correct first-order conditions corresponding to the inequality constrained idiosyncratic variance parameters will be given by the KUHN-TUCKER conditions, which in this case amount to:

$$\sum_{t=1}^{T}\text{vec}d\left[\tilde{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\tilde{\Sigma}_t^{-1} - \tilde{\Sigma}_t^{-1}\right] \leqslant 0$$

$$\tilde{\gamma} \geqslant 0 \tag{7}$$

$$\sum_{t=1}^{T}\text{vec}d\left[\left(\tilde{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\tilde{\Sigma}_t^{-1} - \tilde{\Sigma}_t^{-1}\right)\right] \odot \tilde{\gamma} = 0$$

where $\sim$ denotes maximum likelihood estimates, and $\odot$ the Hadamard (or element by element) product of matrices. Importantly, note that the first line of (7) yields the (minus) KUHN-TUCKER multipliers associated with the N inequality restrictions $\gamma \geqslant 0$.

This means that parameter configurations with one or more zero idiosyncratic variances, which are at the boundary of the admissible space, may well satisfy the first-order conditions for maximisation (7), even though $\sum_{t=1}^{T} \mathbf{s}_\gamma(\mathbf{x}_t | X_{t-1}; \phi) \neq 0$. Such solutions, known in the static factor analysis literature as HEYWOOD cases, after HEYWOOD [1931], actually occur in practice fairly frequently. Given that the maximum number of HEYWOOD cases compatible with $\text{rank}(\Sigma_t) = N$ is k, this implies that on top of interior solutions, there are $\sum_{j=1}^{k} \binom{N}{j}$ potential corner solutions. Therefore, it would be necessary to evaluate the score at those solutions to see whether they constitute local maxima, and also the likelihood function to see if they are the global maximum. [4]

In principle, the conditionally heteroskedastic model (1)-(2) remains well defined even if the true Γ does not have full rank. Strictly speaking, this simply means that some observed variables are perfectly explained by the unobserved factors. But in that case, the conditional distribution of the common factors given the observed variables must be degenerate. For instance, in the limiting case of $\text{rank}(\Gamma) = N - k$, $\Omega_{t|t} = 0$, and all common factors become effectively observable.

Unless there are *a priori*, reasons that suggest so (*e.g.* because we directly observe "*diversified*"' linear combinations of a larger set of $x_{it}'s$), such situations are unlikely to be plausible in many empirical applications. Thus, the occurrence of HEYWOOD cases in a given sample should generally be regarded as the result of sampling variation in the ML estimators of γ. Its incidence, therefore, will be closely related to the degree of left censoring in the asymptotic distribution of $\tilde{\gamma}$, which depends on the true value of Γ, the asymptotic information matrix and the sample size. In this respect, we can use further results from BOLLERSLEV and WOOLDRIDGE [1992] to show that the block of the asymptotic information matrix corresponding to γ will be given by $\frac{1}{2} E\left(\Sigma_t^{-1} \odot \Sigma_t^{-1}\right)$. The fact that the diagonal elements of Σ_t^{-1} are the reciprocals of the residual variances in the (conditional) regressions of each x_{it} on the remaining $N-1$ observed series, is in line with BARTHOLOMEW's [1987] argument that the higher the linear dependence between a variable and the rest, the more likely it is for the corresponding γ_i to be estimated as 0.

---

4. It is possible to come up with examples for which conditions (7) are satisfied by a solution with some HEYWOOD cases, and others for which they are not. As an extreme case, if $N = 2$, $k = 1$ and $\lambda_{1t} = 1 \forall t$, those conditions are always satisfied by the two possible HEYWOOD cases solutions. In fact, both corner solutions will be global maxima, since a static, exact single factor model is underidentified with only two series, but a singular idiosyncratic variance makes it exactly identified.

3.3 EM Algorithm

The EM algorithm provides an alternative way to maximize the log-likelihood function of conditionally heteroskedastic factor models subject to the inequality restrictions $\gamma \geqslant 0$ without explicitly utilizing first order derivatives or penalty functions (see, DEMOS and SENTANA [1998] for details). The algorithm is based on the following identity:

$$
\begin{aligned}
l(\mathbf{x}_t,\mathbf{f}_t|X_{t-1};\phi) &\equiv l(\mathbf{x}_t|\mathbf{f}_t,X_{t-1};\phi)+l(\mathbf{f}_t|X_{t-1};\phi) \\
&\equiv l(\mathbf{x}_t|X_{t-1};\phi)+l(\mathbf{f}_t|\mathbf{x}_t,X_{t-1};\phi)
\end{aligned} \tag{8}
$$

where $l(\mathbf{x}_t,\mathbf{f}_t|X_{t-1};\phi)$ is the joint log-density function of $\mathbf{x}_t$ and $\mathbf{f}_t$, $l(\mathbf{x}_t|\mathbf{f}_t,X_{t-1};\phi)$ is the conditional of $\mathbf{x}_t$ given $\mathbf{f}_t$, $l(\mathbf{f}_t|X_{t-1};\phi)$ the marginal log-density of $\mathbf{f}_t$, and $l(\mathbf{f}_t|\mathbf{x}_t,X_{t-1};\phi)$ the conditional of $\mathbf{f}_t$ given $\mathbf{x}_t$, all given X_{t-1}. At each iteration, the EM algorithm maximizes the expected value of $\sum_t [l(\mathbf{x}_t|\mathbf{f}_t,X_{t-1};\phi)+l(\mathbf{f}_t|X_{t-1};\phi)]$ conditional on X_T and the current parameter estimates, $\phi^{(n)}$. The rationale stems from the fact that the expected value of $\sum_t l(\mathbf{f}_t|\mathbf{x}_t,X_{t-1};\phi)$, conditional on X_T and $\phi^{(n)}$, reaches a maximum at $\phi=\phi^{(n)}$ by virtue of the KULLBACK inequality, which means that *any increase* in the expected value of $\sum_t l(x_t,\mathbf{f}_t|X_{t-1};\phi)$ must represent an increase in $\sum_t l(\mathbf{x}_t|X_{t-1};\phi)$. The EM objective function at iteration $n+1$ is

