Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models[1]

Enrique Sentana (CEMFI)

Gabriele Fiorentini (University of Alicante)

Revised: October 2000

[1]This is an extensively revised version of Sentana (1992), which concentrated on identification issues. Financial support for the earlier version from the Spanish Ministry of Education and Science and the LSE Financial Markets Group as part of the ESRC project "The Efficiency and Regulation of Financial Markets" is gratefully acknowledged, as is support for the current version from IVIE, CICYT through Project PB96-0339, and MURST project "Statistical models for time series analysis". We are also grateful to Manuel Arellano, Antonis Demos, Lars Hansen, Andrew Harvey, Mervyn King, Jan Magnus, Danny Quah, Eric Renault, Esther Ruiz, Mushtaq Shah, Neil Shephard, Sushil Wadhwani and Paolo Zaffaroni, as well as seminar participants at CEMFI, Ente Einaudi (Rome, February 1999), the European Meeting of the Econometric Society (Brussels, August 1992), the ESRC Econometric Study Group Conference (Bristol, July 1997) and the Malinvaud Seminar (Paris, October 1997) for very helpful comments and suggestions. An associate editor and three anonymous referees have also helped us greatly improve the paper. Of course the usual caveat applies. Correspondence should be addressed to E. Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (tel.: +34 91 429 0551, fax: +34 91 429 1056, e-mail: sentana@cemfi.es).

Abstract

We investigate the effects of dynamic heteroskedasticity on statistical factor analysis. We show that identification problems are alleviated when variation in factor variances is accounted for. Our results apply to dynamic APT models and other structural models. We also find that traditional ML estimation of unconditional variance parameters remains consistent if the factor loadings are identified from the unconditional distribution, but their standard errors must be robustified. We develop a simple preliminary LM test for ARCH effects in the common factors, and discuss two-step consistent estimation of the conditional variance parameters. Finally, we conduct a detailed simulation exercise.

Keywords: Volatility, Likelihood Estimation, APT, Simultaneous Equations, Vector Autoregressions

JEL: C32

1 Introduction

One of the most popular approaches to multivariate volatility assumes that each observed series is a linear combination of a small number of dynamic heteroskedastic common factors plus an idiosyncratic term. Such models, which are often compatible with standard factor analysis, are particularly appealing in finance, where multi-index models enjoy a long tradition. Although some of their properties have already been analysed in detail (see Sentana (1998) and the references therein), some crucial aspects have not been fully investigated yet. In this respect, the purpose of the paper is to study in what sense the existence of time-varying heteroskedasticity in the factors affects the usual inference procedures employed in statistical factor analysis. In particular, we thoroughly reassess the identification issue, which has important implications for empirical work related to the Arbitrage Pricing Theory (APT), and it also has some bearing upon the identification of simultaneous equations systems, structural vector autoregressions and oblique factor models with constant conditional covariances. We also study the properties of unconditional maximum likelihood (ML) estimators of the static variance parameters devised for serially independent observations. In addition, we propose a simple preliminary diagnostic test for ARCH effects in the common factors, derive a two-step estimator of the remaining parameters, and investigate the finite sample properties of the different estimators and tests by means of simulation methods.

The rest of the paper is organized as follows. We introduce the model in section 2. Then, in section 3, we analyse in detail its identifiability, discuss the properties of traditional estimators, derive an LM test for ARCH in the common factors, and propose a simple two-step estimator for the conditional variance parameters. Finally, we carry out a Monte Carlo analysis in section 4. Proofs and auxiliary results are gathered in an appendix.

2 Conditionally heteroskedastic factor models

Consider the following multivariate model:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{C}\mathbf{f}_t + \mathbf{w}_t \\ \begin{pmatrix} \mathbf{f}_t \\ \mathbf{w}_t \end{pmatrix} &\mid \mathbf{X}_{t-1} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{pmatrix}\right] \end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is a vector of N observable random variables, $\mathbf{f}_t$ a vector of k unobserved common factors, $\mathbf{C}$ the $N \times k$ matrix of factor loadings, with $N \geq k$ and $\text{rank}\,(\mathbf{C}) = k$, $\mathbf{w}_t$ a vector of N idiosyncratic noises conditionally orthogonal to $\mathbf{f}_t$, $\mathbf{\Gamma} \geq \mathbf{0}$ a $N \times N$ positive semi-definite (p.s.d.) matrix of constant idiosyncratic variances, $\mathbf{\Lambda}_t$ a $k \times k$ diagonal matrix of (possibly) time-varying factor variances, which generally involve some extra parameters $\boldsymbol{\psi}$, with $\underline{\mathbf{\Lambda}} = \inf \mathbf{\Lambda}_t \geq \mathbf{0}$, and $\mathbf{X}_{t-1}$ an information set that contains the values of $\mathbf{x}_t$ up to time $t-1$. Our assumptions imply that the distribution of $\mathbf{x}_t$ conditional on $\mathbf{X}_{t-1}$ is normal with zero mean, and covariance matrix $\mathbf{\Sigma}_t = \mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma}$. For this reason, we refer to the data generation process specified by (1) as a conditionally heteroskedastic factor model. Such a formulation nests several models widely used in the empirical literature, which typically assume that the unobserved factors follow dynamic heteroskedastic processes, but differ in the exact parametrisation of $\mathbf{\Lambda}_t$ and $\mathbf{\Gamma}$ (see Sentana (1998)). Furthermore, if $\mathbf{\Lambda}_t$ is constant, which usually corresponds to $\boldsymbol{\psi} = \mathbf{0}$, it reduces to the static orthogonal factor model. But even if $\mathbf{f}_t$ is conditionally heteroskedastic, provided that it is covariance stationary, the constancy of the factor loadings implies an unconditionally orthogonal k factor structure for $\mathbf{x}_t$, so that the unconditional covariance matrix, $\mathbf{\Sigma} = E(\mathbf{\Sigma}_t)$, can be written as:

$$\mathbf{\Sigma} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}' + \mathbf{\Gamma} \tag{2}$$

where $\mathbf{\Lambda} = V(\mathbf{f}_t) = E(\mathbf{\Lambda}_t)$. This property makes the model considered here compatible with traditional factor analysis (see e.g. Lawley and Maxwell (1971)).

