

INDIRECT ESTIMATION OF CONDITIONALLY HETEROSKEDASTIC FACTOR MODELS

Enrique Sentana, Giorgio Calzolari and
Gabriele Fiorentini

CEMFI Working Paper No. 0409

April 2004

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551. Fax (34) 914 291 056
Internet: www.cemfi.es

We are grateful to Monica Billio, Adrian Pagan, Eric Renault, as well as seminar audiences at 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) and the Symposium on New Frontiers in Financial Volatility Modelling (Florence, 2003) 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" is gratefully acknowledged. Thanks are also due to Javier Mencía for his help in producing Figure 3.

CEMFI Working Paper 0409
April 2004

INDIRECT ESTIMATION OF CONDITIONALLY HETEROSKEDASTIC FACTOR MODELS

Abstract

We derive indirect estimators of multivariate conditionally heteroskedastic factor models in which the volatilities of the latent factors depend on their past values. Specifically, we calibrate the analytical score of a Kalman-filter approximation, taking into account the inequality constraints on the auxiliary model parameters. We also study the determinants of the biases in the parameters of this approximation, and its quality. Moreover, we propose sequential indirect estimators that can handle models with large cross-sectional dimensions. Finally, we analyse the small sample behaviour of our indirect estimators and the approximate maximum likelihood procedures through an extensive Monte Carlo experiment.

JEL Codes: C13, C15, C32.

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

Enrique Sentana
CEMFI
sentana@cemfi.es

Giorgio Calzolari
Università di Firenze
calzolar@ds.unifi.it

Gabriele Fiorentini
Università di Firenze
fiorentini@ds.unifi.it

1 Introduction

Many issues in finance require the analysis of the variances and covariances of a large number of security returns. For instance, asset pricing theories often derive restrictions on risk premia from the cross-sectional correlation structure of a very large, possibly infinite, collection of assets. Similarly, portfolio allocation models exploit the imperfect cross sectional correlation of returns in order to reduce risk by means of diversified portfolios. Traditionally, these issues were considered in a static framework, but over the last two decades, the emphasis has gradually shifted to intertemporal models, in which agents' actions are based on the distribution of returns conditional on their time-varying 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 (see e.g. Shephard (1996) for a survey), and numerous applications have already appeared. By and large, though, most applied work in this area has been on univariate financial time series, as the application of these models in a multivariate context has been hampered by the sheer number of parameters involved. In this respect, 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 parametric restrictions to ensure that the resulting conditional covariance matrices are positive definite (see Bauwens, Laurent and Rombouts (2004) for a recent survey).

Given the strong commonality in volatility movements across different financial assets and markets, it is perhaps 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 ($k < N$) common factors plus an idiosyncratic term, but allowing for conditional heteroskedasticity effects in the underlying factors. 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 (APT) of Ross (1976) (see King, Sentana and Wadhvani (1994) (KSW)). In addition, they automatically guarantee a positive (semi-)definite conditional covariance matrix for the observed series, once we ensure that the variances of common and specific factors are non-negative. Finally, if the factors are covariance stationary, these models imply an unconditional factor structure, which makes them compatible with statistical factor analysis (see e.g. Lawley and Maxwell (1971)).

If the conditional variances of the factors are measurable functions of observed variables, as in the factor ARCH model discussed by Engle, Ng and Rothschild (1990) (ENR), full-information maximum likelihood (ML) estimation is time-consuming, but straightforward. In this respect, the evidence presented in Demos and Sentana (1998) clearly shows that when combined with an efficient maximisation procedure, they can be successfully applied to a very large cross-section of time series. In addition, Sentana and Fiorentini (2001) (SF) propose sequential ML estimators, which, although generally inefficient, can consistently estimate the model parameters in a simpler manner.

However, much care has to be exercised in dealing with conditional variances that depend on past values of the 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 underlying factors introduced by Diebold and Nerlove (1989) (DN) is the best known example. In such models, the form of the distribution and moments of the observed variables conditional on their past values alone is 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, whether lognormal or not (see Pitt and Shephard (1999), Aguilar and West (2000), Doz and Renault (2001), Chib, Nardari and Shephard (2002) and Meddahi and Renault (2004)), 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, it does not necessarily follow that one should abandon the use of GARCH processes in models with latent variables,

especially when the degree of unobservability is small. In this respect, it is important to remember that many macro and finance theories are often specified using one-step ahead moments, although those moments are defined with respect to the economic agents' information set, not the econometrician's. Hence, it is perhaps not surprising that economists have built, and continue to build, many models that involve latent GARCH process in order to tackle a number of important empirical problems.

In a recent paper, Fiorentini, Sentana and Shephard (2003) (FSS) develop computationally feasible Markov Chain Monte Carlo (MCMC) simulation algorithms that can be used to obtain exact likelihood-based estimators of unobserved component time series models with GARCH structures, thereby avoiding the inconsistencies associated with the Kalman filter approximations to the log-likelihood function proposed by DN and HRS. In this paper, we analyse alternative simulation-based estimation methods to correct those asymptotic biases, which belong to the class of indirect estimation procedures proposed by Gallant and Tauchen (1996) (GT), Gouriéroux, Monfort and Renault (1993) (GMR) and Smith (1993). In fact, GMR explicitly considered conditionally heteroskedastic factor models as one of their examples, and suggested the use of 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 alternative indirect estimation procedures). Our approach is more closely related to Dungey, Martin and Pagan (2000) (DMP), who also developed indirect estimators for such models. Specifically, they considered two auxiliary parametric models: a “dual” VAR model for the levels and the 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 DN.¹ Although DMP found in a limited Monte Carlo exercise that the latter yields indirect estimators with substantially smaller root mean square errors (RMSE) than the former, they did not use it in their empirical application because it was considerably more demanding from a computational point of view. These authors also found that the biases associated with the Kalman filter

¹DMP explicitly consider models in which not only the squares but also the levels of the common factors are serially correlated. However, since the econometric complications arise exclusively from the dynamics in the second moments, we shall concentrate on models without mean dynamics for ease of exposition.

as direct estimators were fairly small, and that their sampling variability was often smaller than the variability of the indirect estimators based on the dual VAR model.

In this context, our main contributions are as follows:

1. We provide a thorough analysis of the determinants of the biases in the parameters of the Kalman filter approximation put forward by HRS, who refined the DN approach by including a term in the conditional variance of the latent factors that explicitly accounts for their unobservability.
2. We find that the role of those biases is to ensure that in most parameter configurations the model proposed by HRS provides a rather accurate approximation to the distribution of the observable variables conditional on their past values alone.
3. We show that in some limiting cases, the score of the HRS approximation to the log-likelihood function coincides with the exact log-likelihood score, so that our indirect estimators would be as efficient as ML under those circumstances.
4. We provide fast and numerically reliable analytical expressions for computing the score of the approximate log-likelihood function, which are of paramount importance in the implementation of the score-based indirect estimation procedures introduced by GT.
5. We use the “constrained” indirect estimation procedures developed in our earlier work (see Calzolari, Fiorentini and Sentana (2004)) (CFS) to explicitly account for the fact that the parameters of our auxiliary model are subject to inequality restrictions that guarantee the positive (semi-)definiteness of the conditional covariance matrix,
6. We also use the same CFS procedures to deal with the fact that some of the auxiliary model parameters become very poorly identified, if at all, in certain regions of the auxiliary parameter space.
7. We propose sequential indirect estimation procedures in which the effective dimension of the parameter space does not grow with the number of series under consider-

ation, which allows the application of our methods to practical situations in which the cross-sectional dimension is large.

8. We conduct a detailed Monte Carlo experiment to assess the finite sample performance of our two proposed indirect estimators relative to the full-information and sequential approximate ML methods of HRS and SF, respectively.

The paper is organised as follows. We formally introduce the model in section 2, and study the HRS approximation in section 3. Then in section 4, we include a general discussion on the use of indirect estimation in this context. Monte Carlo evidence on the performance of our proposed estimators relative to the approximations usually employed is included in section 5. Finally, our conclusions can be found in section 6. Proofs and some useful auxiliary results can be found in an appendix

2 Conditionally Heteroskedastic Factor Models

2.1 Definition

Consider the following multivariate model:

$$\mathbf{x}_t = \mathbf{B}\mathbf{f}_t + \mathbf{v}_t, \quad (1)$$

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{v}_t \end{pmatrix} | I_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Delta_t & \mathbf{0} \\ \mathbf{0} & \Psi \end{pmatrix} \right], \quad (2)$$

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 $N \geq k$ and $\text{rank}(\mathbf{B}) = k$, \mathbf{v}_t is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to \mathbf{f}_t , Ψ is a $N \times N$ diagonal positive semidefinite (p.s.d.) matrix of constant idiosyncratic variances, Δ_t is a $k \times k$ diagonal positive definite (p.d.) matrix of time-varying factor variances, and I_{t-1} is an information set that contains the values of \mathbf{x}_t and \mathbf{f}_t up to, and including time $t - 1$. These assumptions imply that the distribution of \mathbf{x}_t conditional on I_{t-1} is $N(\mathbf{0}, \Sigma_t)$, where the conditional covariance matrix $\Sigma_t = \mathbf{B}\Delta_t\mathbf{B}' + \Psi$ 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 (CH).

Such a formulation nests several models widely used in the empirical literature, which typically assume that the unobserved factors follow dynamic heteroskedastic processes, but differ in the particular functional form of the conditional variances of the factors, $\delta_{jt} = V(f_{jt}|I_{t-1})$ ($j = 1, \dots, k$), which generally involve some extra parameters, $\boldsymbol{\nu}$ (see Sentana (1998)). For instance, in the latent factor model with ARCH effects on the underlying factors introduced by DN, the conditional variances of the common factors are parametrised as univariate *strong* ARCH models, in the sense of Drost and Nijman (1993). In particular, for a covariance stationary GARCH(1,1) formulation,

$$\delta_{jt} = (1 - \phi_j - \rho_j)\delta_j + \phi_j f_{jt-1}^2 + \rho_j \delta_{jt-1}, \quad (j = 1, \dots, k) \quad (3)$$

with $\delta_j = E(\delta_{jt}|\boldsymbol{\varrho}) = V(f_{jt}|\boldsymbol{\varrho})$, where $E(\cdot|\boldsymbol{\varrho})$ represents expected values taken with respect to the model (1), (2) and (3) evaluated at the parameter values $\boldsymbol{\varrho} = (\mathbf{b}', \boldsymbol{\psi}', \boldsymbol{\delta}', \boldsymbol{\nu}')'$, $\mathbf{b} = \text{vec}(\mathbf{B}') = (\mathbf{b}'_1, \dots, \mathbf{b}'_N)'$, $\mathbf{b}_i = (b_{i1}, \dots, b_{ik})'$, $\boldsymbol{\psi} = \text{vecd}(\boldsymbol{\Psi}) = (\psi_1, \dots, \psi_N)'$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)'$ and $\boldsymbol{\nu} = (\nu'_1, \dots, \nu'_k)'$, with $\nu_j = (\phi_j, \rho_j)'$. In this respect, since (1) can be equivalently written as

$$\mathbf{x}_t = (\mathbf{B}\boldsymbol{\Delta}^{1/2})(\boldsymbol{\Delta}^{-1/2}\mathbf{f}_t) + \mathbf{v}_t,$$

where $\text{vecd}(\boldsymbol{\Delta}) = \boldsymbol{\delta}$, it is necessary to impose restrictions on either \mathbf{B} or $\boldsymbol{\Delta}$ to eliminate such a scale indeterminacy. For instance, we could impose either $\delta_j = 1$,² or $b_{jj} = 1$ ($j = 1, \dots, k$). As a result, the total number of parameters to estimate, d say, will be $Nk + N + 2k$.

2.2 Properties

A non-trivial advantage of the multivariate model (1)-(2) is that it automatically guarantees a p.d. covariance matrix for \mathbf{x}_t once we ensure that both $\boldsymbol{\Delta}_t$ and $\boldsymbol{\Psi}$ are p.s.d., which obviously requires $\psi_i \geq 0 \forall i$. In this respect, note that $\boldsymbol{\Sigma}_t$ cannot be p.d. if $k + 1$ or more elements of $\boldsymbol{\Psi}$ are zero. Moreover, given that $\boldsymbol{\Psi}$ is diagonal, it is straightforward to see that a necessary and sufficient condition for $\boldsymbol{\Sigma}_t$ to be p.d. when $\boldsymbol{\Psi}$ is p.s.d but not

²If the unconditional variance is unbounded, as in integrated GARCH-type models, other symmetric scaling assumptions can be made. For instance, we could choose $\inf_t \delta_{jt} = 1$, or simply fix to 1 the constant element of δ_{jt} . In any case, note that in principle there is no need to set to zero the strict upper triangle of the factor loading matrix \mathbf{B} in view of the identification results in SF.

p.d. is that the matrix formed with the rows of \mathbf{B} corresponding to the zero elements of Ψ must have full row rank (see Sentana (1998, 2000)). In fact, the rank of Ψ is related to the observability of the factors. If $\text{rank}(\Psi) = N - k$, then the factors would be fully revealed by the \mathbf{x}_t variables, otherwise they are only partially revealed (see KSW). As for δ_{jt} , sufficient conditions can be obtained by re-writing (3) as a rational distributed lag of past squared values of the factors. For instance, since a covariance stationary GARCH(1,1) model can be written as

$$\delta_{jt} = \frac{\varsigma_j}{1 - \rho_j} + \phi_j \sum_{s=1}^{\infty} \rho_j^s f_{jt-s}^2,$$

with $\varsigma_j = (1 - \phi_j - \rho_j)\delta_j$, the positivity restrictions become $\varsigma_j > 0$, $\phi_j \geq 0$ and $\rho_j \geq 0$, together with the stationarity condition $\phi_j + \rho_j < 1$ (see Nelson and Cao (1992) for other higher order models). This implies that the admissible parameter space will be inequality restricted

Another significant characteristic of this model is that it is dimension independent, in the sense that it remains valid for any subset of \mathbf{x}_t . But the most distinctive feature of factor models is that they provide a parsimonious specification of the dynamic and cross-sectional dependence of a vector of observable random variables. In this case in particular, the factor structure, together with the constancy of Ψ , implies that the time-variation of Σ_t is of reduced rank (see ENR, and Engle and Susmel (1993) for a discussion in terms of the common features literature). Specifically, assume for simplicity that Ψ is non-singular, and define the following full rank transformation of the observed series \mathbf{x}_t :

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} = \begin{pmatrix} \Upsilon \mathbf{B}' \Psi^{-1} \\ \mathbf{U}'_1 [\mathbf{I}_N - \mathbf{B} \Upsilon \mathbf{B}' \Psi^{-1}] \end{pmatrix} \mathbf{x}_t = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0} \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \Upsilon \mathbf{B}' \Psi^{-1} \\ \mathbf{U}'_1 [\mathbf{I}_N - \mathbf{B} \Upsilon \mathbf{B}' \Psi^{-1}] \end{pmatrix} \mathbf{v}_t,$$

where $\Upsilon = (\mathbf{B}' \Psi^{-1} \mathbf{B})^{-1}$, and $\mathbf{U}_1 \mathbf{M}_1 \mathbf{U}'_1$ provides the spectral decomposition of the rank $N - k$ matrix $F = \Psi - \mathbf{B} \Upsilon \mathbf{B}'$. Note that if we think of the CH model as a cross-sectional heteroskedastic regression of x_{it} on the “regressors” \mathbf{b}_i with regression parameters \mathbf{f}_t and residual variances ψ_i ($i = 1, \dots, N$), then \mathbf{y}_t corresponds to the generalised least squares (GLS) estimator of \mathbf{f}_t ,³ while \mathbf{z}_t contains the first $N - k$ principal components of the GLS residuals $\mathbf{x}_t - \mathbf{B} \mathbf{y}_t$. Notice also that despite the diagonality of Δ_t , the elements of \mathbf{y}_t are

³These factor mimicking portfolios are usually known as Barlett scores in the multivariate statistical analysis literature (see e.g. Sentana (2004)).

contemporaneously correlated unless Υ is itself diagonal, but with constant conditional covariances. Given that

$$\mathbf{y}_t = \mathbf{f}_t + \Upsilon \mathbf{B}' \Psi^{-1} \mathbf{v}_t = \mathbf{f}_t + \boldsymbol{\eta}_t,$$

$$\begin{pmatrix} \mathbf{f}_t \\ \boldsymbol{\eta}_t \\ \mathbf{z}_t \end{pmatrix} | I_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Delta_t & 0 & 0 \\ 0 & \Upsilon & 0 \\ 0 & 0 & \mathbf{M}_1 \end{pmatrix} \right],$$

(see Gouriéroux, Monfort and Renault (1991)), and that Δ_t is a function of lag values of \mathbf{f}_t only, it is clear that the stochastic process $\{\mathbf{y}_t\}$ can be regarded as a minimal set of “sufficient statistics” for $\{\mathbf{f}_t\}$ because it contains the same information about $\{\mathbf{f}_t\}$ as $\{\mathbf{x}_t\}$ (see FSS). In addition, this means that the degree of unobservability of the factors depends on \mathbf{B} and Ψ exclusively through the magnitude of Υ relative to the unconditional variance of the factors, Δ .