$$
-\frac{TN}{2}\ln 2\pi-\frac{T}{2}\ln|\Gamma|-\frac{1}{2}\sum_{t=1}^{T} tr\left\{\Gamma^{-1}\left[(\mathbf{x}_t-\mathbf{C}\mathbf{f}_{t|t}^{(n)})(\mathbf{x}_t-\mathbf{C}\mathbf{f}_{t|t}^{(n)})'+\mathbf{C}\Omega_{t|t}^{(n)}\mathbf{C}'\right]\right\} \tag{9}
$$

$$
-\frac{Tk}{2}\ln 2\pi-\frac{1}{2}\sum_{t=1}^{T}\ln|\Lambda_t(\psi)|-\frac{1}{2}\sum_{t=1}^{T} tr\left[\Lambda_t^{-1}(\psi)\left(\mathbf{f}_{t|t}^{(n)}\mathbf{f}_{t|t}^{(n)\prime}+\Omega_{t|t}^{(n)}\right)\right] \tag{10}
$$

where $^{(n)}$ means evaluated at $\phi^{(n)}$. Given our assumptions, this implies that:

$$
\mathbf{c}^{(n+1)}=\text{vec}\left\{\left[\sum_{t=1}^{T}\left(\mathbf{f}_{t|t}^{(n)}\mathbf{f}_{t|t}^{(n)\prime}+\Omega_{t|t}^{(n)}\right)\right]^{-1}\left[\sum_{t=1}^{T}\mathbf{f}_{t|t}^{(n)}\mathbf{x}_t'\right]\right\} \tag{11}
$$

$$
\gamma^{(n+1)}=\text{vec}d\left\{\frac{1}{T}\sum_{t=1}^{T}\left[(\mathbf{x}_t-\mathbf{C}\mathbf{f}_{t|t}^{(n)})(\mathbf{x}_t-\mathbf{C}\mathbf{f}_{t|t}^{(n)})'+\mathbf{C}\Omega_{t|t}^{(n)}\mathbf{C}'\right]\right\} \tag{12}
$$

while $\psi_j^{(n+1)}$ can be obtained by numerically maximizing (10). Notice the similarity between expressions (11)-(12) and the multivariate regression ones that would apply to the case in which $\mathbf{f}_t$ were observed. Here, the unobservable factors are replaced by their best (in the conditional mean square error sense) estimates given the available series. The main difference is that in computing the estimates of $\mathbf{C}$, we take into account the fact that $\mathbf{f}_t$ is not really observed. Notice, also, that (11) and (12) coincide with the corresponding expressions for the standard factor analytic model in which the factors have constant variances (see *e.g.* RUBIN and THAYER [1982]). The only difference is that $\Omega_{t|t}^{(n)}$ is time-invariant in that case.

Finally, note that the EM recursions implicitly guarantee that $\gamma^{(n+1)} \geqslant 0$ as long as $\Omega_{t|t}^{(n)}$ is p.d., and therefore, the algorithm will never venture beyond the boundaries of the admissible parameter space. Unfortunately, it may sometimes get stuck at parameter values corresponding to corner solutions, even though the inequality-constrained first-order conditions (7) are not satisfied. For instance, if at the n^{th} EM iteration, the number of HEYWOOD cases is k, it is possible to prove that $(\mathbf{c}^{(n+1)}, \gamma^{(n+1)}) = (\mathbf{c}^{(n)}, \gamma^{(n)})$. In those cases, it would be necessary to restart the computations with different initial values.

4 Evaluation of the Log-Likelihood Function and Score

4.1 A Recursive Algorithm

The state-space representation in (3) allows us to compute the likelihood function based on a cross-sectional analogue of the prediction error decomposition:

$$l(\mathbf{x}_t | X_{t-1}; \phi) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N} \ln |\delta_{it}(\phi)| - \frac{1}{2}\sum_{i=1}^{N} \frac{\varepsilon_{it}^2(\phi)}{\delta_{it}(\phi)} \tag{13}$$

where ε_{it} and δ_{it} are defined in (5).

As usual, such a decomposition may be interpreted as implicitly performing a CHOLESKY-type factorisation of the matrix Σ_t. Specifically, it is possible to prove that ε_t solves $\Phi_t \varepsilon_t = \mathbf{x}_t$, where $\Phi_t \Delta_t \Phi_t'$, with Φ_t unit lower triangular and $\Delta_t = \text{diag}(\delta_{1t}, \ldots, \delta_{Nt})$, is the symmetric factorisation of Σ_t. Therefore, (13) constitutes a rather safe and efficient computational procedure, which avoids unnecessary numerical errors associated with matrix inversions (see, BAUER and REINSCH [1971]). Importantly, this algorithm is not affected by the fact that some $h(\leqslant k)\gamma_i's$ are 0, and thus, remain valid even if the parameter values reach the boundary of the admissible space during the optimization process. Obviously, the symmetric factorisation is not guaranteed to be well defined if Σ_t is not p.s.d., but this will only happen for non-admissible parameter configurations.