3 Inference

3.1 Identification

An observationally equivalent (o.e.) model to (1) up to conditional second moments must satisfy $\mathbf{\Sigma}_t = \mathbf{C}^*\mathbf{\Lambda}_t^*\mathbf{C}^{*\prime} + \mathbf{\Gamma}^*\ \forall t$, with rank$\,(\mathbf{C}^*) = k^*$, $\underline{\mathbf{\Lambda}}^* = \inf \mathbf{\Lambda}_t^* \geq \mathbf{0}$ and $\mathbf{\Gamma}^* \geq \mathbf{0}$. Let us rewrite $\mathbf{\Sigma}_t$ as $\underline{\mathbf{\Sigma}} + \vec{\mathbf{\Sigma}}_t$, with $\underline{\mathbf{\Sigma}} = \inf \mathbf{\Sigma}_t = \mathbf{C}\underline{\mathbf{\Lambda}}\mathbf{C}' + \mathbf{\Gamma}$, which we assume positive definite (p.d.), and $\vec{\mathbf{\Sigma}}_t = \mathbf{C}(\mathbf{\Lambda}_t - \underline{\mathbf{\Lambda}})\mathbf{C}' = \mathbf{C}[diag(\vec{\boldsymbol{\lambda}}_t)]\mathbf{C}' \geq \mathbf{0}$, where $\vec{\boldsymbol{\lambda}}_t = \boldsymbol{\lambda}_t - \underline{\boldsymbol{\lambda}} = vecd(\mathbf{\Lambda}_t) - vecd(\underline{\mathbf{\Lambda}})$, $diag(\mathbf{a})$ indicates the $n \times n$ diagonal matrix containing the elements of the $n \times 1$ vector $\mathbf{a}$ along the main diagonal, $vecd(\mathbf{A})$ denotes the $n \times 1$ vector containing the diagonal elements of the $n \times n$ matrix $\mathbf{A}$, and $dg(\mathbf{A})$ is a diagonal matrix containing the diagonal elements of $\mathbf{A}$, so that $diag[vecd(\mathbf{A})] = dg(\mathbf{A})$. The identifiability of constant covariance matrices of the form $\mathbf{C}\underline{\mathbf{\Lambda}}\mathbf{C}' + \mathbf{\Gamma}$ has been extensively analysed in the literature (see e.g. Anderson and Rubin (1956), Dunn (1973), Jennrich (1978), Bekker (1989), or Wedge (1996)). Consideration of the time-varying term on its own allows us to state the following independent result:

Proposition 1 *If the stochastic processes in $\vec{\boldsymbol{\lambda}}_t$ are linearly independent (i.e. $\nexists \boldsymbol{\delta} \in \mathbb{R}^k, \boldsymbol{\delta} \neq \mathbf{0} : \boldsymbol{\delta}'\vec{\boldsymbol{\lambda}}_t = 0\ \forall t$), all o.e. models to (1) (excluding column permutations and sign changes) satisfy $\mathbf{C}^*\mathbf{D}^* = \mathbf{C}$, $\mathbf{\Lambda}_t^* - \underline{\mathbf{\Lambda}}^* = \mathbf{D}^*(\mathbf{\Lambda}_t - \underline{\mathbf{\Lambda}})\mathbf{D}^*$ and $\mathbf{\Gamma}^* = \underline{\mathbf{\Sigma}} - \mathbf{C}^*\underline{\mathbf{\Lambda}}^*\mathbf{C}^{*\prime}$, where $\mathbf{D}^*$ is a $k^* \times k$ matrix, with $k^* \geq k$, such that $\mathbf{D}^{*\prime} = (\mathbf{D} \mid \mathbf{0})$, with $\mathbf{D}$ p.d. diagonal, and $\underline{\mathbf{\Lambda}}^*$ is any $k^* \times k^*$ p.s.d. diagonal matrix such that the eigenvalues of $\underline{\mathbf{\Lambda}}^*\mathbf{C}^{*\prime}\underline{\mathbf{\Sigma}}^{-1}\mathbf{C}^*$ are less than or equal to 1.*

The matrix $\mathbf{D}$ is related to the scaling of the factors. Since this is largely irrelevant, we impose in what follows that $vecd(\mathbf{\Lambda}) = vecd(\mathbf{I}_k) = \boldsymbol{\iota}_k$ by analogy with the homoskedastic case.[1] Without further restrictions, though, the model is not

[1] If the unconditional variance is unbounded, as in Integrated GARCH-type models, other scaling assumptions can be made. For instance, we can fix $\underline{\mathbf{\Lambda}}$, or the norm of each column of $\mathbf{C}$.

fully identified, as we can transfer unconditional variance from the idiosyncratic terms to the common factors via $\underline{\mathbf{\Lambda}}^*$. The most common assumptions made to differentiate the "common" and "specific" parts of $\mathbf{\Sigma}_t$ are that $k^* = k$ and $\mathbf{\Gamma}$ is diagonal. However, as the following proposition shows, diagonality of $\mathbf{\Gamma}$ is not always sufficient to guarantee identifiability in this context. Let $\mathbf{c}_l$ denote the l^{th} column of $\mathbf{C}$, and define *vecl* as the operator which stacks columnwise the elements of the strict lower triangle of a square matrix.

Proposition 2 *If $\mathbf{\Gamma}$ is diagonal, and $\vec{\mathbf{\lambda}}_t$ linearly independent, then the only admissible $\underline{\mathbf{\Lambda}}^*$ in Proposition 1 when $k^* = k$ and $\mathbf{D} = \mathbf{I}$ is $\underline{\mathbf{\Lambda}}$ iff the $N(N-1)/2 \times k$ matrix with $vecl(\mathbf{c}_l\mathbf{c}_l')$ as l^{th} column $(l = 1, \ldots, k)$ has full column rank k.*

Underidentification with $\mathbf{\Gamma}$ diagonal trivially arises when $N = k = 1$; e.g. if x_{1t} follows the GARCH(1,1) model $\sigma_{11t} = \alpha_0 + \alpha_1 x_{1t-1}^2 + \beta_1\sigma_{11t-1}$, we can write σ_{11t} as $[\alpha_0/(1-\beta_1) - \omega] + [\alpha_1 \sum_{j=0}^{\infty} \beta_1^j x_{1t-j}^2 + \omega]$ for any $\omega \in [0, \alpha_0/(1-\beta_1))$. For larger N, though, identification becomes easier:

Corollary 1 *If $k^* = k = 1$, $\mathbf{\Gamma}$ diagonal, and λ_{11t} not constant, then (1) is identified (up to "scale") iff $N \geq 2$ and at least two factor loadings are nonzero.*

In particular, if $N = 2$ and x_{1t} and x_{2t} are not conditionally uncorrelated, identification is achieved as long as λ_{11t} varies over time (cf. Lemma 5.2 in Anderson and Rubin (1956), or section 5 in Scherrer and Deistler (1998)).