Finally, notice that if \mathbf{f}_t is 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 is conditionally heteroskedastic, provided that it is covariance stationary, it also implies an unconditional exact k factor structure for \mathbf{x}_t . That is, the unconditional covariance matrix, Σ , can be written as:

$$\Sigma = E(\Sigma_t | \boldsymbol{\varrho}) = \mathbf{B} \Delta \mathbf{B}' + \Psi. \quad (4)$$

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

2.3 Estimation issues

From the estimation point of view, it is of the utmost importance to distinguish between $I_{t-1} = \{\mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \mathbf{x}_{t-2}, \mathbf{f}_{t-2}, \dots\}$, and the econometrician’s information set $X_{t-1} = \{\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots\}$, which only includes lagged values of \mathbf{x}_t , where $I_{t-1} \equiv X_{t-1} \cup F_{t-1}$, with $F_{t-1} = \{\mathbf{f}_{t-1}, \mathbf{f}_{t-2}, \dots\}$. If the diagonal elements of Δ_t were measurable functions of X_{t-1} (as in section 3), then the distribution of \mathbf{x}_t conditional on X_{t-1} would be normal, and the parameters of interest, $\boldsymbol{\varrho}$, could be estimated simultaneously on the basis of the log-likelihood function of the observed variables, \mathbf{x}_t . However, when the diagonal elements

of Δ_t are functions of lagged values of \mathbf{f}_t , as in (3), 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 can no longer be obtained explicitly except in simple cases, such as when the conditional variances of all the common factors are in fact constant, which happens when $\boldsymbol{\nu} = \mathbf{0}$ (see HRS).

One attractive solution is to employ the MCMC method recently put forward by FSS, which allows the calculation of a ML estimators via the simulated EM algorithm, as well as a Bayesian approach.

Nevertheless, if we could at least find the first two moments of the conditional distribution of \mathbf{x}_t given X_{t-1} (and $\boldsymbol{\varrho}$), then we could also obtain a root- T consistent, albeit inefficient, estimator of $\boldsymbol{\varrho}$ by maximising the Gaussian pseudo likelihood function of \mathbf{x}_t , as advocated by Bollerslev and Wooldridge (1992) among others. Given our assumptions, it trivially follows that $E(\mathbf{x}_t|X_{t-1}; \boldsymbol{\varrho}) = \mathbf{0}$. As for the conditional variances, it is also clear that

$$\begin{aligned} V(f_{jt}|X_{t-1}; \boldsymbol{\varrho}) &= E(\delta_{jt}|X_{t-1}; \boldsymbol{\varrho}) = \varsigma_j + \phi_j E(f_{jt-1}^2|X_{t-1}; \boldsymbol{\varrho}) + \rho_j E(\delta_{jt-1}|X_{t-1}; \boldsymbol{\varrho}) \\ &= \frac{\varsigma_j}{1 - \rho_j} + \phi_j \sum_{s=1}^{\infty} \rho_j^s [E^2(f_{jt-s}|X_{t-1}; \boldsymbol{\varrho}) + V(f_{jt-s}|X_{t-1}; \boldsymbol{\varrho})]. \end{aligned} \quad (5)$$

Therefore, we would need to compute $E(f_{jt-s}|X_{t-1}; \boldsymbol{\varrho})$ and $V(f_{jt-s}|X_{t-1}; \boldsymbol{\varrho})$, a task for which the Kalman filter seems to be ideally suited. In this respect, it is important to note that the conditionally heteroskedastic factor model in (1)-(2) has a natural time-series state-space representation, with \mathbf{f}_t as the state, (1) as the measurement equation, and

$$\mathbf{f}_t = \mathbf{0} \cdot \mathbf{f}_{t-1} + \mathbf{f}_t$$

as the transition equation. Unfortunately, while the conditional distribution of \mathbf{f}_t given \mathbf{x}_t and I_{t-1} is normal, with mean and variance given by

$$E(\mathbf{f}_t|\mathbf{x}_t, I_{t-1}; \boldsymbol{\varrho}) = \Delta_t \mathbf{B}' (\mathbf{B} \Delta_t \mathbf{B}' + \boldsymbol{\Psi})^{-1} \mathbf{x}_t \quad (6)$$

and

$$V(\mathbf{f}_t|\mathbf{x}_t, I_{t-1}; \boldsymbol{\varrho}) = \Delta_t - \Delta_t \mathbf{B}' (\mathbf{B} \Delta_t \mathbf{B}' + \boldsymbol{\Psi})^{-1} \mathbf{B} \Delta_t, \quad (7)$$

these expressions do not usually give us $E(\mathbf{f}_t|X_t; \boldsymbol{\varrho})$ or $V(\mathbf{f}_t|X_t; \boldsymbol{\varrho})$, which must be generally computed by simulation (see FSS). Hence, Gaussian pseudo-ML estimation is not generally feasible either.

One thing we can do, though, is to estimate the parameters characterising the unconditional covariance matrix $\boldsymbol{\Sigma}$. Specifically, SF show that if (a) $\boldsymbol{\Psi}$ and \mathbf{B} are identified (up to rotation) from unconditional moments (b) $\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} \boldsymbol{\Sigma}^0 = \mathbf{B}^0 \mathbf{B}^{0'} + \boldsymbol{\Psi}^0$, where the superscript 0 denotes the true values of the parameters, (c) $T^{-1/2} \sum_{t=1}^T \text{vech}(\mathbf{x}_t \mathbf{x}_t' - \boldsymbol{\Sigma}^0)$ has a limiting normal distribution and (d) the matrix $(F^0 \odot F^0)$ is nonsingular, where F^0 is the rank $N - k$ covariance matrix of the GLS predictors of the idiosyncratic factors $\mathbf{x}_t - \mathbf{B}\mathbf{y}_t$ described in section 2.2, and \odot denotes Hadamard (or element by element) product of two matrices of the same dimensions, then theorem 12.1 in Anderson and Rubin (1956) and theorem 2 in Kano (1983) imply that

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

are asymptotically normally distributed around $\text{vec}(\mathbf{B}^0 \mathbf{Q}^0)$ and $\text{vecd}(\boldsymbol{\Psi}^0)$, where \mathbf{Q}^0 is the orthogonal matrix that imposes on \mathbf{B}^0 the restrictions used in estimation to avoid the usual rotational indeterminacy.⁴ Note that $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$ correspond to the values of the unconditional variance parameters estimated by a standard factor analytic routine. However, we would need to take into account the serial correlation in $\text{vech}(\mathbf{x}_t \mathbf{x}_t')$ in order to compute the standard errors of $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$.

3 HRS alternative to CH Models

3.1 Definition, Properties and Estimation

Consider now the following closely related model for \mathbf{x}_t proposed by HRS as a user-friendly version of (1), (2) and (3):

$$\mathbf{x}_t = \mathbf{C}\mathbf{g}_t + \mathbf{w}_t \tag{8}$$

⁴Obviously, if the estimators of \mathbf{b} are only consistent for $\text{vec}(\mathbf{B}^0 \mathbf{Q}^0)$ because \mathbf{B} is not uniquely identifiable from the unconditional covariance matrix (e.g. if $k \geq 2$ and \mathbf{B} unrestricted), then consistency is only guaranteed for $\boldsymbol{\psi}$.

$$\begin{pmatrix} \mathbf{g}_t \\ \mathbf{w}_t \end{pmatrix} | I_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{pmatrix} \right] \quad (9)$$

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{w}_t is another $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to \mathbf{g}_t , $\mathbf{\Gamma}$ is the $N \times N$ diagonal p.s.d. matrix of constant idiosyncratic variances, and $\mathbf{\Lambda}_t$ is a $k \times k$ diagonal p.d. matrix of time-varying factor variances. By analogy with (5), HRS assumed that

$$\lambda_{jt}(\boldsymbol{\theta}) = (1 - \alpha_j - \beta_j)\lambda_j + \alpha_j[g_{jt-1|t-1}^2(\boldsymbol{\theta}) + \omega_{jjt-1t-1}(\boldsymbol{\theta})] + \beta_j\lambda_{jt-1}(\boldsymbol{\theta}), \quad (10)$$

with $\lambda_j = E[\lambda_{jt}(\boldsymbol{\theta})|\boldsymbol{\theta}] = V(g_{jt}|\boldsymbol{\theta})$, where $E(\cdot|\boldsymbol{\theta})$ represents expected values taken with respect to the model (8), (9) and (10) evaluated at the parameter values $\boldsymbol{\theta} = (\mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\lambda}', \boldsymbol{\pi}')'$, $\mathbf{c} = \text{vec}(\mathbf{C}') = (\mathbf{c}'_1, \dots, \mathbf{c}'_N)'$, $\mathbf{c}_i = (c_{i1}, \dots, c_{ik})'$, $\boldsymbol{\gamma} = \text{vecd}(\mathbf{\Gamma}) = (\gamma_1, \dots, \gamma_N)'$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}'_1, \dots, \boldsymbol{\pi}'_k)'$, with $\boldsymbol{\pi}_j = (\alpha_j, \beta_j)'$, and where $g_{jt-1|t-1}(\boldsymbol{\theta})$ and $\omega_{jjt-1t-1}(\boldsymbol{\theta})$ are typical elements of

$$\mathbf{g}_{t|t}(\boldsymbol{\theta}) = E(\mathbf{g}_t|X_t; \boldsymbol{\theta}) = \mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'[\mathbf{C}\mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \mathbf{\Gamma}]^{-1}\mathbf{x}_t,$$

and

$$\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) = V(\mathbf{g}_t|X_t; \boldsymbol{\theta}) = \mathbf{\Lambda}_t(\boldsymbol{\theta}) - \mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'[\mathbf{C}\mathbf{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \mathbf{\Gamma}]^{-1}\mathbf{C}\mathbf{\Lambda}_t(\boldsymbol{\theta}).$$

In fact, it is straightforward to see that $E(\mathbf{g}_t|X_T; \boldsymbol{\theta}) = \mathbf{g}_{t|t}(\boldsymbol{\theta})$ and $V(\mathbf{g}_t|X_T; \boldsymbol{\theta}) = \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})$ since smoothing is unnecessary in this model because (conditional on the parameters $\boldsymbol{\theta}$) there are no dynamics in the mean specification of the factors, and therefore, no serial correlation in their filtered estimates (see DN).

Not surprisingly, models (1)-(3) and (8)-(10) share many important features, including the scale indeterminacy of the common factors mentioned in section 2.1. As a result, the total number of parameters to estimate after setting the scaling of the \mathbf{g}_t common factors will also be $Nk + N + 2k = d$. Moreover, it is straightforward to see that the positive definiteness of the conditional covariance of \mathbf{x}_t is guaranteed if $\gamma_i \geq \mathbf{0}$ ($i = 1, \dots, N$) and $\lambda_{jt}(\boldsymbol{\theta}) > 0$ ($j = 1, \dots, k$) (cf. section 2.2). But since we can rewrite (10) as

$$\lambda_{jt}(\boldsymbol{\theta}) = \frac{\varpi_j}{1 - \beta_j} + \alpha_j \sum_{s=1}^{\infty} \beta_j^s [g_{jt-s|t-s}^2(\boldsymbol{\theta}) + \omega_{jjt-s|t-s}(\boldsymbol{\theta})],$$

where $\varpi_j = (1 - \alpha_j - \beta_j)\lambda_j$, it is once more clear that we need $\varpi_j > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, and $\alpha_j + \beta_j < 1$.

The dynamics of the conditional second moments are also governed in this case by k linear transformations of the N observed series \mathbf{x}_t . Specifically, if we assume for simplicity that $\mathbf{\Gamma}$ has full rank, and define the GLS factor representing portfolios and the associated set of GLS idiosyncratic errors as

$$\begin{aligned}\mathbf{g}_t^G(\mathbf{c}, \gamma) &= \mathbf{\Xi}\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{x}_t, \\ \mathbf{w}_t^G(\mathbf{c}, \gamma) &= (\mathbf{I} - \mathbf{\Xi}\mathbf{C}'\mathbf{\Gamma}^{-1})\mathbf{x}_t,\end{aligned}$$

where $\mathbf{\Xi} = (\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})^{-1}$, then we can write

$$\begin{aligned}\mathbf{x}_t &= \mathbf{C}\mathbf{g}_t^G(\mathbf{c}, \gamma) + \mathbf{w}_t^G(\mathbf{c}, \gamma), \\ \mathbf{g}_t^G(\mathbf{c}, \gamma) &= \mathbf{g}_t + \mathbf{\Xi}\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{w}_t = \mathbf{g}_t + \boldsymbol{\zeta}_t,\end{aligned}$$

and

$$\begin{pmatrix} \mathbf{g}_t \\ \boldsymbol{\zeta}_t \\ \mathbf{w}_t^G(\mathbf{c}, \gamma) \end{pmatrix} \mid I_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Lambda}_t & 0 & 0 \\ 0 & \mathbf{\Xi} & 0 \\ 0 & 0 & \mathbf{\Gamma} - \mathbf{C}\mathbf{\Xi}\mathbf{C}' \end{pmatrix} \right].$$

Nevertheless, the crucial difference with the model discussed in the previous section is that in this case the distribution of \mathbf{g}_t conditional in X_{t-1} is normal, with conditional covariance matrix $\boldsymbol{\Lambda}_t(\boldsymbol{\theta})$. Hence, the average log-likelihood function for a sample of size T based on model (8)-(10) can be written in closed form as $\bar{l}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T l_t(\boldsymbol{\theta})$, where

$$l_t(\boldsymbol{\theta}) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \mathbf{\Gamma}| - \frac{1}{2} \mathbf{x}_t' [\mathbf{C}\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}' + \mathbf{\Gamma}^{-1}] \mathbf{x}_t. \quad (11)$$

As a result, ML estimates of $\boldsymbol{\theta}$ will be obtained by numerically maximising (11), where we can set $\lambda_{j1} = \lambda_j \forall j$ to start up the conditional variance recursions.

However, it is important to realise that the usual definition of pseudo-ML estimators $\hat{\boldsymbol{\theta}}_T$ as a root of the unrestricted first order conditions $\sum_{t=1}^T \partial l_t(\hat{\boldsymbol{\theta}}_T) / \partial \boldsymbol{\theta} = \mathbf{0}$ is not necessarily valid in this case, as it ignores the restrictions on γ and the GARCH parameters $\boldsymbol{\pi}$ that we impose to ensure that the conditional covariance matrix remains p.s.d. almost surely. Therefore, it is more appropriate to use constrained optimisation algorithms that take into account such restrictions (see Demos and Sentana (1998) and Sentana (2000) for

alternative optimisation strategies). In fact, it turns out that solutions at the boundary of the admissible parameter space occur frequently in practice (see section 5). An additional problem that (10) 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.

3.2 Relationship between the original CH model and the HRS alternative

HRS explicitly relate the properties of the elements of g_{jt} and f_{jt} when both models are considered as the true DGP's.⁵ In particular, they show that the first two unconditional moments of g_{jt} and f_{jt} will be identical if $\boldsymbol{\theta} = \boldsymbol{\varrho}$. In contrast, they also show that

$$E(g_{jt}^4|\boldsymbol{\theta} = \boldsymbol{\varrho}) \leq E(f_{jt}^4|\boldsymbol{\varrho}).$$

Intuitively, the reason is the following. It is well known from the Kalman filter literature that the properties of $g_{jt|t}(\boldsymbol{\theta})$ are generally different from the time series properties of g_{jt} (see e.g. Pagan (1980)). Although in our case both \mathbf{g}_t and $\mathbf{g}_{t|t}(\boldsymbol{\theta})$ are serially uncorrelated because the transition equation is degenerate, $\mathbf{g}_{t|t}(\boldsymbol{\theta})$ must be smoother than \mathbf{g}_t . Hence, it is not surprising that the variability of λ_{jt} is generally smaller than the variability of δ_{jt} .

Finally, HRS also show that

$$\text{cor}(g_{jt}^2, g_{jt-1}^2|\boldsymbol{\theta} = \boldsymbol{\varrho}) \leq \text{cor}(f_{jt}^2, f_{jt-1}^2|\boldsymbol{\varrho}),$$

and hence that

$$\text{cov}(g_{jt}^2, g_{jt-1}^2|\boldsymbol{\theta} = \boldsymbol{\varrho}) \leq \text{cov}(f_{jt}^2, f_{jt-1}^2|\boldsymbol{\varrho}).$$

Therefore, while the two processes share both static and dynamic second moments, all the fourth moments of g_{jt} are bounded from above by the corresponding moments of f_{jt} .

Nevertheless, there are two circumstances in which both processes give rise to exactly the same log-likelihood functions. The first one trivially arises when $\phi_j = \alpha_j = 0 \forall j$. The

⁵Although HRS only presented formal proofs for the ARCH(1) case, their results could be easily extended to higher order processes.

second one, when there are exactly as many Heywood cases as common factors, which as we saw before would happen when there are k diagonal elements of Ψ that are equal to 0 (see Sentana (2000)), so that the factors are fully revealed by the observable variables. In addition, there is a third instance in which the two processes become arbitrarily close. Specifically, Sentana (2004) shows that both $\mathbf{g}_{t|t}(\boldsymbol{\theta})$ and $\mathbf{g}_t^G(\mathbf{c}, \boldsymbol{\gamma})$ converge in mean square to \mathbf{g}_t as N increases. Although strictly speaking this is a limiting result, in practice, it means that the two models can become almost identical for large enough N . In fact, given that the quantity that effectively measures the degree of observability of the factors is $\Upsilon = (\mathbf{B}'\Psi^{-1}\mathbf{B})^{-1}$, the size of N is effectively irrelevant, and all we need is that Υ be small.

The similarity between the two models is stronger than what the above discussion may suggest, as explained in the following result:

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 (8), (9) and (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), (2) and (3). Then $\mathbf{s}_t(\boldsymbol{\theta}) = \mathbf{q}_t(\boldsymbol{\varrho})$ when $\boldsymbol{\theta} = \boldsymbol{\varrho}$ and $\boldsymbol{\nu} = \boldsymbol{\pi} = \mathbf{0}$.*

In other words, the HRS model provides a first order approximation to the original CH model when δ_{jt} does not vary much.

In addition, we conjecture that an analogous result holds when the factors are fully revealed by the observable variables, regardless of the time-variation in the volatility of the common factors. Unfortunately, we cannot present a second formal proposition because 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 those elements of the two scores are identical in that situation as well.⁶

In other cases, though, the exact relationship between the two models is unknown, and can only be assessed by simulation. For practical purposes, the relevant comparison is between model (8), (9) and (10) evaluated at the values of $\boldsymbol{\theta}$ that provide the best

⁶Specifically, we have numerically checked whether $\partial E[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]/\partial \boldsymbol{\varrho}'$ is equal to $V[\mathbf{s}_t(\boldsymbol{\theta})|\boldsymbol{\varrho}]$ 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.

approximation to the values of $\boldsymbol{\varrho}$. Given that $\boldsymbol{\theta}$ will typically be estimated by maximising (11), then the comparison should be conducted at the pseudo-true ML values, $\boldsymbol{\theta}(\boldsymbol{\varrho})$ say (see e.g. White (1982)).

