Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks

Sentana, Enrique; Calzolari, Giorgio; Fiorentini, Gabriele

Accepted Manuscript

Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks

Enrique Sentana, Giorgio Calzolari, Gabriele Fiorentini

PII: S0304-4076(08)00062-6
DOI: 10.1016/j.jeconom.2008.06.001
Reference: ECONOM 3042

To appear in: *Journal of Econometrics*

Please cite this article as: Sentana, E., Calzolari, G., Fiorentini, G., Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.06.001

Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks*

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

Giorgio Calzolari
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<calzolar@ds.unifi.it>

Gabriele Fiorentini
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<fiorentini@ds.unifi.it>

Revised: April 2008

Abstract

We derive indirect estimators of conditionally heteroskedastic factor models in which the volatilities of common and idiosyncratic factors depend on their past unobserved values by calibrating the score of a Kalman-filter approximation with inequality constraints on the auxiliary model parameters. We also propose alternative indirect estimators for large-scale models, and explain how to apply our procedures to many other dynamic latent variable models. We analyse the small sample behaviour of our indirect estimators and several likelihood-based procedures through an extensive Monte Carlo experiment with empirically realistic designs. Finally, we apply our procedures to weekly returns on the Dow 30 stocks.

Keywords: ARCH, Idiosyncratic risk, Inequality constraints, Kalman filter, Sequential estimators, Simulation estimators, Volatility.

JEL: C13, C15, C32

*Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain. Tel.: +34 91 429 0551, fax: +34 91 429 1056, e-mail: sentana@cemfi.es.

1 Introduction

Many issues in finance, including asset pricing, portfolio allocation and risk management, require the analysis of the variances and covariances of a large number of security returns. Traditionally, these issues were considered in a static framework, but the emphasis has gradually shifted to intertemporal models in which agents' actions are based on the distribution of returns conditional on their information set. Parallel to these theoretical developments, a large family of statistical models for the time variation in conditional variances has grown up following Engle's (1982) work on ARCH processes, and numerous applications have already appeared. By and large, though, most 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 sheer number of parameters involved. In this sense, it is worth mentioning that even with the massive computational power economists have available to them nowadays, the multivariate vec analogue of the ubiquitous univariate GARCH(1,1) model has been hardly ever estimated for more than two series at a time, often with many additional restrictions to ensure that the resulting conditional covariance matrices are positive definite (see Bauwens, Laurent and Rombouts (2006) for a recent survey).

Given the strong commonality in the volatility movements of different financial assets, it is not surprising that 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. That is, it is assumed that each of N observed variables is a linear combination of k ($k \ll N$) common factors plus an idiosyncratic term, but with conditional heteroskedasticity in the underlying variables. Such models are particularly appealing in finance, as the concept of factors plays a fundamental role in major asset pricing theories, such as the Arbitrage Pricing Theory of Ross (1976) (see King, Sentana and Wadhwani (1994)). In addition, they automatically guarantee a positive definite conditional covariance matrix for the observed series, once we ensure that the variances of common and specific factors are non-negative.

If the conditional variances of the latent variables are measurable functions of observed variables, as in Engle, Ng and Rothschild (1990), maximum likelihood (ML) estimation is time-consuming, but straightforward. However, one has to exercise care in dealing with conditional variances that depend on past values of the common or idiosyncratic factors, as their true values do not necessarily belong to the econometrician's information set (see Harvey, Ruiz and Sentana (1992)) (HRS). The original latent factor model with ARCH effects on the common factors introduced by Diebold and Nerlove (1989) is the best known example. In such models, the distribution and moments of the observed variables conditional on their past values alone are unknown. To some degree, this has prompted interest in other parameter driven models (see Andersen (1994) or Shephard (1996)), in which the volatility of the latent factors evolves according to discrete-time stochastic volatility processes (see Pitt and Shephard (1999), Aguilar and West (2000), Meddahi and Renault (2004), Chib, Nardari and Shephard (2006) and Doz and Renault (2006)), or discrete-state Markov chains (see Carter and Kohn (1994), Shephard (1994), Kim and Nelson (1999), and the references therein).

Despite the attractive features of these alternative models, one should not necessarily abandon the use of GARCH processes in latent variable models, especially if one takes into account that many macro and finance theories are often specified using conditional moments, although those moments are defined with respect to the agents' information set, not the econometrician's. For that reason, Fiorentini, Sentana and Shephard (2004) develop computationally feasible Markov Chain Monte Carlo (MCMC) simulation algorithms that can be used to obtain exact likelihood-based estimators of factor models with GARCH structures in the common factors, thereby avoiding the inconsistencies associated with the Kalman filter approximations to the log-likelihood function proposed by Diebold and Nerlove (1989) and HRS. However, they maintain the assumption of constant idiosyncratic variances, which seems at odds with the empirical evidence.

In this paper, we drop that assumption, and analyse alternative simulation-based estimation methods belonging to the class of indirect estimation procedures proposed by Gallant and Tauchen (1996), Gouriéroux, Monfort and Renault (1993) and Smith (1993). In fact, Gouriéroux, Monfort and Renault (1993) explicitly considered conditionally heteroskedastic factor models as one of their examples, and suggested a first order, discrete-state Markov chain as auxiliary model for the case of ARCH(1) dynamics (see also Billio and Monfort (2003), who use non-parametric auxiliary models when N is small). Our approach is more closely related to Dungey, Martin and Pagan (2000), who also developed indirect estimators for such models. Specifically, they considered two auxiliary parametric models: a "dual" VAR model for the levels and squares (but not cross products) of the observed series (see also Zhumabekova and Dungey (2001)), and the Kalman filter-based approximation to the log-likelihood function used by Diebold and Nerlove (1989). Although Dungey, Martin and Pagan (2000) found in a limited Monte Carlo exercise that the latter yields indirect estimators with substantially smaller root mean square errors than the former, they did not use it in their empirical application. Another problem with their VAR approach is that the number of parameters of the auxiliary model increases with the square of the number of asset returns, which rules out its application to large models.

In this context, our methodological contribution is twofold. First, we show that the HRS approximation is an ideal auxiliary model because (a) it has exactly the same number of parameters as the model of interest, and with the same meaning, and it is also easy to estimate;

and (b) it spans the score of the model of interest in some important limiting cases, providing a very good approximation to it in more general situations. Second, we derive an alternative joint indirect estimator on the basis of the sequential estimators of the HRS approximation in Sentana and Fiorentini (2001) (SF), which can be applied to situations with rather large cross-sectional dimensions if we add the empirically plausible assumption that the dynamic variance coefficients of the idiosyncratic terms are common. In addition, we conduct an empirically realistic Monte Carlo experiment to assess the finite sample performance of our two proposed indirect estimators relative to the approximate methods of HRS and SF. We also compare them to the Bayesian estimators of Fiorentini, Sentana and Shephard (2004) in their restricted case.

Importantly, we explain how our proposed estimators can be easily adapted to handle any state space model with GARCH disturbances, which includes many examples that have been used in the empirical economic and finance literatures, such as structural time series models, or regression models with time-varying coefficients (see Harvey (1989) and Kim and Nelson (1999)).

Finally, we apply our estimators to weekly excess returns on the 30 components of the Dow Jones Industrial Average. We also gauge the importance of allowing for time-variation in conditional correlations and idiosyncratic volatilities, as well as for non-normality of returns. In addition, our empirical results shed some light on whether the increase in idiosyncratic risk documented by Campbell et al (2001) continued after the dot-com bubble burst.

In section 2, we define the model of interest and the HRS approximation, and study their relationship in detail. Then in section 3 we introduce our two indirect estimators, and explain how they can be extended to deal with more general models. Our Monte Carlo evidence is included in section 4. Finally, the results of the empirical application are presented in section 5, and our conclusions in section 6. Proofs and auxiliary results can be found in appendices.

2 Conditionally Heteroskedastic Factor Models

2.1 Definition

Consider the following multivariate model:

$$\mathbf{x}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t, \tag{1}$$

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

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, $\mathbf{B}$ is the $N \times k$ matrix of factor loadings, with $N \geq k$ (typically $N \gg k$) and rank $(\mathbf{B}) = k$, $\mathbf{u}_t$ is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to $\mathbf{f}_t$, $\mathbf{\Delta}_t$ is a $k \times k$ diagonal positive definite (p.d.) matrix of time-varying factor variances, $\mathbf{\Psi}_t$ is a $N \times N$ diagonal positive semidefinite (p.s.d.) matrix of time-varying idiosyncratic variances, and I_{t-1} is an information set that contains the values of $\mathbf{x}_t$ and $\mathbf{f}_t$ up to $t-1$. Thus, the distribution of $\mathbf{x}_t$ conditional on I_{t-1} is $N(\mathbf{0}, \mathbf{\Sigma}_t)$, where $\mathbf{\Sigma}_t = \mathbf{B}\mathbf{\Delta}_t\mathbf{B}' + \mathbf{\Psi}_t$ has the usual exact factor structure. For this reason, we shall refer to the data generation process specified by (1) and (2) as a multivariate conditionally heteroskedastic exact factor model. To simplify the exposition, we maintain the normality assumption until the empirical application in section 5.

Such a formulation nests several models for asset returns widely used in the empirical finance literatures on asset pricing, portfolio selection, hedging and risk management. Those applications typically assume that $\mathbf{f}_t$ and $\mathbf{u}_t$ follow dynamic heteroskedastic processes, but differ in the particular specification of their conditional variances $\delta_{jt} = V(f_{jt}|I_{t-1})$ $(j = 1, \ldots, k)$ and $\psi_{it} = V(u_{it}|I_{t-1})$ $(i = 1, \ldots, N)$. For instance, Diebold and Nerlove (1989) parametrised the common factors as univariate *strong* ARCH models, in the sense of Drost and Nijman (1993). For the sake of concreteness, in this paper we study in detail the more realistic covariance stationary GARCH(1,1) formulation:

$$\delta_{jt} = \varsigma_j + \phi_j f_{jt-1}^2 + \rho_j \delta_{jt-1}, \qquad (j = 1, \ldots, k) \tag{3}$$

$$\psi_{it} = \varsigma_i^* + \phi_i^* u_{it-1}^2 + \rho_i^* \psi_{it-1}, \qquad (i = 1, \ldots, N) \tag{4}$$

with $\varsigma_j = (1 - \phi_j - \rho_j)\delta_j$, $\delta_j = E(\delta_{jt}|\boldsymbol{\varrho}) = V(f_{jt}|\boldsymbol{\varrho})$, $\varsigma_i^* = (1 - \phi_i^* - \rho_i^*)\psi_i$ and $\psi_i = E(\psi_{it}|\boldsymbol{\varrho}) = V(u_{it}|\boldsymbol{\varrho})$, where $E(.|\boldsymbol{\varrho})$ denotes expected values taken with respect to model (1)-(4) evaluated at the parameter values $\boldsymbol{\varrho} = (\mathbf{b}', \boldsymbol{\psi}', \boldsymbol{\delta}', \boldsymbol{\nu}', \boldsymbol{\nu}^*)'$, $\mathbf{b} = vec(\mathbf{B}') = (\mathbf{b}_1', \ldots, \mathbf{b}_N')'$, $\mathbf{b}_i' = (b_{i1}, \ldots, b_{ik})$, $\boldsymbol{\psi} = vecd(\mathbf{\Psi}) = (\psi_1, \ldots, \psi_N)'$, $\boldsymbol{\delta} = vecd(\mathbf{\Delta}) = (\delta_1, \ldots, \delta_k)'$, $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \ldots, \boldsymbol{\nu}_k')'$ and $\boldsymbol{\nu}^* = (\boldsymbol{\nu}_1^{*\prime}, \ldots, \boldsymbol{\nu}_N^{*\prime})'$, with $\boldsymbol{\nu}_j' = (\phi_j, \rho_j)$ and $\boldsymbol{\nu}_i^{*\prime} = (\phi_i^*, \rho_i^*)$. To eliminate the usual scale indeterminacy of the common factors, it is necessary to impose restrictions on either $\mathbf{B}$ or $\mathbf{\Delta}$. For instance, we could impose either $\delta_j = 1$,[1] or $b_{jj} = 1$ $(j = 1, \ldots, k)$. As a result, the total number of parameters to estimate, d say, will be $Nk + N + 2k + 2N$.

A non-trivial advantage of the multivariate model (1)-(2) is that it automatically guarantees a p.d. covariance matrix for $\mathbf{x}_t$ when both $\mathbf{\Delta}_t$ and $\mathbf{\Psi}_t$ are p.s.d.[2] In fact, the rank of $\mathbf{\Psi}_t$ is related to the observability of the factors. If rank$(\mathbf{\Psi}_t) = N - k$, then the factors would be fully revealed by the $\mathbf{x}_t$ variables, otherwise they are only partially revealed (see King, Sentana and Wadhwani (1994)). In this second case, note that if we premultiply $\mathbf{x}_t$ by $\mathbf{\Upsilon}\mathbf{B}'\mathbf{\Psi}^{-1}$, where $\mathbf{\Upsilon} = \left(\mathbf{B}'\mathbf{\Psi}^{-1}\mathbf{B}\right)^{-1}$, then we end up with

$$\mathbf{y}_t = \mathbf{f}_t + \mathbf{\Upsilon}\mathbf{B}'\mathbf{\Psi}^{-1}\mathbf{u}_t = \mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{5}$$

[1] See Fiorentini, Sentana and Shephard (2004) for symmetric scaling assumptions for integrated models.

[2] Sufficient conditions for the positivity of δ_{jt} in the GARCH(1,1) model (3) are $\varsigma_j > 0$, $\phi_j \geq 0$ and $\rho_j \geq 0$ (see Nelson and Cao (1992) for other higher order models), while the stationarity condition is $\phi_j + \rho_j < 1$. The same conditions apply to ς_i^*, ϕ_i^* and ρ_i^* in (4).

so that $V(\boldsymbol{\epsilon}_t|\boldsymbol{\varrho}) = \boldsymbol{\Upsilon}$. Therefore, the "signal-to-noise ratio" depends on $\mathbf{B}$ and $\boldsymbol{\Psi}$ through the magnitude of $\boldsymbol{\Upsilon}$ relative to the unconditional variance of the factors, $\boldsymbol{\Delta}$.

Finally, notice that if $\mathbf{f}_t$ and $\mathbf{u}_t$ are conditionally homoskedastic, the model in (1)-(2) reduces to the standard (i.e. static) factor analysis model (see e.g. Lawley and Maxwell (1971)). But even if $\mathbf{f}_t$ or $\mathbf{u}_t$ are conditionally heteroskedastic, provided that they are covariance stationary, it also implies an unconditional exact k factor structure for $\mathbf{x}_t$. That is, the unconditional covariance matrix, $\boldsymbol{\Sigma}$, can be written as:

$$\boldsymbol{\Sigma} = E(\boldsymbol{\Sigma}_t|\boldsymbol{\varrho}) = \mathbf{B}\boldsymbol{\Delta}\mathbf{B}' + \boldsymbol{\Psi}. \tag{6}$$

This property makes the models considered here compatible with traditional factor analysis.

From the estimation point of view, 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 $\mathbf{x}_t$. When the elements of $\boldsymbol{\Delta}_t$ are functions of lagged values of $\mathbf{f}_t$, as in (3), or the elements of $\boldsymbol{\Psi}_t$ are functions of lagged values of $\mathbf{u}_t$, as in (4), the exact form of the conditional density of $\mathbf{x}_t$ given X_{t-1} alone, $p(\mathbf{x}_t|X_{t-1};\boldsymbol{\varrho})$ say, is generally unknown. As a result, the log-likelihood function of the sample cannot be obtained explicitly (see HRS).

In the special case in which $\boldsymbol{\Psi}_t$ is time-invariant, we can employ the MCMC methods mentioned in the introduction, which allow the calculation of ML estimators via the simulated EM algorithm, as well as a Bayesian approach. However, the assumption of constant idiosyncratic variances implies that the time variation of $\boldsymbol{\Sigma}_t$ is of reduced rank (see Engle, Ng and Rothschild (1990), and Engle and Susmel (1993) for a discussion in terms of the common features literature). Therefore, all static trading strategies whose weights are orthogonal to $\mathbf{B}$, a Hilbert space of dimension $N-k$, will have constant volatility, which does not seem empirically plausible.

2.2 HRS alternative

Consider now the closely related model proposed by HRS as a user-friendly version of (1)-(4):

$$\mathbf{x}_t = \mathbf{C}\mathbf{g}_t + \mathbf{v}_t \tag{7}$$

$$\begin{pmatrix} \mathbf{g}_t \\ \mathbf{v}_t \end{pmatrix} |I_{t-1} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Lambda}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_t \end{pmatrix}\right] \tag{8}$$

where $\mathbf{g}_t$ is an alternative $k \times 1$ vector of unobserved common factors, $\mathbf{C}$ is the corresponding $N \times k$ matrix of factor loadings, with $N \geq k$ and $\text{rank}\,(\mathbf{C}) = k$, $\mathbf{v}_t$ is another $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to $\mathbf{g}_t$, $\boldsymbol{\Lambda}_t$ is a $k \times k$ diagonal p.d. matrix of time-varying factor variances, and $\boldsymbol{\Gamma}_t$ is the $N \times N$ diagonal p.s.d. matrix of time-varying idiosyncratic variances. By analogy with (3) and (4), HRS assumed that

$$\lambda_{jt}(\boldsymbol{\theta}) = \varpi_j + \alpha_j[g^2_{jt-1|t-1}(\boldsymbol{\theta}) + \omega_{jjt-1|t-1}(\boldsymbol{\theta})] + \beta_j\lambda_{jt-1}(\boldsymbol{\theta}), \tag{9}$$

$$\gamma_{it}(\boldsymbol{\theta}) = \varpi^*_i + \alpha^*_i[v^2_{it-1|t-1}(\boldsymbol{\theta}) + \xi_{iit-1|t-1}(\boldsymbol{\theta})] + \beta^*_i\gamma_{it-1}(\boldsymbol{\theta}), \tag{10}$$

with $\varpi_j = (1-\alpha_j-\beta_j)\lambda_j$, $\lambda_j = E[\lambda_{jt}(\boldsymbol{\theta})|\boldsymbol{\theta}] = V(g_{jt}|\boldsymbol{\theta})$, $\varpi_i^* = (1-\alpha_i^*-\beta_i^*)\gamma_i$ and $\gamma_i = E[\gamma_{it}(\boldsymbol{\theta})|\boldsymbol{\theta}] = V(v_{it}|\boldsymbol{\theta})$, where $E(.|\boldsymbol{\theta})$ represents expected values taken with respect to the model (7)-(10) evaluated at the parameter values $\boldsymbol{\theta} = (\mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\lambda}', \boldsymbol{\pi}', \boldsymbol{\pi}^{*\prime})'$, $\mathbf{c} = vec(\mathbf{C}') = (\mathbf{c}_1', \ldots, \mathbf{c}_N')'$, $\mathbf{c}_i = (c_{i1}, \ldots, c_{ik})'$, $\boldsymbol{\gamma} = vecd(\boldsymbol{\Gamma}) = (\gamma_1, \ldots, \gamma_N)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)'$, $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \ldots, \boldsymbol{\pi}_k')'$ and $\boldsymbol{\pi}^* = (\boldsymbol{\pi}_1^{*\prime}, \ldots, \boldsymbol{\pi}_N^{*\prime})'$, with $\boldsymbol{\pi}_j = (\alpha_j, \beta_j)'$, $\boldsymbol{\pi}_i^* = (\alpha_i^*, \beta_i^*)'$, and where $g_{jt|t}(\boldsymbol{\theta})$, $v_{it|t}(\boldsymbol{\theta})$, $\omega_{jlt|t}(\boldsymbol{\theta})$ and $\xi_{ilt|t}(\boldsymbol{\theta})$ are typical elements of the outputs of the Kalman filter updating equations (see Appendix B). Intuitively, the HRS trick is to replace the squares of the lag unobserved factors in (3) and (4) with their conditional expectations given the observables.

Not surprisingly, models (1)-(4) and (7)-(10) share many important features, including that the number of parameters to estimate after setting the scaling of $\mathbf{g}_t$ will also be $Nk + N + 2k + 2N = d$. Similarly, the positive definiteness of the conditional covariance of $\mathbf{x}_t$ is guaranteed if $\lambda_{jt}(\boldsymbol{\theta}) \geq 0$ $\forall j$ and $\gamma_{it}(\boldsymbol{\theta}) \geq \mathbf{0}$ $\forall i$. But if we rewrite (9) as an infinite distributed lag, then it is clear that $\varpi_j > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$ and $\alpha_j + \beta_j < 1$ ensure positivity and stationarity of $\lambda_{jt}(\boldsymbol{\theta})$. Likewise, $\varpi_i^* > 0$, $\alpha_i^* \geq 0$, $\beta_i^* \geq 0$, and $\alpha_i^* + \beta_i^* < 1$ achieve the same goals for $\gamma_{it}(\boldsymbol{\theta})$.