A set of analytical recursions can also be developed to compute the score. If we apply the general expression in BOLLERSLEV and WOOLDRIDGE [1992] to (13), we obtain

$$s(\mathbf{x}_t|X_{t-1};\phi) = \tag{14}$$

$$-\sum_{i=1}^{N}\frac{\partial\varepsilon_{it}(\phi)}{\partial\phi}\frac{\varepsilon_{it}(\phi)}{\delta_{it}(\phi)}+\frac{1}{2}\sum_{i=1}^{N}\left[\frac{1}{\delta_{it}(\phi)}\frac{\partial\delta_{it}(\phi)}{\partial\phi}\left(\frac{\varepsilon_{it}^2(\phi)}{\delta_{it}(\phi)}-1\right)\right]$$

where

$$\frac{\partial\varepsilon_{it}(\phi)}{\partial\phi_j}=-\left(\mathbf{c}_i'\frac{\partial\mathbf{f}_{i-1t|i-1t}(\phi)}{\partial\phi_j}+\frac{\partial\mathbf{c}_i'}{\partial\phi_j}\mathbf{f}_{i-1t|i-1t}(\phi)\right)$$

and

$$\frac{\partial\delta_{it}(\phi)}{\partial\phi_j}=2\frac{\partial\mathbf{c}_i'}{\partial\phi_j}\Omega_{i-1t|i-1t}(\phi)\mathbf{c}_i+\mathbf{c}_i'\frac{\partial\Omega_{i-1t|i-1t}(\phi)}{\partial\phi_j}\mathbf{c}_i+\frac{\partial\gamma_i}{\partial\phi_j}$$

A related set of recursions for $\partial\mathbf{f}_{i-1t|i-1t}(\phi)/\partial\phi_j$ and $\partial\Omega_{i-1t|i-1t}(\phi)/\partial\phi_j$ can be obtained by differentiating the updating equations (4), with $\partial\mathbf{f}_{0t|0t}(\phi)/\partial\phi_j=0$ and $\partial\Omega_{0t|0t}\phi)/\partial\phi_j=\partial\Lambda_t(\psi)/\partial\phi_j$ as starting values (see, HARVEY [1989] for details). In this respect, it is worth mentioning that given our assumptions about λ_t, the parameters $\mathbf{c}_i'$ and γ_i do not appear in the cross-sectional prediction error decomposition until x_{it} is processed.

4.2 A Non-Recursive Method

Despite the appeal of the recursive method, in most cases a much more efficient algorithm can be developed from the identity (8). In particular, after grouping terms we can write

$$l(\mathbf{x}_t|X_{t-1};\phi)=-\frac{N}{2}\ln 2\pi- \tag{15}$$

$$\frac{1}{2}\ln\left(|\Gamma|\cdot|\Lambda_t|\cdot\left|\Omega_{t|t}^{-1}\right|\right)-\frac{1}{2}\left(\mathbf{x}_t'\Gamma^{-1}\mathbf{x}_t-\mathbf{f}_{t|t}'\Omega_{t|t}^{-1}\mathbf{f}_{t|t}\right)$$

$$-\frac{1}{2}\mathbf{f}_t'\left(\Lambda_t^{-1}+\mathbf{C}'\Gamma^{-1}\mathbf{C}-\Omega_{t|t}^{-1}\right)\mathbf{f}_t-\mathbf{f}_t'\left(\Omega_{t|t}^{-1}\mathbf{f}_{t|t}-\mathbf{C}'\Gamma^{-1}\mathbf{x}_t\right)$$

provided the necessary inverses exist. But since the last two terms must be identically 0 for any value of $\mathbf{f}_t$, it follows that

$$
\begin{aligned}
\Omega_{t|t} &= (\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C})^{-1} \\
\mathbf{f}_{t|t} &= (\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C})^{-1}\mathbf{C}'\Gamma^{-1}\mathbf{x}_t \\
|\Sigma_t| &= |\Lambda_t| \cdot |\Gamma| \cdot \left|\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C}\right| \\
\mathbf{x}_t'\Sigma_t^{-1}\mathbf{x}_t &= \mathbf{x}_t'\Gamma^{-1}\mathbf{x}_t - \mathbf{x}_t'\Gamma^{-1}\mathbf{C}(\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C})^{-1}\mathbf{C}'\Gamma^{-1}\mathbf{x}_t
\end{aligned}
$$

Note that the factorisation of Σ_t (a $N \times N$ matrix) is replaced by the factorisation of the $k \times k$ matrix $\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C}$ and the $N \times N$ matrix Γ. Therefore, the computational advantages of these formulae arise mainly from the restrictions on Γ. Since Γ is diagonal, its inverse and determinant are trivial to find. Similarly, since it is time-invariant, so is $\mathbf{C}'\Gamma^{-1}\mathbf{C}$. [5]

Expression (8) can also be exploited to simplify the computation of the score. Since the KULLBACK inequality implies that

$$
E\left[\sum_t s(\mathbf{f}_t|\mathbf{x}_t, X_{t-1}; \phi)|X_T, \phi\right] = 0,
$$

it is clear that $\mathbf{s}(\mathbf{x}_t|X_{t-1}; \phi)$ can be obtained as the expected value (given X_T and ϕ) of the sum of the unobservable scores corresponding to $l(\mathbf{f}_t|X_{t-1}; \phi)$ and $l(x_t|\mathbf{f}_t, X_{t-1}; \phi)$. Assuming that $\gamma > 0$, this yields

$$
\mathbf{s}_{\mathbf{c}}(\mathbf{x}_t|X_{t-1}; \phi) = \text{vec}\left\{\left[\mathbf{f}_{t|t}\mathbf{x}_t' - \left(\mathbf{f}_{t|t}\mathbf{f}_{t|t}' + \Omega_{t|t}\right)\mathbf{C}'\right]\Gamma^{-1}\right\}
$$

$$
\mathbf{s}_{\gamma}(x|X_{t-1}; \phi) =
$$

$$
\frac{1}{2}\text{vec}d\left\{\Gamma^{-1}\left[(\mathbf{x}_t - \mathbf{C}\mathbf{f}_{t|t})(\mathbf{x}_t - \mathbf{C}\mathbf{f}_{t|t})' + \mathbf{C}\Omega_{t|t}\mathbf{C}' - \Gamma\right]\Gamma^{-1}\right\} \tag{17}
$$

$$
\mathbf{s}_{\psi}(\mathbf{x}|X_{t-1}; \phi) =
$$

$$
\frac{1}{2}\partial\lambda_t'(\psi)/\partial\psi \cdot \text{vec}d\left\{\Lambda_t^{-1}\left[\mathbf{f}_{t|t}\mathbf{f}_{t|t}' + \Omega_{t|t} - \Lambda_t\right]\Lambda_t^{-1}\right\}
$$