If the unconditional covariance matrix is partly identified even if we ignore the time-variation in the conditional variances (see e.g. Theorems 5.1 - 5.8 in Anderson and Rubin (1956)), a stronger result can be obtained:

Proposition 3 *If $\mathbf{CC}'$ and $\mathbf{\Gamma}$ are uniquely identified from the unconditional covariance matrix, and the stochastic processes in $\mathbf{\lambda}_t$ are linearly independent, $\mathbf{C}$ is unique up to column permutations and sign changes.*

The main difference is that while in Proposition 1 we are implicitly assuming that none of the original factors has constant variance, here we allow for one (but only one) λ_{llt} to be constant $\forall t$. If the processes in $\mathbf{\lambda}_t$ were linearly dependent, though, identification problems would re-appear. Given the parametrisations used in practice, it is difficult to envisage such situations, unless several factor variances are constant. Nevertheless, consider as an example a model in which for all time periods, a group of k_2 factors ($1 < k_2 < k$) is characterised by a scalar covariance matrix $\lambda_{kk,t}\mathbf{I}_{k_2}$, while the others have an unrestricted diagonal covariance matrix $\boldsymbol{\Lambda}_{1t}$. If we partition $\mathbf{C}$ conformably as $\mathbf{C} = (\mathbf{C}_1 \mid \mathbf{C}_2)$, where $\mathbf{C}_1$ and $\mathbf{C}_2$ are $N \times k_1$ and $N \times k_2$ respectively, with $k_1 + k_2 = k$, the following result can be stated:

Proposition 4 *If $\mathbf{CC}'$ and $\boldsymbol{\Gamma}$ are uniquely identified from the unconditional covariance matrix, and the stochastic processes in $(\boldsymbol{\lambda}_{1t}', \lambda_{kk,t})$ are linearly independent, $\mathbf{C}_1$ is unique up to column permutations and sign changes.*

If $k_2 = k$ and $\lambda_{kk,t} = 1$, we return to the static factor model, where even when $\mathbf{CC}'$ and $\boldsymbol{\Gamma}$ are separately identified, $\mathbf{C}$ may be only identified up to rotation by an arbitrary $k \times k$ orthogonal matrix $\mathbf{Q}$. In this respect, note that the imposition of unnecessary restrictions on $\mathbf{C}$ as in traditional factor analysis to ensure that the only admissible $\mathbf{Q}$ is $\mathbf{I}_k$ may produce misleading results. An important implication of our propositions is that if such restrictions are nevertheless made, at least they can be tested. However, the accuracy with which $\mathbf{C}$ can be estimated depends on how much linearly independent variability there is in $\boldsymbol{\lambda}_t$, for if its elements are essentially constant, identifiability problems will reappear.

Our results can be applied to other closely related models, such as $\mathbf{x}_t = \mathbf{C}\boldsymbol{\Lambda}_t\boldsymbol{\tau} + \mathbf{C}\mathbf{f}_t + \mathbf{w}_t$ (see e.g. Engle et al. (1990) or King et al. (1994)), where the columns of $\mathbf{C}$ and the "price of risk" coefficients in $\boldsymbol{\tau}$ corresponding to factors with linearly independent time-varying variances are identifiable (up to sign changes and permutations) from Proposition 3. They also apply to models with N common factors, $\boldsymbol{\Gamma} = \mathbf{0}$ and linear mean dynamics, such as the static simultaneous equation system $\mathbf{A}_0\mathbf{x}_t = \mathbf{f}_t$, or the "structural" VAR process $\mathbf{x}_t = \sum_{j=0}^{\infty} \mathbf{B}_j \mathbf{u}_{t-j}$, $\mathbf{B}_0 = \mathbf{I}$, $\mathbf{u}_t = \mathbf{C}\mathbf{f}_t$, where in both cases $\mathbf{f}_t$ can be understood as the conditionally orthogonal "fundamental" shocks driving $\mathbf{x}_t$. If we estimate these models without considering the time-variation in $\boldsymbol{\Lambda}_t$, neither $\mathbf{A}_0$ nor $\mathbf{C}$ are identifiable without extra restrictions, such as $\mathbf{A}_0, \mathbf{C}$ or $\sum_{j=0}^{\infty} \mathbf{B}_j\mathbf{C}$ lower triangular (see Blanchard and Quah (1989)). But if some elements of $\mathbf{f}_t$ have time-varying variances and this is explicitly recognized in estimation, then Proposition 4 implies that the corresponding columns of $\mathbf{C}$ are identifiable. Alternatively, Proposition 4 says that the set of conditionally uncorrelated disturbances that can be written as a time-invariant linear combination of the innovations in $\mathbf{x}_t$ is unique when $k_2 \leq 1$.

Finally, it turns out that most of the identifiability in fact derives from the conditional covariances of conditionally orthogonal factors being constant over time. Specifically, let $\mathbf{\Lambda}_t$ be a p.d. matrix of possibly time-varying factor variances but constant conditional covariances. Then, given that $\vec{\mathbf{\Lambda}}_t = dg(\vec{\mathbf{\Lambda}}_t)$, Proposition 1 remains valid with $\underline{\mathbf{\Lambda}}^*$ p.s.d. but not necessarily diagonal. In this respect, notice the generality of Proposition 1, which only relies on the constancy of $\mathbf{C}$ and the conditional covariances of the factors, and the linearly independent time-variation of their variances, but not on any particular parametrisation for $\boldsymbol{\lambda}_t$.[2]

3.2 Estimation of unconditional variance parameters

Ignoring initial conditions, the log-likelihood function of a sample of size T of observations generated from model (1) takes the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, where:

$$l_t(\boldsymbol{\phi}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma}| - \frac{1}{2}\mathbf{x}_t'\left(\mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma}\right)^{-1}\mathbf{x}_t \qquad (3)$$

$\boldsymbol{\phi} = (\mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\psi}')'$, $\mathbf{c} = vec(\mathbf{C})$, $\boldsymbol{\gamma} = vecd(\mathbf{\Gamma})$ and $\mathbf{\Lambda}_t = diag\left[\boldsymbol{\lambda}_t(\boldsymbol{\phi})\right]$. Note that we restrict $\mathbf{\Gamma}$ to be diagonal and in principle p.d. in view of the discussion in section 3.1, but allow $\boldsymbol{\lambda}_t$ to depend on $\boldsymbol{\psi}, \mathbf{c}$ and $\boldsymbol{\gamma}$. In this context, a standard factor analytic routine can be regarded as estimating the unconditional variance parameters $\mathbf{c}$ and $\boldsymbol{\gamma}$ as $(\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}}) = \arg\max_{\mathbf{c},\gamma} L_T(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0})$. If (a) $\mathbf{\Gamma}$ and $\mathbf{C}$ are identified (up to rotation) from unconditional moments (b) $\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{\Sigma}_0 = \mathbf{C}_0\mathbf{C}_0' + \mathbf{\Gamma}_0$,

[2] In this respect, Proposition 1 in Sentana (1998) can be obtained as an application of our Proposition 1 to the factor GARCH case. Similarly, Proposition 1 in Rigobon (2000) can also be obtained as a special case of our Proposition 3 when $\boldsymbol{\lambda}_t$ follows a switching regime model.