The crucial difference with the model discussed in the previous section is that the distribution of $\mathbf{x}_t$ conditional in X_{t-1} is normal, with conditional covariance matrix $\mathbf{C}\boldsymbol{\Lambda}_t\mathbf{C}' + \boldsymbol{\Gamma}_t$. As a result, the log-likelihood function for a sample of size T based on the model (7)-(10) can be written in closed form. However, we must use constrained optimisation algorithms to obtain the pseudo-ML estimators $\hat{\boldsymbol{\theta}}_T$ because of the inequality restrictions on $\boldsymbol{\gamma}$ and the GARCH parameters $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^*$ mentioned above. An additional complication that (9) shares with more standard GARCH(1,1) formulations is that if $\alpha_j = 0$, then $\lambda_{jt} = \lambda_j$ irrespective of β_j, which means that $\partial l_t(\boldsymbol{\theta})/\partial \beta_j = 0$ for all t, so that β_j cannot be identified. Similarly, β_i^* cannot be identified either when $\alpha_i^* = 0$.

2.3 Relationship between the original model and the HRS alternative

Both processes yield exactly the same log-likelihood function when $\phi_j = \alpha_j = 0$ $\forall j$ and $\phi_i^* = \alpha_i^* = 0$ $\forall i$, and when there are exactly as many Heywood cases as common factors, so that the factors are fully revealed by the observable variables. In addition, Sentana (2004) shows that $\mathbf{g}_{t|t}(\boldsymbol{\theta})$ converges in mean square to $\mathbf{g}_t$ as N increases, which means that the two models become arbitrarily close in practice for large enough N because in that case $\boldsymbol{\Upsilon}$ will be very small.

The similarity between the two models is stronger than what the preceding discussion may suggest, as the HRS model smoothly embeds (in the sense used by Gallant and Tauchen (1996) in their Theorem 2) the original model in those circumstances:

Proposition 1 *Let $\mathbf{s}_t(\boldsymbol{\theta}) = \partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ denote the log-likelihood score of the model for $\mathbf{x}_t$ given by (7)-(10). Similarly, let $\mathbf{q}_t(\boldsymbol{\varrho}) = \partial p(\mathbf{x}_t|X_{t-1};\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}$ denote the exact log-likelihood score of the model (1)-(4). Then $\mathbf{s}_t(\boldsymbol{\theta}) = \mathbf{q}_t(\boldsymbol{\varrho})$ when $\boldsymbol{\theta} = \boldsymbol{\varrho}$, $\boldsymbol{\nu} = \boldsymbol{\pi} = \mathbf{0}$ and $\boldsymbol{\nu}^* = \boldsymbol{\pi}^* = \mathbf{0}$.*

In other words, the HRS model provides a first-order approximation to the original model when δ_{jt} and ψ_{it} do not vary much. In contrast, Proposition 1 does not hold if we consider instead the approximation put forward by Diebold and Nerlove (1989) and used by Dungey, Martin and Pagan (2000), which omits the terms $\omega_{jjt-1|t-1}(\boldsymbol{\theta})$ and $\xi_{iit-1|t-1}(\boldsymbol{\theta})$ from (9) and (10), respectively, even though its log-likelihood function at $\boldsymbol{\pi} = \boldsymbol{\pi}^* = \mathbf{0}$ coincides.

We conjecture that an analogous result holds when the factors are fully revealed by the observable variables. Unfortunately, we have been unable to obtain analytical expressions for the elements of $\mathbf{q}_t(\boldsymbol{\varrho})$ corresponding to the zero idiosyncratic variances for the reasons explained at the end of the proof of Proposition 1. Nevertheless, extensive numerical simulations suggest that the scores are identical in that situation too.[3]

In other cases, the exact relationship between the two models can only be assessed by simulation. For practical purposes, the relevant comparison is between model (7)-(10) evaluated at the pseudo-true ML values, $\boldsymbol{\theta}(\boldsymbol{\varrho}^0)$ say (see e.g. White (1982)), and model (1)-(4) evaluated at the true values $\boldsymbol{\varrho}^0$. To find out the (asymptotic) binding functions $\boldsymbol{\theta}(\boldsymbol{\varrho})$, we have simulated 400 samples of size 50,000 of the trivariate single factor model $x_{it} = b_i f_t + u_{it}$, $(i = 1, 2, 3)$. We assume that $\phi_i^* = \rho_i^* = 0$ to concentrate on the dynamic variance parameters of the common factor. Given that the effect of $\mathbf{b}$ and $\boldsymbol{\Psi}$ is mostly through the scalar $\upsilon = (\mathbf{b}'\boldsymbol{\Psi}^{-1}\mathbf{b})^{-1}$, we set $\mathbf{b} = (1, 1, 1)'$, $\boldsymbol{\Psi} = \psi\mathbf{I}$ and $\delta_t = (1 - \phi - \rho) + \phi f_{t-1}^2 + \rho\delta_{t-1}$. To minimise experimental error, we use the same underlying random numbers in all designs. For scaling purposes, we set $c_3 = 1$ and unrestrict λ, the unconditional variance of g_t. We used the NAG library E04JBL routine to maximise the log-likelihood of the auxiliary model with respect to $\boldsymbol{\theta} = (c_1,c_2, \gamma_1, \gamma_2, \gamma_3, \lambda, \alpha, \beta)$.

Although we combined ten different values of ψ with ten different pairs of $(\phi,\ \rho)$, we only report the results of four selected configurations, namely: (.2,.6), (.4,.4), (.1,.85) and (.2,.75). The first one corresponds roughly to the values obtained by estimating univariate GARCH models on the basis of monthly data, while the third one to the values obtained with weekly observations. The relevant response surfaces for ϕ and ρ are depicted in the two panels of Figure 1, respectively. Our results suggest that $\alpha(\boldsymbol{\varrho})$ is bigger than ϕ and $\beta(\boldsymbol{\varrho})$ lower than ρ in an attempt to match the static and dynamic fourth moments generated by the HRS approximate model with the true static and dynamic fourth moments of the process, and in this way, provide a better approximation to the true but unknown $V(\mathbf{x}_t|X_{t-1}; \boldsymbol{\varrho}^0)$.[4] At the same time, the asymptotic

[3] Specifically, we have numerically checked the information matrix equality-type efficiency condition $\partial E[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}' = V[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]$ when $\boldsymbol{\theta} = \boldsymbol{\varrho}$ and the factors are effectively observed, where $E[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]$ and $V[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]$ refer to the first two moments of the pseudo-score with respect to the true distribution of the data.

[4] These biases contrast with those in Nijman and Sentana (1996). Given that y_t is the sum of a strong GARCH process, f_t, whose squares follow an ARMA(1,1) process with AR coefficient $\phi + \rho$ and MA coefficient ρ, and an independent white noise process, η_t, their results show that y_t^2 will also follow an ARMA(1,1) process with the same AR coefficient, but with an MA coefficient which is an explicit function of ϕ, ρ, and the unconditional

bias $[\alpha(\boldsymbol{\varrho}) + \beta(\boldsymbol{\varrho})] - (\phi + \rho)$, although not exactly zero, is fairly small. However, this bias can be substantial for other parameter configurations. The most important determinant of the asymptotic biases $\alpha(\boldsymbol{\varrho}) - \phi$ and $\beta(\boldsymbol{\varrho}) - \rho$ is the noise to signal ratio, as measured by the variance of ϵ_t relative to the variance of f_t. However, the second most important determinant of the biases is not the so-called persistence parameter, $\phi + \rho$, but rather, the unconditional coefficient of variation of the unobserved conditional variance δ_{jt}, i.e.

$$\varkappa^2(\boldsymbol{\varrho}) = \frac{V(\delta_t|\boldsymbol{\varrho})}{\delta^2} = \frac{2\phi^2}{(1 - 3\phi^2 - 2\phi\rho - \rho^2)},$$

(cf. Jacquier, Polson and Rossi (1994)).

In order to assess how closely the HRS model approximates the model of interest, we have generated realisations of the portfolio y_t in (5) for the trivariate single factor models discussed above, and compared the Gaussian distribution of y_t given Y_{t-1} and $\boldsymbol{\varrho}^0$ implied by the HRS model with the true conditional distribution. The former is simply $N\{0, \lambda_t[\boldsymbol{\theta}(\boldsymbol{\varrho}^0)] + tr[\boldsymbol{\Gamma}(\boldsymbol{\varrho}^0)]/N^2\}$. To obtain the latter, we have simulated $y_t|Y_{t-1}; \boldsymbol{\varrho}^0$ by drawing 100,000 random numbers from a Gaussian distribution with 0 mean and variance $\delta_t + \upsilon^0$, where δ_t had been previously drawn from its distribution conditional on Y_{t-1} and $\boldsymbol{\varrho}^0$ by using an exact MCMC sampler. Although the resulting distribution will be necessarily leptokurtic because it is a scale mixture of normals, the degree of leptokurtosis will depend on the variability of $\delta_t + \upsilon^0$ given Y_{t-1} and $\boldsymbol{\varrho}^0$, which would be 0 if either δ_t is constant or $\upsilon^0 = 0$. The results of such a comparison clearly show that the approximate Kalman filter provides extremely reliable results.

We have also computed the probability integral transform (PIT) with respect to the approximate model $N\{0, \lambda_t[\boldsymbol{\theta}(\boldsymbol{\varrho}^0)] + tr[\boldsymbol{\Gamma}(\boldsymbol{\varrho}^0)]/N^2\}$ of the aforementioned realisations of the equally weighted portfolio y_t. If the conditional distribution of y_t given Y_{t-1} and $\boldsymbol{\varrho}^0$ were indeed Gaussian, such PIT sequences would be independently and identically distributed uniformly between 0 and 1 (see e.g. Diebold, Gunther and Tay (1998)). We find that the maximum difference between the empirical cumulative distribution of 4,000,000-long PIT sequences and the 45° degree line are very small (<.005), with the approximate Kalman filter providing more reliable results the closer the unconditional distribution of the latent factors is to the normal ($\phi = .2$, $\rho = .6$), and the larger the signal to noise ratio ($\upsilon^0 = 1/9$). Therefore, given that in many empirical applications the signal to noise ratio is likely to be high, and the conditional variance a fairly smooth process, we would expect the model proposed by HRS to provide a very accurate approximation to the conditional distribution of the observations given their past values.

variances of f_t and η_t. Given that the size of the MA coefficient of f_t^2 unequivocally increases as a result of adding uncorrelated noise, their results also imply that the weak GARCH parameter would become larger, while the weak ARCH parameter would become smaller.

3 Indirect estimation

3.1 A natural indirect estimator

The impossibility of writing down the log-likelihood function of the model of interest in closed-form, combined with the ease with which we can simulate from it, suggest that the indirect estimation procedures of Gallant and Tauchen (1996), Gouriéroux, Monfort and Renault (1993) and Smith (1993) should be ideally suited for our case (see also Gouriéroux and Monfort (1996)). In this context, the approximate model proposed by HRS is an ideal candidate for auxiliary model. On the one hand, it has exactly the same number of parameters as the model of interest, and with the same meaning, and it is also easy to estimate, especially if one exploits the fast and numerically reliable procedure for the computation of its score in Appendix B. On the other hand, and unlike the Diebold and Nerlove (1989) extended Kalman filter considered by Dungey, Martin and Pagan (2000), model (7)-(10) spans the score of the model of interest in some important limiting cases (see Proposition 1), and it provides a rather good approximation to it in more general situations, as the discussion at the end of section 2.3 shows.

Given that the HRS auxiliary model must be estimated subject to some inequality constraints, which are sometimes binding, we must use the constrained indirect estimation procedures proposed in our earlier work (see Calzolari, Fiorentini and Sentana (2004)), which can handle a mix of equality and inequality restrictions on $\boldsymbol{\theta}$. More specifically, the Generalised Method of Moments (GMM) version of the inequality constrained indirect estimator of $\boldsymbol{\varrho}$ is

$$\tilde{\boldsymbol{\varrho}}(\mathbf{P}) = \arg\min_{\boldsymbol{\varrho}} \mathfrak{m}'(\boldsymbol{\varrho};\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\mu}}) \cdot \mathbf{P} \cdot \mathfrak{m}(\boldsymbol{\varrho};\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\mu}}), \quad (11)$$

$$\mathfrak{m}(\boldsymbol{\varrho};\boldsymbol{\theta},\boldsymbol{\mu}) = E\left[\bar{m}_T(\boldsymbol{\theta},\boldsymbol{\mu})|\,\boldsymbol{\varrho}\right] = E\left[\bar{s}_T(\boldsymbol{\theta})|\,\boldsymbol{\varrho}\right] + \left[\partial h'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]\boldsymbol{\mu}, \quad (12)$$

where $\mathbf{P}$ is a p.d. weighting matrix of order d, $\bar{s}_T(\boldsymbol{\theta}) = \partial\bar{l}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ the average score of the auxiliary model, $\boldsymbol{\mu}$ the Kuhn-Tucker multipliers associated with the s constraints implicitly characterised by the restrictions $h(\boldsymbol{\theta}) \geq \mathbf{0}$ described in section 2.2, and ˆ indicates constrained pseudo ML estimates. Since it is often impossible to obtain $\mathfrak{m}(\boldsymbol{\varrho};\boldsymbol{\theta},\boldsymbol{\mu})$ in closed form, we can exploit the strict stationarity and ergodicity of $\mathbf{x}_t$ to approximate arbitrarily well the required expectations by their sample analogues in a single but very long realisation of the process $\{\mathbf{x}_n(\boldsymbol{\varrho}), n = 1,\ldots,T\cdot H\}$, as explained by Gallant and Tauchen (1996). In particular:

$$\mathfrak{m}(\boldsymbol{\varrho};\boldsymbol{\theta},\boldsymbol{\mu}) \simeq \frac{1}{T\cdot H}\sum_{n=1}^{T\cdot H}\partial\ln f[\mathbf{x}_n(\boldsymbol{\varrho})|X_{n-1}(\boldsymbol{\varrho});\boldsymbol{\theta}]/\partial\boldsymbol{\theta} + \left[\partial h'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]\boldsymbol{\mu} = \mathfrak{m}_{TH}(\boldsymbol{\varrho};\boldsymbol{\theta},\boldsymbol{\mu}).$$

But since the term $[\partial h'(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}]\hat{\boldsymbol{\mu}}$ is fixed across simulations, the only thing we need to do is to minimise with respect to $\boldsymbol{\varrho}$ the distance between $\bar{s}_T(\hat{\boldsymbol{\theta}})$, which is no longer zero if the some of the constraints on $\hat{\boldsymbol{\theta}}$ are binding, and the average score in the simulated samples.

Another problem with the GARCH(1,1)-type formulation in (9) is that $\lambda_{jt} = \lambda_j$ when $\alpha_j = 0$ irrespective of β_j, which means that $\partial l_t(\boldsymbol{\theta})/\partial \beta_j = 0$ for all j and t. Similarly, $\gamma_{it} = \gamma_i$ when $\alpha_i^* = 0$ irrespective of β_i^*, so $\partial l_t(\boldsymbol{\theta})/\partial \beta_i^* = 0$ for all i and t. However, if both $\hat{\alpha}_j = 0\ \forall j$ and $\hat{\alpha}_i^* = 0\ \forall i$, we know from Proposition 1 that the binding function is the identity matrix, and the approximate and true scores coincide, which means that we can safely set $\tilde{\boldsymbol{\varrho}} = \hat{\boldsymbol{\theta}}$. In contrast, if $\hat{\alpha}_j$ is strictly positive but very small, then β_j will be very poorly estimated, which in turn implies that there will be very little information in the auxiliary model about ρ_j. In those cases, we re-estimate the auxiliary model subject to the additional restriction $\alpha_j = \alpha_j^{\min} > 0$, where $\alpha_j^{\min}$ is some small, but not too small, number. Finally, since we are assuming a covariance stationary auxiliary model, then we will also impose the restriction $\alpha_j + \beta_j \leq (\alpha_j + \beta_j)^{\max} \leq 1$.[5] We follow an analogous procedure for α_i^* and β_i^*.

Given that the number of auxiliary model parameters coincides with the number of parameters of the model of interest, d, $\boldsymbol{\varrho}$ is exactly identified from the moments implied by the modified first-order conditions of the auxiliary model. Therefore, when the inequality constraints on $\boldsymbol{\varrho}$ do not bind, the indirect estimators are numerically invariant (for large enough T) to the choice of weighting matrix, $\mathbf{P}$. Nevertheless, since the inequality restrictions that apply to $\boldsymbol{\varrho}$ occasionally bind, the choice of $\mathbf{P}$ could matter. In those circumstances, we optimally choose $\mathbf{P}$ to be equal to the inverse of the asymptotic covariance matrix of $\sqrt{T}\bar{m}_T(\boldsymbol{\theta}^0, \boldsymbol{\mu}^0)$, which can be consistently estimated by means of standard techniques.[6]

3.2 A numerically efficient indirect estimator for large models

The practical implementation of the indirect estimator described in the previous section in large-scale multivariate models must overcome two problems. First, the estimation of the pseudo-ML parameters $\boldsymbol{\theta}$ can become a time consuming procedure when N is quite large because the dimension of the auxiliary parameter space d grows linearly with N. Second, the indirect estimation of $\boldsymbol{\varrho}$ is even more time consuming because the dimension of the true parameter space is exactly the same and the d moment functions $\mathfrak{m}(\boldsymbol{\varrho}; \boldsymbol{\theta}, \boldsymbol{\mu})$ must be computed by simulation.

For that reason, we propose an alternative *joint* indirect estimator of $\boldsymbol{\varrho}$ in which we only need to numerically minimise the GMM criterion function over a subset of this parameter vector. The trick is to replace the HRS estimators of the auxiliary model parameters $\mathbf{c}$, $\boldsymbol{\gamma}$, $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^*$ by the sequential estimators proposed by SF, which are related to Engle and Mezrich (1996)

[5] After some Monte Carlo experimentation, we recommend $\alpha_j^{\min} = .05$ and $(\alpha_j + \beta_j)^{\max} = .999$. Importantly, note that these choices do not impair the consistency of the constrained indirect estimators of $\boldsymbol{\varrho}$.

[6] By using this optimal weighting matrix we ensure that the objective function is evenly scaled across parameters, which improves the numerical properties of the optimisation algorithm even in the typical case in which the inequality restrictions on $\boldsymbol{\varrho}$ do not bind. But if its condition number were too big, it would be convenient to regularise the estimated asymptotic covariance matrix of $\sqrt{T}\bar{m}_T(\boldsymbol{\theta}^0, \boldsymbol{\mu}^0)$ by adding a positive scalar matrix.

variance targeting approach, but imposing the unconditional factor structure (6). Specifically, SF estimate the Nk factor loadings and the N unconditional idiosyncratic variances by means of standard factor analytic routines, for which the EM algorithm provides good initial values (see Rubin and Thayer (1982)). Under certain regularity conditions, SF show that

$$(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}) = \arg\max_{\mathbf{b},\boldsymbol{\psi}} \sum_{t=1}^{T} p(\mathbf{x}_t|X_{t-1}; \mathbf{b}, \boldsymbol{\psi}, \mathbf{0}, \mathbf{0}) = \arg\max_{\mathbf{c},\boldsymbol{\gamma}} \sum_{t=1}^{T} l_t(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0}, \mathbf{0})$$

are asymptotically normally distributed around $\mathbf{b}^0$ and $\boldsymbol{\psi}^0$ when $\boldsymbol{\Psi}$ and $\mathbf{B}$ are identified from unconditional moments. On this basis, they define sequential estimators of the auxiliary model conditional variance parameters $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^*$ as $(\ddot{\boldsymbol{\pi}}, \ddot{\boldsymbol{\pi}}^*) = \arg\max_{\pi,\pi^*} \bar{l}_T(\mathbf{c} = \dot{\mathbf{b}}; \boldsymbol{\gamma} = \dot{\boldsymbol{\psi}}; \boldsymbol{\pi}, \boldsymbol{\pi}^*)$.

Assuming for ease of exposition that there are no binding constraints on either $\dot{\boldsymbol{\psi}}$, $\ddot{\boldsymbol{\pi}}$ or $\ddot{\boldsymbol{\pi}}^*$, the relevant first-order conditions that characterise these sequential estimators of $\boldsymbol{\theta}$ are:

$$\left[\begin{array}{cccc} \bar{s}'_{\mathbf{c}T}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) & \bar{s}'_{\boldsymbol{\gamma}T}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) & \bar{s}'_{\boldsymbol{\pi}T}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}, \ddot{\boldsymbol{\pi}}^*) & \bar{s}'_{\boldsymbol{\pi}^*T}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}, \ddot{\boldsymbol{\pi}}^*) \end{array}\right] = \mathbf{0}', \qquad (13)$$

where $\bar{s}_{\mathbf{c}T}(\boldsymbol{\theta})$, $\bar{s}_{\boldsymbol{\gamma}T}(\boldsymbol{\theta})$, $\bar{s}_{\boldsymbol{\pi}T}(\boldsymbol{\theta})$ and $\bar{s}_{\boldsymbol{\pi}^*T}(\boldsymbol{\theta})$ are the average pseudo log-likelihood scores corresponding to $\mathbf{c}$, $\boldsymbol{\gamma}$, $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^*$, respectively. Importantly, note that both $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^*$ must be set to 0 in the first two blocks of first-order conditions, and to $\ddot{\boldsymbol{\pi}}$ and $\ddot{\boldsymbol{\pi}}^*$ in the last two ones.