In order to find out how the (asymptotic) binding functions $\boldsymbol{\theta}(\boldsymbol{\varrho})$ behaves, we have carried out the following Monte Carlo experiments. We have simulated 400 samples of size 50,000 of the following trivariate single factor model:

$$x_{it} = b_i f_t + v_{it} \quad (i = 1, 2, 3).$$

Given that the effect of \mathbf{b} and $\boldsymbol{\Psi}$ is mostly through the scalar quantity $v = (\mathbf{b}'\boldsymbol{\Psi}^{-1}\mathbf{b})^{-1}$, as previously discussed, all the models considered have $\mathbf{b} = (1, 1, 1)'$, $\boldsymbol{\Psi} = \psi\mathbf{I}$ and $\delta_t = (1 - \phi - \rho) + \phi f_{t-1}^2 + \rho\delta_{t-1}$. In order to minimise experimental error, we use the same set of underlying random numbers in all designs. For scaling purposes, we set $c_3 = 1$ and unrestricted the unconditional variance of the common factor λ . Maximisation of the pseudo log-likelihood (11) with respect to $\boldsymbol{\theta} = (c_1, c_2, \lambda, \gamma_1, \gamma_2, \gamma_3, \alpha, \beta)$ was carried out using the NAG library E04JBL routine.

We have combined 10 different values of ψ with ten different pairs of (ϕ, ρ) . For the sake of brevity, though, 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 ones obtained with weekly observations. The relevant response surfaces for ϕ and ρ are depicted in the two panels of Figure 1, respectively.

As can be clearly seen from Figures 1a and 1b, 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 η_t relative to the variance of f_t . However, these figures also illustrate the fact that 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.

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

(cf. Jacquier, Polson and Rossi (1994)), which is related to the unconditional coefficient

of kurtosis of the latent factor

$$\kappa(\boldsymbol{\varrho}) = \frac{E(f_{jt}^4|\boldsymbol{\varrho})}{\delta_j^2} = \frac{3[1 - (\phi + \rho)^2]}{(1 - 3\phi^2 - 2\phi\rho - \delta^2)}.$$

In this respect, it is worth mentioning that for the four reported configurations of (ϕ, ρ) , we obtain the following values:

$$\begin{array}{llllll} \varkappa(\boldsymbol{\varrho}) = 0.51 & \kappa(\boldsymbol{\varrho}) = 3.77 & \phi = .1 & \rho = .85 & \phi + \rho = .95 \\ \varkappa(\boldsymbol{\varrho}) = 0.54 & \kappa(\boldsymbol{\varrho}) = 3.86 & \phi = .2 & \rho = .6 & \phi + \rho = .8 \\ \varkappa(\boldsymbol{\varrho}) = 2.14 & \kappa(\boldsymbol{\varrho}) = 16.7 & \phi = .2 & \rho = .75 & \phi + \rho = .95 \\ \varkappa(\boldsymbol{\varrho}) = 2.83 & \kappa(\boldsymbol{\varrho}) = 27.0 & \phi = .4 & \rho = .4 & \phi + \rho = .8 \end{array}$$

On the other hand, the asymptotic 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. For instance, when $\phi = .4$ and $\rho = 0$, there is no asymptotic downward bias in $\beta(\boldsymbol{\varrho})$, while $\alpha(\boldsymbol{\varrho})$ is still upward biased. In contrast, when $\phi = .05$ and $\rho = .55$, we find that the downward bias in $\beta(\boldsymbol{\varrho})$ is bigger than the upward bias in $\alpha(\boldsymbol{\varrho})$.

The asymptotic biases found imply that the HRS approximation moves $\alpha(\boldsymbol{\varrho})$ and $\beta(\boldsymbol{\varrho})$ in the opposite direction to what the weak GARCH results in Nijman and Sentana (1996) (NS) indicate. Specifically, given that in a single factor model the GLS factor 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 , NS show that y_t^2 will also follow an ARMA(1,1) process with the same AR coefficient, i.e. $\phi + \rho$, but with an MA coefficient which is an explicit function not only of ϕ and ρ , but also of the (conditional) kurtosis of f_t and η_t (which are both 3 in our case) and the unconditional variances of f_t and η_t , which are 1 and v , respectively. Given that the size of the MA coefficient of f_t^2 unequivocally increases as a result of adding uncorrelated noise, NS found that the weak GARCH parameter would become larger, while the weak ARCH parameter would become smaller. Although the weak GARCH parameters are not generally consistently estimated by the Gaussian PML estimators of univariate GARCH(1,1) models fitted to y_t , the Monte Carlo evidence in NS suggests that the asymptotic biases of those estimators go in the same direction.

In contrast, the main effect of the HRS procedure, which explicitly acknowledges that the GLS factor y_t is the sum of a latent factor plus white noise with variance v , is to

smooth out the conditional variance of the factor. As a consequence, the estimation procedure seems to force $\alpha(\boldsymbol{\varrho})$ to be bigger than ϕ in an attempt to match the static and dynamic fourth moments of y_t generated by the HRS procedure with the true static and dynamic fourth moments of the process, and in this way, provide a better approximation to the true but unknown $V(\mathbf{x}_t|X_{t-1}; \boldsymbol{\varrho}^0)$.

In order to determine to what extent this intuition is correct, we have generated realisations of the GLS factor representing portfolios that would correspond to the trivariate single factor models discussed above, and compared the Gaussian distribution of y_t given Y_{t-1} and $\boldsymbol{\varrho}^0$ that the HRS model implies with the true conditional distribution. The former is simply $N\{0, \lambda_t[\boldsymbol{\theta}(\boldsymbol{\varrho}^0)] + \xi(\boldsymbol{\varrho}^0)\}$, where $\xi(\boldsymbol{\varrho}^0) = [\mathbf{c}'(\boldsymbol{\varrho}^0)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\varrho}^0)\mathbf{c}(\boldsymbol{\varrho}^0)]^{-1}$. In contrast, 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 $\delta_t + v^0$, where δ_t had been previously drawn from its distribution conditional on Y_{t-1} and $\boldsymbol{\varrho}^0$ by using the exact MCMC samplers developed in FSS. 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 $\delta_t + v^0$ given Y_{t-1} and $\boldsymbol{\varrho}^0$, which would be 0 if either δ_t is constant or $v^0 = 0$. Figure 2, which presents the results of such a comparison for two arbitrarily chosen observations, clearly shows that the approximate Kalman filter provides extremely reliable results, despite the fact that the true distribution of $y_t|Y_{t-1}; \boldsymbol{\varrho}^0$ is not exactly Gaussian, as the reported excess kurtosis statistics confirm.

To gain some further insight on the relationship between both models, we have also computed the probability integral transform (PIT) with respect to the approximate model $N\{0, \lambda_t[\boldsymbol{\theta}(\boldsymbol{\varrho}^0)] + \xi(\boldsymbol{\varrho}^0)\}$ of the aforementioned realisations of the GLS factor representing 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)). As can be seen from Figure 3, which shows the difference between the empirical cumulative distribution of 4,000,000-long PIT sequences and the 45° degree line, the approximate Kalman filter provides 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 ($v^0 = 1/9$). Nevertheless, note

that the maximum differences are very small ($<.005$) even in the other cases. Therefore, given that in many empirical applications it is likely that the signal to noise ratio will be high, and the conditional variance a fairly smooth process, in practice we would expect the model proposed by HRS to provide a rather accurate approximation to the conditional distribution of the observations given their past values.

3.3 Sequential pseudo-ML estimators

Although $\mathbf{c}(\boldsymbol{\varrho}) - \mathbf{b}$ and $\boldsymbol{\gamma}(\boldsymbol{\varrho}) - \boldsymbol{\psi}$, which are the asymptotic biases of the PML estimators of \mathbf{c} and $\boldsymbol{\gamma}$ when regarded as estimators of \mathbf{b} and $\boldsymbol{\psi}$, are generally rather small, they are not exactly zero.⁷ In fact, it is possible to find parameter configurations in which those biases become noticeable. To some extent, this prompted SF to suggest sequential estimators of the conditional variance parameters $\boldsymbol{\pi}$, defined as

$$\ddot{\boldsymbol{\pi}} = \arg \max_{\boldsymbol{\pi}} \bar{l}_T(\mathbf{c} = \dot{\mathbf{b}}; \boldsymbol{\gamma} = \dot{\boldsymbol{\psi}}; \boldsymbol{\pi}),$$

where $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$ are the consistent estimators described at the end of section 2. Although such a sequential estimator of $\boldsymbol{\pi}$ will continue to be inconsistent for $\boldsymbol{\nu}$, the biases could be potentially lower since they are not contaminated by the biases in $\mathbf{c}(\boldsymbol{\varrho})$ and $\boldsymbol{\gamma}(\boldsymbol{\varrho})$. Paradoxically, it is worth mentioning that if we were to iterate the sequential procedure, by alternating between the maximisation of (11) with respect to $(\mathbf{c}, \boldsymbol{\gamma})$ on the one hand, and $\boldsymbol{\pi}$ on the other, and achieved convergence, then we would destroy the consistency of the estimators of the static factor parameters because we would recover $\hat{\boldsymbol{\theta}}_T$.

In this context, it is worth mentioning that $\dot{\mathbf{b}}$, $\dot{\boldsymbol{\psi}}$ and $\ddot{\boldsymbol{\pi}}$ have a very interesting interpretation in terms of the GLS factors $\mathbf{g}_t^G(\mathbf{c}, \boldsymbol{\gamma})$. First of all, we can use the results in Grinblatt and Titman (1987) to prove that the first stage ML estimates of the factor loadings $\dot{\mathbf{b}}$ can be obtained from the OLS regression of \mathbf{x}_t on the estimated GLS factor representing portfolios $\mathbf{g}_t^G(\mathbf{c} = \dot{\mathbf{b}}, \boldsymbol{\gamma} = \dot{\boldsymbol{\psi}})$, which suggests an obvious iterative procedure to estimate the factor loadings. Moreover, since the pseudo log-likelihood function of the observed series \mathbf{x}_t (given X_{t-1}) can be factorised as the sum of the log-likelihood function

⁷These binding functions, which we do not report for the sake of brevity, are available from the authors on request.

of the GLS factors, $\mathbf{g}_t^G(\mathbf{c}, \gamma)$, conditional on X_{t-1} , which depends on all the elements of $\boldsymbol{\theta}$, plus the log-likelihood function of the first $N - k$ principal components of $\mathbf{w}_t^G(\mathbf{c}, \gamma)$, which does not depend on $\boldsymbol{\pi}$, it is clear that the sequential pseudo-ML estimators of $\boldsymbol{\pi}$ can also be obtained by maximising the k -variate log-likelihood function of $\mathbf{g}_t^G(\mathbf{c}, \gamma)$ in which we set $\mathbf{c} = \dot{\mathbf{b}}$ and $\gamma = \dot{\boldsymbol{\psi}}$ as if they were the true values of the parameters. Specifically, the objective function takes the following form:

$$-\frac{k}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) + \boldsymbol{\Phi}| - \frac{1}{2} \mathbf{g}_t^{G'}(\mathbf{c}, \gamma) [\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) + \boldsymbol{\Phi}]^{-1} \mathbf{g}_t^G(\mathbf{c}, \gamma),$$

which explicitly acknowledges that $\mathbf{g}_t^G(\mathbf{c}, \gamma)$ is generally a noisy estimate of \mathbf{g}_t . We shall recall this interpretation in section 4.2.

4 Indirect estimation

The impossibility of writing the log-likelihood function of the model of interest in closed-form, combined with the ease with which we can simulate drawings from it, suggest that the indirect estimation procedures of GT, GMR and Smith (1993) should be ideally suited for our case (see Gouriéroux and Monfort (1996) for an advanced textbook discussion). In this context, the most important decision that we have to make is the choice of auxiliary model. Ideally, our choice should take us close to the situation covered by Theorem 2 in GT, which loosely speaking says that if the generally unknown score of the true model, $\mathbf{q}_t(\boldsymbol{\varrho})$, is spanned by the pseudo log-likelihood score of the auxiliary model, $\mathbf{s}_t(\boldsymbol{\theta})$, then indirect estimation will be as efficient as ML. More generally, the lower the asymptotic residual covariance matrix in the limiting least squares regression of the (average) log-likelihood score of the true model, $\sqrt{T} \bar{\mathbf{q}}_T(\boldsymbol{\varrho}^0)$, on the (average) modified score of the auxiliary model, $\sqrt{T} \bar{\mathbf{s}}_T[\boldsymbol{\theta}(\boldsymbol{\varrho}^0)]$, the closer the indirect estimator will be to achieving the asymptotic efficient of ML (see Proposition 7 in CFS and the references therein). In principle, one way to reach the asymptotic Cramer-Rao bound is to allow the number of parameters of the auxiliary model to go to infinity at a suitable but as yet unknown rate, as the semi non-parametric procedures recommended by GT are designed to do. Unfortunately, their methods are not designed to work with large-scale multivariate models involving many parameters of interest, like the ones we analyse in this paper.

But since the approximate model proposed by HRS also spans the score of the model of interest in some important limiting cases (see Proposition 1), and it should provide a rather good approximation to it in more general situations, as the evidence at the end of section 3.2 suggests, it looks like an ideal candidate for auxiliary model.

4.1 Constrained indirect estimation

Given that the HRS auxiliary model must be estimated subject to inequality constraints, which are often binding in practice, we must use the constrained indirect estimation procedures proposed in our earlier work (see CFS), which can handle a mix of equality and inequality restrictions on $\boldsymbol{\theta}$. More specifically, let $\boldsymbol{\mu}$ denote the Kuhn-Tucker multipliers associated with s constraints implicitly characterised by the vector of functions $h(\boldsymbol{\theta})$, and define the (scaled) Lagrangian function as

$$Q_T(\boldsymbol{\beta}) = \bar{l}_T(\boldsymbol{\theta}) + h'(\boldsymbol{\theta})\boldsymbol{\mu},$$

where $\boldsymbol{\beta} = (\boldsymbol{\theta}', \boldsymbol{\mu}')$ is an augment parameter vector of dimension $d + s$. Assuming that both the average pseudo-log likelihood function $\bar{l}_T(\boldsymbol{\theta})$, and the vector of functions $h(\boldsymbol{\theta})$ are continuously differentiable with respect to $\boldsymbol{\theta}$, the latter with a Jacobian matrix $\partial h'(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ whose rank coincides with the number of effective constraints at $\boldsymbol{\theta}$, the first-order conditions that take into account the inequality constraints $h(\boldsymbol{\theta}) \geq \mathbf{0}$ will be given by:

$$\frac{\partial Q_T(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\theta}} = \bar{m}_T(\hat{\boldsymbol{\beta}}) = \mathbf{0},$$

where $\bar{m}_T(\boldsymbol{\beta})$ is the sample mean of

$$m_t(\boldsymbol{\beta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\boldsymbol{\mu},$$

which is the contribution of the t^{th} observation to the modified score of the auxiliary model. In addition, $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\theta}}', \hat{\boldsymbol{\mu}}')$ must satisfy the complementary slackness restrictions

$$h(\hat{\boldsymbol{\theta}}_T^r) \odot \hat{\boldsymbol{\mu}}_T^r = \mathbf{0},$$

plus the inequality restrictions $h(\hat{\boldsymbol{\theta}}) \geq \mathbf{0}$ and $\hat{\boldsymbol{\mu}} \geq \mathbf{0}$. Note that the main difference with the usual unrestricted case is that $m_t(\boldsymbol{\beta})$ not only depends on the d auxiliary model parameters $\boldsymbol{\theta}$, but also on the s multipliers $\boldsymbol{\mu}$ associated with the restrictions.

Let us now define

$$\mathbf{m}(\boldsymbol{\varrho}; \boldsymbol{\beta}) = E[\bar{m}_T(\boldsymbol{\beta}) | \boldsymbol{\varrho}] = E\left[\frac{\partial \bar{l}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\varrho}\right] + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\mu}.$$

In this context, the Generalised Method of Moments (GMM) version of the inequality constrained indirect estimator of $\boldsymbol{\varrho}$ is

$$\tilde{\boldsymbol{\varrho}}(\mathbf{P}) = \arg \min_{\boldsymbol{\varrho}} \mathbf{m}'(\boldsymbol{\varrho}; \hat{\boldsymbol{\beta}}) \cdot \mathbf{P} \cdot \mathbf{m}(\boldsymbol{\varrho}; \hat{\boldsymbol{\beta}}),$$

where \mathbf{P} is a p.d. weighting matrix of order d . Therefore, our moment conditions nest the ones employed by GT when there are no constraints, or when they are not binding, but remain valid even if they are.⁸

In practice, of course, it is often impossible to obtain $\mathbf{m}(\boldsymbol{\varrho}; \boldsymbol{\beta})$ in closed form (but see section 4.2 below). As explained by GT and CFS, though, 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 $\{\mathbf{x}_n(\boldsymbol{\varrho}), n = 1, \dots, T \cdot H\}$. In particular, we will have:

$$\begin{aligned} \mathbf{m}(\boldsymbol{\varrho}; \boldsymbol{\beta}) &= E\left(\frac{\partial \bar{l}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\varrho}\right) + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\mu} \\ &\simeq \frac{1}{T \cdot H} \sum_{n=1}^{T \cdot H} \frac{\partial \ln f[\mathbf{x}_n(\boldsymbol{\varrho}) | X_{n-1}(\boldsymbol{\varrho}); \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} + \frac{\partial h'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\mu} = \mathbf{m}_{TH}(\boldsymbol{\varrho}; \boldsymbol{\beta}), \end{aligned}$$

where we can make left and right hand sides arbitrarily close in a numerical sense as $H \rightarrow \infty$. Importantly, 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 in practice is to minimise with respect to $\boldsymbol{\varrho}$ the distance between the average score in the actual sample, $\bar{s}_T(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}$, 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. Given that in practice it is of paramount importance to have available a fast and numerically reliable

⁸Alternatively, we could combine the constrained parameter estimators $\hat{\boldsymbol{\theta}}$ and Kuhn-Tucker multipliers $\hat{\boldsymbol{\mu}}$ to obtain a constrained classical minimum distance indirect estimator of $\boldsymbol{\varrho}$, which would reduce to the one proposed by GMR in the unrestricted case. Nevertheless, since we proved in CFS that we can always find restricted classical minimum distance indirect estimators that are asymptotically equivalent to the constrained GMM estimators by an appropriate choice of weighting matrix, we shall only consider GMM-based indirect estimators hereinafter because the computations required to obtain the indirect estimators are far more costly for the former than for the latter.

procedure for the computation of the score of the auxiliary model, in Appendix B we derive analytical formulae for the score of model (8)-(10) that satisfy those requirements.