As in the recursive algorithm, these expressions can be alternatively obtained by applying standard matrix results. In particular, we can exploit the form of Σ_t by means of the so-called WOODBURY formulae (see the appendix for details).

Unfortunately, the advantages in computing $l(\mathbf{x}_t|X_{t-1}; \phi)$ and $\mathbf{s}(\mathbf{x}_t|X_{t-1}; \phi)$ with (15) and (17) are lost even if there is only one zero element in γ, however large N is. For that reason, we propose a third algorithm, which advantageously combines the other two.

---

5. In order to avoid unnecessary numerical errors associated with matrix inversions, it is convenient to perform the symmetric factorization of $\Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C} = \mathbf{F}_{Lt}\mathbf{F}_{Dt}\mathbf{F}_{Lt}'$ first, so that we can compute the required expressions as $|\Sigma_t| = |\Lambda_t| \cdot |\Gamma| \cdot |\mathbf{F}_{Dt}|$, $\mathbf{x}_t'\Sigma_t^{-1}\mathbf{x}_t = \mathbf{x}_t'\Gamma^{-1}\mathbf{x}_t - \mathbf{y}_t'\mathbf{F}_{Dt}^{-1}\mathbf{y}_t$, $\Omega_{t|t} = \mathbf{F}_{Lt}^{-1\prime}\mathbf{F}_{Dt}^{-1}\mathbf{F}_{Lt}^{-1}$ and $\mathbf{f}_{t|t} = \mathbf{z}_t$, with $\mathbf{y}_t, \mathbf{z}_t$ and $\mathbf{F}_{Lt}^{-1}$ obtained as the solutions to the unit triangular systems of linear equations $\mathbf{F}_{Lt}\mathbf{y}_t = \mathbf{C}'\Gamma^{-1}\mathbf{x}_t$, $\mathbf{F}_{Lt}'\mathbf{z}_t = \mathbf{F}_{Dt}^{-1}\mathbf{y}_t$ and $\mathbf{F}_{Lt}\mathbf{F}_{Lt}^{-1} = \mathbf{I}_k$ respectively.

4.3 A Block-Recursive Algorithm

For simplicity, we shall only consider two blocks of sizes N_a and N_b. [6] Let us partition all vectors and matrices accordingly so that we can write (1) as:

$$\begin{aligned} \mathbf{x}_{at} &= \mathbf{C}_a \mathbf{f}_t + \mathbf{w}_{at} \\ \mathbf{x}_{bt} &= \mathbf{C}_b \mathbf{f}_t + \mathbf{w}_{bt} \end{aligned} \tag{18}$$

and define $\mathbf{c}_a$, $\mathbf{c}_b$, γ_a and γ_b as the corresponding elements of $\mathbf{c}$ and γ.

We can factorize the joint log-density of $\mathbf{x}_{at}$ and $\mathbf{x}_{bt}$ as:

$$l(\mathbf{x}_{at}, \mathbf{x}_{bt} | X_{t-1}; \phi) = l(\mathbf{x}_{at} | X_{T-1}; \phi) + l(\mathbf{x}_{bt} | \mathbf{x}_{at}, X_{t-1}; \phi) \tag{19}$$

If $\gamma_a > 0$, $l(\mathbf{x}_{at} | X_{t-1}; \phi)$ can be computed in one go on the basis of the expressions in section 4.2. If not, we can reorder the variables so that the h null elements of γ appear in the last N_b positions. This is equivalent to premultiplying (1) by a permutation matrix U'. [7] On the other hand,

$$l(\mathbf{x}_{bt} | \mathbf{x}_{at}, X_{t-1}; \phi) = -\frac{N_b}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_{b.at}| - \frac{1}{2} \eta'_{bt} \Sigma^{-1}_{b.at} \eta_{bt}$$

where

$$\eta_{bt} = \mathbf{x}_{bt} - \mathbf{C}_b \Lambda_t \mathbf{C}'_a \Sigma^{-1}_{at} \mathbf{x}_{at} = \mathbf{x}_{bt} - \mathbf{C}_b \mathbf{f}_{at|at}$$

and

$$\Sigma_{b.at} = \mathbf{C}_b \Lambda_t \mathbf{C}'_b + \Gamma_b - \mathbf{C}_b \Lambda_t \mathbf{C}'_a \Sigma^{-1}_{at} \mathbf{C}_a \Lambda_t \mathbf{C}'_b = \mathbf{C}_b \Omega_{at|at} \mathbf{C}'_b + \Gamma_b$$

Note that since $\eta_{bt} = \mathbf{C}_b(\mathbf{f}_t - \mathbf{f}_{at|at}) + \mathbf{w}_{bt}$, the conditional covariance matrix of η_{bt} has the same factor structure as the conditional covariance matrix of $\mathbf{x}_{bt}$, but with Λ_t replaced with $\Omega_{at|at}$. Also, note that the positive definiteness of Σ_t implies that the inverse of $\Sigma_{b.at}$ exist even in the extreme case of $\gamma_b = 0$, provided that $N_b \leqslant k$.