But the results in Appendix B of SF imply that we can compute the expected values of the scores for $\mathbf{c}$ and $\boldsymbol{\gamma}$ without resorting to simulations as

$$\begin{aligned} \mathfrak{m}_{\mathbf{c}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) &= vec[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}(\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}} - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}}], \\ \mathfrak{m}_{\boldsymbol{\gamma}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) &= \frac{1}{2} vecd[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}(\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1} - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}]. \end{aligned}$$

Then the consistency of $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$ implies that

$$\left[\begin{array}{c} \mathfrak{m}_{\mathbf{c}}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \boldsymbol{\nu}, \boldsymbol{\nu}^*; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) \\ \mathfrak{m}_{\boldsymbol{\gamma}}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \boldsymbol{\nu}, \boldsymbol{\nu}^*; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0}) \end{array}\right] = \mathbf{0}$$

regardless of the values of the common and idiosyncratic GARCH parameters $\boldsymbol{\nu}$ and $\boldsymbol{\nu}^*$ in the true model, as long as these moments are well defined. As a result, the binding functions for the factor loadings and unconditional idiosyncratic variances, which constitute a large fraction of the parameters, are given by the identity mappings $\mathbf{c}(\boldsymbol{\varrho}) = \mathbf{b}$ and $\boldsymbol{\gamma}(\boldsymbol{\varrho}) = \boldsymbol{\psi}$. This, together with the common dimension of $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$, implies that the *joint* indirect estimators of $\mathbf{b}$ and $\boldsymbol{\psi}$ based on (13) will in fact coincide with $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$. In this sense, note that although $s_{\mathbf{c}t}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0})$ and $s_{\boldsymbol{\gamma}t}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}, \mathbf{0})$ are serially correlated, which will require the use of a computationally costly HAC-estimator of the variance of their sample average, the fact that the auxiliary model exactly identifies $\boldsymbol{\varrho}$ renders the GMM weigthing matrix $\mathbf{P}$ irrelevant.

Thus, the only task left is to minimise with respect to $\boldsymbol{\nu}$ and $\boldsymbol{\nu}^*$ any norm of

$$\left[\begin{array}{c} \mathfrak{m}_{\boldsymbol{\pi}TH}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \boldsymbol{\nu}, \boldsymbol{\nu}^*; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}, \ddot{\boldsymbol{\pi}}^*) \\ \mathfrak{m}^*_{\boldsymbol{\pi}TH}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \boldsymbol{\nu}, \boldsymbol{\nu}^*; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}, \ddot{\boldsymbol{\pi}}^*) \end{array}\right].$$

As a result, the parameter space over which we must numerically conduct indirect estimation has been reduced in $(k+1)N$ dimensions, despite the fact that this second indirect estimator is indeed minimising the GMM criterion (11) over the whole parameter vector $\boldsymbol{\varrho}$.

Although the dimension of $\boldsymbol{\nu}$ does not depend on N, the number of parameters in $\boldsymbol{\nu}^*$ still grows linearly with N when there are no restrictions on ϕ_i^* and ρ_i^*. However, given that estimated GARCH coefficients typically lie in a fairly narrow range, it makes sense to deal with this incidental parameter problem by imposing the cross-sectional homogeneity restriction $\phi_i^* = \phi^*$ and $\rho_i^* = \rho^*$ $\forall i$ by analogy with the dynamic panel data literature. Such an assumption will allow us to consider models with large cross-sectional dimensions, as sections 4 and 5 illustrate.

3.3 Extensions to general state space models

Many of the issues discussed so far are relevant for any unobserved component time series model with GARCH disturbances. Specifically, consider the following state-space model:

$$\left.\begin{array}{l} \mathbf{x}_t = \mathbf{B}(\mathbf{z}_t;\boldsymbol{\varrho})\mathbf{w}_t + \mathbf{e}(\mathbf{z}_t;\boldsymbol{\varrho}) + \mathbf{u}_t, \\ \mathbf{w}_t = \mathbf{A}(\mathbf{z}_t;\boldsymbol{\varrho})\mathbf{w}_{t-1} + \mathbf{d}(\mathbf{z}_t;\boldsymbol{\varrho}) + \mathbf{f}_t, \end{array}\right\} \tag{14}$$

where $\mathbf{w}_t$ is a vector of k unobserved state variables, $\mathbf{z}_t$ are m weakly exogenous or lagged dependent explanatory variables, $\mathbf{A}(\mathbf{z}_t;\boldsymbol{\varrho})$ the $k \times k$ transition matrix of the state variables, $\mathbf{B}(\mathbf{z}_t;\boldsymbol{\varrho})$ an $N \times k$ matrix that links observed variables and unobserved states, $\mathbf{d}(\mathbf{z}_t;\boldsymbol{\varrho})$ and $\mathbf{e}(\mathbf{z}_t;\boldsymbol{\varrho})$ are k and N regression functions, and the remaining elements are as in (1)-(4).

This formulation is very general (see Kim and Nelson (1999)), and covers many models that have been extensively used in the empirical economic and finance literatures, such as dynamic factor models (see Dungey, Martin and Pagan (2000)) or conditionally heteroskedastic generalisations of the structural time series models in Harvey (1989) (see Broto and Ruiz (2006)). It also includes conditionally heteroskedastic versions of traditional time-varying regression coefficient models, such as the GARCH in mean model put forward by Chou, Engle and Kane (1992), or the US monetary growth function considered by Kim and Nelson (1999).

Given that HRS devoted their paper to the general state space model (14), it is straightforward to extend the indirect estimator considered in section 3.1 to this more general context by using as auxiliary model the general conditionally Gaussian likelihood approximation that they proposed. Analogously, it is also straightforward to extend the numerically efficient indirect estimator discussed in section 3.2 to model (14) because as HRS explain, any unknown parameters in $\mathbf{A}(\mathbf{z}_t;\boldsymbol{\varrho})$, $\mathbf{B}(\mathbf{z}_t;\boldsymbol{\varrho})$, $\mathbf{d}(\mathbf{z}_t;\boldsymbol{\varrho})$ and $\mathbf{e}(\mathbf{z}_t;\boldsymbol{\varrho})$, together with the unconditional covariance matrices of the disturbances $\mathbf{u}_t$ and $\mathbf{f}_t$, will be estimated consistently if the standard linear Kalman filter is applied to (14) ignoring the GARCH effects in those disturbances.

4 Monte Carlo Evidence

4.1 Designs and main results

In this section, we assess the finite sample performance of the two indirect estimators that we have proposed in the previous section relative to the approximate ML methods of HRS and SF on which they are based. We consider two cross-sectional sample sizes, $N = 3$ and $N = 30$, the second of which matches the one in our empirical application.

We have used the NAG library G05DDF routine to generate 1,600 samples of 1,000 observations each (plus another 100 for initialisation) of single factor models. This sample size corresponds roughly to 20 years of weekly data, as in our empirical analysis, or 4 years of daily data. Since the performance of the different estimators depends on $\mathbf{b}$ and $\mathbf{\Psi}$ mostly through $\upsilon = (\mathbf{b}'\mathbf{\Psi}^{-1}\mathbf{b})^{-1}$, all the examples of the DGP in (1)-(4) considered are of the form:

$$\begin{aligned} x_{it} &= b_i f_t + u_{it} \qquad (i = 1, \ldots, N) \\ \delta_t &= (1 - \phi - \rho)\delta + \phi f_{t-1}^2 + \rho\delta_{t-1} \\ \psi_{it} &= \psi_i(1 - \phi_i^* - \rho_i^*) + \phi_i^* u_{it-1}^2 + \rho_i^*\psi_{it-1} \end{aligned}$$

with $\delta = 1$, $b_i = 1$, $\psi_i = \psi$, $\phi_i^* = \phi^*$ and $\rho_i^* = \rho^*$ $\forall i$. Therefore, the values of ϕ, ρ, ϕ^*, ρ^* and ψ fully characterise all our Monte Carlo designs. In this version we only report results for (ϕ, ρ) equal to (.1, .85) and (.2, .75) and exactly the same values for (ϕ^*, ρ^*). As for ψ, we have set $\psi = N/9$ and N. In order to understand such values, note that if we regress the y_t in (5), which due to our balanced design is simply the equally weighted portfolio of the $x'_{it}s$, on the latent factor f_t, we would obtain theoretical R^2 coefficients of .9, and .5, respectively, regardless of the cross-sectional dimension N. All in all, we have considered 4 different parameter configurations. However, we use the same underlying pseudo-random numbers in all designs to minimise experimental error.[7]

When we estimate the HRS auxiliary model (7)-(10) that corresponds to the DGP described in the previous paragraph, we set $c_N = 1$ for scaling purposes, and leave λ free. We also set λ_1 and γ_{i1} to the unconditional variance of the respective factors to start up the recursions. Finally, we impose the restriction $\alpha_i^* = \alpha^*$ and $\beta_i^* = \beta^*$ $\forall i$ in estimation by analogy with what we do in the empirical section. In order to guarantee the positivity and stationarity restrictions, we first optimise the pseudo log-likelihood function in terms of some unrestricted parameters $\boldsymbol{\theta}^\dagger$, where $\lambda = (\lambda^\dagger)^2$, $\gamma_i = (\gamma_i^\dagger)^2$ $(i = 1, \ldots, N)$, $\alpha = \dot{1}\sin^2(\alpha^\dagger)$, $\beta = (\dot{1} - \alpha)\sin^2(\beta^\dagger)$, $\alpha^* = \dot{1}\sin^2(\alpha^{*\dagger})$ and $\beta^* = (\dot{1} - \alpha^*)\sin^2(\beta^{*\dagger})$, where $\dot{1} = .999$ acts as our effective upper bound on

[7] In their Monte Carlo experiments, Dungey, Martin and Pagan (2000) considered a single signal to noise ratio and two conditional variance parameter configurations: $\phi = \rho = 0$, and $\phi = .2$ and $\rho = .7$.

$\alpha+\beta$ and $\alpha^*+\beta^*$. Then, we compute the score in terms of the $2N+4$ original parameters $\boldsymbol{\theta} = (c_1,\ldots,c_{N-1},\gamma_1,\ldots,\gamma_N,\lambda,\alpha,\beta,\alpha^*,\beta^*)$ using the analytical expressions derived in Appendix B to avoid numerical errors, and introduce one multiplier for each of the first order conditions, which take away any slack left.[8] As we explained before, though, if the maximum of the log-likelihood function happens at $\alpha=0$ and $\alpha^*=0$, then there is no need to resort to indirect estimation in view of Proposition 1, and we simply set $\tilde{\boldsymbol{\varrho}} = \hat{\boldsymbol{\theta}}$. If, on the other hand, the pseudo ML estimate of α is strictly positive but less than $\alpha^{\min}$, then we re-estimate the auxiliary model subject to the restriction $\alpha=\alpha^{\min}$. We use an analogous procedure for α^*.

Since there are no closed-form expressions for the expected value of the modified score, we compute them on the basis of single path simulations of length TH. In order to reduce the estimation error, we choose $H=100$, which implies that all the required moments have been effectively computed on the basis of 100,000 simulated observations. An even larger value of H should in theory slightly reduce the Monte Carlo variability of the indirect estimators according to the relation $(1+H^{-1})$, but at the cost of a significant increase in the computational burden. Then, we obtain constrained indirect estimators of the $2N+4$ parameters of interest in $\boldsymbol{\varrho} = (b_1,\ldots,b_{N-1},\psi_1,\ldots,\psi_N,\varsigma,\phi,\rho,\phi^*,\rho^*)$ by numerically minimising the GMM criterion function (11) in terms of some unrestricted parameters $\boldsymbol{\varrho}^\dagger$, with b_i free for $i<N$, $b_N=1$, $\varsigma=(\varsigma^\dagger)^2$, $\psi_i=(\psi_i^\dagger)^2$ $(i=1,\ldots,N)$, $\phi=\ddot{1}\sin^2(\phi^\dagger)$, $\rho=(\ddot{1}-\phi)\sin^2(\rho^\dagger)$, $\phi^*=\ddot{1}\sin^2(\phi^{*\dagger})$, and $\rho^*=(\ddot{1}-\phi^*)\sin^2(\rho^{*\dagger})$, where $\ddot{1}=.999999$, so as to ensure that $\delta\geq 0$, $\psi_i\geq 0\ \forall i$, $\phi,\rho,\phi^*,\rho^*\geq 0$, $\phi+\rho\leq 1$ and $\phi^*+\rho^*\leq 1$. In contrast, in the case of the alternative indirect estimator described in section 3.2, the numerical minimisation takes place over $\phi^\dagger,\rho^\dagger,\phi^{\dagger*}$ and $\rho^{\dagger*}$ only, as we set $\dot{b}_i=\dot{c}_i$ for $i<N$, $\dot{\psi}_i=\gamma_i\ \forall i$ and $\varsigma=\dot{\lambda}(1-\phi-\rho)$.

Given that the auxiliary model fits the simulated data rather well, in the sense that the relevant scores of the auxiliary model are essentially a vector martingale difference sequence, we have estimated the optimal weighting matrix as the variance in the original data of the elements of the modified score of the auxiliary model evaluated at the pseudo-ML parameter estimates. Note that by including a multiplier in each first order condition, we automatically centre the scores around their sample mean. Nevertheless, since we are in an exactly identified context, the weighting matrix only really matters in those circumstances in which the parameters of the structural model are inequality constrained, which does not happen very often.

Figures 2a-b display box-plots with the sampling distributions of the two GMM-based indirect estimators of the structural parameters ϕ and ρ that we have proposed, together with the

[8] As explained in footnote 1 of Calzolari, Fiorentini and Sentana (2004), one should not conduct unrestricted indirect estimation in terms of the unrestricted parameters $\boldsymbol{\theta}^\dagger$ because the Jacobian of the transformation is singular when $\boldsymbol{\theta}$ is at the boundary of its parameter space, which violates the usual regularity conditions required for the consistency of the indirect estimators.

two pseudo-ML estimators on which they are based. The upper four distributions correspond to $N = 3$, while the lower two to $N = 30$, where we only analyse the SF based estimators because the HRS-based estimators become too costly for a full set of Monte Carlo experiments (but not for our empirical application; see also footnote 10). The central boxes describe the first and third quartiles of the sampling distributions, as well as their median. The maximum length of the whiskers is one interquartile range. Finally, we also report the fraction of estimates outside those whiskers to complement the information on the distributions tails. In this sense, the pseudo ML estimators of the auxiliary parameters α and β reach their lower bounds fairly frequently in some designs, especially when ψ is large. For instance, the estimated value of α was less than $\alpha^{\min}$ 8.63% of the time when $\phi = .1$, $\rho = .85$ and $\psi = N = 3$. Thus, these results clearly show that the constrained indirect estimation procedures that we have used are highly relevant in practice.

The small sample behaviour of the HRS and SF estimators is very much in accordance with what we have seen in section 2.3. When the signal to noise ratio is high (i.e. $\psi = N/9$) and the unconditional coefficient of variation of the unobserved conditional variance is low ($\phi = .1$, $\rho = .85$), the biases in those two estimators are both very small and indistinguishable from each other. In contrast, when the signal to noise ratio is low (i.e. $\psi = N$) and the unconditional coefficient of variation of the unobserved conditional variance is high ($\phi = .2$ and $\rho = .75$), their biases become rather noticeable, with neither estimator dominating the other.

It is precisely in those cases that the systematic elimination of the biases achieved by our two indirect estimators is more pronounced.[9] At the same time, it seems that the variability of the indirect estimators does not increase much with respect to the approximate pseudo-ML ones, which suggests that we have accurately estimated the required expectations by simulating $100,000$ observations. When $N = 3$, the SF-based indirect estimators of ϕ and ρ often outperform the corresponding HRS-based estimators in terms of sampling variability but not in terms of bias, although the differences are rather minor. In any case, the important message is that both indirect estimators are consistent across all Monte Carlo designs. Finally, note that the only effect of increasing N from 3 to 30 while maintaining the signal to noise ratio is to slightly reduce the dispersion of the estimators.[10]

In contrast, the increase in N has dramatic effects on the sampling distributions of the estimators of ϕ^* and ρ^*, which are displayed in Figures 3a and 3b. Not surprisingly, the dispersion of those distributions is greatly reduced when we effectively pool 30 series instead of 3. Another

[9] The finite sample bias might be reduced even further by using the implicit bias adjustment procedures discussed by Gouriéroux, Renault and Touzi (2000) (see also Arvanitis and Demos (2006)).

[10] The results from the experimental designs with $\psi = N/9$ suggest that the small differences between HRS-based and SF-based indirect estimators observed with $N = 3$ continue to hold with $N = 30$.

differential result that we obtain in estimating the idiosyncratic dynamic variance parameters is that the effect of the signal to noise ratio is the reverse of the one seen for ϕ and ρ. Intuitively, the reason is that the roles of signal and noise are somewhat exchanged when we are considering the estimation of the idiosyncratic variance parameters. Once again, though, the differences between our two indirect estimators are very small, although the means of the HRS-based estimators of ϕ^* and ρ^* tend to be slightly closer to the true values.

As for the parameters of the unconditional covariance matrix (i.e. factor loadings $\mathbf{B}$ and idiosyncratic unconditional variances $\mathbf{\Psi}$), our results, which we do not report for the sake of brevity, indicate that the HRS estimators of the auxiliary model are very similar to the indirect estimators based on them, even in a case in which the signal to noise ratio is low and the unconditional coefficient of variation of the unobserved conditional variances are high. At the same time, the SF estimators, which coincide with the corresponding indirect estimators, are less efficient than the other two, which is in line with the Monte Carlo results reported by SF.

4.2 Efficiency comparisons

Proposition 1 indicates that our indirect estimators are fully efficient in the limiting case of conditional homoskesdasticity, and the same is likely to be true when the factors are fully revealed. There is another instance in which we can compute exact likelihood-based estimators, although we cannot obtain analytical relative efficiency results. As we mentioned in section 2.1, it is possible to obtain either classical maximum likelihood estimators via a simulated EM algorithm, or Bayesian estimators via MCMC in the special case in which idiosyncratic variances are constant. Given that the simulated EM algorithm is very slow, in this section we shall compare the posterior means of 100,000 simulated values of their Bayesian estimators with our indirect ones for restricted versions of the designs in the previous subsection in which $\phi^* = \rho^* = 0$. To maintain the computations within reasonable limits, we only consider $N = 3$. Once again, we reutilise the underlying random numbers to facilitate comparisons across designs.

Following Fiorentini, Sentana and Shephard (2004), we use independent beta priors on $\phi+\rho$ and $\rho/(\phi+\rho)$ with parameters (6,1.25) and (2.5,1.25) for the $\phi^0+\rho^0 = .95$ designs in an attempt to reflect the typical estimates obtained with weekly data. We also use a standard inverted gamma prior for the unconditional variance of the common factor, δ, with mean 1 and standard deviation 1/2. Finally, given that we normalise b_N to 1, we use normal priors for the remaining factor loadings with unit mean and variance $\psi_i/1.6$, and the usual marginal inverted gamma priors for ψ_i with mean 1.6 and standard deviation 3.2. The results of our Monte Carlo experiments are reported in Figure 4.

The first thing to note is that the sampling distributions of our indirect estimators are very similar to the corresponding sampling distributions reported in Figures 2a-b. If anything, the presence of time-varying idiosyncratic variances leads to indirect estimators of ϕ and ρ with slightly lower sampling variability. The second observation is that the Bayesian estimators exhibit the usual trade-off between bias and variance: prior information leads to lower variation in the estimated parameters, but unless the prior is centred around the true values, it introduces a finite sample bias. In principle, the effect of a non-degenerate prior should vanish asymptotically. However, our experience suggests that the strong dependence in higher order moments introduced by a GARCH model implies that it takes many more than 1,000 observations for those asymptotic results to provide a reliable guide to the finite sample distribution of these Bayesian point estimators. The third conclusion is that the signal to noise ratio is also a decisive factor for the behaviour of the Bayesian estimators. When the signal to noise is low, the sampling information tends to be dominated by the prior, while the opposite happens when it is high. It is precisely in those cases that the our indirect estimators have similar dispersions to the Bayesian estimators, but without their biases.