As we mentioned before, another problem with the GARCH(1,1)-type formulation in (10) 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 . If $\hat{\alpha}_j = 0 \forall j$, this is actually a computational blessing in disguise because 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 $\tilde{\boldsymbol{\rho}} = \hat{\boldsymbol{\theta}}$. In contrast, if $\hat{\alpha}_j$ is strictly positive but 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 $\alpha_j \geq \alpha_j^{\min} > 0$, where α_j^{\min} is some small number. Finally, since the pseudo-ML estimators of $\boldsymbol{\theta}$ may not be well behaved when $\alpha_j + \beta_j > 1$ (cf. Lumsdaine (1996)), and we are assuming a covariance stationary auxiliary model, then we will also impose the restriction $\alpha_j + \beta_j \leq (\alpha_j + \beta_j)^{\max} \leq 1$.⁹

Given that the number of auxiliary model parameters coincides with the number of parameters of the model of interest, d , $\boldsymbol{\rho}$ is exactly identified from the moments implied by the modified first-order conditions of the auxiliary model. Therefore, in the absence of inequality constraints on $\boldsymbol{\rho}$, the indirect estimators would be numerically invariant (for large enough T) to the choice of weighting matrix, \mathbf{P} . Nevertheless, since the same inequality restrictions that apply to $\boldsymbol{\theta}$ apply to $\boldsymbol{\rho}$ too, the choice of \mathbf{P} will sometimes matter. For that reason, we will optimally choose \mathbf{P} to be equal to the inverse of the asymptotic covariance matrix of $\sqrt{T}\bar{m}_T(\boldsymbol{\beta}^0)$, which can be consistently estimated on the basis of $\bar{m}_t(\hat{\boldsymbol{\beta}})$ by means of standard techniques.¹⁰

⁹After some 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 $\boldsymbol{\rho}$.

¹⁰In addition, 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 those cases in which the inequality restrictions on $\boldsymbol{\rho}$ do not bind.

4.2 Sequential indirect estimators

The practical application of the indirect estimators described in the previous section to large-scale multivariate model must overcome two problems. First, the estimation of the pseudo-ML parameters θ can become a very time consuming procedure, especially when N is large, because the dimension of the auxiliary parameter space grows linearly with N .¹¹ Second, the indirect estimation of ϱ is at least as time consuming because the dimension of the true parameter space is the same. In some important cases, though, we can exploit the SF sequential estimators of \mathbf{c} , γ and π described before so that the dimension of the parameter space over which we effectively minimise the GMM criterion function does not grow with the number of series under consideration.

To see how, it is convenient to fully characterise those sequential estimators. Assuming for ease of exposition that there are no binding inequality constraints on either $\dot{\psi}$ or $\ddot{\pi}$, the relevant first-order conditions are:

$$\begin{bmatrix} \bar{l}_{\mathbf{c}T}(\dot{\mathbf{b}}, \dot{\psi}, \mathbf{0}) \\ \bar{l}_{\gamma T}(\dot{\mathbf{b}}, \dot{\psi}, \mathbf{0}) \\ \bar{l}_{\pi T}(\dot{\mathbf{b}}, \dot{\psi}, \ddot{\pi}) \end{bmatrix} = \mathbf{0},$$

where $l_{\mathbf{c}t}(\theta)$, $l_{\gamma t}(\theta)$ and $l_{\pi t}(\theta)$ are the period t contributions to pseudo log-likelihood scores corresponding to \mathbf{c} , γ and π , respectively. Importantly, note that π must be set to 0 in the first two blocks of first-order conditions, and to $\ddot{\pi}$ in the last one. By analogy with the indirect estimator described in the previous section, we could then estimate ϱ by minimising the norm of

$$\frac{1}{T \cdot H} \sum_{n=1}^{T \cdot H} \left\{ \begin{array}{l} \partial \ln f[\mathbf{x}_n(\varrho) | X_{n-1}(\varrho); \dot{\mathbf{b}}, \dot{\psi}, \mathbf{0}] / \partial \mathbf{c} \\ \partial \ln f[\mathbf{x}_n(\varrho) | X_{n-1}(\varrho); \dot{\mathbf{b}}, \dot{\psi}, \mathbf{0}] / \partial \gamma \\ \partial \ln f[\mathbf{x}_n(\varrho) | X_{n-1}(\varrho); \dot{\mathbf{b}}, \dot{\psi}, \ddot{\pi}] / \partial \pi \end{array} \right\}. \quad (12)$$

¹¹The most computationally efficient way to increase the pseudo log-likelihood function in (11) starting from an arbitrary set of parameter values is by means of the EM algorithm in Demos and Sentana (1998) (see also Sentana (2000)). These authors showed that after just a few very fast iterations, the EM algorithm takes θ much closer to their pseudo-ML estimators than a quasi-Newton method after many very slow iterations, especially when the cross-sectional dimension N is large. But since the EM algorithm slows down substantially when it gets very close to the optimum, they suggested to shift to a quasi-Newton algorithm after a few iterations. In that context, EM arguments can also be used to obtain convenient expressions for the value of the score.

In fact, given that we can use the results in appendix B of SF to write

$$\begin{aligned} \frac{\partial \ln f[\mathbf{x}_n(\boldsymbol{\varrho})|X_{n-1}(\boldsymbol{\varrho}); \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}]}{\partial \mathbf{c}} &= \text{vec}[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\mathbf{x}_n(\boldsymbol{\varrho})\mathbf{x}'_n(\boldsymbol{\varrho})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}} \\ &\quad - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}}] \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ln f[\mathbf{x}_n(\boldsymbol{\varrho})|X_{n-1}(\boldsymbol{\varrho}); \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}]}{\partial \boldsymbol{\gamma}} &= \frac{1}{2}\text{vecd}[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\mathbf{x}_n(\boldsymbol{\varrho})\mathbf{x}'_n(\boldsymbol{\varrho})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1} \\ &\quad - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}], \end{aligned}$$

we can easily compute the expected values of these scores without resorting to simulations as

$$\begin{aligned} \mathbf{m}_{\mathbf{c}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}) &= E \left\{ \frac{\partial \ln f[\mathbf{x}_n(\boldsymbol{\varrho})|X_{n-1}(\boldsymbol{\varrho}); \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}]}{\partial \mathbf{c}} \middle| \boldsymbol{\varrho} \right\} \\ &= \text{vec}[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}(\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}} - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}\dot{\mathbf{B}}] \end{aligned}$$

and

$$\begin{aligned} \mathbf{m}_{\boldsymbol{\gamma}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}) &= E \left\{ \frac{\partial \ln f[\mathbf{x}_n(\boldsymbol{\varrho})|X_{n-1}(\boldsymbol{\varrho}); \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}]}{\partial \boldsymbol{\gamma}} \middle| \boldsymbol{\varrho} \right\} \\ &= \text{vec}[(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}(\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi})(\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1} - (\dot{\mathbf{B}}\dot{\mathbf{B}}' + \dot{\boldsymbol{\Psi}})^{-1}]. \end{aligned}$$

However, while it is true that such an alternative indirect estimation procedure would benefit from the simplified estimation of $\boldsymbol{\theta}$, and a faster calculation of $\mathbf{m}_{\mathbf{c}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0})$ and $\mathbf{m}_{\boldsymbol{\gamma}}(\boldsymbol{\varrho}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0})$, we would still have to minimise the GMM criterion function over $\boldsymbol{\varrho}$, which would be very time consuming for large N . Nevertheless, the consistency of $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$ implies that

$$\begin{bmatrix} \mathbf{m}_{\mathbf{c}}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{v}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}) \\ \mathbf{m}_{\boldsymbol{\gamma}}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{v}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{0}) \end{bmatrix} = \mathbf{0}$$

regardless of the GARCH parameters \mathbf{v} , as long as these moments are well defined. This fact, together with the exact identification of $\boldsymbol{\varrho}$, implies that the indirect estimators of \mathbf{b} and $\boldsymbol{\psi}$ based on (12) will in fact coincide with $\dot{\mathbf{b}}$ and $\dot{\boldsymbol{\psi}}$. As a result, the only task left is to minimise with respect to \mathbf{v} the norm of

$$\mathbf{m}_{\boldsymbol{\pi}}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{v}; \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}) \simeq \frac{1}{T \cdot H} \sum_{n=1}^{T \cdot H} \partial \ln f[\mathbf{x}_n(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{v})|X_{n-1}(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \mathbf{v}); \dot{\mathbf{b}}, \dot{\boldsymbol{\psi}}, \ddot{\boldsymbol{\pi}}] / \partial \boldsymbol{\pi}.$$

The interpretation of the sequential estimators of the auxiliary model parameter that we gave at the end of section 3.3 can also be exploited to enhance our understanding of the sequential indirect estimators that we have just presented. In effect, what we do is to use the static factor model parameter estimates to construct k portfolios of the original N assets, $\mathbf{g}_t^G(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}})$, which, conditional on those parameter values, contain as much information about the latent factors as the original series \mathbf{x}_t . Then, we estimate GARCH-type models for $\mathbf{g}_t^G(\dot{\mathbf{b}}, \dot{\boldsymbol{\psi}})$, and use indirect estimation procedures to adjust for the fact that $\hat{\boldsymbol{\pi}}$ is generally inconsistent for \boldsymbol{v} . Given that the dimension of \boldsymbol{v} does not depend on N , we can use this sequential indirect estimation procedure to handle practical situations in which the cross-sectional dimension is large (as in Engle (2002)).

5 Monte Carlo Evidence

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 full-information and sequential approximate ML methods of HRS and SF, respectively. Unfortunately, given that the estimation of these models is computationally rather intensive, we are forced to consider here a smaller number of series than in many empirical applications. However, we select the parameter values, and in particular the signal-to-noise ratio, so as to reflect empirically relevant situations.

We have used the NAG library G05DDF routine to generate 1,600 samples of 1,000 observations each (plus another 100 for initialisation) of a trivariate single factor model, which is the smallest possible cross-sectional dimension for which the static variance parameters can be identified. This sample size corresponds roughly to twenty years of weekly data or four years of daily data. Since the performance of the different estimators depends on \mathbf{b} and $\boldsymbol{\Psi}$ mostly through the scalar quantity $v = (\mathbf{b}'\boldsymbol{\Psi}^{-1}\mathbf{b})^{-1}$, the model considered is:

$$x_{it} = b_i f_t + v_{it} \quad (i = 1, 2, 3)$$

with $\mathbf{b} = (1, 1, 1)'$, $\delta_t = (1 - \phi - \rho) + \phi f_{t-1}^2 + \rho \lambda_{t-1}$ and $\boldsymbol{\Psi} = \psi \mathbf{I}$. Therefore, ϕ , ρ and ψ are the only relevant parameters as far as the Monte Carlo designs are concerned. In this

respect, we have considered the same four different combinations of ϕ and ρ documented in section 3.2. As for ψ , we have considered three different values for each of those four combinations. Specifically, we have chosen $\psi = 1/3, 1$ and 3 . In order to understand these values, it is convenient to look at three different measures of the strength of the signal relative to the size of the noise. In particular, if we regress the GLS factor, y_t , which due to our balanced design is simply the cross-sectional mean of the x'_{it} s, on the latent factor f_t , we would obtain theoretical R^2 coefficients of $.9, .75$ and $.5$, respectively. Similarly, the ratios of the unconditional variance of the residual in the previous regression to the unconditional variance of the latent factor, which trivially coincide with ν , would be equal to $1/9, 1/3$ and 1 , respectively. Finally, the correlation between any two of the three observed series would be $.75, .5$ and $.25$. All in all, we have considered 12 different parameter configurations.¹²

For scaling purposes, we use $c_3 = 1$, and leave the unconditional variance free. We also set λ_1 to the unconditional variance of the common factor to start up the recursions. In order to guarantee the positivity and stationarity restrictions, we first optimise the pseudo log-likelihood function in terms of some unrestricted parameters θ^* , where $\lambda = (\lambda^*)^2$, $\gamma_i = (\gamma_i^*)^2$ ($i = 1, 2, 3$), $\alpha = .999 \sin^2(\alpha^*)$ and $\beta = (.999 - \alpha) \sin^2(\beta^*)$, where $.999$ acts as our effective upper bound on $\alpha + \beta$. Then, we compute the score in terms of the original parameters θ using the analytical expressions derived in Appendix B to avoid large numerical errors, and introduce one multiplier for each of the first order conditions, which take away any slack left. As we explained before, though, if the maximum of the log-likelihood function happens at $\alpha = 0$, then there is no need to resort to indirect estimation procedures in view of Proposition 1, and we simply set $\tilde{\theta} = \hat{\theta}$. If, on the other hand, the ML estimate of α is strictly positive but less than α^{\min} , we re-estimate the auxiliary model subject to the restriction $\alpha = \alpha^{\min}$ by using $\alpha = \alpha^{\min} + (.999 - \alpha^{\min}) \sin^2(\alpha^*)$.

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

¹²In their Monte Carlo experiments, DMP considered a single signal to noise ratio and two conditional variance parameter configurations: $\phi = \rho = 0$, and $\phi = .2$ and $\rho = .7$.

have been effectively computed on the basis of 100,000 simulated observations. A 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. Finally, we minimise numerically the GMM criterion function in terms of some unrestricted parameters ϱ^* , with $b_3 = 1$, b_1 and b_2 free, $\delta = (\delta^*)^2$, $\psi_i = \psi_i^{*2}$ ($i = 1, 2, 3$), $\phi = .999999 \sin^2(\phi^*)$ and $\rho = (.999999 - \phi) \sin^2(\rho^*)$, so as to ensure that $\delta \geq 0$, $\psi_i \geq 0 \forall i$, $\phi, \rho \geq 0$ and $\phi + \rho \leq 1$. Given that the auxiliary model fits the simulated data rather well, in the sense that the score of the auxiliary model is close to being a vector martingale difference sequence, the optimal weighting matrix has been estimated as the variance in the original data of the modified score of the auxiliary model evaluated at the pseudo-ML parameter estimates. In this respect, note that by including a multiplier in each first order condition, we automatically centre the scores around their sample mean.

The first thing to note is that the pseudo ML estimators of the auxiliary parameters α and β reached their lower bounds fairly frequently in some designs, especially when the true value of ψ was large. For instance, the estimated value of α was less than α^{\min} 8.75% of the time when $\phi = .1$, $\rho = .85$ and $\psi = 3$, while β was estimated as 0 in 10% of the samples when $\phi = \rho = .4$ and $\psi = 3$. Hence, these results clearly show that the constrained indirect estimation procedures that we have used are highly relevant in practice.

Figures 4a to 4d display kernel estimates of the sampling distributions of the joint and sequential GMM-based indirect estimators of the structural parameters ϕ and ρ , together with the two pseudo-ML estimators on which they are based. As for bandwidth, we have used the automatic choice given in expression (3.29) in Silverman (1986).

The small sample behaviour of the HRS and SF estimators is very much in accordance with what we have seen in section 3.2. When the signal to noise ratio is high (i.e. $\psi = 1/3$) and the unconditional coefficient of variation of the unobserved conditional variance is low (i.e. $\phi = .2$, $\rho = .6$, or $\phi = .1$, $\rho = .85$), 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. $\psi = 3$) and the unconditional coefficient of variation of the unobserved

conditional variance is high (i.e. $\phi = .4$, $\rho = .4$, or $\phi = .2$ and $\rho = .75$), their biases become rather noticeable, with neither dominating the other.

It is precisely in those cases that the systematic elimination of the biases achieved by the joint and sequential indirect estimators is more pronounced.¹³ At the same time, it seems that the variability of the two 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.

Although the sequential indirect estimator of ϕ often seems to outperform the corresponding joint estimator, at least in terms of sampling variability, the joint indirect estimator of ρ tends to outperform the sequential estimator across many Monte Carlo designs. Nevertheless, the differences are rather minor, except when the signal to noise ratio is low. In any case, the important message is that both indirect estimators are consistent across all Monte Carlo designs.

As for the estimators of the parameters characterising the unconditional covariance matrix (i.e. factor loadings and idiosyncratic variances), the results presented in Figure 5 clearly indicate that the HRS estimators are very similar to the joint indirect estimators, even in a case in which the signal to noise ratio is low and the unconditional coefficient of variation of the unobserved conditional variance is high. At the same time, the sequential indirect estimators, which are numerically identical to the SF estimators, seem to be less efficient than the other two, as one would expect from the Monte Carlo results reported by SF.