where the 0 subscripts indicate true values, (c) $T^{-1/2}\sum_{t=1}^{T} vech(\mathbf{x}_t\mathbf{x}_t' - \mathbf{\Sigma}_0)$ has a limiting normal distribution and (d) the matrix $(\mathbf{\Gamma}_0^G \odot \mathbf{\Gamma}_0^G)$ is nonsingular, where $\mathbf{\Gamma}_0^G = \mathbf{\Gamma}_0 - \mathbf{C}_0(\mathbf{C}_0'\mathbf{\Gamma}_0^{-1}\mathbf{C}_0)^{-1}\mathbf{C}_0'$ is the rank $N-k$ covariance matrix of the GLS estimates of the idiosyncratic factors $\mathbf{w}_t^G = \mathbf{x}_t - (\mathbf{C}_0'\mathbf{\Gamma}_0^{-1}\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{\Gamma}_0^{-1}\mathbf{x}_t$, and $\odot$ denotes Hadamard product, then theorem 12.1 in Anderson and Rubin (1956) and theorem 2 in Kano (1983) imply that $(\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}})$ are asymptotically normally distributed around $vec(\mathbf{C}_0\mathbf{Q}_0)$ and $vecd(\mathbf{\Gamma}_0)$, where $\mathbf{Q}_0$ is the orthogonal matrix that imposes on $\mathbf{C}_0$ the restrictions used in estimation to avoid the usual rotational indeterminacy.[3] However, even though the expected value of the score of the estimated model evaluated at $\boldsymbol{\phi}_0$ is 0 under our assumptions, it does not preserve the martingale difference property when there are ARCH effects in the common factors because the first derivatives are proportional to $vech(\mathbf{x}_t\mathbf{x}_t')$ (see the appendix). Hence, standard errors computed assuming conditional homoskedasticity will be wrong, and it is necessary to robustify them taking into account the serial correlation in $vech(\mathbf{x}_t\mathbf{x}_t')$.

3.3 A simple test for ARCH in the common factors

If the factors were observable, we could carry out standard ARCH tests on them. We can derive similar tests using their expected values evaluated at parameter configurations consistent with the null, i.e. $\mathbf{f}_{t|t}(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0}) = E(\mathbf{f}_t|\mathbf{X}_t; \mathbf{c}, \boldsymbol{\gamma}, \mathbf{0}) = \mathbf{C}'\mathbf{\Sigma}^{-1}\mathbf{x}_t$. If we regard $\mathbf{f}_{t|t}(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0})$ as k particular linear combinations of the elements in $\mathbf{x}_t$, their true distribution will be given by $\mathbf{f}_{t|t}(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0})|\mathbf{X}_{t-1}, \boldsymbol{\phi}_0 \sim N\left\{\mathbf{0}, \mathbf{C}'\mathbf{\Sigma}^{-1}\left[\mathbf{C}_0 diag\left[\boldsymbol{\lambda}_t(\boldsymbol{\phi}_0)\right]\mathbf{C}_0' + \mathbf{\Gamma}_0\right]\mathbf{\Sigma}^{-1}\mathbf{C}\right\}$, so that $\mathbf{f}_{t|t}(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0})$ will be homoskedastic if $\boldsymbol{\lambda}_t(\boldsymbol{\phi}_0)$ is constant over time. Hence, we can test whether moment con-

[3] Primitive conditions for (b) and (c) in *univariate* dynamic heteroskedastic contexts are only beginning to emerge. In particular, theorem 5.1 in Giraitis et al. (2000) implies that if x_{1t} is a strictly stationary strong ARCH(∞) process with zero mean and a bounded unconditional fourth moment, $T^{-1/2}\sum_{t=1}^{T}[x_{1t}^2 - V(x_{1t})] \xrightarrow{d} N[0, \sum_{j=-\infty}^{\infty} cov(x_{1t}^2, x_{1t-j}^2)]$.

ditions such as $cov\left[f^2_{jt|t}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0}), f^2_{jt-1|t-1}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})\right] = 0$, $j = 1,\ldots,k$ hold. Moreover, given that $f_{jt|t}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})$ will follow a weak GARCH process as long as the j^{th} row of $\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}_0$ is not $\mathbf{0}$ (see Nijman and Sentana (1996)), such tests have non-trivial power because $f_{jt|t}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})$ will show serial correlation in the squares under the alternative. In practice, we shall base the tests on $\hat{\mathbf{f}}_{t|t} = \mathbf{f}_{t|t}(\hat{\mathbf{c}},\hat{\boldsymbol{\gamma}},\mathbf{0})$, where $\hat{\mathbf{c}}$ and $\hat{\boldsymbol{\gamma}}$ are the estimators discussed in the previous section. Nevertheless, when conditions (a)-(d) hold, it is straightforward to show that (i) the asymptotic null distribution is unaffected because the covariance matrix of the moment restrictions defining the static variance parameters and the moment restrictions being tested is block diagonal under the null (see the appendix), and (ii) $\hat{\mathbf{C}}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C}_0 \xrightarrow{p} \mathbf{Q}_0'\mathbf{C}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0$, a full rank matrix. We also show that if $\hat{\mathbf{c}}$ is not merely root-T consistent for $vec(\mathbf{C}_0\mathbf{Q}_0)$, but actually for $\mathbf{c}_0$, then the aforementioned moment test is precisely the standard LM test of conditional homoskedasticity vs ARCH-type behaviour in the common factors based on the score of (3) evaluated under H_0. In particular, we can compute a two-sided LM test against ARCH(1) in each common factor as T times the uncentred R^2 from the regression of either 1 on $(\hat{f}^2_{jt|t} + \hat{\omega}_{jj,t|t} - 1)$ times $(\hat{f}^2_{jt-1|t-1} + \hat{\omega}_{jj,t-1|t-1} - 1)$ (outer-product version), or $(\hat{f}^2_{jt|t} + \hat{\omega}_{jj,t|t} - 1)$ on $(\hat{f}^2_{jt-1|t-1} + \hat{\omega}_{jj,t-1|t-1} - 1)$ (Hessian version), where $\boldsymbol{\Omega}_{t|t}(\boldsymbol{\phi}) = V(\mathbf{f}_t|\mathbf{X}_t;\boldsymbol{\phi}) = \left[\boldsymbol{\Lambda}_t(\boldsymbol{\phi})^{-1} + (\mathbf{C}'\boldsymbol{\Gamma}^{-1}\mathbf{C})\right]^{-1}$. More powerful variants of these tests can be obtained by taking into account the sign of the relevant regression coefficient (see Demos and Sentana (1998b)).

3.4 Estimation of conditional variance parameters

On the basis of well-known results from Durbin (1970), we can show that if $\hat{\mathbf{c}}$ and $\hat{\boldsymbol{\gamma}}$ are root-T consistent for $\mathbf{c}_0$ and $\boldsymbol{\gamma}_0$, we can obtain root-T consistent but inefficient estimates of the conditional variance parameters by maximising (3) with respect to $\boldsymbol{\psi}$ keeping $\mathbf{c}$ and $\boldsymbol{\gamma}$ fixed at those consistent estimates. However, since the asymptotic covariance matrix is not generally block-diagonal between static and dynamic variance parameters when $\boldsymbol{\psi}_0 \neq \mathbf{0}$, standard errors will be underestimated by the usual expressions. Asymptotically correct standard errors can be computed from an estimate of the inverse information matrix corresponding to (3) evaluated at these two-step estimators (cf. Lin (1992)). Note that if we were to iterate the two-step procedure and achieved convergence, we would obtain the fully efficient conditional ML estimates of all model parameters.[4]

Obviously, if the initial estimates of $\mathbf{c}$ are only consistent for $vec(\mathbf{C}_0\mathbf{Q}_0)$ because $\mathbf{C}$ is not uniquely identifiable from the unconditional covariance matrix (e.g. if $k \geq 2$ and $\mathbf{C}$ unrestricted), then the two-step estimator of $\boldsymbol{\psi}$ will be inconsistent. One possibility would be to replace $\hat{\mathbf{c}}$ by a consistent estimator based on an alternative objective function that took into account the autocorrelation in $vech(\mathbf{x}_t\mathbf{x}_t')$. Unfortunately, the evidence from univariate ARCH models suggests that the resulting estimators are likely to be rather inefficient.