In Figure 4 we also include the usual normal approximation to the finite sample distribution of the estimators of the auxiliary model obtained with an asymptotic information matrix computed by Monte Carlo integration with 100 million drawings under the assumption that the auxiliary model is the correct one. By construction, those asymptotic approximations are centred around the true value, but their dispersion gives us an alternative efficiency benchmark, which is the more relevant the closer the auxiliary model is to the true one. These results confirm that the efficiency loss of our indirect estimators is likely to be very small.

5 Empirical application to the Dow 30 components

5.1 Parameter estimates

In this section we estimate a conditionally heteroskedastic single factor model for the weekly (arithmetic) excess returns on the thirty individual components of the Dow Jones Industrial Average (DJIA) stock index as of November 21st, 2005 for the period July 10th, 1986 to January 19th, 2007 (1070 observations).[11] To concentrate on the modelling of the conditional covariance matrix, all the data were demeaned before estimation. Although we could allow the

[11] The historical evolution of the index composition is described in http://djindexes.com/mdsidx/downloads/DJIA_Hist_Comp.pdf. The closing price of the last working day of each week, adjusted for splits and dividends, has been obtained from http://finance.yahoo.com/q/hp?s=%5EDJI. Given that the US stock markets closed for the rest of the week after the terrorist attacks on September 11, 2001, we have taken September 17 (Monday) as the last day of the previous week. As for the interest rate, we have used the (Friday) 1-month US $ Eurodeposit rate from www.federalreserve.gov/releases/h15/data/Weekly_Friday_/H15_ED_M1.txt to avoid the abnormal movements due to transitory liquidity squeezes occasionaly observed in the overnight and 7-day series.

idiosyncratic conditional variance parameters to differ across series, it makes sense to impose the cross-sectional homogeneity restriction $\phi_i = \phi$ and $\rho_i = \rho \ \forall i$ by analogy with the dynamic panel data literature. Therefore, the total number of parameters is $2 \times 30 + 4 = 64$. To compute the expected value of the scores by simulation, we use $H = 100$ throughout.

The static variance parameter estimates are reported in Table 1, while Table 2 reports the ones corresponding to the conditional variances. An important fact that we can learn from Table 1 is that the signal to noise ratio is quite high. Specifically, the R^2 of the theoretical regression of y_t in (5) on the underlying factor evaluated at those estimates is almost 95%. Consequently, the degree of observability of the common factor is somewhat higher than in the Monte Carlo designs studied in the previous section. As a result, it is not surprising that the relationship between the auxiliary model estimators and the indirect estimators that can be observed from Table 2 is essentially what our simulations predict. Specifically, in the case of the common factor the indirect estimators of the ARCH coefficient ϕ are lower than the corresponding auxiliary model estimators, while the opposite tends to be true for the GARCH parameter ρ. In contrast, the adjustment is very small for the idiosyncratic variance parameter estimates ϕ^* and ρ^*.

To learn more about the nature of the common factor, we regress the factor estimates on the 30 individual stock returns. Although our model implies that the Kalman-filter based factor representing portfolios would change from period to period, we can understand the results reported in Table 3 as providing average weights for the HRS and SF basis portfolios. Given the high correlation between their respective weights (99.5%), it is perhaps not surprising that the correlation between those two basis portfolios is almost perfect. For comparison purposes, we also include the analogous weights for the DJIA index, whose composition also changes over time. Although the correlation in weights is only 26%, the correlation between the returns to this index and our estimated factors is still rather high (96.2%), which confirms that the cross-sectional dependence between the individual components of the Dow 30 is strong.

An even higher correlation is observed between our estimated factor and an equally weighted portfolio of the thirty stocks (98.7%). In this sense, Figure 5a compares the conditional standard deviation that our estimated multivariate model implies for such equally weighted portfolio with the one obtained by estimating a univariate Gaussian GARCH(1,1) model for its excess returns. In this respect, note that conditional variance of $\boldsymbol{\iota}'\mathbf{r}_t/N$ implied by our single factor model is $(\mathbf{c}'\boldsymbol{\iota}/N)^2\lambda_t + tr(\boldsymbol{\Gamma}_t)/N^2$, where $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones. The correlation between both series is again very high (98.4%), which is reassuring given that our estimation criterion is multivariate, and not targeted to this particular portfolio. Importantly, the univariate parameter estimates reported in the last column of Table 2 are very similar to our indirect estimators, which confirms once again that the latter are achieving their goal.

5.2 The importance of idiosyncratic risk

Unlike our multivariate model, however, a univariate GARCH(1,1) model cannot distinguish common from idiosyncratic shocks. And although the equally weighted portfolio is reasonably well diversified by construction, it is obviously not fully diversified. An important characteristic of our model is that the conditional variances of the idiosyncratic factors vary over time. This can also be seen from Figure 5a, which shows the time series evolution of (the square root of) $tr(\mathbf{\Gamma}_t)/N^2$. It is worth noticing that such a series more than doubled between the beginning of 1995 and the end of 2000, which is in line with the empirical results reported by Campbell et al (2001). Since the bursting of the dot-com bubble, though, idiosyncratic risk seems to have decreased again, reaching its 1994 values at the beginning of 2004.

Despite the fact that constant idiosyncratic variances imply that we could easily construct portfolios with constant volatility, which does not seem empirically plausible, one could argue that the variability in idiosyncratic risk that we observe in Figure 5a could be simply due to estimation error. To assess the empirical relevance of such a claim we have estimated a restricted version of our model in which we set both α^* and β^* to 0. The deterioration in fit is dramatic, with the Gaussian pseudo-log likelihood function dropping by almost 1747 points!

It is also interesting to look at the temporal changes in the mix of common and idiosyncratic risk for some specific stocks. Figures 5b and 5c present such a decomposition for GM and IBM. In this sense, the first thing that one observes is that for these individual stocks, idiosyncratic risk is quite important. Once again, though, its importance changes over time. For instance, common risk was the dominant component in both cases around the October 87 crash, whose effects were widespread. In contrast, our model suggests that the increase in volatility that GM shares experienced in 2005, when the debt of the world's biggest automaker was downgraded, can be attributed almost entirely to firm-specific characteristics, such as poor sales and rampant health care and pension costs. On the other hand, idiosyncratic volatility became very important for IBM during the first half of the 1990's, as this firm struggled to grasp the era of personal computers before eventually giving up and selling that line of business to a Chinese company. These results emphasise the need to allow for time-varying idiosyncratic variances.

5.3 The evolution of conditional correlations

The decomposition between common and idiosyncratic risks that we have just studied also has very interesting implications for correlations. A stylised fact that has been noted before is that periods when markets are increasingly correlated are also times when markets are volatile (see King, Sentana and Wadhwani (1994)). Since our empirical evidence suggests that the unobservable factor leads to individual stocks move in the same direction, it is the case in our model that periods when the volatility of the unobservable factor rises are also those when, ceteris paribus, individual stocks appear to exhibit greater inter-correlation. To see this, note that the conditional correlation coefficient between any two stocks is given by

$$\rho_{12t} = \frac{b_1 b_2 \delta_t}{\sqrt{b_1^2\delta_t + \psi_{1t}}\sqrt{b_2^2\delta_t + \psi_{2t}}}.$$

Hence, ρ_{12t} will be increasing in δ_t if $b_1 b_2 > 0$ and decreasing in ψ_{1t} and ψ_{1t}.

Figure 5d presents the evolution of the GM-IBM conditional correlation. As we have just seen, an increase in the volatility of the common factor will be associated with an increase in their correlation. This is consistent with the observed rise in both volatilities and interdependences around the 1987 crash. In contrast, an increase in the volatility of one of the idiosyncratic factors will be associated with a decrease in the correlation. Therefore, it is not entirely surprising that the correlation between these two stocks reached its minimum values precisely in the two periods in which either IBM idiosyncratic risks, or GM idiosyncratic risks were highest.

5.4 Non-normality of returns

Nowadays it is widely accepted that the distribution of stock returns has fat tails even after controlling for time-variation in volatilities and correlations. This fact is confirmed in our dataset by the Fiorentini, Sentana and Calzolari (2003) test of multivariate normal versus multivariate t innovations applied to the HRS estimates of the auxiliary model, whose (information matrix) t-ratio version takes the value 111.95. For that reason, we have also obtained indirect estimators of all the model parameters under the assumption that the joint distribution of common and idiosyncratic factors given their past values is multivariate t with the same mean and covariance matrix as before, but degrees of freedom given by the reciprocal of an additional parameter η, with $0 \leq \eta < \frac{1}{2}$. Specifically, we have considered two different joint indirect estimators of $\boldsymbol{\varrho}$ and η based on the same HRS and SF estimators of the auxiliary model parameters $\boldsymbol{\theta}$ that we have been considering so far. To identify the tail thickness parameter η, we have added as an extra influence function the modified score of the t-based log-likelihood function of the auxiliary model evaluated under normality, which is given by

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\varsigma_t^2(\boldsymbol{\theta})}{4} - \frac{(N+2)}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{N(N+2)}{4}\right] + \mu,$$

where

$$\varsigma_t(\boldsymbol{\theta}) = \mathbf{x}_t'[\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}_t(\boldsymbol{\theta})]^{-1}\mathbf{x}_t$$

and μ is the corresponding Kuhn-Tucker multiplier. Therefore, we are considering the equality restricted indirect estimators studied in Calzolari, Fiorentini and Sentana (2004). In this sense, note that since the consistency of the SF estimators of $\mathbf{b}$ and $\boldsymbol{\psi}$ does not depend on the normality assumption, the results discussed in section 3.2 continue to apply as long as the fourth moments of the innovations remain bounded, which happens whenever $\eta < \frac{1}{4}$.

In order to obtain the expected value of the modified score of the auxiliary model, we generate the standardised multivariate t random numbers as $\sqrt{1-2\eta}/\sqrt{\eta}$ times the ratio of a standard multivariate normal to the square root of an independent gamma variate with parameters $1/2\eta$ and 2. An important point to bear in mind in this context is that the usual routines for simulating gamma random variables involve some degree of rejection, which unfortunately can change for different values of η. For that reason, we use the slower but smooth inversion method based on the NAG G01FFF gamma quantile function so that we can keep the underlying uniform variates fixed across simulations. The results reported in the last columns of Tables 1 and 2 indicate that the degrees of freedom of the multivariate student t are estimated as being slightly above 10. The remaining parameters remain by and large unchanged.

6 Conclusions

In this paper, we propose two indirect estimators for conditionally heteroskedastic factor models in which the conditional variances of both common and idiosyncratic factors are continuous functions of their own past values, and as such, unobserved by the econometrician.

Our first estimator uses the Kalman filter-based approximation proposed by Harvey, Ruiz and Sentana (1992) with analytical derivatives as auxiliary model. We show the superiority of our choice over other alternatives previously considered in the literature on the grounds that (a) it has exactly the same number of parameters as the model of interest, and with the same meaning, and it is also easy to estimate; and (b) it spans the score of the model of interest in some important limiting cases, providing a very accurate approximation to it in more general situations. To tackle the inequality restrictions on the auxiliary model parameters, we employ the constrained indirect estimation procedures introduced in our earlier work.

Our second joint indirect estimator of the parameters of interest is based on sequential estimators of the HRS approximation, which we show can be successfully applied to situations in which the cross-sectional dimension is rather large if we make the empirically plausible assumption that the dynamic variance coefficients of the idiosyncratic terms are common across series, or at least across a limited number of homogenous groups of series.

We compare the finite sample performance of our proposed indirect estimators relative to the inconsistent HRS and SF estimators on which they are based by means of Monte Carlo methods in empirically realistic situations involving up to 30 series. Our results suggest that the estimators that we propose consistently estimate the parameters of the conditional variances of the latent factors, eliminating the biases of the approximate ML methods without simultaneously increasing their sampling variability. However, it should also be noted that the performance of the pseudo ML estimators is itself rather good, except when the signal to noise ratio is low, or the unconditional coefficient of variation of the volatility of the factors is high.

In addition, we assess the efficiency of our estimators relative to the Bayesian estimators proposed by Fiorentini, Sentana and Shephard (2004) under the assumption of constant idiosyncratic variances. Interestingly, we find that their estimators exhibit the usual trade-off between bias and variance: prior information leads to lower variation in the estimated parameters, but unless the prior is centred around the true values, it introduces a finite sample bias. We also find that when the sample information dominates the prior, our indirect estimators have similar dispersions to the Bayesian estimators, but without their biases.

Importantly, we also explain how both our proposed estimators can be easily adapted to handle any state space model with GARCH disturbances, which includes many examples that have been used in the empirical economic and finance literatures, such as structural time series models, or regression models with time-varying coefficients.

Finally, we apply our estimators to weekly excess returns on the thirty individual components of the Dow Jones Industrial Average over the last two decades. Our empirical results confirm that conditional correlations tend to increase in turbulent market episodes, while the opposite happens when there are important firm-specific news. In addition, we document that the strategic changes that IBM underwent during the 90's, and the more recent financial problems that affected GM, clearly left their mark in the temporal evolution of their idiosyncratic volatilities. On the other hand, our results also suggest that the increase in idiosyncratic risk documented by Campbell et al (2001) was partly reversed after the dot-com bubble burst.

Our empirical results also confirm the need to consider leptokurtic distributions for financial returns even after controlling for volatility clustering effects. In this sense, we explain in detail how to modify our estimators to handle a multivariate t instead of a multivariate normal without re-estimating the HRS auxiliary model by adding the Lagrange multiplier associated with the nesting restriction. In this sense, an interesting avenue for research would be to develop analogous procedures for more flexible, possibly asymmetric distributions.

Acknowledgements

This is an abridged version of Sentana, Calzolari and Fiorentini (2007), which in turn was a thoroughly revised version of Sentana, Calzolari and Fiorentini (2004). We are grateful to Monica Billio, Adrian Pagan, Eric Renault, Neil Shephard, George Tauchen, an associate editor and three anonymous referees, as well as seminar audiences at Copenhagen, ECARES-ULB, Monash, the North Carolina Research Triangle, the ASSET Conference (Bologna, 1998), the III LSE FMG Workshop on Empirical Finance (London, 2001), the AUEB Financial Econometrics Conference (Delphi, 2001), the European Meeting of the Econometric Society (Lausanne, 2001), the Statistica Computazionale Conference (Bressanone, 2001), the Symposium on New Frontiers in Financial Volatility Modelling (Florence, 2003) and the Second Italian Congress on Econometrics and Empirical Economics (Rimini, 2007) for helpful comments and suggestions. Of course, the usual caveat applies. Financial support from MIUR through the project "Specification, estimation and testing of latent variable models. Applications to the analysis and forecasting of economic and financial time series" (Calzolari and Fiorentini) and the Spanish Ministry of Education and Science through grants SEJ 2005-08880 (Sentana) is gratefully acknowledged. Thanks are also due to Javier Mencía for his help in computing the PIT sequences and the asymptotic information matrices, and to Gian Piero Aielli for collecting the data and compiling Figures 2, 3 and 4.

Appendix

A Proof of Proposition 1

For the sake of brevity, the proof will be developed for the following univariate model:

$$y_t = f_t + \epsilon_t,$$

$$\begin{pmatrix} f_t \\ \epsilon_t \end{pmatrix} | I_{t-1} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1+\phi(f_{t-1}^2 - 1) & 0 \\ 0 & \upsilon \end{bmatrix} \right\},$$

where $\phi \geq 0$, $\upsilon \geq 0$. Nevertheless, it can be tediously extended to cover the general case.

Let $p(y_t|Y_{t-1};\varrho)$ denote the conditional density of y_t given $Y_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ and the parameters $\boldsymbol{\varrho} = (\upsilon, \phi)'$. The log-likelihood function of a sample of size T on y_t, $\mathbf{y} = (y_1, \ldots, y_T)'$ cannot be written in closed-form except when $\phi = 0$ and/or $\upsilon = 0$. In particular, when $\phi = 0$, we just have the log-likelihood function of an *i.i.d.* $N(0, 1+\upsilon)$ process, while when $\upsilon = 0$, we will have the log-likelihood function of a univariate ARCH(1) model with unit unconditional variance.

In contrast, the joint log-likelihood function of $\mathbf{y}$ and the latent factors $\mathbf{f} = (f_1, \ldots, f_T)'$ can always be trivially written as the sum of the marginal log-likelihood function of $\mathbf{f}$ and the conditional log-likelihood of $\mathbf{y}$ given $\mathbf{f}$, where (ignoring initial conditions)

$$\ln p(\mathbf{y}|\mathbf{f};\boldsymbol{\varrho}) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln \upsilon - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - f_t)^2}{\upsilon}, \tag{A1}$$

and

$$\ln(\mathbf{f}|\boldsymbol{\varrho}) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left\{\ln[1+\phi(f_{t-1}^2 - 1)] + \frac{f_t^2}{1+\phi(f_{t-1}^2-1)}\right\}. \tag{A2}$$

Let

$$\bar{q}_T(\boldsymbol{\varrho}) = \frac{1}{T}\sum_{t=1}^{T} q(y_t|Y_{t-1};\boldsymbol{\varrho})$$

denote the sample average of the score of the marginal log-likelihood function of $\mathbf{y}$, where

$$q(y_t|Y_{t-1};\boldsymbol{\varrho}) = \frac{\partial \ln p(y_t|Y_{t-1};\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}}$$

represents the contribution to the score function from observation t. Although $q(y_t|Y_{t-1};\boldsymbol{\varrho})$ cannot generally be obtained in closed form, the well-known Kullback inequality implies that

$$E\left[\left.\frac{\partial \ln p(\mathbf{f}|\mathbf{y};\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}}\right|\mathbf{y};\boldsymbol{\varrho}\right] = \mathbf{0}.$$

As a result, $\bar{q}_T(\boldsymbol{\varrho})$ can be obtained as the expected value given $\mathbf{y}$ and $\boldsymbol{\varrho}$ of the sample average of the unobservable scores corresponding to $\ln p(\mathbf{y}|\mathbf{f};\boldsymbol{\varrho})$ and $\ln p(\mathbf{f}|\boldsymbol{\varrho})$. Specifically, assuming that $\upsilon > 0$, this yields

$$\bar{q}_{\upsilon T}(\boldsymbol{\varrho}) = \frac{1}{2T}\frac{1}{\upsilon}\sum_{t=1}^{T} E\left[\left.\frac{(y_t - f_t)^2}{\upsilon} - 1\right|\mathbf{y};\boldsymbol{\varrho}\right],$$

and

$$\bar{q}_{\phi T}(\boldsymbol{\varrho}) = \frac{1}{2T}\sum_{t=1}^{T} E\left\{\left.\frac{f_{t-1}^2 - 1}{1+\phi(f_{t-1}^2-1)}\left[\frac{f_t^2}{1+\phi(f_{t-1}^2-1)} - 1\right]\right|\mathbf{y};\boldsymbol{\varrho}\right\}.$$

Then, we can use the MCMC procedures proposed by Fiorentini, Sentana and Shephard (2004), which draw samples of $\mathbf{f}$ given $\mathbf{y}$ and $\boldsymbol{\varrho}$, to compute these expected values by simulation. However, it is straightforward to prove that when $\phi = 0$

$$\bar{q}_{\upsilon T}(\upsilon, \phi{=}0) = \frac{1}{2T}\frac{1}{(1+\upsilon)}\sum_{t=1}^{T}\left(\frac{y_t^2}{1+\upsilon} - 1\right)$$

and

$$\bar{q}_{\phi T}(\upsilon, \phi{=}0) = \frac{1}{2T}\frac{1}{(1+\upsilon)^2}\sum_{t=1}^{T}\left(\frac{y_{t-1}^2}{1+\upsilon} - 1\right)\left(\frac{y_t^2}{1+\upsilon} - 1\right),$$

because

$$f_t|\mathbf{y};\upsilon,\phi{=}0 \sim i.i.d.\ N\left(\frac{y_t}{1+\upsilon}, \frac{\upsilon}{1+\upsilon}\right).$$

Consider now the following HRS-style auxiliary model

$$y_t = g_t + \zeta_t,$$

$$\begin{aligned}
\begin{pmatrix} g_t \\ \zeta_t \end{pmatrix} |I_{t-1} &\sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1+\alpha\left[g^2_{t-1|t-1}(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta})-1\right] & 0 \\ 0 & \varphi \end{bmatrix} \right\}, \\
g_{t|t}(\boldsymbol{\theta}) &= \frac{1+\alpha[g^2_{t-1|t-1}(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta})-1]}{1+\alpha[g^2_{t-1|t-1}(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta})-1]+\varphi}\cdot y_t, \\
\omega_{t|t}(\boldsymbol{\theta}) &= \frac{1+\alpha[g^2_{t-1|t-1}(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta})-1]}{1+\alpha[g^2_{t-1|t-1}(\boldsymbol{\theta})+\omega_{t-1|t-1}(\boldsymbol{\theta})-1]+\varphi}\cdot \varphi,
\end{aligned}$$

and $\alpha \geq 0$, $\varphi \geq 0$. Since this model is a rather special case of (7)-(10) we can use the expressions in Appendix B to compute its score. Tedious but straightforward algebra then shows that

$$\bar{s}_{\varphi T}(\varphi,\alpha{=}0) = \frac{1}{2T}\frac{1}{(1+\varphi)}\sum_{t=1}^{T}\left(\frac{y_t^2}{1+\varphi}-1\right),$$

and

$$\bar{s}_{\alpha T}(\varphi,\alpha{=}0) = \frac{1}{2T}\frac{1}{(1+\varphi)^2}\sum_{t=1}^{T}\left(\frac{y_{t-1}^2}{1+\varphi}-1\right)\left(\frac{y_t^2}{1+\varphi}-1\right),$$

which confirms that the pseudo log-likelihood score of the auxiliary model evaluated at $\alpha{=}0$ coincides with the score of the true model evaluated at $\phi{=}0$ for $\varphi = \upsilon$. Hence, the HRS auxiliary model smoothly embeds the true model at $\alpha = \phi = 0$ and $\varphi = \upsilon$. As a result, Theorem 2 in Gallant and Tauchen (1996) implies that indirect estimation will be as efficient as maximum likelihood in that case.