Finally, it is worth mentioning that we also considered a third indirect estimation procedure, which effectively uncouples the modified score generator $m_t(\boldsymbol{\theta})$ from the estimator of $\boldsymbol{\theta}$ at which the score is evaluated. In particular, we computed indirect estimators based on the full score of the HRS auxiliary model evaluated at the sequential pseudo-ML estimators $\mathbf{c} = \hat{\mathbf{b}}$, $\boldsymbol{\gamma} = \hat{\boldsymbol{\psi}}$ and $\boldsymbol{\pi}$. However, since their sampling distribution always lied between the distributions of the other two indirect estimators, we decided not to report

¹³The finite sample bias might be reduced even further by using the implicit bias adjustment procedures discussed by Gouriéroux, Renault and Touzi (2000), either on their own (see Arvanitis and Demos (2003)), or together with the control variates techniques developed by Calzolari, Di Iorio and Fiorentini (1998).

them to avoid cluttering unnecessarily the graphs (see section 2.3 of CFS for a theoretical discussion of such “fully non-optimised” indirect estimators).

6 Conclusions

In this paper, we discuss the application of the indirect estimation procedures popularised by GMR and GT to conditionally heteroskedastic factor models, in which the conditional variances of the latent variables are continuous functions of their own past values, and as such, unobserved by the econometrician. We use the Kalman filter-based approximation proposed by HRS with analytical derivatives as auxiliary model. Importantly, we employ the constrained indirect estimation procedures introduced in our earlier work to take into account inequality restrictions on the auxiliary model parameters. We compare the finite sample performance of our proposed estimators relative to the HRS approximation by means of Monte Carlo methods. Our results suggest that the indirect estimators that we propose consistently estimate the parameters of the conditional variances of the heteroskedastic 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 estimator of HRS 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.

We have also proposed sequential indirect estimation procedures that can handle large multivariate models because the effective dimension of the parameter space does not grow with the number of series under consideration. In this respect, our simulation evidence suggests that these sequential estimators are another serious competitor for estimating the conditional variance parameters. In contrast, while the HRS and joint indirect estimators perform very similarly as far as the unconditional variance parameters (i.e. factor loadings and idiosyncratic variances) are concerned, they outperform the sequential indirect estimators.

We have shown that the main determinants of the asymptotic biases in the approximate ML estimators proposed by HRS are the signal to noise ratio, and the coefficient of

variation of the unobserved conditional volatility of the latent factors. Therefore, given that in many empirical applications it is likely that the signal to noise ratio will be high, and the conditional variance a fairly smooth process, in practice we would expect those biases to be small.

Importantly, we also find that even in less realistic parameter configurations the model proposed by HRS provides a rather accurate approximation to the distribution of the observable variables conditional on their past values alone. In any case, the MCMC algorithms in FSS can be easily combined with our indirect estimators to obtain the exact distribution and moments of the latent variables conditional upon the observed data.

Finally, it is important to emphasise that although we have concentrated on factor models for the sake of concreteness, our indirect estimation procedures can be easily applied to the general class of unobserved component time series models with GARCH disturbances analysed by HRS (see Kim and Nelson (1999) for a textbook description).

One potential drawback of our approach is that our auxiliary model exactly identifies the parameters of interest because it was tailor-made to approximate the true one. As a result, we cannot use the optimum value of the GMM estimation criterion to assess the adequacy of the true model to the data. In this context, one attractive way to generate testable overidentifying restrictions without re-estimating the HRS auxiliary model would be to artificially nest it into an augmented model, and to add the Lagrange multipliers associated with the implicit equality restrictions, as suggested by CFS. For instance, we could replace the Gaussian-based log-likelihood in (11) by a generalised hyperbolic-based function, and use the scores underlying the multivariate asymmetric and kurtosis tests in Fiorentini, Sentana and Calzolari (2003) and Mencía and Sentana (2004) as additional moment conditions. Although we know from Proposition 8 in CFS that the resulting indirect estimators would be at least as asymptotically efficient as the ones that we have considered, their finite sample behaviour constitute an interest topic for further research.

Appendix

A Proof of Proposition 1

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

$$y_t = f_t + \eta_t$$

where

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

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

Let $p(y_t | Y_{t-1}; \boldsymbol{\varrho})$ denote the conditional density of y_t given $Y_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ and the parameters $\boldsymbol{\varrho} = (v, \phi)'$. The log-likelihood function of a sample of size T on y_t , $\mathbf{y} = (y_1, \dots, y_T)'$ cannot be written in closed except when $\phi = 0$ and/or $v = 0$. In particular, when $\phi = 0$, we just have the log-likelihood function of an *i.i.d.* $N(0, 1 + v)$ process, while when $v = 0$, 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 $\mathbf{f} = (f_1, \dots, f_T)'$ 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)

$$\ln p(\mathbf{y} | \mathbf{f}; \boldsymbol{\varrho}) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln v - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - f_t)^2}{v}, \quad (13)$$

and

$$\ln(\mathbf{f} | \boldsymbol{\varrho}) = -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \left\{ \ln[1 + \phi(f_{t-1}^2 - 1)] + \frac{f_t^2}{1 + \phi(f_{t-1}^2 - 1)} \right\}. \quad (14)$$

Let

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

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

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

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

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

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 $v > 0$, this yields

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

and

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

Then, we can use the MCMC procedures proposed by FSS, 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 $\phi = 0$

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

and

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

because

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

Consider now the following auxiliary model

$$y_t = g_t + \zeta_t,$$

where

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

and $\alpha \geq 0$, $\varphi \geq 0$. Since this model is a rather special case of (8)-(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

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

Hence, it is clear that this auxiliary model smoothly embeds the true model at $\alpha = \phi = 0$ and $\varphi = v$. \square

A similar argument can be used to show that

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

which is not very surprising given that

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

for every possible value of ϕ .

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

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

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

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

B The score of the HRS approximate likelihood function

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:

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

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

$$d(\mathbf{C}\boldsymbol{\Lambda}_t\mathbf{C}' + \boldsymbol{\Gamma}) = d(\mathbf{C})\boldsymbol{\Lambda}_t\mathbf{C}' + \mathbf{C}d(\boldsymbol{\Lambda}_t)\mathbf{C}' + \mathbf{C}\boldsymbol{\Lambda}_td(\mathbf{C}') + d(\boldsymbol{\Gamma}), \quad (15)$$

(cf. Magnus and Neudecker (1988)), we have that Jacobian of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ corresponding to $\mathbf{c}, \boldsymbol{\gamma}$ and $\boldsymbol{\pi}$ will be:

$$\frac{\partial \text{vec}'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} = \begin{pmatrix} [\mathbf{I}_N \otimes \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'](\mathbf{I}_{N^2} + \mathbf{K}_{NN}) \\ \mathbf{E}'_N \\ \mathbf{0} \end{pmatrix} + \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{E}'_k(\mathbf{C}' \otimes \mathbf{C}'),$$

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

After some straightforward algebraic manipulations, we get:

$$\begin{aligned} \mathbf{s}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \text{vec} [\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}'_t\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] \\ \frac{1}{2}\text{vecd} [\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}'_t\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})] \\ \mathbf{0} \end{pmatrix} \\ &\quad + \frac{1}{2} \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{vecd} [\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{x}_t\mathbf{x}'_t\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\mathbf{C}]. \end{aligned}$$

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

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

and

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

In view of (10), the expressions for $\partial \lambda_{jt}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ will be:

$$\begin{aligned} \frac{\partial \lambda_{jt}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \alpha_j \left[2g_{jt-1|t-1}(\boldsymbol{\theta}) \frac{\partial g_{jt-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \omega_{jjt-1|t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] + \beta_j \frac{\partial \lambda_{jt-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &\quad + \frac{\partial \varpi_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + [g_{jt-1|t-1}^2(\boldsymbol{\theta}) + \omega_{jjt-1|t-1}(\boldsymbol{\theta})] \frac{\partial \alpha_j}{\partial \boldsymbol{\theta}} + \lambda_{jt-1}(\boldsymbol{\theta}) \frac{\partial \beta_j}{\partial \boldsymbol{\theta}}, \end{aligned}$$

where $\varpi_j(\boldsymbol{\theta}) = (1 - \alpha_j - \beta_j)\lambda_j$.

Now, given that

$$\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) = (\mathbf{C}'\boldsymbol{\Gamma}^{-1}\mathbf{C} + \boldsymbol{\Lambda}_t^{-1})^{-1}$$

when $\boldsymbol{\Gamma}$ has full rank, the differential of $\boldsymbol{\Omega}_{t|t}$ will be $-\boldsymbol{\Omega}_{t|t}d(\mathbf{C}'\boldsymbol{\Gamma}^{-1}\mathbf{C} + \boldsymbol{\Lambda}_t^{-1})\boldsymbol{\Omega}_{t|t}$, where

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

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

$$\frac{\partial \boldsymbol{\omega}'_{t|X_t}}{\partial \boldsymbol{\theta}} = \left[\begin{array}{c} -2[\boldsymbol{\Gamma}^{-1}\mathbf{C}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) \otimes \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})] \\ \mathbf{E}'_N[\boldsymbol{\Gamma}^{-1}\mathbf{C}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) \otimes \boldsymbol{\Gamma}^{-1}\mathbf{C}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})] \\ \mathbf{0} \end{array} \right] + \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{E}'_k[\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) \otimes \boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})] \mathbf{D}_k^{+'}.$$

Similarly, since

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

when $\boldsymbol{\Gamma}$ has full rank, the differential of $\mathbf{g}_{t|t}$ is given by

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

As a result, we will have that

$$\frac{\partial \mathbf{g}'_{t|t}}{\partial \boldsymbol{\theta}} = \left(\begin{array}{c} \boldsymbol{\Gamma}^{-1}\mathbf{x}_t \otimes \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) \\ -\mathbf{E}'_N[\boldsymbol{\Gamma}^{-1}\mathbf{x}_t \otimes \boldsymbol{\Gamma}^{-1}\mathbf{C}\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})] \\ \mathbf{0} \end{array} \right) + \frac{\partial \boldsymbol{\omega}'_{t|t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{D}'_k(\mathbf{C}'\boldsymbol{\Gamma}^{-1}\mathbf{x}_t \otimes \mathbf{I}_k).$$

If $\boldsymbol{\gamma} \neq \mathbf{0}$, though, the above expressions become invalid. Nevertheless, appropriately modified expressions can be developed along the lines of Sentana (2000). For the sake of brevity, we only obtain the score when 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 original model as

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

where

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

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

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

and

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

where

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

and

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

Hence, if we partition \mathbf{c} and $\boldsymbol{\gamma}$ as $(\mathbf{c}'_a, \mathbf{c}'_b)'$ and $(\boldsymbol{\gamma}'_a, \boldsymbol{\gamma}'_b)$ respectively, where $\mathbf{c}_a = \text{vec}(\mathbf{C}'_a)$, $\mathbf{c}_b = \text{vec}(\mathbf{C}'_b)$, $\boldsymbol{\gamma}_a = \text{vecd}(\Gamma_a)$, and $\boldsymbol{\gamma}_b = \text{vecd}(\Gamma_b)$, then we can use the expressions derived

before to find

$$\mathbf{s}_{at}(\boldsymbol{\theta}) = \begin{pmatrix} \text{vec} [\boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'_a\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at}\mathbf{x}'_{at}\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_t(\boldsymbol{\theta})\mathbf{C}'_a\boldsymbol{\Sigma}_{at}(\boldsymbol{\theta})] \\ \mathbf{0} \\ \frac{1}{2}\text{vecd} [\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at}\mathbf{x}'_{at}\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})] \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \frac{1}{2}\frac{\partial\boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\text{vecd} [\mathbf{C}'_a\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at}\mathbf{x}'_{at}\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a - \mathbf{C}'_a\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a].$$

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 $\partial\text{vec}[\boldsymbol{\Sigma}_{bt|at}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$. Straightforward algebra shows that

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

and

$$\frac{\partial\text{vec}' [\boldsymbol{\Sigma}_{bt|at}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{0} \\ [\mathbf{I}_N \otimes \boldsymbol{\Omega}_{bt|at}(\boldsymbol{\theta})\mathbf{C}'_b](\mathbf{I}_{(N-k)^2k^2} + \mathbf{K}_{(N-k)k,(N-k)k}) \\ \mathbf{0} \\ \mathbf{E}'_{N-k} \\ \mathbf{0} \end{pmatrix} + \frac{\partial\boldsymbol{\omega}'_{bt|at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{D}'_k(\mathbf{C}'_b \otimes \mathbf{C}'_b).$$

Hence,

$$\begin{aligned} \mathbf{s}_{bt|at}(\boldsymbol{\theta}) &= \frac{\partial\mathbf{g}'_{t|at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{C}'_b\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta}) \\ &+ \begin{pmatrix} \mathbf{0} \\ \text{vec} \left[\begin{array}{c} \boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta})\mathbf{g}'_{t|at}(\boldsymbol{\theta}) \\ +\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{bt|at}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\mathbf{C}_b\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\mathbf{C}_b\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta}) \end{array} \right] \\ \mathbf{0} \\ \frac{1}{2}\text{vecd} \left[\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{bt|at}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta}) \right] \\ \mathbf{0} \end{pmatrix} \\ &+ \frac{1}{2}\frac{\partial\boldsymbol{\omega}'_{bt|at}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{D}'_k\text{vec} \left[\mathbf{C}'_b\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{bt|at}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{bt|at}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\mathbf{C}_b - \mathbf{C}'_b\boldsymbol{\Sigma}_{bt|at}^{-1}(\boldsymbol{\theta})\mathbf{C}_b \right]. \end{aligned}$$

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

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

where $d(\boldsymbol{\Sigma}_{at})$ is analogous to (15). As a result,

$$\frac{\partial \mathbf{g}'_{t|at}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} [\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at} \otimes \boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})] - [\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) \otimes \mathbf{g}_{t|at}(\boldsymbol{\theta})] \\ 0 \\ -\mathbf{E}'_k[\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{x}_{at} \otimes \boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\boldsymbol{\Lambda}_t(\boldsymbol{\theta})] \\ 0 \\ 0 \end{pmatrix} + \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{E}'_k [[\mathbf{g}_{t|at}(\boldsymbol{\theta}) \otimes \boldsymbol{\Lambda}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})].$$

Similarly, the differential of $\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})$ 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,

$$\frac{\partial \boldsymbol{\omega}'_{t|at}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -2[\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})] \\ \mathbf{0} \\ -\mathbf{E}'_k[\boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\boldsymbol{\Lambda}_t \otimes \boldsymbol{\Sigma}_{at}^{-1}(\boldsymbol{\theta})\mathbf{C}_a\boldsymbol{\Lambda}_t] \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{E}'_k \{ \{ [\boldsymbol{\Lambda}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta}) \otimes \mathbf{I}_k] - [\mathbf{C}'_a\boldsymbol{\Sigma}_{at}^{-1}\mathbf{C}_a\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) \otimes \boldsymbol{\Lambda}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{t|at}(\boldsymbol{\theta})] \} \} \mathbf{D}'_k.$$

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

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

and

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

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

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

$$\boldsymbol{\Sigma}_{at}(\check{\boldsymbol{\theta}}) = \mathbf{C}_a\boldsymbol{\Lambda}_t\mathbf{C}'_a,$$

$$\mathbf{g}_{t|at}(\check{\boldsymbol{\theta}}) = \mathbf{C}_a^{-1}\mathbf{x}_{at},$$

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

Moreover,

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

As a result, we can write

$$\mathbf{s}_{at}(\dot{\boldsymbol{\theta}}) = \begin{pmatrix} \text{vec} \left[\left(\mathbf{g}_{t|at} \mathbf{g}'_{t|at} - \boldsymbol{\Lambda}_t \right) \boldsymbol{\Lambda}_t^{-1} \mathbf{C}_a^{-1} \right] \\ \mathbf{0} \\ \frac{1}{2} \text{vecd} \left[\mathbf{C}_a^{-1} \boldsymbol{\Lambda}_t^{-1} \left(\mathbf{g}_{t|at} \mathbf{g}'_{t|at} - \boldsymbol{\Lambda}_t \right) \boldsymbol{\Lambda}_t^{-1} \mathbf{C}_a^{-1} \right] \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \frac{1}{2} \frac{\partial \boldsymbol{\lambda}'_t(\dot{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \text{vecd} \left[\boldsymbol{\Lambda}_t^{-1} \left(\mathbf{g}_{t|at} \mathbf{g}'_{t|at} - \boldsymbol{\Lambda}_t \right) \boldsymbol{\Lambda}_t^{-1} \right]$$

and

$$\mathbf{s}_{bt|at}(\dot{\boldsymbol{\theta}}) = \begin{pmatrix} -\text{vec}[\mathbf{C}_b^{*'} \boldsymbol{\Gamma}_b^{-1} \boldsymbol{\varepsilon}_{bt|at}(\dot{\boldsymbol{\theta}}) \mathbf{g}'_{t|at}(\dot{\boldsymbol{\theta}})] \\ \text{vec}[\boldsymbol{\Gamma}_b^{-1} \boldsymbol{\varepsilon}_{bt|at}(\dot{\boldsymbol{\theta}}) \mathbf{g}'_{t|at}(\dot{\boldsymbol{\theta}})] \\ \left\{ \frac{1}{2} \text{vecd} \{ \mathbf{C}_b^{*'} \boldsymbol{\Gamma}_b^{-1} [\boldsymbol{\varepsilon}_{bt|at}(\dot{\boldsymbol{\theta}}) \boldsymbol{\varepsilon}'_{bt|at}(\dot{\boldsymbol{\theta}}) - \boldsymbol{\Gamma}_b] \boldsymbol{\Gamma}_b^{-1} \mathbf{C}_b^{*'} \} \right. \\ \left. - \mathbf{E}'_k [\mathbf{C}_b^{*'} \boldsymbol{\Gamma}_b^{-1} \boldsymbol{\varepsilon}_{bt|at}(\dot{\boldsymbol{\theta}}) \mathbf{g}'_{t|at}(\dot{\boldsymbol{\theta}}) \boldsymbol{\Lambda}_t^{-1} \mathbf{C}_a^{-1}] \right\} \\ \frac{1}{2} \text{vecd} \{ \boldsymbol{\Gamma}_b^{-1} [\boldsymbol{\varepsilon}_{bt|at}(\dot{\boldsymbol{\theta}}) \boldsymbol{\varepsilon}'_{bt|at}(\dot{\boldsymbol{\theta}}) - \boldsymbol{\Gamma}_b] \boldsymbol{\Gamma}_b^{-1} \} \\ \mathbf{0} \end{pmatrix}.$$

Finally, we obtain

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

and

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

where

$$\frac{\partial \mathbf{g}'_{t|at}(\dot{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -[\mathbf{C}_a^{-1'} \otimes \mathbf{g}'_{t|at}(\dot{\boldsymbol{\theta}})] \\ \mathbf{0} \\ \mathbf{E}'_k [\mathbf{C}_a^{-1'} \boldsymbol{\Lambda}_t^{-1} \mathbf{g}_{t|at}(\dot{\boldsymbol{\theta}}) \otimes \mathbf{C}_a^{-1'}] \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

and

$$\frac{\partial \boldsymbol{\omega}'_{t|at}(\dot{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{E}'_k (\mathbf{C}_a^{-1'} \otimes \mathbf{C}_a^{-1'}) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \mathbf{D}_K.$$

Although these expressions are strictly speaking only valid when the idiosyncratic variances are identically 0, in practice, we recommend their use whenever the γ'_j s are less than .0001 because the expressions obtained for $\boldsymbol{\gamma} > \mathbf{0}$ become numerically unreliable for smaller values.