4 Monte Carlo evidence

We generated 8000 samples of 240 observations (plus 100 initial ones) of a trivariate single factor model. Since the performance of the estimators depends on $\mathbf{C}_0$ and $\mathbf{\Gamma}_0$ mostly through $\left(\mathbf{C}_0'\mathbf{\Gamma}_0^{-1}\mathbf{C}_0\right)$, we set $\mathbf{c}_0=(1,1,1)'$, $\mathbf{\Gamma}_0=\gamma_0\mathbf{I}$, $\gamma_0 = 2$ or $1/2$, corresponding to low and high signal to noise ratios, $\lambda_t = (1-\alpha_0-\beta_0) + \alpha_0(f_{t-1|t-1}^2+\omega_{t-1|t-1})+\beta_0\lambda_{t-1}$, and $(\alpha_0,\beta_0) = (0,0), (.2,.6)$ or $(.4,.4)$, which represent constant variances, persistent but smooth GARCH behaviour, and persistent but volatile conditional variances respectively ($\gamma_0 = 1/2, \alpha_0 = .2, \beta_0 = .6$ matches roughly what we tend to see in the empirical literature). Note that λ_t differs from

[4] Such an iterated estimation procedure is closely related to the zig-zag estimation method suggested in Demos and Sentana (1992), which combined the EM algorithm to estimate the static factor parameters conditional on the values of the conditional variance parameters, followed by the direct maximisation of (3) with respect to $\boldsymbol{\psi}$ holding $\mathbf{c}$ and $\boldsymbol{\gamma}$ fixed.

a standard GARCH(1,1) model in that the unobserved factors are replaced by their best (in the conditional mean square error sense) estimates $f_{t-1|t-1}$, and the term $\omega_{t-1|t-1}$ is included to reflect the uncertainty in the factor estimates (see Harvey et al., 1992). We use the same underlying random numbers in all designs to minimise experimental error, and maximise the log-likelihood (3), with initial values obtained via the EM algorithm in Demos and Sentana (1998a). For scaling purposes, we use $c_1^2 + c_2^2 + c_3^2 = 1$. We use the re-parametrisation $\gamma_i = (\gamma_i^*)^2, \alpha = \sin^2(\theta_1)$ and $\beta = \sin^2(\theta_2)(1-\alpha)$ to guarantee $\gamma_i \geq 0$ and $0 \leq \beta \leq 1-\alpha \leq 1$. We also set λ_1 to $E(\lambda_t)$ to start up the recursions, but since this implies that β is not identified if $\alpha = 0$, we set $\hat{\beta} = 0$ whenever $\hat{\alpha} = 0$. Given that these parameter values lie on the boundary of the admissible range, to compute standard errors we use case 2, theorem 2 of Self and Liang (1987), which implies that the asymptotic distribution of the ML estimators of $(\beta, \alpha, \mathbf{c}', \boldsymbol{\gamma}')$ when $\alpha_0 = \beta_0 = 0$ should be a $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ mixture of (i) the usual asymptotic distribution, (ii) the asymptotic distribution of a restricted ML estimator which sets $\alpha = \beta = 0$, and (iii) the asymptotic distribution of a restricted ML estimator which only sets $\beta = 0$.

Table 1 presents mean biases and standard deviations across replications for conditional (C) and unconditional (U) ML estimates of the static factor model parameters. Note that although they all are very mildly downward biased, the more variable λ_t is, the better CML estimates are relative to UML ones. Nevertheless, the differences are minor for the sample size used. Given the large number of parameters, we summarise the performance of the estimates of the asymptotic covariance matrix of the estimators by computing the experimental distribution of some simple test statistics. In particular, we test $c_1{=}c_2{=}c_3$, and $\gamma_1{=}\gamma_2{=}\gamma_3$, and should obtain asymptotic χ_2^2 distributions under the null. Standard errors for CML estimates are computed from the Hessian, while the usual sandwich estimator with a 4-lag triangular window is employed for UML estimates. The results, not reported for conciseness, suggest that the size distortions are not very large.

Our experimental design also allows us to analyse the performance of the LM tests for ARCH. The size properties under the null of the one-sided and two-sided versions of the outer-product and Hessian-based forms are summarised in Figure 1 using Davidson and MacKinnon's (1998) **p-value discrepancy plots**. As expected, the outer-product versions have much larger distortions than the Hessian-based ones, whose sizes are fairly accurate. The evidence on power for the Hessian-based one-sided and two-sided tests is presented in Figure 2 using Davidson and MacKinnon's (1998) **size-power curves**. As can be seen, power is an increasing function of both the value of α, and the signal-to-noise ratio. Also, our results confirm that one-sided versions are always more powerful than two-sided ones, although not overwhelmingly so (cf. Demos and Sentana (1998b)).

Asymptotically, the proportions of $\hat{\alpha} = 0$ and $\hat{\alpha} \neq 0, \hat{\beta} = 0$ should be $(\frac{1}{2}, \frac{1}{4})$ if α_0=β_0=0, and $(0, 0)$ otherwise. But Table 2 shows that $\hat{\alpha}$=0 and $\hat{\beta}$=0 occur more frequently in finite samples, especially when the signal-to-noise ratio is small. These results are confirmed in Table 3, which presents mean biases and standard deviations across replications for the conditional and two-step (2S) ML estimates of section 3.4 (the figures for $\hat{\beta}$ correspond to $\hat{\alpha} \neq 0$). Note that the $\hat{\alpha}'s$ are rather more accurate than the $\hat{\beta}'s$. Also note that the biases for the CML estimates of α are smaller than for the 2S ones, although the latter have smaller Monte Carlo variability. In contrast, the downward biases in β are larger for CML estimates, which, to some extent, reflects the larger proportion of zero $\hat{\beta}'s$ in Table 2.