In this sense, the inclusion of the term $\omega_{t-1|t-1}(\boldsymbol{\theta})$ in the conditional variance specification of the HRS auxiliary model is crucial for the result, for otherwise the average score with respect to α at $\alpha = 0$ will be equal to

$$\frac{1}{2T}\frac{1}{(1+\varphi)}\sum_{t=1}^{T}\left[\frac{y_{t-1}^2}{(1+\varphi)^2}-1\right]\left(\frac{y_t^2}{1+\varphi}-1\right),$$

which no longer coincides with $\bar{q}_{\phi T}(\upsilon,\phi{=}0)$. □

A similar argument can be used to show that

$$\bar{s}_{\alpha T}(\varphi{=}0,\alpha{=}\phi) = \bar{q}_{\phi T}(\upsilon{=}0,\phi),$$

which is not very surprising given that

$$\ln p(y_t|Y_{t-1};\upsilon = 0,\phi) = l_t(\varphi = 0, \alpha = \phi)$$

for every possible value of ϕ.

Unfortunately, it is not possible to obtain closed-form expressions for

$$\bar{q}_{\upsilon T}(\upsilon{=}0,\phi) = \lim_{\upsilon\to 0}\bar{q}_{\upsilon T}(\boldsymbol{\varrho}) = \lim_{\upsilon\to 0}\frac{1}{2T}\frac{1}{\upsilon}\sum_{t=1}^{T}E\left[\left.\frac{(y_t-f_t)^2}{\upsilon}-1\right|\mathbf{y};\upsilon,\phi\right]$$

except when $\phi = 0$, so we cannot prove in this way whether or not

$$\bar{s}_{\varphi T}(\varphi{=}0,\alpha{=}\phi) = \bar{q}_{\upsilon T}(\upsilon{=}0,\phi).$$

References

Aguilar, O. and M. West, 2000, Bayesian dynamic factor models and variance matrix discounting for portfolio allocation, Journal of Business and Economic Statistics 18, 338-357.

Andersen, T.G., 1994, Volatility, unpublished paper, Finance Department, Northwestern University.

Arvanitis, S. and A. Demos, 2006, Bias properties of three indirect estimators, mimeo, Athens University of Economics and Business.

Bauwens, L., S. Laurent and J.V.K. Rombouts, 2006, Multivariate GARCH models: a survey, Journal of Applied Econometrics 21, 79-109.

Billio, M. and A. Monfort, 2003, Kernel-based indirect inference, Journal of Financial Econometrics 1, 297-326.

Broto, C. and E. Ruiz, 2006, Unobserved component models with asymmetric conditional variances, Computational Statistics and Data Analysis 50, 2146-2166.

Calzolari, G., G. Fiorentini and E. Sentana, 2004, Constrained indirect estimation, Review of Economic Studies 71, 945-973.

Campbell, J.Y., M. Lettau, B.G. Malkiel and Y. Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, Journal of Finance 56, 1-43.

Carter, C.K. and R. Kohn, 1994, On Gibbs sampling for state space models, Biometrika 81, 541-553.

Chib, S., F. Nardari and N. Shephard, 2006, Analysis of high dimensional multivariate stochastic volatility models, Journal of Econometrics 134, 341-371.

Chou, R., R.F. Engle and A. Kane, 1992, Measuring risk aversion from excess returns on a stock index, Journal of Econometrics 52, 201-224.

Diebold, F.X., A.G. Gunther and A.S. Tay, 1998, Evaluating density forecasts, with applications to financial risk management, International Economic Review 39, 863-883.

Diebold, F.X. and M. Nerlove, 1989, The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, Journal of Applied Econometrics 4, 1-21.

Doz, C. and E. Renault, 2006, Factor stochastic volatility in mean models: a GMM approach, Econometric Reviews 25, 275-309

Drost, F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica 61, 909-927.

Dungey, M., V.L. Martin and A.R. Pagan, 2000, A multivariate latent factor decomposition of international bond yield spreads, Journal of Applied Econometrics 15, 697-715.

Engle, R.F. , 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987-1007.

Engle, R.F. and J. Mezrich, 1996, GARCH for groups, Risk 9:8, 36-40.

Engle, R.F., V.M. Ng and M. Rothschild, 1990, Asset pricing with a factor–ARCH structure: Empirical estimates for Treasury Bills, Journal of Econometrics 45, 213-237.

Engle, R.F. and R. Susmel, 1993, Common volatility in international equity markets, Journal of Business and Economic Statistics 11, 167-176.

Fiorentini, G., E. Sentana and G. Calzolari, 2003, Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with student t innovations, Journal of Business and Economic Statistics 21, 532-546.

Fiorentini, G., E. Sentana and N. Shephard, 2004, Likelihood estimation of latent generalised ARCH structures, Econometrica 72, 1481-1517.

Gallant, A.R. and G. Tauchen, 1996, Which moments to match?, Econometric Theory 12, 657-681.

Gouriéroux, C. and A. Monfort, 1996, Simulation-based econometric methods, Oxford University Press, Oxford.

Gouriéroux, C., A. Monfort and E. Renault, 1993, Indirect inference, Journal of Applied Econometrics 8, S85-S118.

Gouriéroux, C., E. Renault and N. Touzi, 2000, Calibration by simulation for small sample bias correction, in: R. Mariano, T. Shuermann and M. Weeks, (EDS.), Simulation-based inference in econometrics. Methods and applications, Cambridge University Press, Cambridge.

Harvey, A.C. , 1989, Forecasting, structural time series models and the Kalman filter, Cambridge University Press, Cambridge.

Harvey, A.C., E. Ruiz and E. Sentana, 1992, Unobservable component time series models with ARCH disturbances, Journal of Econometrics 52, 129-158.

Jacquier, E., N.G. Polson and P.E. Rossi, 1994, Bayesian analysis of stochastic volatility models (with discussion), Journal of Business and Economic Statistics 12, 371-417.

Kim, C.-J., and C.R. Nelson, 1999, State-space models with regime switching. Classical and Gibbs-sampling approaches with applications, MIT Press, Cambridge.

King, M., E. Sentana and S. Wadhwani, 1994, Volatility and links between national stock markets, Econometrica 62, 901-933.

Lawley, D.N. and A.E. Maxwell, 1971, Factor analysis as a statistical method, 2nd edition, London Butterworths, London.

Magnus, J.R., 1988, Linear structures, Oxford University Press, New York.

Magnus, J.R. and H. Neudecker, 1999, Matrix differential calculus with applications in Statistics and Econometrics, 2nd edition, Wiley, Chichester.

Meddahi, N. and E. Renault, 2004, Temporal aggregation of volatility models, Journal of Econometrics 119, 355-379.

Nelson, D.B. and C.Q. Cao, 1992, Inequality constraints in the univariate GARCH model, Journal of Business and Economic Statistics 10, 229-235.

Nijman, T. and E. Sentana, 1996, Marginalization and contemporaneous aggregation of multivariate GARCH processes, Journal of Econometrics 71, 71-87.

Pitt, M.K. and N. Shephard, 1999, Time varying covariances: a factor stochastic volatility approach (with discussion), in: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, (EDS.), Bayesian Statistics 6, 547-570, Oxford University Press, Oxford.

Ross, S.A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341-360.

Rubin, D. and D. Thayer, 1982, EM algorithms for ML factor analysis, Psychometrika 47, 69-76.

Sentana, E., 2000, The likelihood function of conditionally heteroskedastic factor models, Annales d'Economie et de Statistique 58, 1-19.

Sentana, E., 2004, Factor representing portfolios in large asset markets, Journal of Econometrics 119, 257-289.

Sentana, E., G. Calzolari and G. Fiorentini, 2004, Indirect estimation of conditionally heteroskedastic factor models, CEMFI Working paper 0409, ftp://ftp.cemfi.es/wp/04/0409.pdf

Sentana, E., G. Calzolari and G. Fiorentini, 2007, Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks, Rimini Centre for Economic Analysis Working paper 40-07, http://www.rcfea.org/RePEc/pdf/wp40_07.pdf

Sentana, E. and G. Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, Journal of Econometrics 102, 143-164.

Shephard, N., 1994, Partial non-Gaussian state space, Biometrika 81, 115-131.

Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, in: D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, (EDS.), Time series models with econometric, finance and other applications, 1-67, Chapman and Hall, London.

Smith, A.A. 1993, Estimating nonlinear time series models using simulated vector autoregressions, Journal of Applied Econometrics 8, S63-S84.

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1-28.

Zhumabekova, D. and M. Dungey, 2001, Factor analysis of a model of stock market returns using simulation-based estimation techniques, Federal Reserve Bank of San Francisco Pacific Basin Working Paper PB01-08.

Table 1

Estimates of static variance parameters

	Factor loadings				Idiosyncratic variances			
Company	SF	HRS Aux	HRS IE (G)	HRS IE (t)	SF	HRS Aux	HRS IE (G)	HRS IE (t)
AA	1.76	1.91	1.92	1.90	14.1	15.9	16.3	16.0
AIG	1.85	1.98	1.98	1.96	7.4	8.2	7.9	8.5
AXP	2.18	2.24	2.24	2.23	8.5	6.4	7.0	6.8
BA	1.55	1.61	1.63	1.60	11.9	11.5	12.2	12.0
C	2.28	2.46	2.48	2.44	11.8	13.1	13.5	14.0
CAT	1.70	1.83	1.85	1.81	14.1	20.0	18.8	19.7
DD	1.72	1.81	1.82	1.79	8.2	7.3	7.5	7.8
DIS	1.73	1.89	1.89	1.88	10.6	10.4	10.0	9.8
GE	1.90	1.94	1.94	1.92	4.8	4.5	4.3	4.6
GM	1.58	1.63	1.63	1.61	13.9	16.6	18.3	18.3
HD	2.05	2.18	2.18	2.16	14.1	13.0	13.2	13.3
HON	1.76	1.90	1.91	1.90	12.0	9.8	9.5	9.6
HPQ	1.86	2.02	2.03	2.01	21.4	25.0	23.7	25.2
IBM	1.29	1.38	1.38	1.38	12.1	9.3	9.8	10.0
INTC	2.04	2.26	2.25	2.25	25.1	28.0	26.0	27.4
JNJ	1.28	1.41	1.41	1.40	7.7	7.9	8.4	8.4
JPM	2.20	2.19	2.20	2.18	14.1	8.5	9.5	9.4
KO	1.37	1.52	1.53	1.51	8.4	7.5	7.5	7.5
MCD	1.24	1.32	1.33	1.32	9.0	10.2	10.4	10.4
MMM	1.45	1.55	1.56	1.55	5.9	6.3	6.7	6.7
MO	1.12	1.37	1.38	1.35	13.0	12.6	13.8	14.2
MRK	1.35	1.53	1.54	1.52	11.0	14.8	15.2	15.9
MSFT	1.76	1.83	1.83	1.83	19.8	18.9	15.6	16.3
PFE	1.44	1.53	1.53	1.53	11.0	13.8	17.2	16.9
PG	1.25	1.34	1.34	1.32	8.8	7.7	7.9	7.8
T	1.16	1.27	1.28	1.26	10.7	8.4	9.1	8.9
UTX	1.84	1.82	1.82	1.80	7.6	8.9	8.8	9.2
VZ	1.05	1.21	1.21	1.20	9.4	8.4	8.9	8.8
WMT	1.78	1.88	1.89	1.87	10.0	9.7	9.7	9.9
XOM	*1*	*1*	*1*	*1*	5.9	6.8	6.8	7.1
δ	1.82	3.35	1.97	2.15				

Notes: SF refers to the Sentana - Fiorentini (2001) sequential estimators, while HRS Aux denotes the estimator that uses the Harvey, Ruiz and Sentana (1992) approximation to the log-likelihood function. In contrast, HRS IE represents the indirect estimators based on the score of the HRS Aux estimator, with G and t referring to Gaussian and multivariate t innovations, respectively. The expectations of the scores required for the indirect estimators use $H = 100$.

Table 2

Estimates of dynamic variance parameters

Parameter	SF Aux	SF IE (G)	SF IE (t)	HRS Aux	HRS IE(G)	HRS IE (t)	EW
ϕ	.110	.107	.103	.121	.113	.111	.110
ρ	.875	.877	.881	.873	.874	.878	.865
ϕ^*	.045	.044	.045	.046	.046	.046	
ρ^*	.944	.944	.944	.942	.942	.942	
η			.098			.097	

Notes: HRS Aux denotes the estimator that uses the Harvey, Ruiz and Sentana (1992) approximation to the log-likelihood function, while SF Aux refers to the Sentana - Fiorentini (2001) sequential estimator of the same model. In contrast, SF IE and HRS IE represent the indirect estimators based on those two estimators, with G and t referring to Gaussian and multivariate t innovations, respectively. Finally, EW corresponds to the equally weighted portfolio of the 30 individual stocks that make the DJIA index. The expectations of the scores required for the indirect estimators use $H = 100$.

Table 3

Unobservable factor score and DJIA weights (%)

Company	SF	HRS	DJIA
AA	2.17	2.14	4.63
AIG	6.58	6.51	2.06
AXP	5.20	5.29	4.41
BA	2.54	2.47	3.04
C	3.82	3.80	1.84
CAT	1.80	1.61	3.61
DD	4.02	4.10	4.94
DIS	3.79	3.91	2.79
GE	9.52	9.18	5.24
GM	2.49	2.31	4.66
HD	3.21	3.21	2.27
HON	2.54	2.75	3.53
HPQ	1.74	1.67	2.42
IBM	2.23	2.33	5.77
INTC	2.02	2.05	1.91
JNJ	3.63	3.79	2.94
JPM	3.43	3.53	2.06
KO	3.08	3.35	2.41
MCD	3.08	3.03	2.10
MMM	4.59	4.50	6.13
MO	1.46	1.77	3.80
MRK	1.84	1.80	3.03
MSFT	1.66	1.64	1.30
PFE	3.24	3.06	0.32
PG	2.97	3.08	4.24
T	1.88	2.04	2.73
UTX	5.16	4.72	5.75
VZ	2.34	2.64	0.69
WMT	4.15	4.20	2.42
XOM	3.83	3.53	6.95

Notes: HRS Aux denotes the estimator that uses the Harvey, Ruiz and Sentana (1992) approximation to the log-likelihood function, while SF Aux refers to the Sentana - Fiorentini (2001) sequential estimator of the same model. Finally, DJIA corresponds to the Dow 30 index.

FIGURE 1: Asymptotic biases in HRS estimators of conditional variance parameters

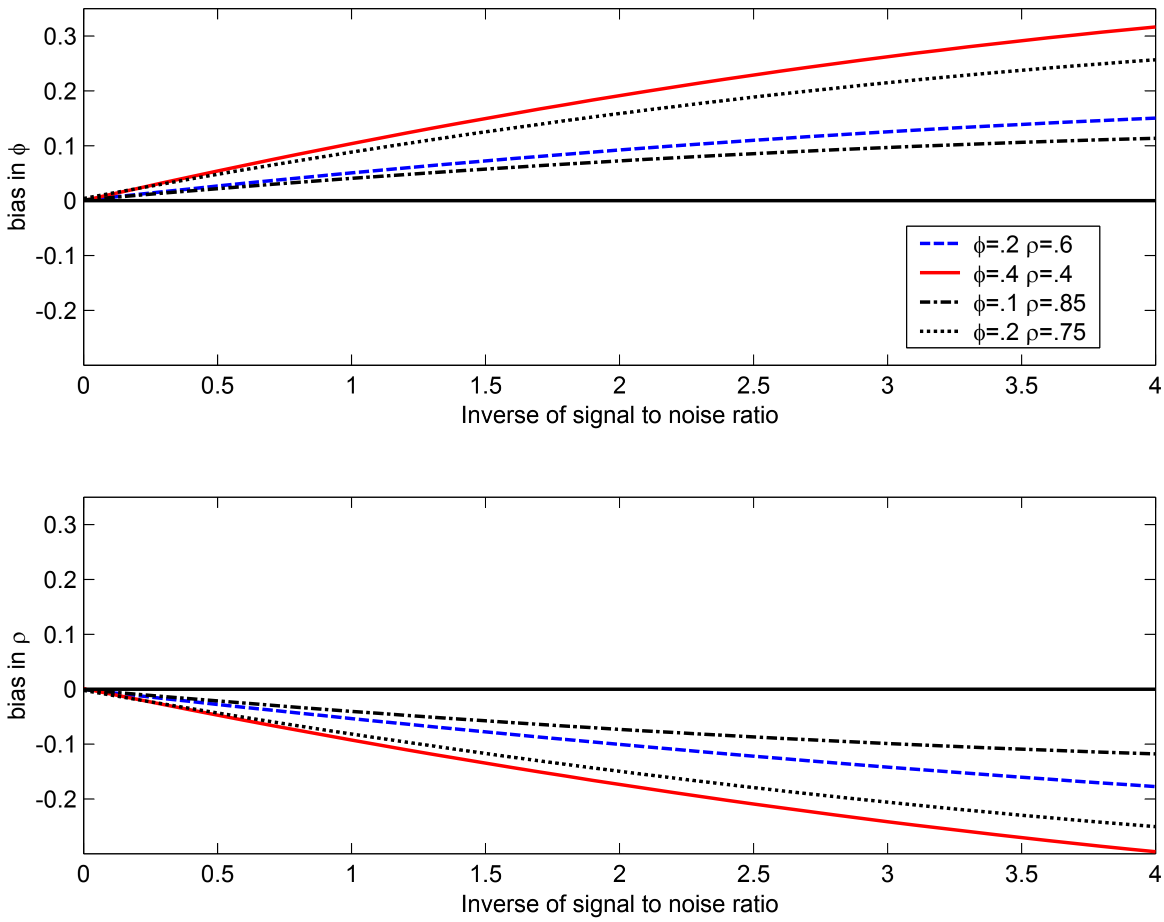
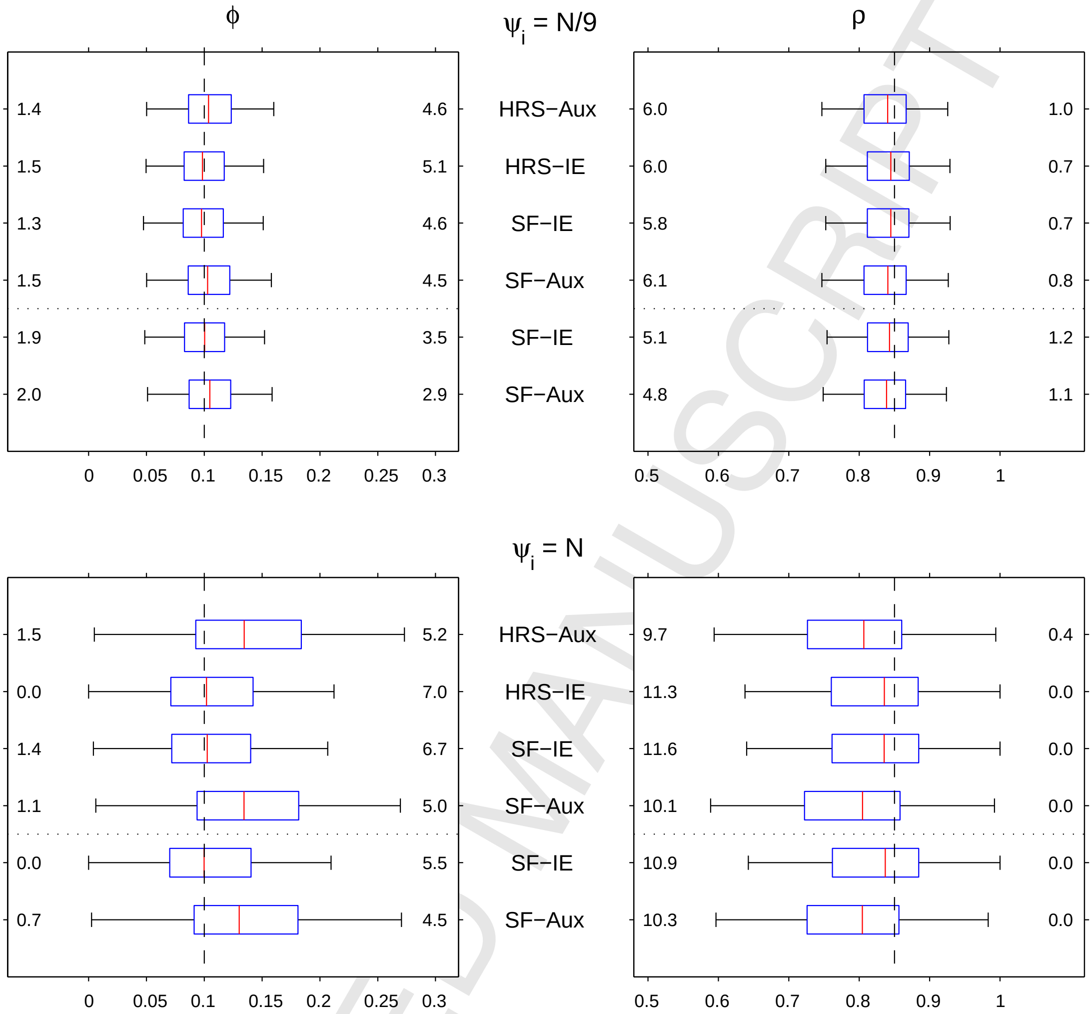


Fig. 2A: Monte Carlo distributions of estimators of conditional variance parameters for common factors.