Finally, it is worth mentioning that if we fix the factor scales by setting $c_{jj} = 1$ as opposed to $\lambda_j = 1$ for $j = 1, \dots, k$, 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 (11) through $\varpi_j(\boldsymbol{\theta})$, which is the constant term in the conditional variance expressions. Either way, since we initialise the conditional variances of the factors with $\boldsymbol{\Lambda}_1 = E(\boldsymbol{\Lambda}_t) = \boldsymbol{\Lambda}$, then we must always start up the derivative recursions with $\partial\lambda_{j1}/\partial\boldsymbol{\theta}' = \partial\lambda/\partial\boldsymbol{\theta}'$.

References

- Aguilar, O. and West, M. (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.
- Anderson, T.W. and Rubin, H. (1956): “Statistical inference in factor analysis”, in J. Neymann (ed.), *Proceedings of the III Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley.
- Arvanitis, S. and Demos, A. (2003): “Aspects of the indirect inference theory”, mimeo, Athens University of Economics and Business.
- Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2004): “Multivariate GARCH models: a survey”, revised CORE Discussion Paper 2003/31.
- Billio, M. and Monfort, A. (2003): “Kernel-based indirect inference”, *Journal of Financial Econometrics* 1, 297-326.
- Bollerslev, T. and Wooldridge, J.M. (1992): “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying variances”, *Econometric Reviews* 11, 143-172.
- Calzolari G., Di Iorio, F. and Fiorentini, G. (1998): “Control variates for variance reduction in indirect inference: interest rate models in continuous time”, *Econometrics Journal* 1, C100-C112.
- Calzolari, G., Fiorentini, G. and Sentana, E. (2004): “Constrained indirect estimation”, forthcoming in the *Review of Economic Studies*.
- Carter, C.K. and Kohn, R. (1994): “On Gibbs sampling for state space models”, *Biometrika* 81, 541-553.
- Chib, S., Nardari, F. and Shephard, N. (2002): “Analysis of high dimensional multivariate stochastic volatility models”, mimeo, Nuffield College, Oxford.
- Demos, A. and Sentana, E. (1998): “An EM algorithm for conditionally heteroskedastic factor models”, *Journal of Business and Economic Statistics* 16, 357-361.

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

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

Doz, C. and Renault, E. (2001): “Conditionally heteroskedastic factor models: identification and instrumental variable estimation”, mimeo, University of Cergy-Pontoise.

Drost, F.C. and Nijman, T.E. (1993): “Temporal aggregation of GARCH processes”, *Econometrica* 61, 909-927.

Dungey, M., Martin, V.L. and Pagan, A.R. (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. (2002): “Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models”, *Journal of Business and Economic Statistics* 12, 339-350.

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

Engle, R.F. and Susmel, R. (1993): “Common volatility in international equity markets”, *Journal of Business and Economic Statistics* 11, 167-176.

Fiorentini, G., Sentana, E. and Calzolari, G. (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., Sentana, E. and Shephard N. (2003): “Likelihood estimation of latent generalised ARCH structures”, forthcoming in *Econometrica*.

Gallant, A.R. and Tauchen, G. (1996): “Which moments to match?”, *Econometric Theory* 12, 657-681.

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

Gouriéroux, C., A. Monfort and E. Renault (1991). A general framework for factor models. mimeo, INSEE.

Gouriéroux, C., A. Monfort and E. Renault (1993): “Indirect inference”, *Journal of Applied Econometrics* 8, S85-S118.

Gouriéroux, C., Renault, E. and Touzi, N. (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.

Grinblatt, M. and Titman, S. (1987) “The relation between mean-variance efficiency and arbitrage pricing”, *Journal of Business* 60, 97-112.

Harvey, A., Ruiz, E. and Sentana, E. (1992): “Unobservable component time series models with ARCH disturbances”, *Journal of Econometrics* 52, 129-158.

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

Kano, Y. (1983): “Consistency of estimators in factor analysis”, *Journal of the Japan Statistical Society* 13, 137-144.

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

King, M., Sentana, E. and Wadhvani, S. (1994): “Volatility and links between national stock markets”, *Econometrica* 62, 901-933.

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

Lumsdaine, R.L. (1996): “Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models”, *Econometrica* 64, 575-596.

Magnus, J.R. (1988): *Linear structures*, Oxford University Press, New York.

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, Chichester.

Meddahi, N. and Renault, E. (2004): “Temporal aggregation of volatility models”, *Journal of Econometrics* 119, 355-379.

Mencía, F.J. and Sentana, E. (2004): “Estimation and testing of dynamic models with generalised hyperbolic innovations”, mimeo, CEMFI.

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

Nijman, T. and Sentana, E. (1996): “Marginalization and contemporaneous aggregation of multivariate GARCH processes”, *Journal of Econometrics* 71, 71-87.

Pagan, A. (1980): “Some identification and estimation results for regression models with stochastically varying coefficients”, *Journal of Econometrics* 13, 341-364.

Pitt, M.K. and Shephard, N. (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.

Sentana, E. (1998): “The relation between conditionally heteroskedastic factor models and factor GARCH models”, *Econometrics Journal* 1, 1-9.

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. and Fiorentini, G. (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.

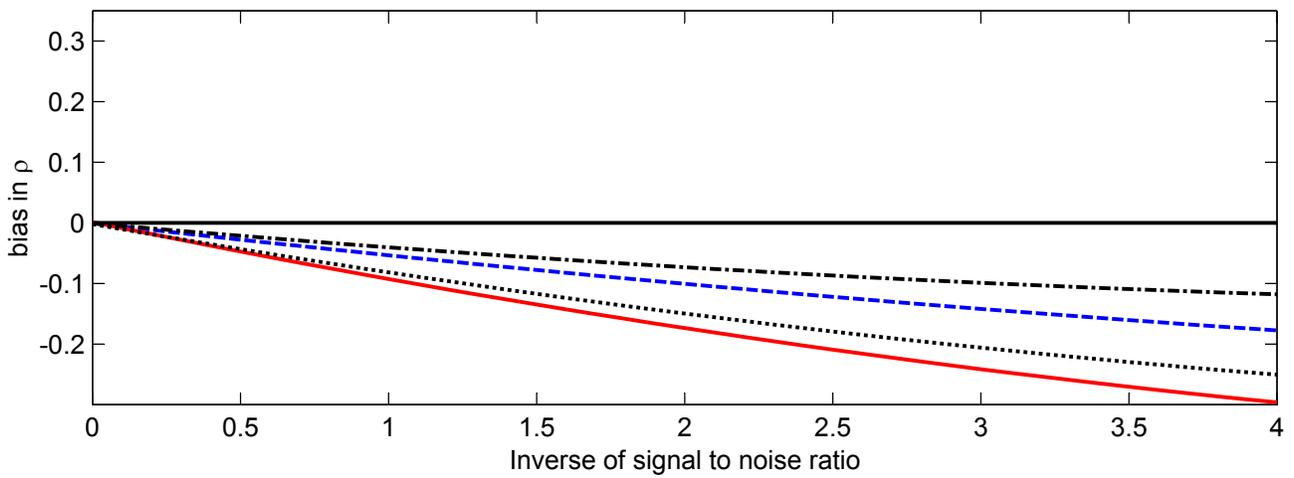
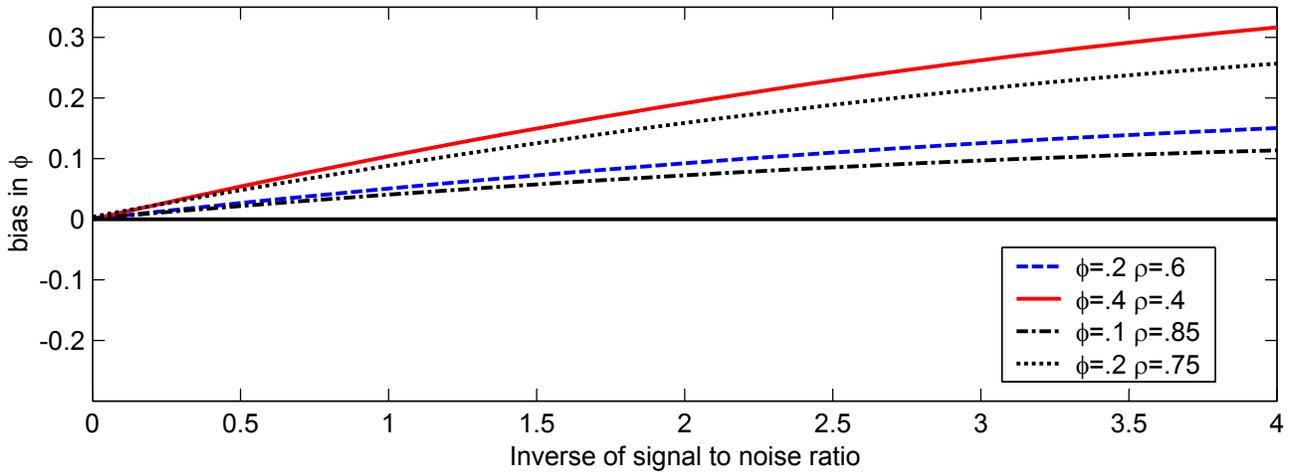
Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

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 Dungey, M. (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.

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



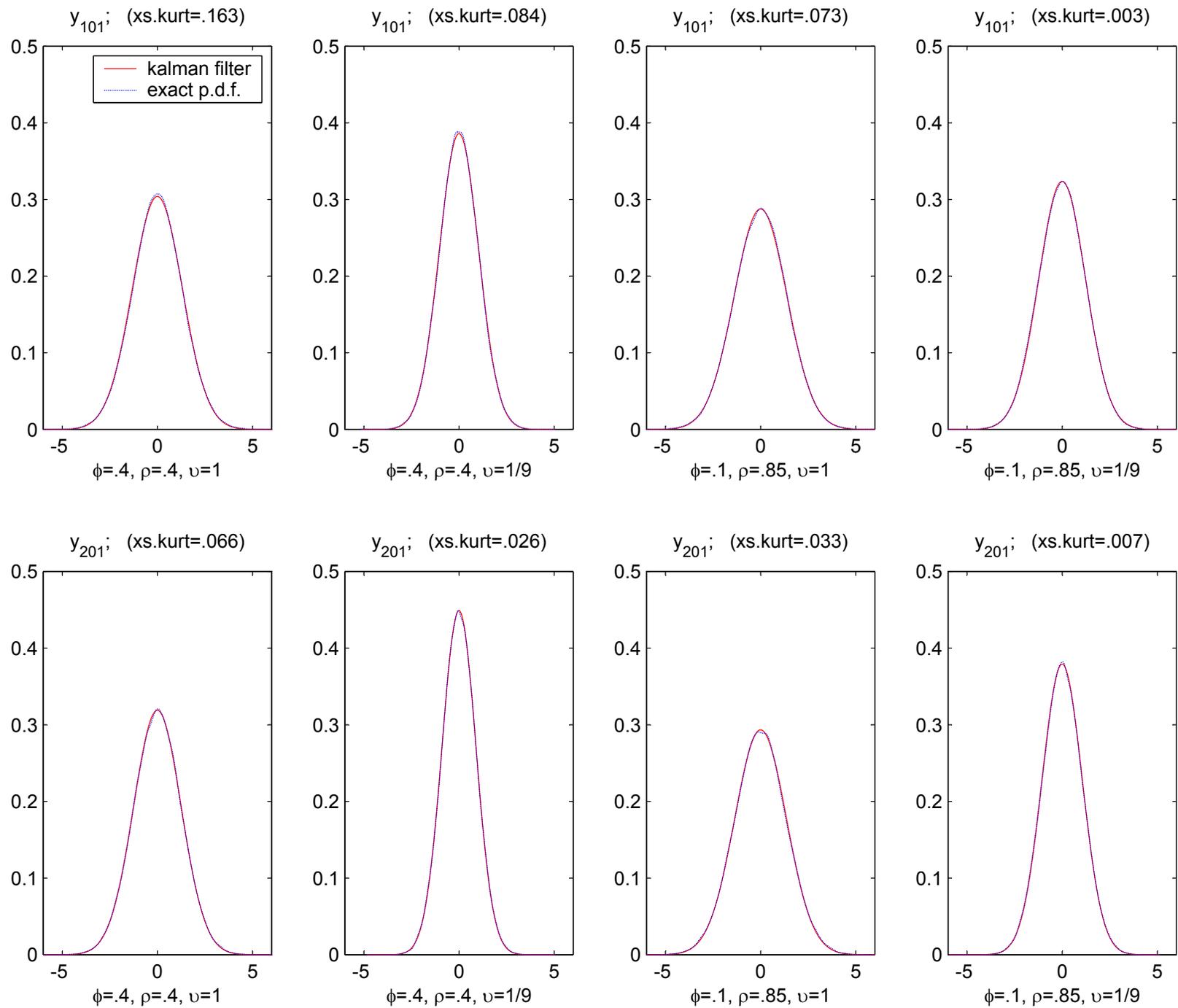


FIGURE 2: Comparison of the p.d.f. of the GLS factor representing portfolios given their past values and parameters with the HRS Kalman filter-based Gaussian approximation for different parameter configurations

FIGURE 3: Differences between the c.d.f. of the Probability Integral Transforms of the GLS representing portfolios generated by the HRS Kalman filter-based Gaussian approximation and their true distribution for different parameter configurations (T=4,000,000)

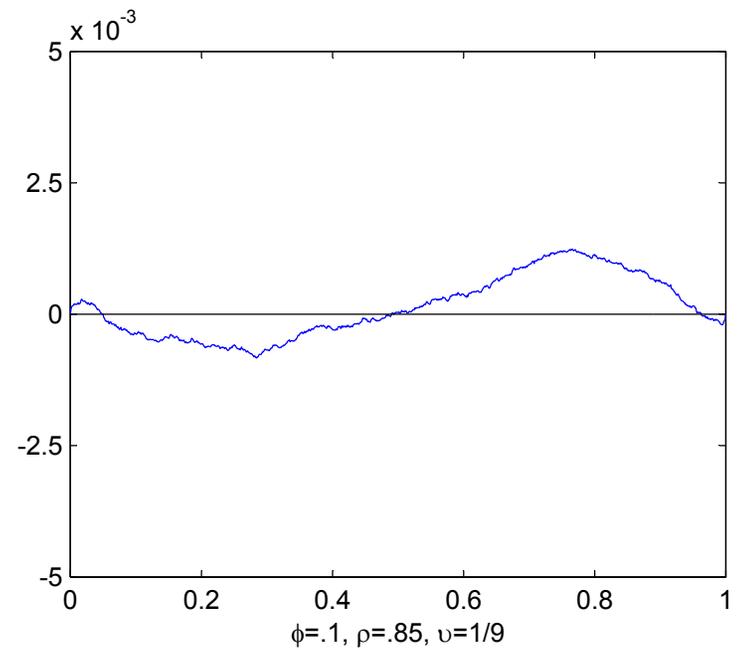
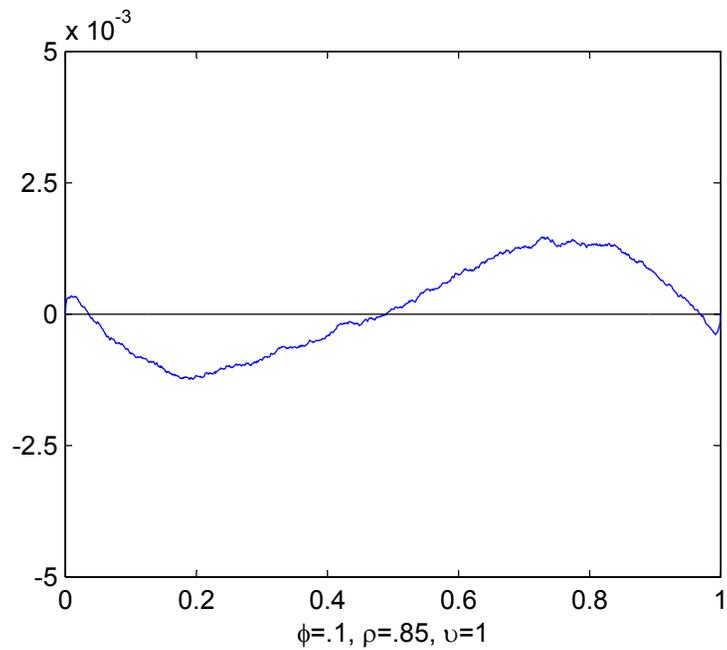
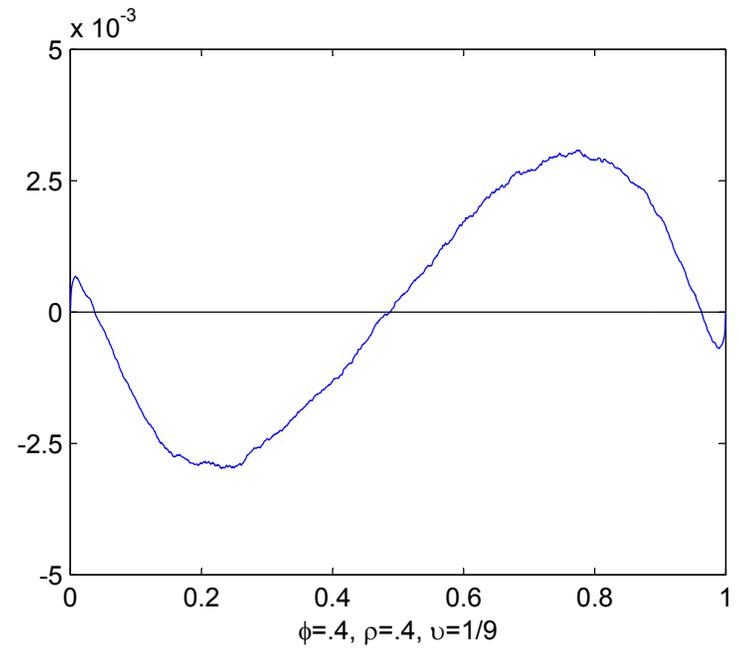
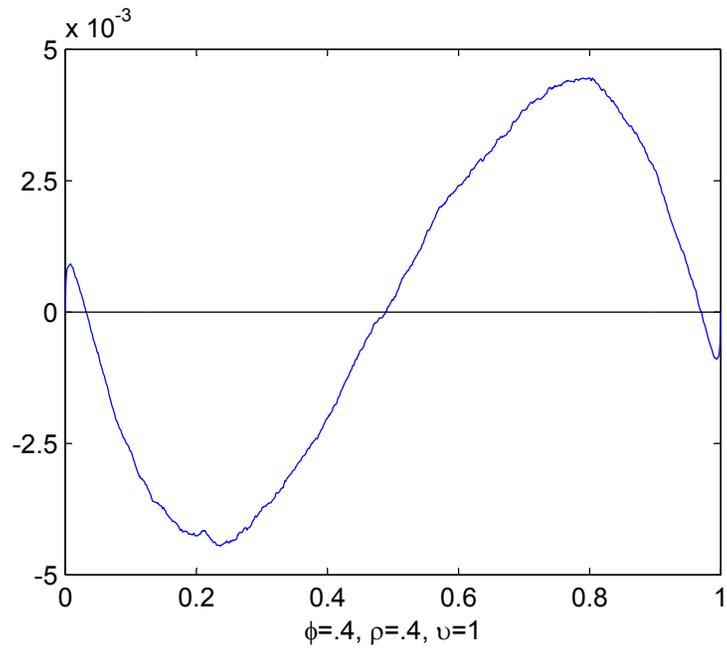