We have also simulated a six-variate model with two factors in which $\lambda_{11,t} = (1 - \alpha_0 - \beta_0) + \alpha_0(f^2_{1t-1|t-1} + \omega_{11,t-1|t-1}) + \beta_0\lambda_{11,t-1}$, $\lambda_{22,t} = 1$ and $\mathbf{\Gamma}_0$=$\gamma_0\mathbf{I}$. Note from Proposition 3 that $\mathbf{c}$ is identified without further restrictions, provided that $\alpha \neq 0$ and we take into account the time-variation in conditional variances. We have selected $\gamma_0 = 2$ or $1/2$, and $(\alpha_0, \beta_0) = (.4, .4)$ or $(.2, .6)$, and

$\mathbf{c}_0'$=$(0, 0, 0, 1, 1, 1; 1, 1, 1, 0, 0, 0)$, which corresponds to two models like the one considered in the previous subsection put together, or $\mathbf{c}_0'$= $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, 1, 1; 1, 1, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, which introduces "correlation" in the columns of $\mathbf{C}$. Given that this model is four times as costly to estimate as the previous one, we only generated 2000 samples of 240 observations each. The remaining details are as in section 4.1.[5]

Table 4 presents mean biases and standard deviations across replications for CML and UML estimates, as well as a restricted ML estimator which imposes the same identifying restriction as the UML estimator, i.e. $c_{62} = 0$. Such an estimator is efficient when the overidentifying restriction is true, but inconsistent if it is false. More precisely, RML and UML estimators of $\mathbf{c}$ are consistent for a rotation of $\mathbf{c}_0$ such that $[\mathbf{C}_0\mathbf{Q}_0]_{62} = 0$. The first panel of Table 4 contains the results for the designs with $\mathbf{C}_0$ "orthogonal". Not surprisingly, the RML estimator is clearly the best as far as estimates of the factor loadings are concerned, but the UML estimator performs very similarly, except when there is significant variability in $\lambda_{11,t}$, which is in line with the single factor model results. In contrast, the CML estimator is the worst performer when the signal to noise ratio and the variability in $\lambda_{11,t}$ are low, but comes very close to the RML in the opposite case.[6] This behaviour is not unexpected, given that the identifiability of the CML estimator comes from $\lambda_{11,t}$ changing over time, while for the other two estimators it comes from the restriction $c_{62} = 0$. It seems, though, that the latter identifiability condition is more informative than the former, which should be borne in mind in empirical work. There are only minor differences, though, between the estimates of the idiosyncratic variance parameters. Obviously, their Monte Carlo standard deviations increase with γ_0, but the coefficients of variation remain approximately

[5] One additional issue was that occasionally some idiosyncratic variances were estimated as 0 (see Sentana (2000)). The incidence of Heywood cases increased with γ_0, and especially c_{62}. Nevertheless, since at worst only 35 out of 2000 replications had this problem, we replaced them by new ones.

[6] Since the CML estimates of $\mathbf{c}$ are not identified if $\alpha = 0$, the reported values are for $\hat{\alpha} \neq 0$.

the same. The second panel of Table 4 contains the results for the other designs. Note that the different estimates of γ_j are hardly affected. As expected, though, the behaviour of both RML and UML estimators of $\mathbf{C}$ radically changes, as they clearly become inconsistent. In contrast, the performance of the CML estimator of $\mathbf{c}$ is basically the same as in the first panel. We computed the experimental distribution of some simple test statistics to summarise the performance of the estimates of the asymptotic covariance matrix of these estimators. In particular, we test $c_{11}=c_{21}=c_{31}$; $c_{41}=c_{51}=c_{61}$; $c_{12}=c_{22}=c_{32}$; $\gamma_1=\gamma_2=\gamma_3$ and $\gamma_4=\gamma_5=\gamma_6$. Given that our choices of $\boldsymbol{\phi}_0$ imply that the plims of all the estimators satisfy these restrictions even when $c_{62} \neq 0$, they should all have asymptotic χ^2_2 distributions. The results, not reported for conciseness, suggest that the size distortions associated with the UML estimator, are small, but larger than for the others.

Our design also allows us to consider the finite sample distribution of the likelihood ratio (LR) test for the restriction $c_{62} = 0$. The p-value discrepancy plot in Figure 3 shows that nominal sizes are fairly accurate at the 5% level, though less so when γ_0 is small. For very large significance levels, however, the size distortions are higher, as the test takes the value 0 whenever $\hat{\alpha} = 0$. Its distribution under the alternative is far more interesting, and provides a summary indicator of the determinants of the information in our identifiability restrictions. Figure 4 present the size-power curves for the four experimental designs in which $c_{62} \neq 0$, with the required implicit size-corrections based on the closest match (cf. Davidson and MacKinnon (1998)). The absolute power of the test is small, since the Monte Carlo variability in the joint estimator of c_{62} is large relative to the re-scaled value of this parameter ($\simeq .14$) for the sample size considered (see Table 4). Nevertheless, the power of the test increases with the signal-to-noise ratio, and especially, with the variability of the conditional variance of the factor. This confirms the crucial role that changes in $\lambda_{11,t}$ play in the identifiability of the model.

Table 5 presents the proportion of estimates of α and β at the boundary of the parameter space, which should be 0 asymptotically. But as before, $\hat{\alpha} = 0$, and especially $\hat{\beta} = 0$ occur more frequently in finite samples, especially when the signal-to-noise ratio is small. These results are confirmed in Table 6, which presents mean biases and standard deviations across replications for CML, RML, and the 2S estimators of α and β of section 3.4. Once more, the $\alpha's$ are estimated rather more accurately than the $\beta's$, which reflects the larger proportion of zero $\hat{\beta}'s$ in Table 5. As in Table 4, though, there are significant differences between the two panels. While the performance of the CML estimators is by and large independent of whether or not $c_{62} = 0$, the behaviour of RML and 2S estimators radically changes, and they clearly become inconsistent (see section 3.4).

5 Conclusions

We investigate the effects of dynamic heteroskedasticity on inference procedures in factor analysis. We find that if the variation of conditional moments is explicitly recognised in estimation, identification problems are often alleviated. We also find that the ML estimators of the unconditional variance parameters derived for *i.i.d.* observations remain consistent if the factor loadings are identified from the unconditional distribution, but their asymptotic covariance matrix has to be estimated taking into account the serial correlation in $vech(\mathbf{x}_t\mathbf{x}_t')$. We develop a simple moment test for the presence of ARCH in the common factors, relate it to the standard LM test, and propose more powerful one-sided versions. We also discuss two-step ML estimators of the conditional variance parameters that keep the static variance parameters fixed at some initial consistent estimates, and explain how to compute correct standard errors. Finally, we investigate the finite sample properties of the different estimators and hypothesis tests by simulation methods. Our results suggest that: (i) the relative efficiency of conditional versus unconditional ML estimators of $\mathbf{c}$ and $\boldsymbol{\gamma}$ increases with the variability in conditional variances; (ii) standard errors of the estimates are fairly accurate; (iii) size distortions of the LM test for ARCH are far smaller for Hessian-based versions than for outer-product ones; (iv) the power of this test is an increasing function of α and the signal-to-noise ratio, with one-sided versions being preferred; (v) ARCH and GARCH parameters are estimated as 0 more frequently than they should, especially when the signal-to-noise ratio is small; and (vi) although time-variation in factor variances ensures identification in practice, traditional conditions are more informative, as long as they are correct.

References

Anderson, T.W. and H. Rubin, 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.