Finally, the cross-sectional state-space representation of the model in (3) allows us to write the block-updating equations

$$\begin{aligned} \mathbf{f}_{t|t} &= \mathbf{f}_{at|at} + \Omega_{at|at} \mathbf{C}'_b \Sigma^{-1}_{b.at} \eta_{bt} \\ \Omega_{t|t} &= \Omega_{at|at} - \Omega_{at|at} C'_b \Sigma^{-1}_{b.at} C_b \Omega_{at|at} \end{aligned} \tag{20}$$

---

6. With only one block, the algorithm collapses to the non-recursive one in section 4.2. On the other hand, with N blocks, we obtain the fully recursive algorithm in section 4.1.

7. Such a matrix can be found by interchanging the first zero element found moving down γ from the top with the first positive element found moving up from the bottom, and then repeating the same procedure for the remaining elements in the centre. This sorting scheme not only involves a single pass over γ, but also produces a symmetric $\mathbf{U}$, simplifying the subsequent permutation of $\mathbf{x}_t$ and $\mathbf{C}$.

Again, these expressions can be alternatively obtained by applying a partitioned inverse version of the WOODBURY formula (see the appendix for details). [8] Despite their apparent complexity, numerical simulations confirm that the computational advantages of this block-recursive algorithm over that based on the CHOLESKY decomposition of Σ_t are still substantial, especially if N_b is small relative to N. For instance, if $N_b = 1$, $\Sigma_{b.at}$ is a scalar. Importantly, the expressions remain valid when the parameter values move to the boundary of the admissible space.

As far as the gradient is concerned, note that (19) implies that

$$\mathbf{s}(\mathbf{x}_{at}, \mathbf{x}_{bt} | X_{t-1}; \phi) = \mathbf{s}(\mathbf{x}_{at} | X_{t-1}; \phi) + \mathbf{s}(\mathbf{x}_{bt} | \mathbf{x}_{at}, X_{t-1}; \phi)$$

Since given our assumptions about λ_t, $\mathbf{c}_b$ and γ_b only affect the second component, the derivatives of the log-likelihood function with respect to these parameters will be given by

$$s_{\mathbf{c}_b}(\mathbf{x}_t | X_{t-1}; \phi) =$$
$$\text{vec}\left[\mathbf{f}_{at|at}\eta'_{bt}\Sigma^{-1}_{b.at} + \Omega_{at|at}\mathbf{C}'_b\Sigma^{-1}_{b.at}\eta_{bt}\eta'_{bt}\Sigma^{-1}_{b.at} - \Omega_{at|at}\mathbf{C}'_b\Sigma^{-1}_{b.at}\right]$$

$$s_{\gamma_b}(\mathbf{x}_t | X_{t-1}; \phi) = \frac{1}{2}\text{vec}d\left[\Sigma^{-1}_{b.at}\eta_{bt}\eta'_{bt}\Sigma^{-1}_{b.at} - \Sigma^{-1}_{b.at}\right]$$

In this respect, it is worth recalling that $\sum_t s_{\gamma_b}(\mathbf{x}_t | X_{t-1}; \phi)$ are precisely the derivatives whose sign has to be checked to decide whether a boundary solution satisfies the first-order KUHN-TUCKER conditions (7). On the other hand, since $\mathbf{c}_a$ and γ_a affect the first component directly, and the second through $\mathbf{f}_{at|at}$ and $\Omega_{at|at}$, we can combine the expressions in section 4.2 (*cf.* (17)), with the block-updating equations (20) to find $s_{\mathbf{c}_a}(\mathbf{x}_t | X_{t-1}; \phi)$ and $s_{\gamma_a}(\mathbf{x}_t | X_{t-1}; \phi)$. Finally, note that the expression for $s\ (\mathbf{x}_t | X_{t-1}; \phi)$ in (17) is unaffected by zeros in γ_b.

If the number of HEYWOOD cases h is equal to the number of factors, it is in fact more convenient to move them to the first positions, with $N_a = k$. The reason is that since $\mathbf{C}_a$ must have full rank in that case to guarantee a p.d. Σ_t,

---

8. Once more, it is numerically convenient to perform the symmetric factorization of $\Lambda_t^{-1} + \mathbf{C}'_a\Gamma_a^{-1}\mathbf{C}_a = \mathbf{F}_{aLt}\mathbf{F}_{aDt}\mathbf{F}'_{aLt}$ first, so that we can compute the required expressions as

$$|\Sigma_t| = |\Lambda_t| \cdot |\Gamma_a| \cdot |\mathbf{F}_{aDt}| \cdot |\mathbf{G}_{Dt}|, \quad \mathbf{x}'_t\Sigma_t^{-1}\mathbf{x}_t = \mathbf{x}'_{at}\Gamma_a^{-1}\mathbf{x}_{at} - \mathbf{y}'_{at}\mathbf{F}_{aDt}^{-1}\mathbf{y}_{at} + \mathbf{w}'_{bt}\mathbf{G}_{Dt}^{-1}\mathbf{w}_{bt},$$