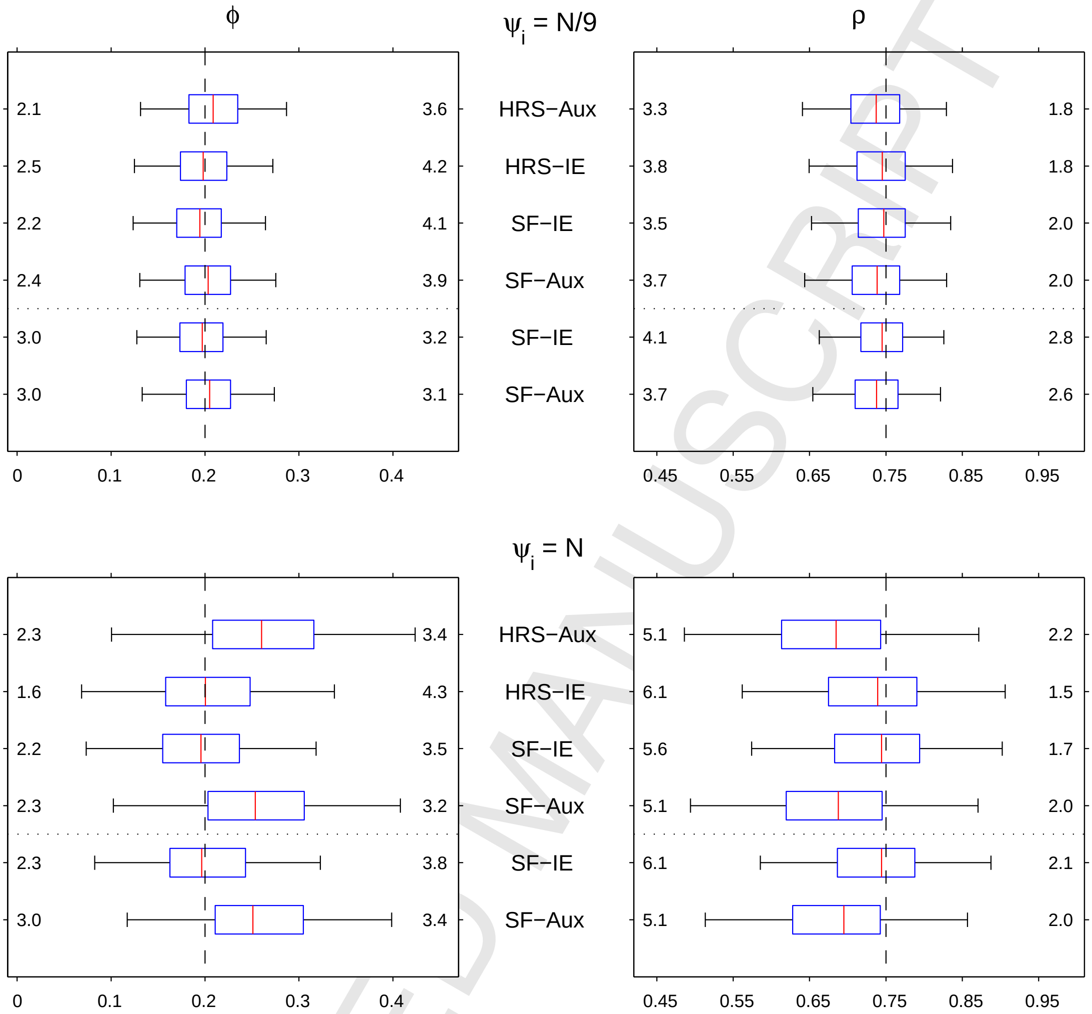
$b_i = 1; \phi = \phi^* = 0.1; \rho = \rho^* = 0.85.$



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. The upper four distributions correspond to N=3, while the lower two to N=30. IE means indirect estimators, and Aux estimators of the auxiliary model. HRS denotes estimators based on the Harvey, Ruiz and Sentana (1992) approximation to the log−likelihood, while SF refers to the sequential version in Sentana & Fiorentini (2001).

Fig. 2B: Monte Carlo distributions of estimators of conditional variance parameters for common factors.

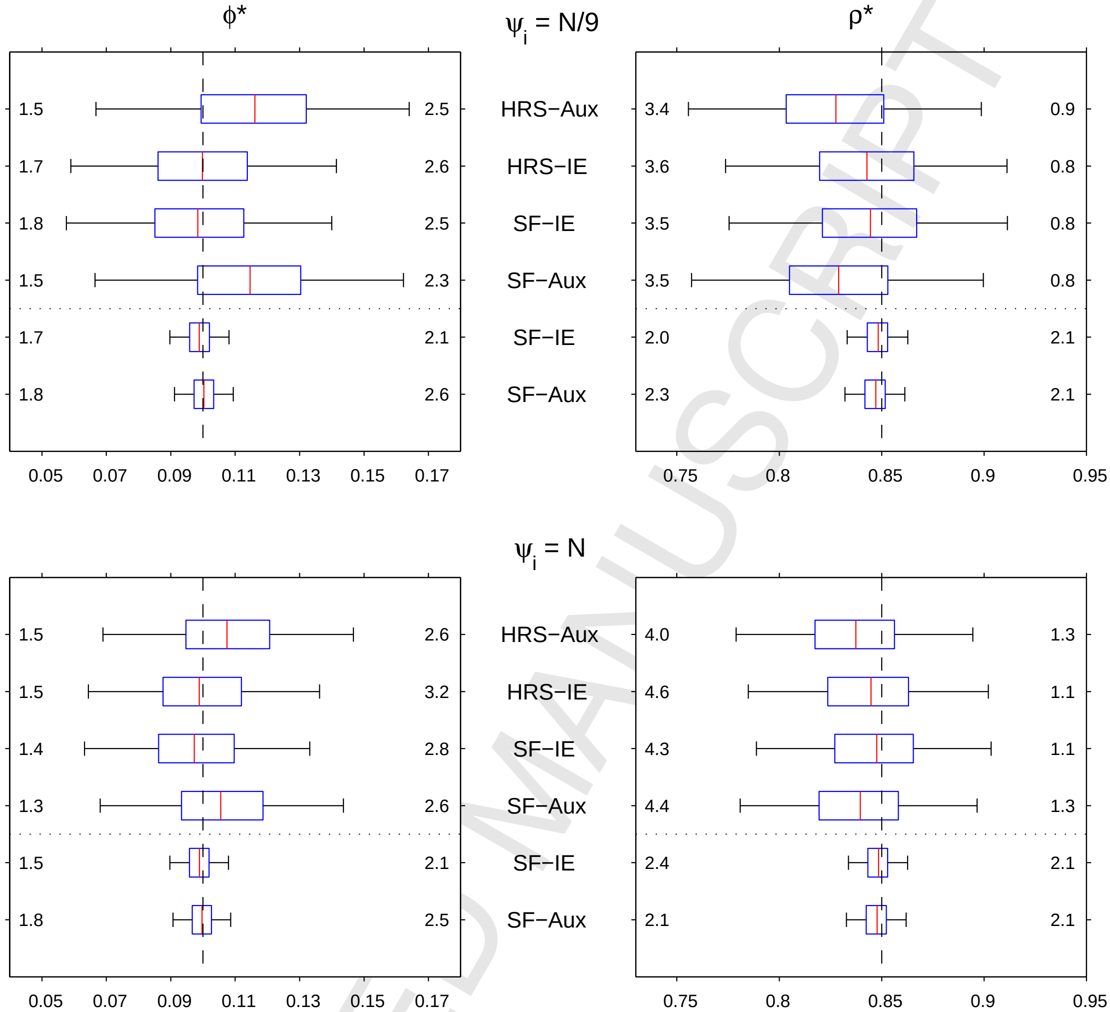
$b_i = 1; \phi = \phi^* = 0.2; \rho = \rho^* = 0.75$.



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. The upper four distributions correspond to N=3, while the lower two to N=30. IE means indirect estimators, and Aux estimators of the auxiliary model. HRS denotes estimators based on the Harvey, Ruiz and Sentana (1992) approximation to the log−likelihood, while SF refers to the sequential version in Sentana & Fiorentini (2001).

Fig. 3A: Monte Carlo distributions of estimators of conditional variance parameters for idiosyncratic factors.

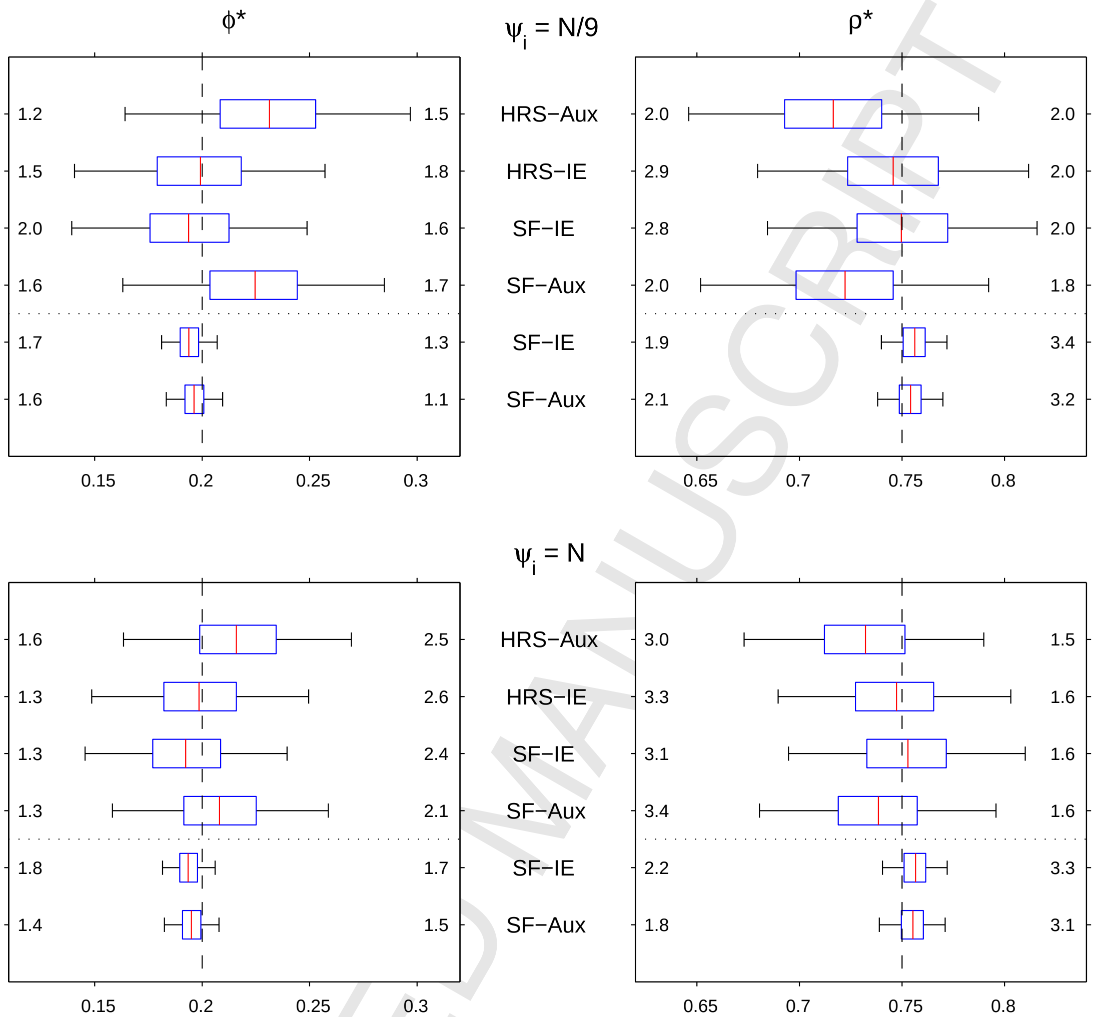
$b_i = 1$; $\phi = \phi^* = 0.1$; $\rho = \rho^* = 0.85$.



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. The upper four distributions correspond to N=3, while the lower two to N=30. IE means indirect estimators, and Aux estimators of the auxiliary model. HRS denotes estimators based on the Harvey, Ruiz and Sentana (1992) approximation to the log−likelihood, while SF refers to the sequential version in Sentana & Fiorentini (2001).

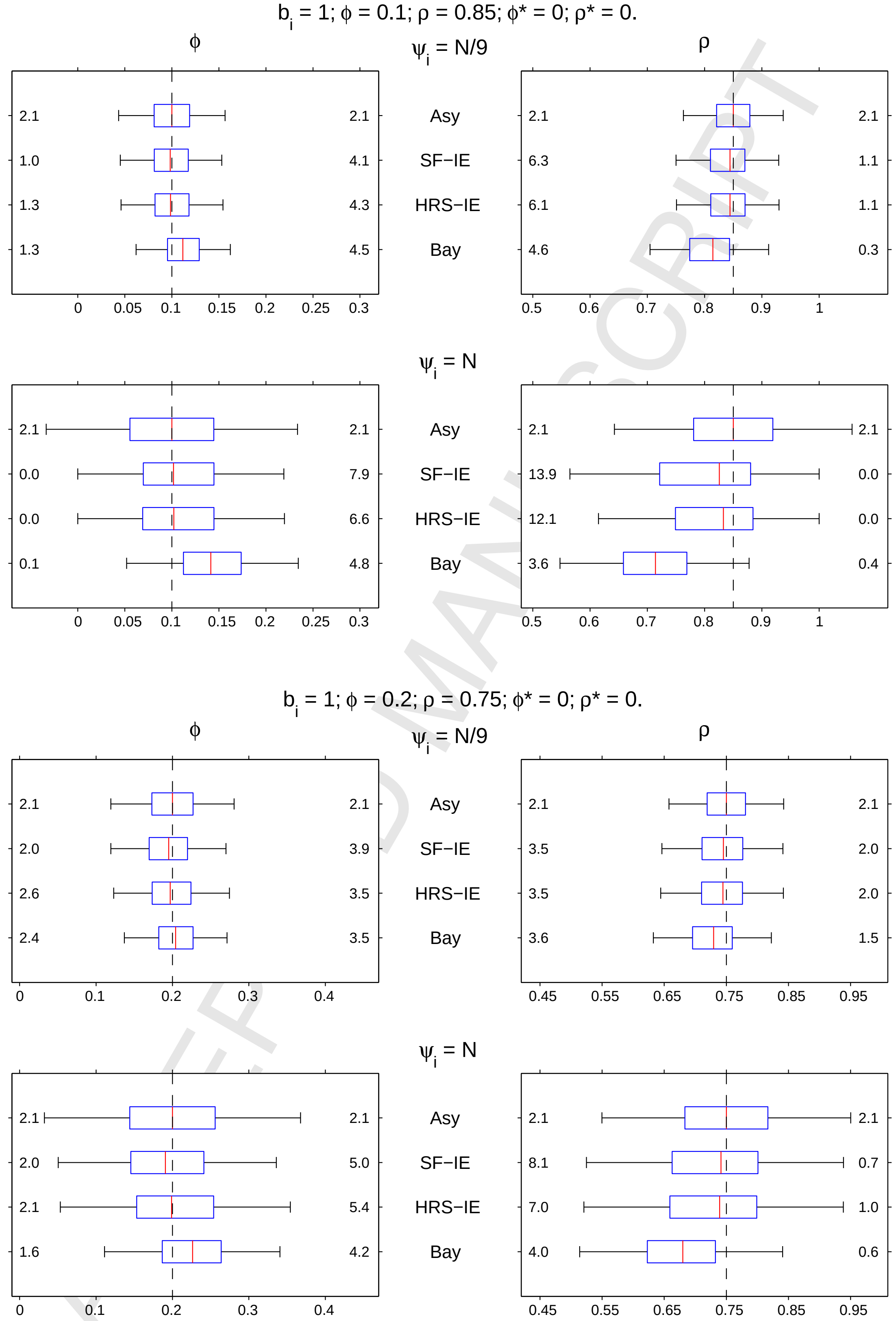
Fig. 3B: Monte Carlo distributions of estimators of conditional variance parameters for idiosyncratic factors.

$b_i = 1;\ \phi = \phi^* = 0.2;\ \rho = \rho^* = 0.75.$



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one interquartile range. We also report the fraction of replications outside those whiskers. The upper four distributions correspond to N=3, while the lower two to N=30. IE means indirect estimators, and Aux estimators of the auxiliary model. HRS denotes estimators based on the Harvey, Ruiz and Sentana (1992) approximation to the log−likelihood, while SF refers to the sequential version in Sentana & Fiorentini (2001).

Fig. 4: Monte Carlo distributions of estimators of conditional variance parameters for common factors.



The central boxes describe the 1st and 3rd quartiles of the sampling distributions, and their median. The maximum length of the whiskers is one inter-quartile range. We also report the fraction of replications outside those whiskers. IE means indirect estimators based on either the Harvey, Ruiz and Sentana (1992) approximation to the log−likelihood (HRS), or the sequential version in Sentana & Fiorentini (2001) (SF). Asy refers to distributions based on the asymptotic approximation, while Bay denotes the Bayesian MCMC estimators in Fiorentini, Sentana and Shephard (2004).