Bekker, P.A., 1989, Identification in restricted factor models and the evaluation of rank conditions, Journal of Econometrics 41, 5-16.

Blanchard, O.J. and D. Quah, 1989, The dynamic effects of aggregate demand and supply disturbances, American Economic Review 79, 655-673.

Bollerslev, T. and J. Wooldridge, 1992, Quasi-ML estimation and inference in dynamic models with time-varying variances, Econometric Reviews 11, 143-172.

Davidson, R. and J.G. MacKinnon, 1998, Graphical methods for investigating the size and power of tests statistics, The Manchester School 66, 1-26.

Demos, A. and E. Sentana, 1992, An EM-based algorithm for conditionally heteroskedastic factor models, LSE FMG Discussion Paper 140.

Demos, A. and E. Sentana, 1998a, An EM algorithm for conditionally heteroskedastic factor models, Journal of Business and Economic Statistics 16, 357-361.

Demos, A. and E. Sentana, 1998b, Testing for GARCH effects: a one-sided approach, Journal of Econometrics 86, 97-127.

Dunn, J.E., 1973, A note on a sufficiency condition for uniqueness of a restricted factor matrix, Psychometrika 38, 141-143.

Durbin, J., 1970, Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, Econometrica 38, 410-421.

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

Giraitis, L., P. Kokoszka and R. Leipus, 2000, Stationary ARCH models: dependence structure and central limit theorem, Econometric Theory 16, 3-22.

Harvey, A.C., E. Ruiz, and E. Sentana, 1992, Unobservable component time series models with ARCH disturbances, Journal of Econometrics 52, 129-157.

Jennrich, R.I., 1978, Rotational equivalence of factor loading matrices with specified values, Psychometrika 43, 421-426.

Kano, Y., 1983, Consistency of estimators in factor analysis, Journal of the Japan Statistical Society 13, 137-144.

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

Lawley, D.N. and A.E. Maxwell, 1971, Factor Analysis as a Statistical Method, 2nd edition, Butterworths, London.

Lin, W.L., 1992, Alternative estimators for factor GARCH models: A Monte Carlo comparison, Journal of Applied Econometrics 7, 259-279.

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

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

Rigobon, R., 2000, Identification through heteroskedasticity, mimeo, MIT.

Scherrer, W. and M. Deistler, 1998, A structure theory for linear dynamic errors-in-variables models, SIAM Journal of Control and Optimization 36, 2148-2175.

Self, S.G. and K.Y. Liang, 1987, Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, Journal of the American Statistical Association 82, 605-610.

Sentana, E., 1992, Identification of multivariate conditionally heteroskedastic factor models, LSE FMG Discussion Paper 139.

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.

Wegge, L.E., 1996, Local identifiability of the factor analysis and measurement error model parameter, Journal of Econometrics 70, 351-382.

Appendix

Proof of Proposition 1

Given that by assumption rank$(\mathbf{C}^*) = k^*$, if we regard $\mathbf{C}\vec{\mathbf{\Lambda}}_t\mathbf{C}' = \mathbf{C}^*\vec{\mathbf{\Lambda}}_t^*\mathbf{C}^{*\prime}$ as a system of linear equations in $\vec{\mathbf{\Lambda}}_t^*$, the only solution is $\vec{\mathbf{\Lambda}}_t^* = \mathbf{D}^*\vec{\mathbf{\Lambda}}_t\mathbf{D}^{*\prime}$, with $\mathbf{D}^*= (\mathbf{C}^{*\prime}\mathbf{C}^*)^{-1}\mathbf{C}^{*\prime}\mathbf{C}$, $\mathbf{C}^*\mathbf{D}^*\vec{\mathbf{\Lambda}}_t\mathbf{D}^*\mathbf{C}^{*\prime} = \mathbf{C}\vec{\mathbf{\Lambda}}_t\mathbf{C}'$ (see Magnus and Neudecker (1986), theorem 2.13). But since $\boldsymbol{\delta}'\vec{\boldsymbol{\lambda}}_t = 0\ \forall t$ iff $\boldsymbol{\delta} = \mathbf{0}$, the necessary and sufficient condition above is equivalent to $\mathbf{C}^*\mathbf{D}^* = \mathbf{C}$. Diagonality of $\vec{\mathbf{\Lambda}}_t^*$ then requires $\sum_{l=1}^{k}\vec{\lambda}_{ll,t}d_{il}^*d_{jl}^* = 0$ for $j > i, i = 1, \ldots, k^*$ and $t = 1, \ldots, T$. For a given i, j $(j > i)$, these restrictions can be expressed as $\vec{\mathbf{\Lambda}}_T\mathbf{d}_{ij}^* = \mathbf{0}_T$, where $\vec{\mathbf{\Lambda}}_T = \left(\vec{\boldsymbol{\lambda}}_1, \ldots, \vec{\boldsymbol{\lambda}}_T\right)'$ is a $T \times k$ matrix with typical row $\vec{\boldsymbol{\lambda}}_t'$, $\mathbf{0}_T$ a vector of T zeros and $\mathbf{d}_{ij}^* = \left(d_{i1}^*d_{j1}^*, \ldots, d_{ik}^*d_{jk}^*\right)'$ a $k \times 1$ vector. Given that the rank of $\vec{\mathbf{\Lambda}}_T$ is k when the processes in $\vec{\boldsymbol{\lambda}}_t$ are linearly independent, the only solution to such a set of T homogenous linear equations is $\mathbf{d}_{ij}^* = \mathbf{0}_k$ irrespective of i and j. Therefore, there cannot be two elements in any column of $\mathbf{D}^*$ which are different from 0, which means that each column of $\mathbf{C}$ is proportional to some column of $\mathbf{C}^*$. Given that rank$(\mathbf{C}) = k$, we can find a permutation matrix $\mathbf{P}^*$ of order $k^* \geq k$ such that the first k rows of $\mathbf{P}^*\mathbf{D}^*$ are a full rank diagonal matrix, $\mathbf{D}$ say, its last $k^* - k$ rows are zero, $\mathbf{C}^*\mathbf{P}^{*\prime} = (\mathbf{C}_I^* \mid \mathbf{C}_{II}^*)$, with $\mathbf{C}_I^* = \mathbf{C}\mathbf{D}^{-1}$, and $\mathbf{P}^*\vec{\mathbf{\Lambda}}_t^*\mathbf{P}^{*\prime} = \begin{pmatrix} \vec{\mathbf{\Lambda}}_{It}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, with $\vec{\mathbf{\Lambda}}_{It}^* = \mathbf{D}\vec{\mathbf{\Lambda}}_t\mathbf{D}$. If we now add and subtract $\mathbf{C}^*\underline{\mathbf{\Lambda}}^*\mathbf{C}^{*\prime}$ to $\mathbf{\Sigma}_t$ and group terms, we can equate $\mathbf{\Gamma}^*$ with $\underline{\mathbf{\Sigma}} - \mathbf{C}^*\underline{\mathbf{\Lambda}}^*\mathbf{C}^{*\prime}$, $\mathbf{\Lambda}_{It}^*$ with $\vec{\mathbf{\Lambda}}_{It}^* + \underline{\mathbf{\Lambda}}_I^*$, and $\mathbf{\Lambda}_{IIt}^*$ with $\underline{\mathbf{\Lambda}}_{II}^* \geq 0$ $\forall t$. Finally, the upper limit condition on $\underline{\mathbf{\Lambda}}^*$ follows from Lemma 2a in Sentana (1998), which provides a necessary and sufficient condition for $\mathbf{\Gamma}^*$ to remain p.s.d.