$\Omega_{at|at} = \mathbf{F}_{aLt}^{-1\prime}\mathbf{F}_{aDt}^{-1}\mathbf{F}_{aLt}^{-1}$, $\Omega_{t|t} = \Omega_{at|at} - \mathbf{W}'_{bt}\mathbf{G}_{Dt}^{-1}\mathbf{W}_{bt}$, $\mathbf{f}_{at|at} = \mathbf{z}_{at}$ and $\mathbf{f}_{t|t} = \mathbf{f}_{at|at} + \mathbf{W}'_{bt}\mathbf{G}_{Dt}^{-1}\mathbf{w}_{bt}$, where $\mathbf{G}_{Lt}\mathbf{G}_{Dt}\mathbf{G}'_{Lt}$ denotes the symmetric factorization of $\mathbf{G}_t = \mathbf{Y}'_{bt}\mathbf{F}_{aDt}^{-1}\mathbf{Y}_{bt} + \Gamma_b = \Sigma_{b.at}$, and $\mathbf{y}_{at}, \mathbf{z}_{at}, \mathbf{F}_{aLt}^{-1}, \mathbf{Y}_{bt}, \mathbf{Z}_{bt}, \mathbf{w}_{bt}$ and $\mathbf{W}_{bt}$ are obtained as the solutions to the unit triangular systems of linear equations

$$\mathbf{F}_{aLt}\mathbf{y}_{at} = \mathbf{C}'_a\Gamma_a^{-1}\mathbf{x}_{at},\ \mathbf{F}'_{aLt}\mathbf{z}_{at} = \mathbf{F}_{aDt}^{-1}\mathbf{y}_{at},\ \mathbf{F}_{aLt}\mathbf{F}_{aLt}^{-1} = \mathbf{I}_k,\ \mathbf{F}_{aLt}\mathbf{Y}_{bt} = \mathbf{C}'_b,\ \mathbf{F}'_{aLt}\mathbf{Z}_{bt} = \mathbf{F}_{aDt}^{-1}\mathbf{Y}_{bt},$$

$\mathbf{G}_{Lt}\mathbf{w}_{bt} = \eta_{bt}$ and $\mathbf{G}_{Lt}\mathbf{W}_{bt} = \mathbf{Z}'_{bt}$ respectively.

the factors become effectively observable as $\mathbf{C}_a^{-1}\mathbf{x}_{at}$, and the conditional distribution of $\mathbf{x}_{bt}$ given $\mathbf{x}_{at}$ and X_{t-1} has mean $\mathbf{C}_b^*\mathbf{x}_{at}$, with $\mathbf{C}_b^* = \mathbf{C}_b\mathbf{C}_a^{-1}$, and diagonal covariance matrix Γ_b. Since the induced reparameterization is one-to-one, by the invariance property of ML, one can combine the estimates of $\mathbf{c}_a$ and ψ obtained from the marginal model for $\mathbf{x}_{at}$ with the multivariate regression estimates of $\mathbf{c}_b^*$ and γ_b obtained from the OLS regression of each element of $\mathbf{x}_{bt}$ on $\mathbf{x}_{at}$ (see, SENTANA [1997] for details). Unless the KUHN-TUCKER conditions (7) are satisfied, however, note that the resulting parameter values can only be regarded as ML estimates of an equality restricted model in which k $\gamma_j's$ are set to 0, and that if we unrestrict some of those idiosyncratic variance parameters in the admissible positive direction, the joint log-likelihood function will increase. A related approach can be used if the number of HEYWOOD cases is strictly less than k, provided that the subset of variables with zero idiosyncratic variances depends only on h factors (see, LAWLEY and MAXWELL [1971]). Although this condition can always be achieved in static factor models through orthogonal rotations due to the indeterminacy of the factor loading matrix, in general, it is no longer feasible in models with time-varying factor variances (see, SENTANA and FIORENTINI [1997]).

5 Conclusions

In this paper, we discuss the estimation by full information maximum likelihood of a class of factor models analysed by DEMOS and SENTANA [1998], in which the common factors are subject to ARCH-type effects. We derive the likelihood function and score, as well as the KUHN-TUCKER first-order conditions which define the inequality restricted maximum likelihood estimators. These conditions are relevant because the usual definition of ML estimates as a root of the unrestricted first-order conditions ignores that the idiosyncratic variances must be non-negative for the model to make sense. In this respect, we show that the incidence of HEYWOOD cases (*i.e.* zero estimates for some idiosyncratic variances) is related not only to the sample size and the true value of the specific variances, but also to the linear dependence of a given variable with the rest.

We also express conditionally heteroskedastic factor models as random fields, with both a cross-sectional state space representation, and a more standard time-series representation. In this framework, we explain the application of the KALMAN filter to obtain the best (in the mean square error sense) estimates of the factors.

Finally, we present three methods to compute the likelihood function, its gradient, and the best filtered estimates of the factors, which are not only numerically efficient and reliable, but also intuitive from a statistical point of view. The evidence presented in DEMOS and SENTANA [1998] shows that when combined with an efficient maximisation procedure, they make feasible the empirical application of this class of models to a very large cross-section of time-series.

• References

BARTHOLOMEW D. J. (1987). – *Latent Variable Models and Factor Analysis*, Charles Griffin, London.

BAUER F.L., REINSCH C. (1971). – « Inversion of Positive Definite Matrices by the Gauss-Jordan Method », in Wilkinson, J.H. and Reinsch, C. eds., *Handbook for Automatic Computation vol. II: Linear Algebra*, Springer-Verlag, Berlin.

BOLLERSLEV T., ENGLE R.F. (1993). – « Common Persistence in Conditional Variances », *Econometrica* 61, pp. 166-187.

BOLLERSLEV T., WOOLDRIDGE J.M. (1992). – « Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Variances », *Econometric Reviews* 11, pp. 143-172.

CALZOLARI G., FIORENTINI G., SENTANA E. (1999). – « Indirect Inference Estimation of Conditionally Heteroskedastic Factor Models », *mimeo*, CEMFI.

CHAMBERLAIN G., ROTHSCHILD M. (1983). – « Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets », *Econometrica* 51, pp. 1281-1304.

DEMOS A., SENTANA E. (1998). – « An EM Algorithm for Conditionally Heteroskedastic Factor Models », *Journal of Business and Economic Statistics* 16, pp. 357-361.

DIEBOLD F.X., NERLOVE M. (1989). – « The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model », *Journal of Applied Econometrics* 4, pp. 1-21.