Figure 5a: Conditional Standard Deviation of Equally Weighted Portfolio

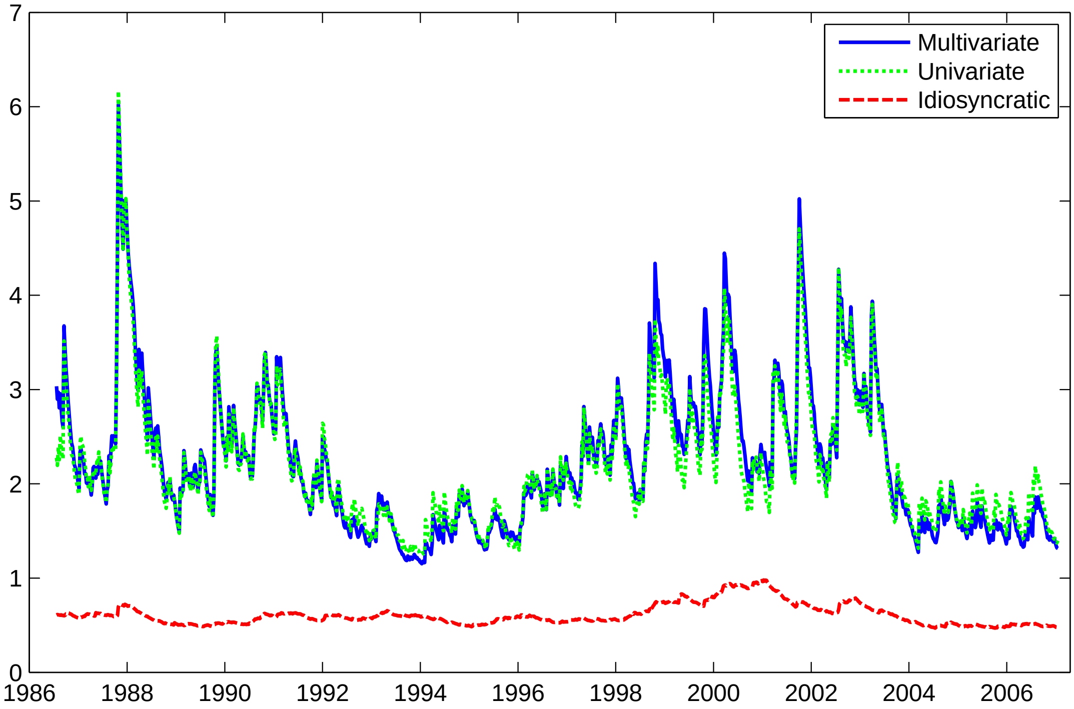


Figure 5b: Decomposition of the Conditional Standard Deviation of GM Stock Returns

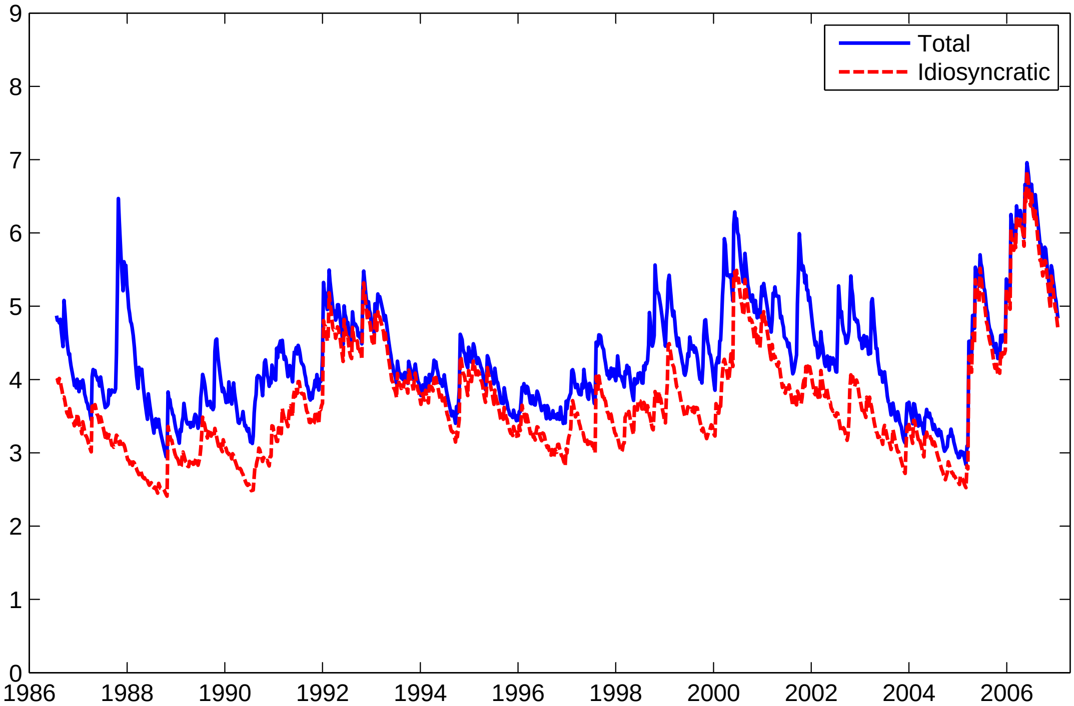


Figure 5c: Decomposition of the Conditional Standard Deviation of IBM Stock Returns

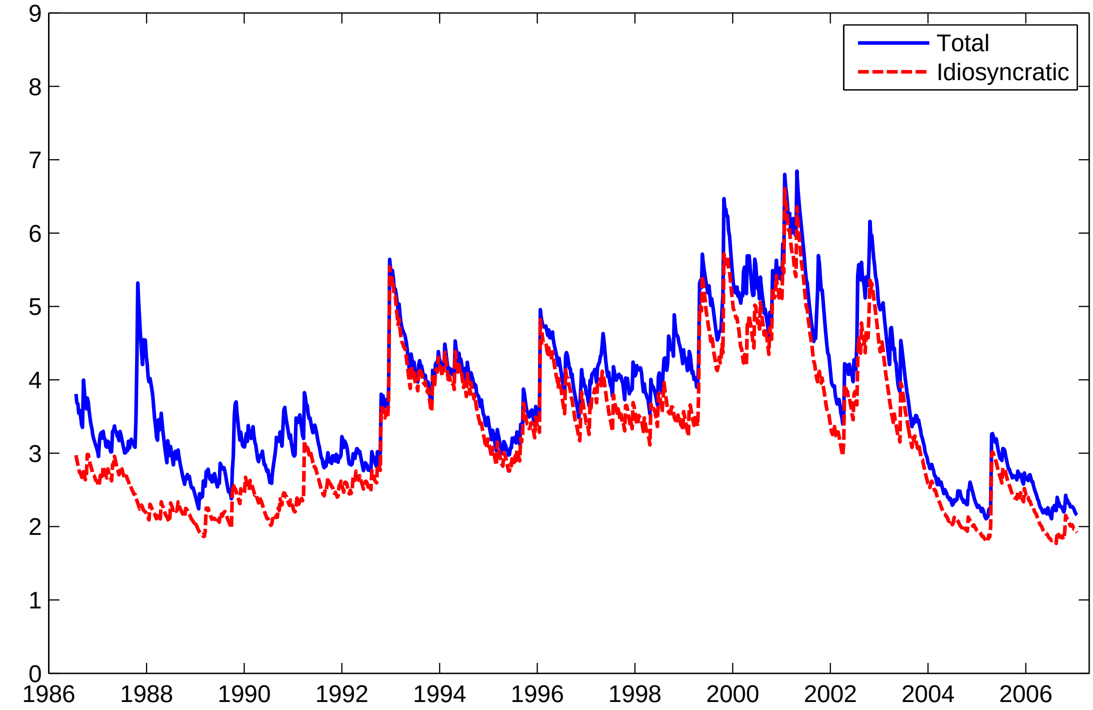
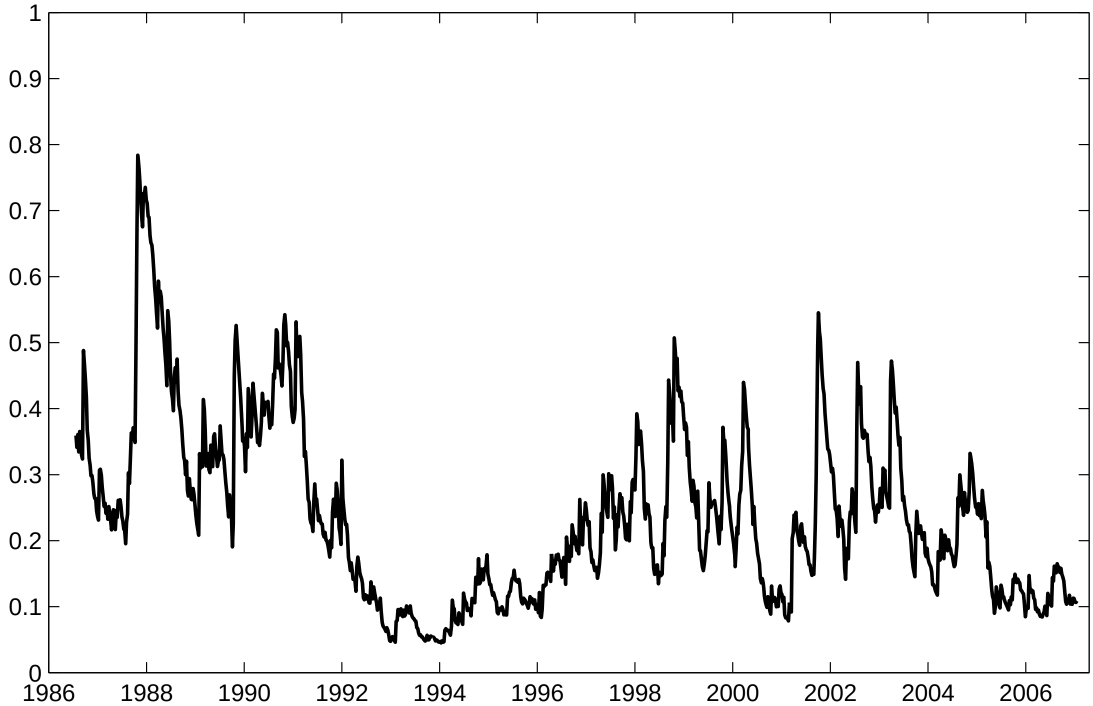


Figure 5d: Conditional Correlation between GM and IBM Stock Returns



Supplemental appendix for

Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

Giorgio Calzolari
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<calzolar@ds.unifi.it>

Gabriele Fiorentini
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
<fiorentini@ds.unifi.it>

Revised: April 2008

B The score of the HRS approximate likelihood function

The log-likelihood function of the HRS model that we consider is given by $\bar{l}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T} l_t(\boldsymbol{\theta})$, where

$$l_t(\boldsymbol{\theta}) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log\left|\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}_t(\boldsymbol{\theta})\right| - \frac{1}{2}\mathbf{x}_t'[\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}_t(\boldsymbol{\theta})]^{-1}\mathbf{x}_t, \quad \text{(B3)}$$

$\boldsymbol{\Lambda}_t(\boldsymbol{\theta})$ is a $k \times k$ diagonal matrix with typical element

$$\lambda_{jt}(\boldsymbol{\theta}) = \varpi_j + \alpha_j[g^2_{jt-1|t-1}(\boldsymbol{\theta}) + \omega_{jjt-1|t-1}(\boldsymbol{\theta})] + \beta_j\lambda_{jt-1}(\boldsymbol{\theta}),$$

with $\varpi_j = (1-\alpha_j-\beta_j)\lambda_j$, $\boldsymbol{\Gamma}_t(\boldsymbol{\theta})$ is a $N \times N$ diagonal matrix with typical element

$$\gamma_{it}(\boldsymbol{\theta}) = \varpi_i^* + \alpha_i^*[v^2_{it-1|t-1}(\boldsymbol{\theta}) + \xi_{iit-1|t-1}(\boldsymbol{\theta})] + \beta_i^*\gamma_{it-1}(\boldsymbol{\theta}),$$

with $\varpi_i^* = (1-\alpha_i^*-\beta_i^*)\gamma_i$ and where $g_{jt|t}(\boldsymbol{\theta})$, $v_{it|t}(\boldsymbol{\theta})$, $\omega_{jlt|t}(\boldsymbol{\theta})$ and $\xi_{ilt|t}(\boldsymbol{\theta})$ are typical elements of the outputs of the Kalman filter updating equations:

$$\begin{aligned}
\mathbf{g}_{t|t}(\boldsymbol{\theta}) &= E(\mathbf{g}_t|X_t;\boldsymbol{\theta}) = \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'[\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}_{\mathbf{t}}(\boldsymbol{\theta})]^{-1}\mathbf{x}_t,\\
\mathbf{v}_{t|t}(\boldsymbol{\theta}) &= E(\mathbf{v}_t|X_t;\boldsymbol{\theta}) = \mathbf{x}_t - \mathbf{C}\mathbf{g}_{t|t}(\boldsymbol{\theta}),\\
\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) &= V(\mathbf{g}_t|X_t;\boldsymbol{\theta}) = \boldsymbol{\Lambda}_t(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'[\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \boldsymbol{\Gamma}_{\mathbf{t}}(\boldsymbol{\theta})]^{-1}\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta}),\\
\boldsymbol{\Xi}_{t|t}(\boldsymbol{\theta}) &= V(\mathbf{v}_t|X_t;\boldsymbol{\theta}) = \mathbf{C}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})\mathbf{C}'.
\end{aligned}$$

Bollerslev and Wooldridge (1992) show that the score function $\mathbf{s}_t(\boldsymbol{\theta}) = \partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ of any multivariate conditionally heteroskedastic dynamic regression model with conditional mean vector $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and conditional covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is given by the following expression:

$$\begin{aligned}
\mathbf{s}_t(\boldsymbol{\theta}) = &\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})[\mathbf{x}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})]\\
&+\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] vec\left\{[\mathbf{x}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})][\mathbf{x}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})]' - \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right\}.
\end{aligned}$$

In our case the first term disappears because $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mathbf{0}$. As for the second, given that the differential of $\boldsymbol{\Sigma}_t$ is

$$d(\mathbf{C}\boldsymbol{\Lambda}_t\mathbf{C}' + \boldsymbol{\Gamma}_t) = (d\mathbf{C})\boldsymbol{\Lambda}_t\mathbf{C}' + \mathbf{C}(d\boldsymbol{\Lambda}_t)\mathbf{C}' + \mathbf{C}\boldsymbol{\Lambda}_t(d\mathbf{C}') + d\boldsymbol{\Gamma}_t, \quad \text{(B4)}$$

(cf. Magnus and Neudecker (1999)), we have that Jacobian of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ will be:

$$\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'} = (\mathbf{I}_{N^2} + \mathbf{K}_{NN})[\mathbf{I}_N \otimes \mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})]\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}'} + \mathbf{E}_N\frac{\partial\boldsymbol{\gamma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + (\mathbf{C}\otimes\mathbf{C})\mathbf{E}_k\frac{\partial\boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

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

After some straightforward algebraic manipulations, we get:

$$\begin{aligned}
\mathbf{s}_t(\boldsymbol{\theta}) &= \frac{\partial \mathbf{c}'}{\partial \boldsymbol{\theta}} vec\left[\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \\
&+ \frac{1}{2}\frac{\partial \boldsymbol{\gamma}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} vecd\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \\
&+ \frac{1}{2}\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} vecd\left[\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{C}\right].
\end{aligned}$$

In view of (9), we will have that:

$$\begin{aligned}
\frac{\partial \lambda_{jt}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \alpha_j\left[2g_{jt-1|t-1}(\boldsymbol{\theta})\frac{\partial g_{jt-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \omega_{jjt-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] + \beta_j\frac{\partial \lambda_{jt-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
&+ \frac{\partial \varpi_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \alpha_j}{\partial \boldsymbol{\theta}}\left[g^2_{jt-1|t-1}(\boldsymbol{\theta}) + \omega_{jjt-1|t-1}(\boldsymbol{\theta})\right] + \frac{\partial \beta_j}{\partial \boldsymbol{\theta}'}\lambda_{jt-1}(\boldsymbol{\theta}),
\end{aligned}$$

where $\varpi_j(\boldsymbol{\theta}) = (1-\alpha_j-\beta_j)\lambda_j$. Similarly, (10) implies that

$$\begin{aligned}
\frac{\partial \gamma_{it}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \alpha_i^*\left[2v_{it-1|t-1}(\boldsymbol{\theta})\frac{\partial v_{it-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \xi_{iit-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] + \beta_i^*\frac{\partial \gamma_{it-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
&+ \frac{\partial \varpi_i^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \alpha_i^*}{\partial \boldsymbol{\theta}}\left[v^2_{it-1|t-1}(\boldsymbol{\theta}) + \xi_{iit-1|t-1}(\boldsymbol{\theta})\right] + \frac{\partial \beta_i^*}{\partial \boldsymbol{\theta}'}\gamma_{it-1}(\boldsymbol{\theta}),
\end{aligned}$$

where $\varpi_i^*(\boldsymbol{\theta}) = (1-\alpha_i^*-\beta_i^*)\gamma_i$. If we impose the restriction $\alpha_i^* = \alpha^*$ and $\beta_i^* = \beta^*$ $\forall i$, then the usual chain rule implies that $s_{\alpha^*t}(\boldsymbol{\theta}) = \sum_{i=1}^N s_{\alpha_i^*t}(\boldsymbol{\theta})$ and $s_{\beta^*t}(\boldsymbol{\theta}) = \sum_{i=1}^N s_{\beta_i^*t}(\boldsymbol{\theta})$.

Finally, it is worth mentioning that if we fix the factor scales by setting $c_{jj} = 1$ instead of $\lambda_j = 1$ for $j = 1, \ldots, k$, then we must exclude the elements of the score corresponding to those factor loadings, and replace them with the derivatives with respect to λ_j, which can be trivially found from the previous expressions because the unconditional variance parameters only appear directly in the expression for the pseudo log-likelihood function $l_t(\boldsymbol{\theta})$ in (B3) through the constant term in the conditional variance expressions, $\varpi_j(\boldsymbol{\theta})$. Either way, since we initialise the conditional variances with $\lambda_{j1}(\boldsymbol{\theta}) = \lambda_j$ and $\gamma_{i1}(\boldsymbol{\theta}) = \gamma_i$, then we must always start up the derivative recursions with $\partial\lambda_{j1}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \partial\lambda_j/\partial\boldsymbol{\theta}$ and $\partial\gamma_{i1}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \partial\gamma_i/\partial\boldsymbol{\theta}$.

If $\boldsymbol{\Gamma}_t > \mathbf{0}$, then we can use the Woodbury formula to prove that

$$\begin{aligned}
\mathbf{g}_{t|t} &= \boldsymbol{\Omega}_{t|t}\mathbf{C}'\boldsymbol{\Gamma}_t^{-1}\mathbf{x}_t, \\
\boldsymbol{\Omega}_{t|t} &= \left(\mathbf{C}'\boldsymbol{\Gamma}_t^{-1}\mathbf{C}+\boldsymbol{\Lambda}_t^{-1}\right)^{-1}, \\
\boldsymbol{\Lambda}_t\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1} - \mathbf{C}\boldsymbol{\Sigma}_t^{-1}\mathbf{C}\boldsymbol{\Lambda}_t &= [\mathbf{g}_{t|t}\mathbf{x}_t - (\mathbf{g}_{t|t}\mathbf{g}_{t|t}' + \boldsymbol{\Omega}_{t|t})\mathbf{C}']\boldsymbol{\Gamma}_t^{-1}, \\
\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1} &= \boldsymbol{\Gamma}_t^{-1}[(\mathbf{x}_t - \mathbf{C}\mathbf{g}_{t|t})(\mathbf{x}_t - \mathbf{C}\mathbf{g}_{t|t})' + \mathbf{C}\boldsymbol{\Omega}_{t|t}\mathbf{C}' - \boldsymbol{\Gamma}_t]\boldsymbol{\Gamma}_t^{-1},
\end{aligned}$$

and

$$\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} = \boldsymbol{\Lambda}_t^{-1}[(\mathbf{g}_{t|t}\mathbf{g}_{t|t}' + \boldsymbol{\Omega}_{t|t}) - \boldsymbol{\Lambda}_t]\boldsymbol{\Lambda}_t^{-1},$$

which greatly simplifies the computations (see Sentana (2000)).