FIGURE 4A: Monte Carlo Distribution of conditional variance parameters estimators. ($b_i=1$; $\phi=.2$; $\rho=.6$)

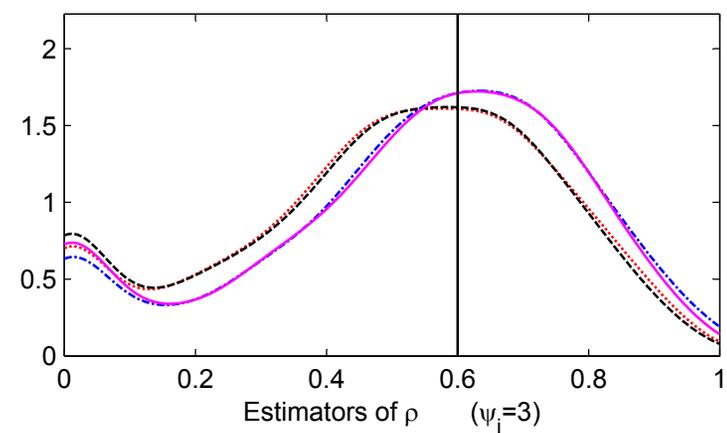
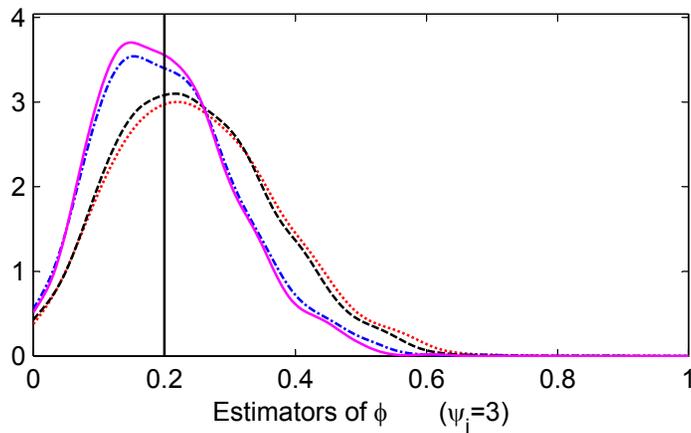
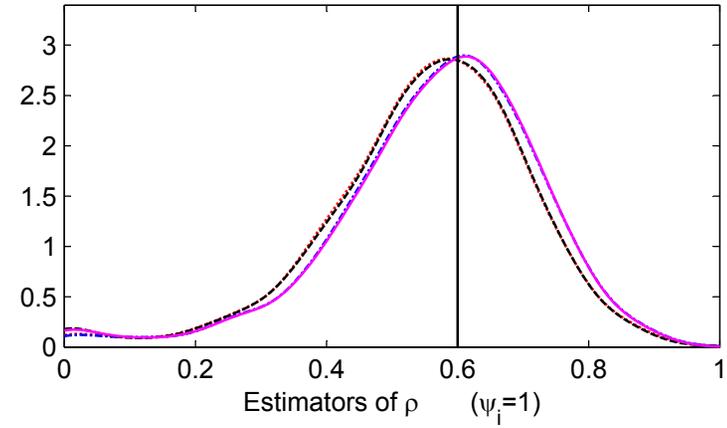
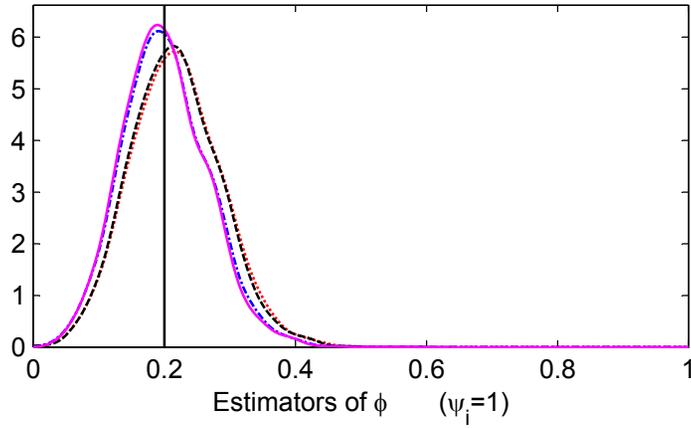
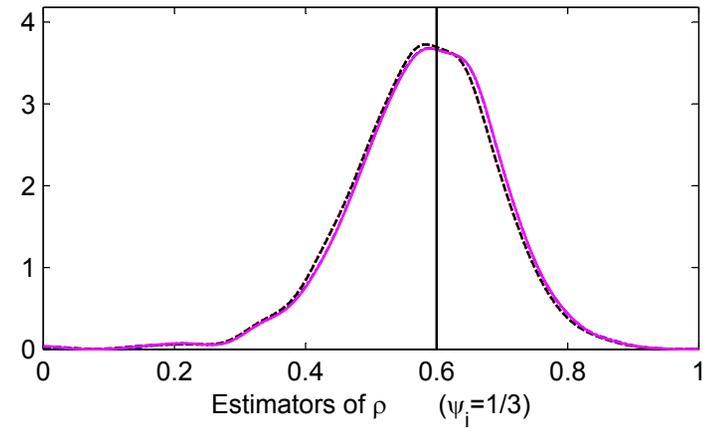
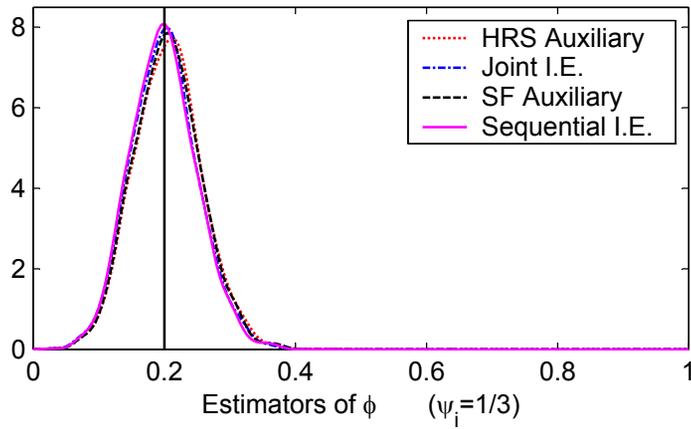


FIGURE 4B: Monte Carlo Distribution of conditional variance parameters estimators. ($b_i=1$; $\phi=.4$; $\rho=.4$)

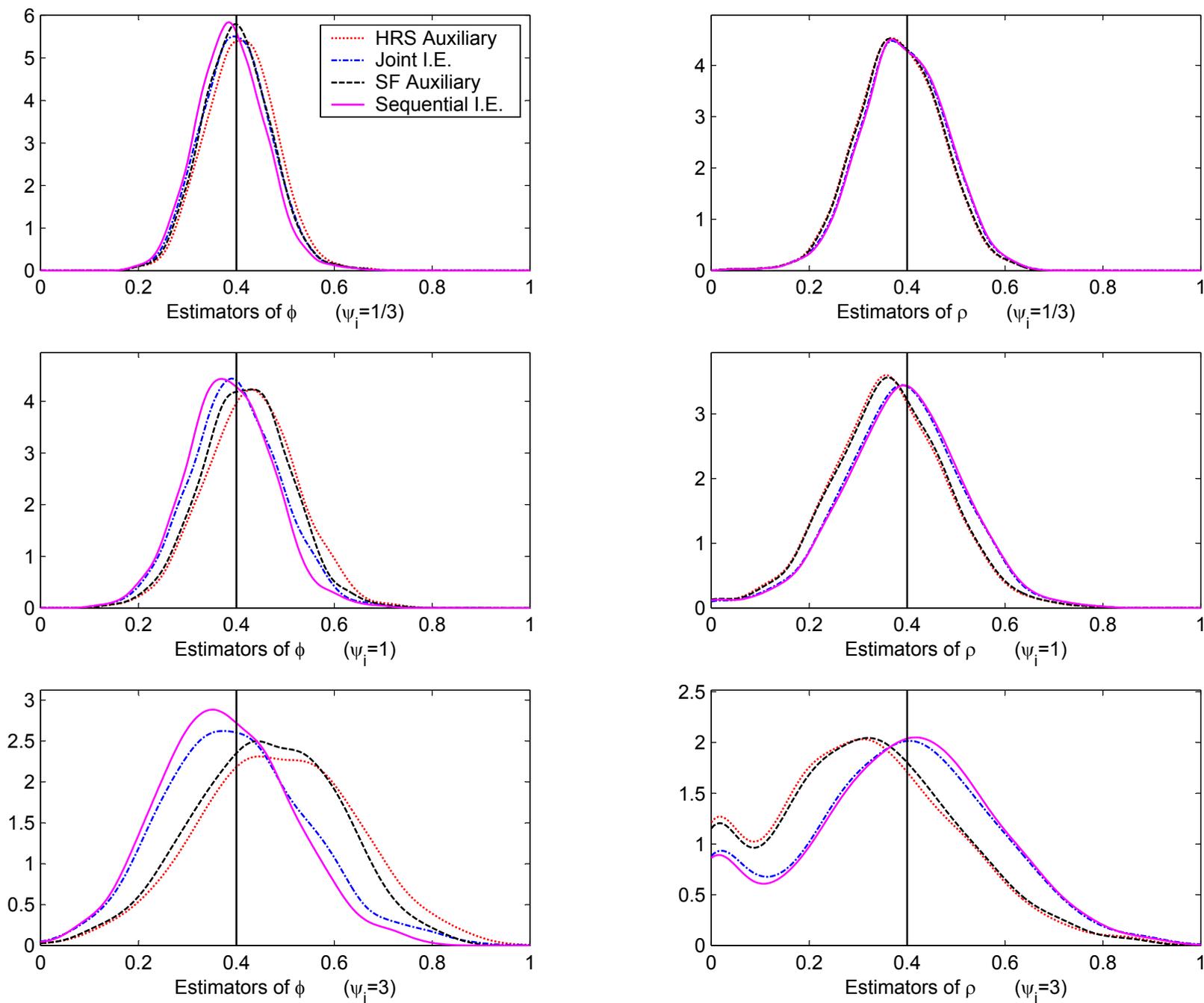


FIGURE 4C: Monte Carlo Distribution of conditional variance parameters estimators. ($b_i=1$; $\phi=.1$; $\rho=.85$)

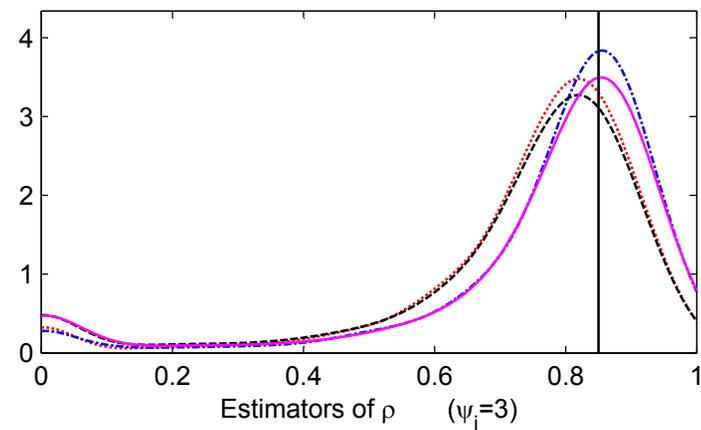
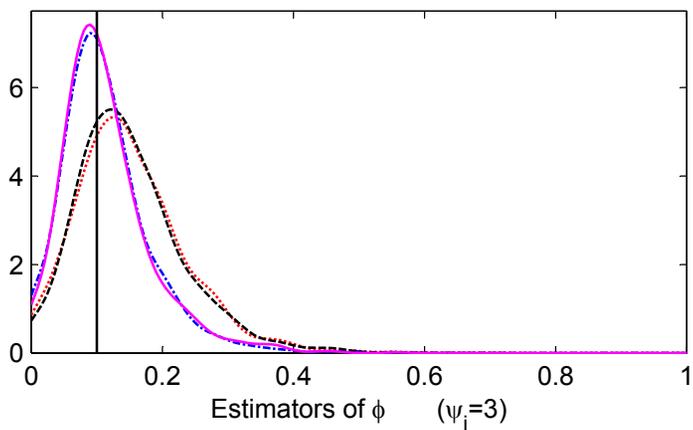
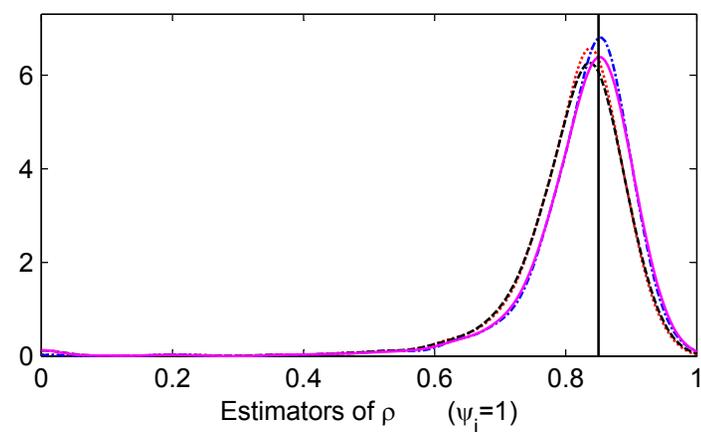
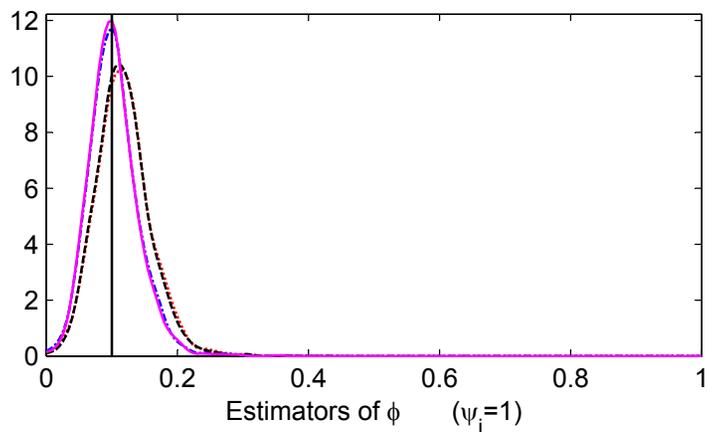
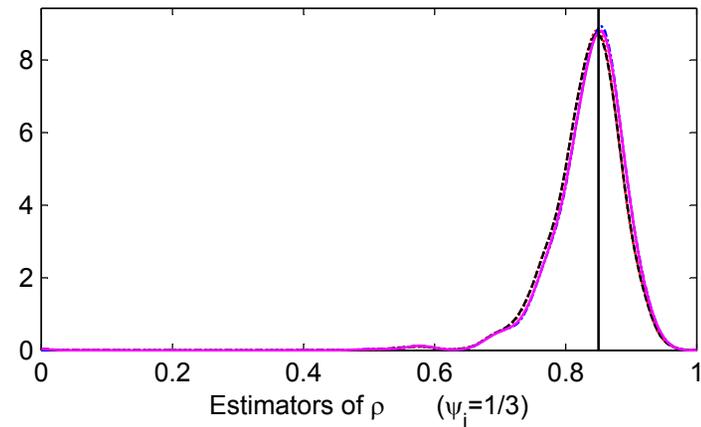
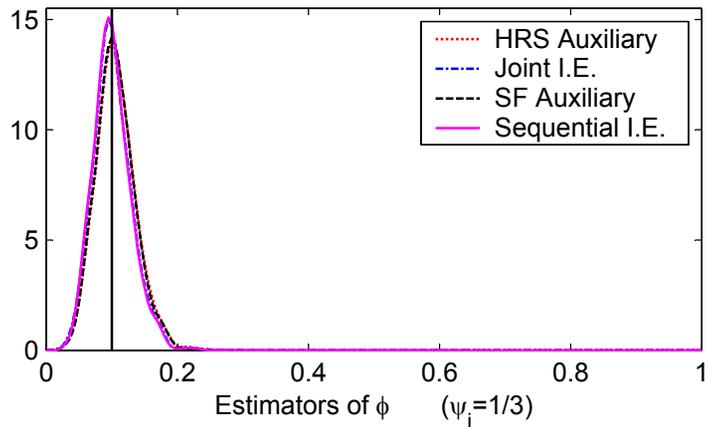