ENGLE R.F., NG V.M., ROTHSCHILD M. (1990). – « Asset Pricing with a Factor-ARCH Structure: Empirical Estimates for Treasury Bills », *Journal of Econometrics* 45, pp. 213-237.

FIORENTINI G., SENTANA E., SHEPHARD N. (1998). – « Exact Likelihood-Based Estimation of Conditionally Heteroskedastic Factor Models », *mimeo*, CEMFI.

GOURIÉROUX C., MONFORT A., Renault E. (1993). – « Indirect Inference », *Journal of Applied Econometrics* 8, pp. S85-S118.

HANSEN L.P., RICHARD S.F. (1987). – « The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models », *Econometrica* 55, pp. 587-613.

HARVEY A.C. (1989). – *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

HARVEY A.C., RUIZ E., SENTANA E. (1992). – « Unobservable Component Time-Series Models with ARCH Disturbances », *Journal of Econometrics* 52, pp. 129-157.

HEYWOOD H.B. (1931). – « On Finite Sequences of Real Numbers », *Proceedings of the Royal Society, Series A* 134, pp. 486-510.

HOUSEHOLDER A.S. (1964). – *The Theory of Matrices in Numerical Analysis*, Blaisdell Publishing Company, London.

KING M.A., SENTANA E., WADHWANI S.B. (1994). – « Volatility and Links Between National Stock Markets », *Econometrica* 62, pp. 901-933.

LAWLEY D.N., MAXWELL A.E. (1971). – *Factor Analysis as a Statistical Method*, 2nd edition, London Butterworths.

MAGNUS J.R. (1988). – *Linear Structures*, Oxford University Press, New York.

MAGNUS J.R., NEUDECKER H. (1988). – *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.

ROSS S.A. (1976). – « The Arbitrage Theory of Capital Asset Pricing », *Journal of Economic Theory*, 13, pp. 341-360.

RUBIN D., THAYER D. (1982). – « EM Algorithms for ML Factor Analysis », *Psychometrika* 47, pp. 69-76.

SENTANA E. (1994). – « The Likelihood Function of Conditionally Heteroskedastic Factor Models with Heywood Cases », CEMFI, *Working Paper* 9420.

SENTANA E. (1997). – « Risk and Return in the Spanish Stock Market: Some Evidence from Individual Assets », *Investigaciones Económicas* 21, pp. 297-359.

SENTANA E. (1998). – « The Relation Between Conditionally Heteroskedastic Factor Models and Factor GARCH Models », *Econometrics Journal* 1, pp. 1-9.

SENTANA E., FIORENTINI G. (1997). – « Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models », CEMFI, *Working Paper*, 9709.

APPENDIX

Generalized WOODBURY Formulae

Let $\mathbf{A}_{N\times N}, \mathbf{B}_{N\times k}, \mathbf{L}_{k\times k}$ and $\mathbf{D}_{N\times k}$ be general complex matrices, and let

$$\mathbf{E}_{N\times N} = \mathbf{A} + \mathbf{BLD}^H \tag{A1}$$

where $\mathbf{D}^H$ denotes the conjugate transpose of $\mathbf{D}$. Then, provided that the necessary inverses exist, the WOODBURY formulae say that:

$$\mathbf{E}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BF}^{-1}\mathbf{D}^H\mathbf{A}^{-1} \tag{A2}$$

$$|\mathbf{E}| = |\mathbf{A}| \cdot |\mathbf{L}| \cdot |\mathbf{F}| \tag{A3}$$

where

$$\mathbf{F}_{k\times k} = \mathbf{L}^{-1} + \mathbf{D}^H\mathbf{A}^{-1}\mathbf{B}$$

(see *e.g.* HOUSEHOLDER [1964]). For the case of $k = 1$, the WOODBURY formulae reduce to the SHERMAN-MORRISON expressions. These formulae are particularly useful when $\mathbf{A}^{-1}$ is readily available, and k is much smaller than N, so that $\mathbf{F}$ is easy to invert.

Since (A2) also implies that $\mathbf{F}^{-1} = \mathbf{L} - \mathbf{LD}^H\mathbf{E}^{-1}\mathbf{BL}$, it follows that the inverse of $\mathbf{F}$ exists if and only if the inverse of $\mathbf{E}$ does. Hence, we just need to look at the case of $\mathbf{A}$ or $\mathbf{L}$ singular. Let's write their singular value decompositions as

$$\mathbf{A} = \mathbf{U}\Delta\mathbf{V}^H = (\mathbf{U}_1 \quad \mathbf{U}_2)\begin{pmatrix}\Delta_1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{V}_1^H\\ \mathbf{V}_2^H\end{pmatrix}$$

$$\mathbf{L} = \mathbf{P}\Lambda\mathbf{Q}^H = (\mathbf{P}_1 \quad \mathbf{P}_2)\begin{pmatrix}\Lambda_1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{Q}_1^H\\ \mathbf{Q}_2^H\end{pmatrix}$$

with $\mathbf{U},\mathbf{V},\mathbf{P},\mathbf{Q}$ unitary, $\Delta_1, \Lambda_1 > 0$, $\text{rank}(\Delta_1) = N_1$ and $\text{rank}(\Lambda_1) = k_1$. It is clear that a necessary condition for $\mathbf{A} + \mathbf{BLD}^H$ to have full rank N is that $N_1 + k_1 \geqslant N$.