Under the same assumption, the differential of $\mathbf{\Omega}_{t|t}$ will be $-\mathbf{\Omega}_{t|t} d\left(\mathbf{C}'\mathbf{\Gamma}_t^{-1}\mathbf{C}+\mathbf{\Lambda}_t^{-1}\right)\mathbf{\Omega}_{t|t}$, where

$$d\left(\mathbf{C}'\mathbf{\Gamma}_t^{-1}\mathbf{C}+\mathbf{\Lambda}_t^{-1}\right) = (d\mathbf{C}')\mathbf{\Gamma}_t^{-1}\mathbf{C} + \mathbf{C}'\mathbf{\Gamma}_t^{-1}(d\mathbf{C}) - \mathbf{C}'\mathbf{\Gamma}_t^{-1}(d\mathbf{\Gamma}_t)\mathbf{\Gamma}_t^{-1}\mathbf{C} - \mathbf{\Lambda}_t^{-1}(d\mathbf{\Lambda}_t)\mathbf{\Lambda}_t^{-1}.$$

If we call $\boldsymbol{\omega}_{t|t} = vech(\mathbf{\Omega}_{t|t}) = \mathbf{D}_k^+ vec(\mathbf{\Omega}_{t|t})$, where $\mathbf{D}_k$ is the duplication matrix of order k and $\mathbf{D}_k^+$ its Moore-Penrose inverse, then we will have that

$$\begin{aligned}\frac{\partial \boldsymbol{\omega}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \left[-2\frac{\partial \mathbf{c}'}{\partial \boldsymbol{\theta}}(\mathbf{\Gamma}_t^{-1}\mathbf{C}\mathbf{\Omega}_{t|t}\otimes\mathbf{\Omega}_{t|t}) + \frac{\partial \boldsymbol{\gamma}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}'_N(\mathbf{\Gamma}_t^{-1}\mathbf{C}\mathbf{\Omega}_{t|t}\otimes\mathbf{\Gamma}_t^{-1}\mathbf{C}\mathbf{\Omega}_{t|t})\right. \\ &\left. + \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}'_k(\mathbf{\Lambda}_t^{-1}\mathbf{\Omega}_{t|t}\otimes\mathbf{\Lambda}_t^{-1}\mathbf{\Omega}_{t|t})\right]\mathbf{D}_k^{+\prime}.\end{aligned}$$

In addition, the differential of $\mathbf{g}_{t|t}$ when $\mathbf{\Gamma}_t$ has full rank will be given by

$$d\mathbf{g}_{t|t} = (d\mathbf{\Omega}_{t|t})\mathbf{C}'\mathbf{\Gamma}_t^{-1}\mathbf{x}_t + \mathbf{\Omega}_{t|t}d(\mathbf{C}')\mathbf{\Gamma}_t^{-1}\mathbf{x}_t - \mathbf{\Omega}_{t|t}\mathbf{C}'\mathbf{\Gamma}_t^{-1}d(\mathbf{\Gamma}_t)\mathbf{\Gamma}_t^{-1}\mathbf{x}_t.$$

As a result, we will have that

$$\frac{\partial \mathbf{g}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{c}'}{\partial \boldsymbol{\theta}}(\mathbf{\Gamma}_t^{-1}\mathbf{x}_t\otimes\mathbf{\Omega}_{t|t}) - \frac{\partial \boldsymbol{\gamma}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}'_N(\mathbf{\Gamma}_t^{-1}\mathbf{x}_t\otimes\mathbf{\Gamma}_t^{-1}\mathbf{C}\mathbf{\Omega}_{t|t}) + \frac{\partial \boldsymbol{\omega}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{D}'_k(\mathbf{C}'\mathbf{\Gamma}_t^{-1}\mathbf{x}_t\otimes\mathbf{I}_k).$$

Similarly, given that $\mathbf{v}_{t|t} = \mathbf{x}_t - \mathbf{C}\mathbf{g}_{t|t}$, we will have that

$$d\mathbf{v}_{t|t} = -(d\mathbf{C})\mathbf{g}_{t|t} - \mathbf{C}(d\mathbf{g}_{t|t}),$$

whence

$$\frac{\partial \mathbf{v}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial \mathbf{c}'}{\partial \boldsymbol{\theta}}(\mathbf{I}_N\otimes\mathbf{g}_{t|t}) - \frac{\partial \mathbf{g}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{C}'.$$

In addition, since $\mathbf{\Xi}_{t|t} = \mathbf{C}\mathbf{\Omega}_{t|t}\mathbf{C}'$, then

$$d\mathbf{\Xi}_{t|t} = (d\mathbf{C})\mathbf{\Omega}_{t|t}\mathbf{C}' + \mathbf{C}(d\mathbf{\Omega}_{t|t})\mathbf{C}' + \mathbf{C}\mathbf{\Omega}_{t|t}(d\mathbf{C}').$$

Hence, if we call $\boldsymbol{\xi}_{t|t} = vech(\mathbf{\Xi}_{t|t}) = \mathbf{D}_N^+ vec(\mathbf{\Xi}_{t|t})$, then after some algebraic manipulations we will have that

$$\frac{\partial \boldsymbol{\xi}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[2\frac{\partial \mathbf{c}'}{\partial \boldsymbol{\theta}}(\mathbf{I}_N\otimes\mathbf{\Omega}_{t|t}\mathbf{C}') + \frac{\partial \boldsymbol{\omega}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{D}'_k(\mathbf{C}'\otimes\mathbf{C}')\right]\mathbf{D}_N^{+\prime}.$$

If some $\gamma_{it} = 0$, though, the above expressions become invalid. Nevertheless, appropriately modified expressions can be developed along the lines of Sentana (2000). For the sake of brevity, though, we only obtain the score when $rank(\mathbf{\Gamma}_t) = N - k$, so that there are as many Heywood cases as factors. To do so, let us partition the original set of variables in two subsets, say $\mathbf{x}_{at}$ and $\mathbf{x}_{bt}$, of dimensions k and $N - k$ respectively. With this notation, we can re-write the auxiliary model as

$$\begin{pmatrix}\mathbf{x}_{at}\\ \mathbf{x}_{bt}\end{pmatrix} = \begin{pmatrix}\mathbf{C}_a\\ \mathbf{C}_b\end{pmatrix}\mathbf{g}_t + \begin{pmatrix}\mathbf{v}_{at}\\ \mathbf{v}_{bt}\end{pmatrix},$$

where

$$\begin{pmatrix} \mathbf{g}_t \\ \mathbf{v}_{at} \\ \mathbf{v}_{bt} \end{pmatrix} |X_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{at} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}_{bt} \end{pmatrix} \right].$$

In this context, it is convenient to factorise the joint log-likelihood function of $\mathbf{x}_{at}$ and $\mathbf{x}_{bt}$ (given X_{t-1}) as the marginal log-likelihood function of $\mathbf{x}_{at}$ (given X_{t-1}) plus the conditional log-likelihood function of $\mathbf{x}_{bt}$ given $\mathbf{x}_{at}$ (and X_{t-1}). More formally, we can write

$$l_t(\boldsymbol{\theta}) = l_{at}(\boldsymbol{\theta}) + l_{bt|at}(\boldsymbol{\theta}),$$

so that

$$\mathbf{s}_t(\boldsymbol{\theta}) = \mathbf{s}_{at}(\boldsymbol{\theta}) + \mathbf{s}_{bt|at}(\boldsymbol{\theta}).$$

The two log-likelihood components will be given by

$$l_{at}(\boldsymbol{\theta}) = -\frac{k}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma}_{at}| - \frac{1}{2}\mathbf{x}'_{at}\mathbf{\Sigma}_{at}^{-1}\mathbf{x}_{at}$$

and

$$l_{bt|at}(\boldsymbol{\theta}) = -\frac{N-k}{2}\log 2\pi - \frac{1}{2}\log\left|\mathbf{\Sigma}_{bt|at}(\boldsymbol{\theta})\right| - \frac{1}{2}\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta})'\mathbf{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta}),$$

where

$$\begin{aligned}
\mathbf{\Sigma}_{at}(\boldsymbol{\theta}) &= V(\mathbf{x}_{at}|X_{t-1};\boldsymbol{\theta}) = \mathbf{C}_a\mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'_a + \mathbf{\Gamma}_{at}(\boldsymbol{\theta}), \\
\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta}) &= \mathbf{x}_{bt} - \boldsymbol{\mu}_{bt|at}(\boldsymbol{\theta}), \\
\boldsymbol{\mu}_{bt|at}(\boldsymbol{\theta}) &= E(\mathbf{x}_{bt}|\mathbf{x}_{at}, X_{t-1};\boldsymbol{\theta}) = \mathbf{C}_{bt}\mathbf{g}_{t|at}(\boldsymbol{\theta}), \\
\mathbf{g}_{t|at}(\boldsymbol{\theta}) &= E(\mathbf{g}_t|\mathbf{x}_{at}, X_{t-1};\boldsymbol{\theta}) = \mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'_a\mathbf{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at}, \\
\mathbf{\Sigma}_{bt|at}(\boldsymbol{\theta}) &= E(\mathbf{g}_t|\mathbf{x}_{at}, X_{t-1};\boldsymbol{\theta}) = \mathbf{C}_b\mathbf{\Omega}_{t|at}(\boldsymbol{\theta})\mathbf{C}'_b + \mathbf{\Gamma}_{bt}(\boldsymbol{\theta}),
\end{aligned}$$

and

$$\mathbf{\Omega}_{t|at}(\boldsymbol{\theta}) = V(\mathbf{g}_t|\mathbf{x}_{at}, X_{t-1};\boldsymbol{\theta}) = \mathbf{\Lambda}_t(\boldsymbol{\theta}) - \mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'_a\mathbf{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\mathbf{\Lambda}_t(\boldsymbol{\theta}).$$

Therefore, if we partition $\mathbf{c}$ and $\boldsymbol{\gamma}$ as $(\mathbf{c}'_a, \mathbf{c}'_b)'$ and $(\boldsymbol{\gamma}'_a, \boldsymbol{\gamma}'_b)'$, respectively, where $\mathbf{c}_a = vec(\mathbf{C}'_a)$, $\mathbf{c}_b = vec(\mathbf{C}'_b)$, $\boldsymbol{\gamma}_{at} = vecd(\mathbf{\Gamma}_{at})$, and $\boldsymbol{\gamma}_{bt} = vecd(\mathbf{\Gamma}_{bt})$, then we can use the expressions derived before to find

$$\begin{aligned}
\mathbf{s}_{at}(\boldsymbol{\theta}) = {} & \frac{\partial\mathbf{c}'_a}{\partial\boldsymbol{\theta}} vec(\mathbf{\Lambda}_t\mathbf{C}'_a\mathbf{\Sigma}_{at}^{-1}\mathbf{x}_{at}\mathbf{x}'_{at}\mathbf{\Sigma}_{at}^{-1} - \mathbf{\Lambda}_t\mathbf{C}'_a\mathbf{\Sigma}_{at}) + \frac{1}{2}\frac{\partial\boldsymbol{\gamma}'_{at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} vecd(\mathbf{\Sigma}_{at}^{-1}\mathbf{x}_{at}\mathbf{x}'_{at}\mathbf{\Sigma}_{at}^{-1} - \mathbf{\Sigma}_{at}^{-1}) \\
& + \frac{1}{2}\frac{\partial\boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} vecd(\mathbf{C}'_a\mathbf{\Sigma}_{at}^{-1}\mathbf{x}_{at}\mathbf{x}'_{at}\mathbf{\Sigma}_{at}^{-1}\mathbf{C}_a - \mathbf{C}'_a\mathbf{\Sigma}_{at}^{-1}\mathbf{C}_a).
\end{aligned}$$

In order to obtain $\mathbf{s}_{bt|at}(\boldsymbol{\theta})$, though, we first need to find the Jacobian matrices $\partial\boldsymbol{\mu}_{bt|at}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec[\Sigma_{bt|at}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$. Straightforward algebra shows that

$$\frac{\partial\boldsymbol{\mu}'_{bt|at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial\mathbf{c}'_b}{\partial\boldsymbol{\theta}}(\mathbf{I}_{N-k}\otimes\mathbf{g}_{t|at}) + \frac{\partial\mathbf{g}'_{t|at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{C}'_b$$

and

$$\frac{\partial vec'\left[\boldsymbol{\Sigma}_{bt|at}(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{c}_b'}{\partial \boldsymbol{\theta}}(\mathbf{I}_{N-k} \otimes \boldsymbol{\Omega}_{bt|at}\mathbf{C}_b')(\mathbf{I}_{(N-k)^2k^2} + \mathbf{K}_{(N-k)k,(N-k)k}) + \frac{\partial \boldsymbol{\gamma}_{bt}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_{N-k}' + \frac{\partial \boldsymbol{\omega}_{bt|at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{D}_k'(\mathbf{C}_b' \otimes \mathbf{C}_b').$$

Hence,

$$\mathbf{s}_{bt|at}(\boldsymbol{\theta}) = \frac{\partial \mathbf{g}_{t|at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{C}_b'\boldsymbol{\Sigma}_{bt|at}^{-1}\boldsymbol{\varepsilon}_{bt|at} + \frac{\partial \mathbf{c}_b'}{\partial \boldsymbol{\theta}} vec\left(\boldsymbol{\Sigma}_{bt|at}^{-1}\boldsymbol{\varepsilon}_{bt|at}\mathbf{g}_{t|at} + \boldsymbol{\Sigma}_{bt|at}^{-1}\boldsymbol{\varepsilon}_{bt|at}'\boldsymbol{\varepsilon}_{bt|at}\boldsymbol{\Sigma}_{bt|at}^{-1}\mathbf{C}_b\boldsymbol{\Omega}_{t|at} - \boldsymbol{\Sigma}_{bt|at}^{-1}\mathbf{C}_b\boldsymbol{\Omega}_{t|at}\right) + \frac{1}{2}\frac{\partial \boldsymbol{\gamma}_{bt}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} vecd\left(\boldsymbol{\Sigma}_{bt|at}^{-1}\boldsymbol{\varepsilon}_{bt|at}\boldsymbol{\varepsilon}_{bt|at}'\boldsymbol{\Sigma}_{bt|at}^{-1} - \boldsymbol{\Sigma}_{bt|at}^{-1}\right) + \frac{1}{2}\frac{\partial \boldsymbol{\omega}_{bt|at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{D}_k' vec\left(\mathbf{C}_b'\boldsymbol{\Sigma}_{bt|at}^{-1}\boldsymbol{\varepsilon}_{bt|at}\boldsymbol{\varepsilon}_{bt|at}'\boldsymbol{\Sigma}_{bt|at}^{-1}\mathbf{C}_b - \mathbf{C}_b'\boldsymbol{\Sigma}_{bt|at}^{-1}\mathbf{C}_b\right).$$

In this case, the differential of $\mathbf{g}_{t|at}(\boldsymbol{\theta})$ will be

$$d\mathbf{g}_{t|at} = (d\boldsymbol{\Lambda}_t)\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}\mathbf{x}_{at} + \boldsymbol{\Lambda}_t(d\mathbf{C}_a')\boldsymbol{\Sigma}_{at}^{-1}\mathbf{x}_{at} - \boldsymbol{\Lambda}_t\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}(d\boldsymbol{\Sigma}_{at})\boldsymbol{\Sigma}_{at}^{-1}\mathbf{x}_{at},$$

where $d\boldsymbol{\Sigma}_{at}$ is analogous to (B4). As a result,

$$\frac{\partial \mathbf{g}_{t|at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{c}_a'}{\partial \boldsymbol{\theta}}\left[(\boldsymbol{\Sigma}_{at}^{-1}\mathbf{x}_{at} \otimes \boldsymbol{\Omega}_{t|at}) - (\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \mathbf{g}_{t|at})\right] - \frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'(\boldsymbol{\Sigma}_{at}^{-1}\mathbf{x}_{at} \otimes \boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t) + \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'(\boldsymbol{\Lambda}_t^{-1}\mathbf{g}_{t|at} \otimes \boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Omega}_{t|at}).$$

Similarly, the differential of $\boldsymbol{\Omega}_{t|at}$ will be given by

$$d\boldsymbol{\Omega}_{t|at} = (d\boldsymbol{\Lambda}_t) - (d\boldsymbol{\Lambda}_t)\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t(d\mathbf{C}_a')\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t + \boldsymbol{\Lambda}_t\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}(d\boldsymbol{\Sigma}_{at})\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}(d\mathbf{C}_a)\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a(d\boldsymbol{\Lambda}_t).$$

Hence,

$$\frac{\partial \boldsymbol{\omega}_{t|at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left\{-2\frac{\partial \mathbf{c}_a'}{\partial \boldsymbol{\theta}}(\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \boldsymbol{\Omega}_{t|at}) - \frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'(\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t) + \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'[(\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Omega}_{t|at} \otimes \mathbf{I}_k) - (\mathbf{C}_a'\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Omega}_{t|at})]\right\}\mathbf{D}_k'.$$

Finally, we need to obtain $\partial \mathbf{g}_{t|t}'(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $\partial \boldsymbol{\omega}_{t|t}'(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$. But since

$$\mathbf{g}_{t|t}(\boldsymbol{\theta}) = \mathbf{g}_{t|at}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})\mathbf{C}_b'\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}$$

and

$$\boldsymbol{\Omega}_{t|t} = \boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta}) - \boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})\mathbf{C}_b'\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\mathbf{C}_b\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta}),$$

we can obtain the required derivatives by combining the previous expressions.

Fortunately, all the above formulae simplify considerably when $\boldsymbol{\Gamma}_{at} = \mathbf{0}$. Specifically, let $\mathring{\boldsymbol{\theta}}$ denote the value of $\boldsymbol{\theta}$ when $\boldsymbol{\Gamma}_{at} = \mathbf{0}$. Then, it is immediate to see that

$$\begin{aligned}\boldsymbol{\Sigma}_{at}(\mathring{\boldsymbol{\theta}}) &= \mathbf{C}_a\boldsymbol{\Lambda}_t\mathbf{C}_a',\\ \mathbf{g}_{t|at}(\mathring{\boldsymbol{\theta}}) &= \mathbf{C}_a^{-1}\mathbf{x}_{at},\end{aligned}$$

and $\boldsymbol{\Omega}_{t|at}(\mathring{\boldsymbol{\theta}}) = \mathbf{0}$, so that $\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}}) = \mathbf{x}_{bt} - \mathbf{C}_b^*\mathbf{x}_{at}$, with $\mathbf{C}_b^* = \mathbf{C}_b\mathbf{C}_a^{-1}$, and $\boldsymbol{\Sigma}_{bt|at}(\mathring{\boldsymbol{\theta}}) = \boldsymbol{\Gamma}_{bt}$. Moreover,

$$\boldsymbol{\Sigma}_{at}(\mathring{\boldsymbol{\theta}})\mathbf{x}_{at}\mathbf{x}_{at}'\boldsymbol{\Sigma}_{at}(\mathring{\boldsymbol{\theta}}) - \boldsymbol{\Sigma}_{at}(\mathring{\boldsymbol{\theta}}) = \mathbf{C}_a'^{-1}\boldsymbol{\Lambda}_t^{-1}(\mathring{\boldsymbol{\theta}})\left[\mathbf{g}_{t|at}(\mathring{\boldsymbol{\theta}})\mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}}) - \boldsymbol{\Lambda}_t(\mathring{\boldsymbol{\theta}})\right]\boldsymbol{\Lambda}_t^{-1}(\mathring{\boldsymbol{\theta}})\mathbf{C}_a^{-1}.$$

As a result, we can write

$$\mathbf{s}_{at}(\mathring{\boldsymbol{\theta}}) = \frac{\partial \mathbf{c}_a'}{\partial \boldsymbol{\theta}} vec\left[\left(\mathbf{g}_{t|at}\mathbf{g}_{t|at}' - \boldsymbol{\Lambda}_t\right)\boldsymbol{\Lambda}_t^{-1}\mathbf{C}_a^{-1}\right]$$
$$+\frac{1}{2}\frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} vecd[\mathbf{C}_a'^{-1}\boldsymbol{\Lambda}_t^{-1}(\mathbf{g}_{t|at}\mathbf{g}_{t|at}' - \boldsymbol{\Lambda}_t)\boldsymbol{\Lambda}_t^{-1}\mathbf{C}_a^{-1}] + \frac{1}{2}\frac{\partial \boldsymbol{\lambda}_t'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} vecd[\boldsymbol{\Lambda}_t^{-1}(\mathbf{g}_{t|at}\mathbf{g}_{t|at}' - \boldsymbol{\Lambda}_t)\boldsymbol{\Lambda}_t^{-1}]$$

and

$$\mathbf{s}_{bt|at}(\mathring{\boldsymbol{\theta}}) = -\frac{\partial \mathbf{c}_a'}{\partial \boldsymbol{\theta}} vec[\mathbf{C}_b^{*\prime}\boldsymbol{\Gamma}_{bt}^{-1}\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}})\mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})] + \frac{\partial \mathbf{c}_b'}{\partial \boldsymbol{\theta}} vec[\boldsymbol{\Gamma}_{bt}^{-1}\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}})\mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})]$$
$$+\frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left[\frac{1}{2} vecd\{\mathbf{C}_b^{*\prime}\boldsymbol{\Gamma}_{bt}^{-1}[\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}})\boldsymbol{\varepsilon}_{bt|at}'(\mathring{\boldsymbol{\theta}}) - \boldsymbol{\Gamma}_{bt}]\boldsymbol{\Gamma}_{bt}^{-1}\mathbf{C}_b^{*\prime}\} - \mathbf{E}_k'\mathbf{C}_b^{*\prime}\boldsymbol{\Gamma}_{bt}^{-1}\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}})\mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})\boldsymbol{\Lambda}_t^{-1}\mathbf{C}_a^{-1}\right]$$
$$+\frac{1}{2}\frac{\partial \boldsymbol{\gamma}_{bt}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} vecd\{\boldsymbol{\Gamma}_{bt}^{-1}[\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}})\boldsymbol{\varepsilon}_{bt|at}'(\mathring{\boldsymbol{\theta}}) - \boldsymbol{\Gamma}_{bt}]\boldsymbol{\Gamma}_{bt}^{-1}\}.$$

Finally, we obtain

$$\frac{\partial \mathbf{g}_{t|t}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\omega}_{t|at}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\mathbf{D}_k[\mathbf{C}_b^{*\prime}\boldsymbol{\Gamma}_{bt}^{-1}\boldsymbol{\varepsilon}_{bt|at}(\mathring{\boldsymbol{\theta}}) \otimes \mathbf{I}_k]$$

and

$$\frac{\partial \boldsymbol{\omega}_{t|t}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\omega}_{t|at}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}},$$

where

$$\frac{\partial \mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = -\frac{\partial \mathbf{c}_a'}{\partial \boldsymbol{\theta}}[\mathbf{C}_a^{-1\prime} \otimes \mathbf{g}_{t|at}'(\mathring{\boldsymbol{\theta}})] + \frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'[\mathbf{C}_a^{-1\prime}\boldsymbol{\Lambda}_t^{-1}\mathbf{g}_{t|at}(\mathring{\boldsymbol{\theta}}) \otimes \mathbf{C}_a^{-1\prime}],$$

and

$$\frac{\partial \boldsymbol{\omega}_{t|at}'(\mathring{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\gamma}_{at}'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{E}_k'(\mathbf{C}_a^{-1\prime} \otimes \mathbf{C}_a^{-1\prime})\mathbf{D}_k.$$

Although these expressions are strictly speaking only valid when an idiosyncratic variance is identically 0, we recommend their use whenever some γ_{it} is less than .0001 because the expressions for $\boldsymbol{\gamma}_t > \mathbf{0}$ become numerically unreliable for smaller values.