FIGURE 4D: Monte Carlo Distribution of conditional variance parameters estimators. ($b_i=1$; $\phi=.2$; $\rho=.75$)

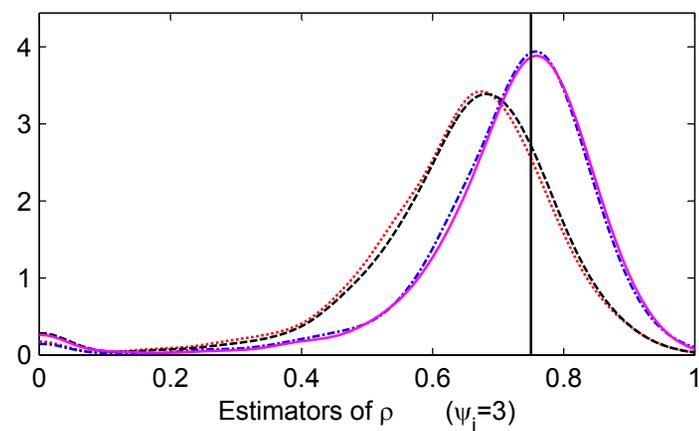
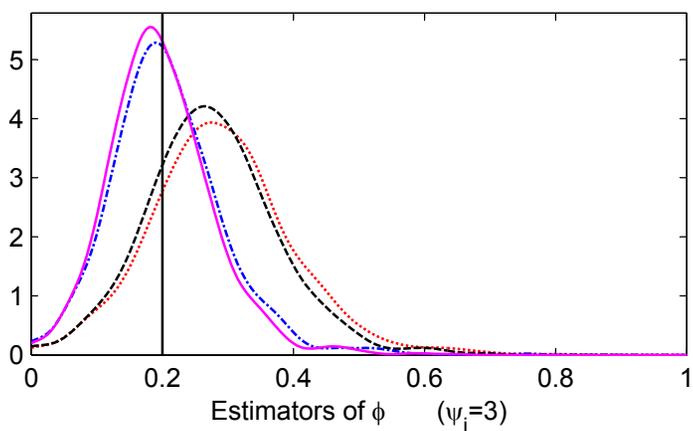
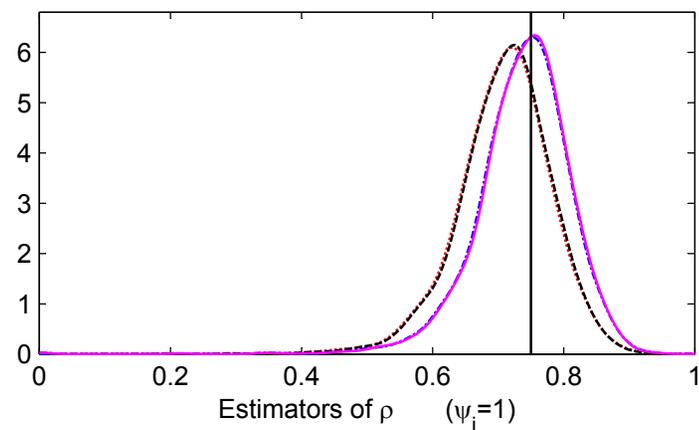
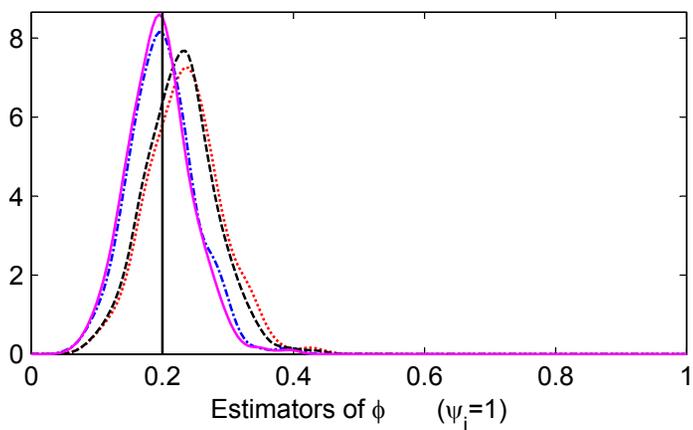
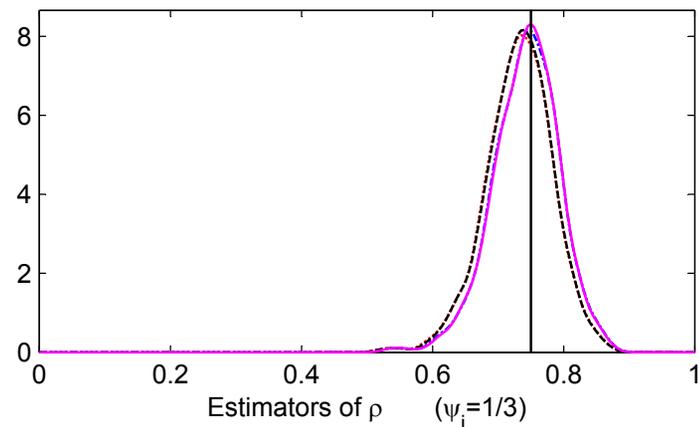
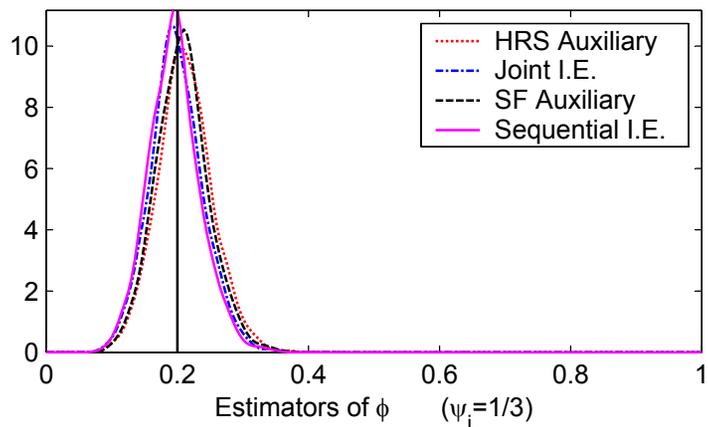
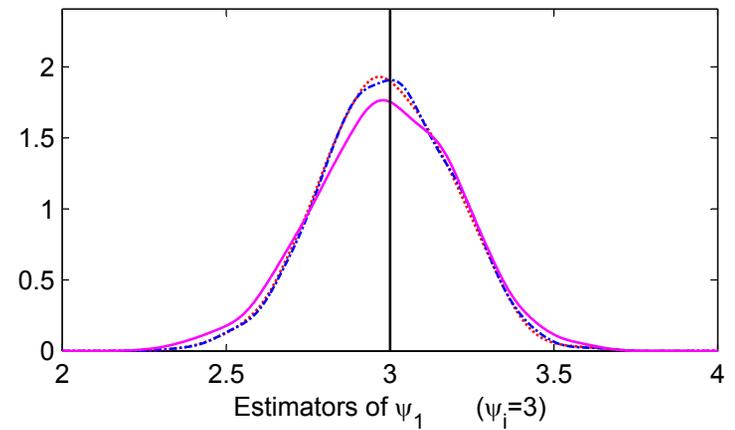
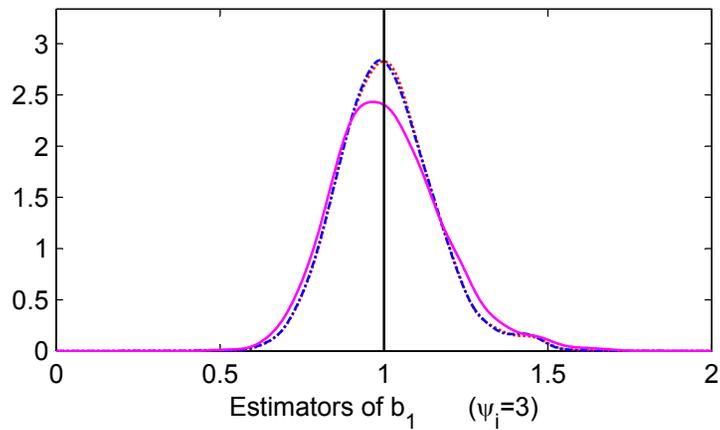
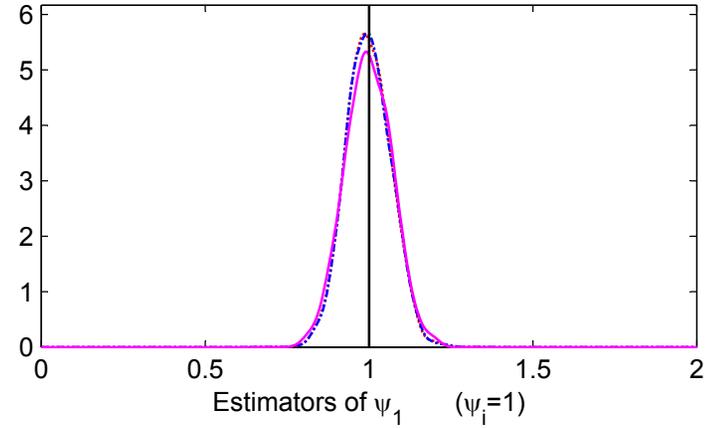
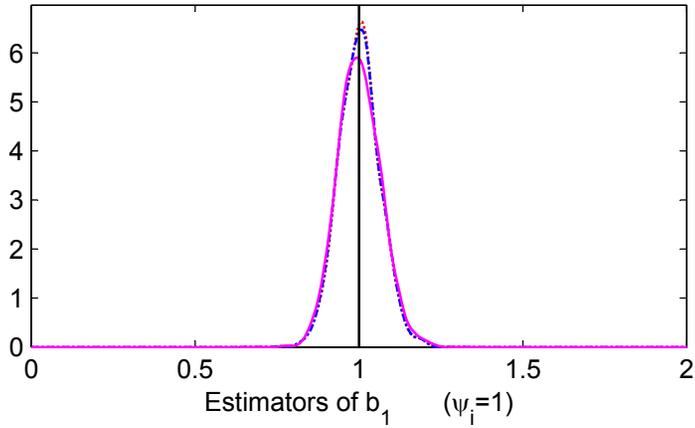
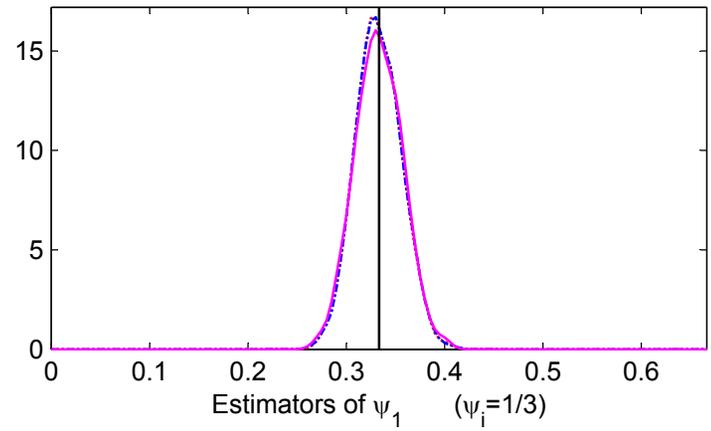
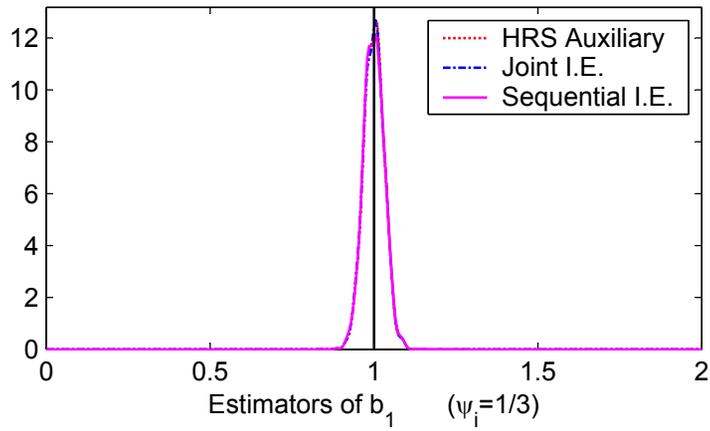


FIGURE 5: Monte Carlo Distribution of factor loading and idiosyncratic variance estimators. ($b_i=1$; $\phi=.4$; $\rho=.4$)



CEMFI WORKING PAPERS

- 0101 *Manuel Arellano*: "Discrete choices with panel data".
- 0102 *Gerard Llobet*: "Patent litigation when innovation is cumulative".
- 0103 *Andres Almazán and Javier Suarez*: "Managerial compensation and the market reaction to bank loans".
- 0104 *Juan Ayuso and Rafael Repullo*: "Why did the banks overbid? An empirical model of the fixed rate tenders of the European Central Bank".
- 0105 *Enrique Sentana*: "Mean-Variance portfolio allocation with a Value at Risk constraint".
- 0106 *José Antonio García Martín*: "Spot market competition with stranded costs in the Spanish electricity industry".
- 0107 *José Antonio García Martín*: "Cournot competition with stranded costs".
- 0108 *José Antonio García Martín*: "Stranded costs: An overview".
- 0109 *Enrico C. Perotti and Javier Suárez*: "Last bank standing: What do I gain if you fail?".
- 0110 *Manuel Arellano*: "Sargan's instrumental variable estimation and GMM".
- 0201 *Claudio Michelacci*: "Low returns in R&D due to the lack of entrepreneurial skills".
- 0202 *Jesús Carro and Pedro Mira*: "A dynamic model of contraceptive choice of Spanish couples".
- 0203 *Claudio Michelacci and Javier Suarez*: "Incomplete wage posting".
- 0204 *Gabriele Fiorentini, Enrique Sentana and Neil Shephard*: "Likelihood-based estimation of latent generalised ARCH structures".
- 0205 *Guillermo Caruana and Marco Celentani*: "Career concerns and contingent compensation".
- 0206 *Guillermo Caruana and Liran Einav*: "A theory of endogenous commitment".
- 0207 *Antonia Díaz, Josep Pijoan-Mas and José-Víctor Ríos-Rull*: "Precautionary savings and wealth distribution under habit formation preferences".
- 0208 *Rafael Repullo*: "Capital requirements, market power and risk-taking in banking".
- 0301 *Rubén Hernández-Murillo and Gerard Llobet*: "Patent licensing revisited: Heterogeneous firms and product differentiation".
- 0302 *Cristina Barceló*: "Housing tenure and labour mobility: A comparison across European countries".
- 0303 *Víctor López Pérez*: "Wage indexation and inflation persistence".
- 0304 *Jesús M. Carro*: "Estimating dynamic panel data discrete choice models with fixed effects".
- 0305 *Josep Pijoan-Mas*: "Pricing risk in economies with heterogenous agents and incomplete markets".
- 0306 *Gabriele Fiorentini, Enrique Sentana and Giorgio Calzolari*: "On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models".
- 0307 *Samuel Bentolila and Juan F. Jimeno*: "Spanish unemployment: The end of the wild ride?".

- 0308 *Rafael Repullo and Javier Suarez*: “Loan pricing under Basel capital requirements”.
- 0309 *Matt Klaeffling and Victor Lopez Perez*: “Inflation targets and the liquidity trap”.
- 0310 *Manuel Arellano*: “Modelling optimal instrumental variables for dynamic panel data models”.
- 0311 *Josep Pijoan-Mas*: “Precautionary savings or working longer hours?”.
- 0312 *Meritxell Albertí, Ángel León and Gerard Llobet*: “Evaluation of a taxi sector reform: A real options approach”.
- 0401 *Andres Almazan, Javier Suarez and Sheridan Titman*: “Stakeholders, transparency and capital structure”.
- 0402 *Antonio Diez de los Rios*: “Exchange rate regimes, globalisation and the cost of capital in emerging markets”.
- 0403 *Juan J. Dolado and Vanessa Llorens*: “Gender wage gaps by education in Spain: Glass floors vs. glass ceilings”.
- 0404 *Sascha O. Becker, Samuel Bentolila, Ana Fernandes and Andrea Ichino*: “Job insecurity and children’s emancipation”.
- 0405 *Claudio Michelacci and David Lopez-Salido*: “Technology shocks and job flows”.
- 0406 *Samuel Bentolila, Claudio Michelacci and Javier Suarez*: “Social contacts and occupational choice”.
- 0407 *David A. Marshall and Edward Simpson Prescott*: “State-contingent bank regulation with unobserved actions and unobserved characteristics”.
- 0408 *Ana Fernandes*: “Knowledge, technology adoption and financial innovation”.
- 0409 *Enrique Sentana, Giorgio Calzolari and Gabriele Fiorentini*: “Indirect estimation of conditionally heteroskedastic factor models”.