Let's now call

$$\mathbf{R} = \mathbf{U}^H\mathbf{BP} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{pmatrix} \quad \mathbf{S} = \mathbf{V}^H\mathbf{DQ} = \begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}$$

Then, if we use the partitioned inverse formula, and provided the necessary inverses exist

$$
\begin{aligned}
\mathbf{E}^{-1} &= [\mathbf{U}(\Delta + \mathbf{R}\Lambda_1\mathbf{S}^H)\mathbf{V}^H]^{-1} \\
&= (\mathbf{V}_1 \quad \mathbf{V}_2)\begin{pmatrix} \Delta_1 + \mathbf{R}_{11}\Lambda_1\mathbf{S}_{11}^H & \mathbf{R}_{11}\Lambda_1\mathbf{S}_{21}^H \\ \mathbf{R}_{21}\Lambda_1\mathbf{S}_{11}^H & \mathbf{R}_{21}\Lambda_1\mathbf{S}_{21}^H \end{pmatrix}\begin{pmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{pmatrix} \\
&= (\mathbf{V}_1 \quad \mathbf{V}_2) \\
&\begin{pmatrix} \mathbf{E}_{11}^{-1} + \mathbf{E}_{11}^{-1}\mathbf{R}_{11}\Lambda_1\mathbf{S}_{21}^H\mathbf{G}^{-1}\mathbf{R}_{21}\Lambda_1\mathbf{S}_{11}^H\mathbf{E}_{11}^{-1} & -\mathbf{E}_{11}^{-1}\mathbf{R}_{11}\Lambda_1\mathbf{S}_{21}^H\mathbf{G}^{-1} \\ -\mathbf{G}^{-1}\mathbf{R}_{21}\Lambda_1\mathbf{S}_{11}^H\mathbf{E}_{11}^{-1} & \mathbf{G}^{-1} \end{pmatrix} \\
&\begin{pmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{pmatrix} = \mathbf{V}_1\mathbf{E}_{11}^{-1}\mathbf{U}_1^H \\
&+ (\mathbf{V}_2 - \mathbf{V}_1\mathbf{E}_{11}^{-1}\mathbf{R}_{11}\Lambda_1\mathbf{S}_{21}^H)\mathbf{G}^{-1}(\mathbf{U}_2^H - \mathbf{R}_{21}\Lambda_1\mathbf{S}_{11}^H\mathbf{E}_{11}^{-1}\mathbf{U}_1^H)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{G} &= \mathbf{R}_{21}\Lambda_1\mathbf{S}_{21}^H - \mathbf{R}_{21}\Lambda_1\mathbf{S}_{11}^H\mathbf{E}_{11}^{-1}\mathbf{R}_{11}\Lambda_1\mathbf{S}_{21}^H \\
\mathbf{E}_{11} &= \Delta_1 + \mathbf{R}_{11}\Lambda_1\mathbf{S}_{11}^H
\end{aligned}
$$

Repeated application of the standard WOODBURY formula

$$
\begin{aligned}
\mathbf{E}_{11}^{-1} &= \Delta_1^{-1} - \Delta_1^{-1}\mathbf{R}_{11}\mathbf{F}_{11}^{-1}\mathbf{S}_{11}^H\Delta_1^{-1} \\
\mathbf{F}_{11} &= (\Lambda_1^{-1} + \mathbf{S}_{11}^H\Delta_1^{-1}\mathbf{R}_{11})
\end{aligned}
$$

gives after some simplification

$$
\text{(A4)} \qquad \mathbf{E}^{-1} = \mathbf{V}_1(\Delta_1^{-1} - \Delta_1^{-1}\mathbf{R}_{11}\mathbf{F}_{11}^{-1}\mathbf{S}_{11}^H\Delta_1^{-1})\mathbf{U}_1^H +
$$

$$
+(\mathbf{V}_2 - \mathbf{V}_1\Delta_1^{-1}\mathbf{R}_{11}\mathbf{F}_{11}^{-1}\mathbf{S}_{21}^H)(\mathbf{R}_{21}\mathbf{F}_{11}^{-1}\mathbf{S}_{21}^H)^{-1}(\mathbf{U}_2^H - \mathbf{R}_{21}\mathbf{F}_{11}^{-1}\mathbf{S}_{11}^H\Delta_1^{-1}\mathbf{U}_1^H)
$$

Notice that if $m_1 = m$ and $k_1 = k$, the second term of (A4) disappears, and the first one reduces to (A2).

Similarly, the partitioned formula for determinants and (A3) yield:

$$
\text{(A5)} \qquad |\mathbf{E}| = |\mathbf{E}_{11}| \cdot |\mathbf{G}| = |\Delta_1| \cdot |\Lambda_1| \cdot |\mathbf{F}_{11}| \cdot |\mathbf{G}|
$$

This generalization could be done in terms of alternative factorisations of $\mathbf{A}$ and $\mathbf{L}$, which may be more convenient if these matrices have special forms.

Since the conditional covariance matrix of conditionally heteroskedastic factor models has the form (A1), with $\mathbf{A} = \Gamma$, $\mathbf{B} = \mathbf{D} = \mathbf{C}$ and $\mathbf{L} = \Lambda_t$, the

WOODBURY formulae can be directly applied to compute the log-likelihood function and score. In this respect, it is straightforward but rather tedious to prove that if Γ has full rank, expressions (15) and (17) for $l(\mathbf{x}_t|X_{t-1};\phi)$ and $s(\mathbf{x}_t|X_{t-1};\phi)$ can be obtained algebraically by using (A2) and (A3), with $\mathbf{F} = \Lambda_t^{-1} + \mathbf{C}'\Gamma^{-1}\mathbf{C}$. Similarly, it is also possible to prove that the expressions derived in section 4.3 coincide with those obtained if we apply the modified WOODBURY formulae (A4) and (A5) instead, where $\mathbf{P} = \mathbf{Q} = \mathbf{I}_k$, $\mathbf{U} = \mathbf{V}$ is a permutation matrix that moves the series with zero $\gamma_i's$ to the last positions, and $\mathbf{G} = \Sigma_{b.at}$.