□

Proof of Proposition 2

Since from Proposition 1 all o.e. models satisfy $\mathbf{\Gamma}^* = \mathbf{\Gamma} + \mathbf{C}(\underline{\mathbf{\Lambda}} - \underline{\mathbf{\Lambda}}^*)\mathbf{C}'$ when $k^* = k$ and $\mathbf{D} = \mathbf{I}_k$, diagonality of $\mathbf{\Gamma}^*$ requires $\sum_{l=1}^{k}(\underline{\lambda}_{ll} - \underline{\lambda}_{ll}^*)c_{il}c_{jl} = 0$ for $j > i, i = 1, \ldots, k$. If we express these restrictions as $\bar{\mathbf{C}}vecd(\underline{\mathbf{\Lambda}} - \underline{\mathbf{\Lambda}}^*) = \mathbf{0}$, where $\bar{\mathbf{C}}$ is a $N(N-1)/2 \times k$ matrix with typical column $vecl(\mathbf{c}_l\mathbf{c}_l')$ $(l = 1, \ldots, k)$, the result follows from the fact $vecd(\underline{\mathbf{\Lambda}} - \underline{\mathbf{\Lambda}}^*) = \mathbf{0}$ iff the rank of $\bar{\mathbf{C}}$ is k. □

Proof of Proposition 3

It is well known that if $\mathbf{CC}'$ is identified and $vecd(\mathbf{\Lambda}) = \boldsymbol{\iota}_k$, the only potential o.e. models must satisfy $\mathbf{C}^* = \mathbf{CQ}$, where $\mathbf{Q}$ is an $k \times k$ orthogonal matrix with typical element q_{ij}. Since the covariance matrix of the transformed factors is $\mathbf{\Lambda}_t^* = \mathbf{Q}\mathbf{\Lambda}_t\mathbf{Q}'$, conditional orthogonality requires $\sum_{l=1}^{k}\lambda_{ll,t}q_{il}q_{jl} = 0$ for $j > i, i = 1, \ldots, k$ and $t = 1, \ldots, T$. Applying the same argument as in the proof of Proposition 1, we can show that if $\boldsymbol{\lambda}_t$ is linearly independent, there cannot be two elements in any column of $\mathbf{Q}$ which are different from 0. Given that $\mathbf{Q}$ is orthogonal, the only admissible matrices are permutations of Cholesky square roots of $\mathbf{I}_k$, where $[\mathbf{I}_{kij}^{1/2}] = \pm 1$ for $i = j$ and 0 otherwise. □

Proof of Proposition 4

The only difference with Proposition 3 is that since $\boldsymbol{\lambda}_t' = (\boldsymbol{\lambda}_{1t}', \lambda_{kk,t}\boldsymbol{\iota}_{k_2}')$, the relevant equation system is $\bar{\mathbf{\Lambda}}_T\bar{\mathbf{q}}_{ij} = \mathbf{0}_T$, where $\bar{\mathbf{\Lambda}}_T = \left(\bar{\boldsymbol{\lambda}}_1, \ldots, \bar{\boldsymbol{\lambda}}_T\right)'$ is a $T \times (k_1+1)$ matrix with typical row $\bar{\boldsymbol{\lambda}}_t' = (\lambda_{11,t}, \ldots, \lambda_{k_1k_1,t}, \lambda_{kk,t})$, and $\bar{\mathbf{q}}_{ij} = (q_{i1}q_{j1}, \ldots, q_{ik_1}q_{jk_1}, \sum_{l=k_1+1}^{k} q_{il}q_{jl}')$ a $(k_1 + 1) \times 1$ vector. Given that the rank of $\bar{\mathbf{\Lambda}}_T$ is $k_1 + 1$ by assumption, then for all $j > i, i = 1, \ldots, k$ we must have $q_{il}q_{jl} = 0$ for $l = 1, \ldots, k_1$ and also $\sum_{l=k_1+1}^{k} q_{il}q_{jl} = 0$. The first set of restrictions implies that there cannot be two elements in the first k_1 columns of $\mathbf{Q}$ which are different from 0. If we partition $\mathbf{Q}$ in four blocks conformably, then, given that $\mathbf{Q}$ is orthogonal, if we exclude mere permutations of the factors, it must be the case that $\mathbf{Q}_{11} = \mathbf{I}_{k_1}^{1/2}$, $\mathbf{Q}_{21} = \mathbf{0}$, $\mathbf{Q}_{12} = \mathbf{0}$ and $\mathbf{Q}_{22}$ is orthogonal. $\square$

The score and information matrix of a conditionally heteroskedastic factor model

Bollerslev and Wooldridge (1992) show that the score $\mathbf{s}_t(\boldsymbol{\phi}) = \partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ of any conditionally heteroskedastic multivariate model with zero conditional mean is given by $\frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t\right]/\partial\boldsymbol{\phi}\left[\boldsymbol{\Sigma}_t^{-1}\otimes\boldsymbol{\Sigma}_t^{-1}\right] vec\left[\mathbf{x}_t\mathbf{x}_t' - \boldsymbol{\Sigma}_t\right]$, which in our case yields

$$\begin{bmatrix} vec\left[\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C}\boldsymbol{\Lambda}_t - \boldsymbol{\Sigma}_t^{-1}\mathbf{C}\boldsymbol{\Lambda}_t\right] \\ \frac{1}{2}vecd\left[\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}\right] \\ \mathbf{0} \end{bmatrix} + \frac{1}{2}\frac{\partial\boldsymbol{\lambda}_t'(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}} vecd\left[\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{C}\right]$$

Assuming that rank$(\boldsymbol{\Gamma}) = N$, we can use the Woodbury formula to prove that

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

$$\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Gamma}^{-1}\left[(\mathbf{x}_t - \mathbf{C}\mathbf{f}_{t|t})(\mathbf{x}_t - \mathbf{C}\mathbf{f}_{t|t})' + \mathbf{C}\boldsymbol{\Omega}_{t|t}\mathbf{C}' - \boldsymbol{\Gamma}\right]\boldsymbol{\Gamma}^{-1}$$

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

As a simple yet important example, consider the ARCH(1)-type conditional variance specification $\lambda_{jj,t} = (1-\alpha_{j1}) + \alpha_{j1}(f_{jt-1|t-1}^2 + \omega_{jj,t-1|t-1})$, in which $\boldsymbol{\psi}' = (\alpha_{11}, \alpha_{21}, ..., \alpha_{k1})$ (see Harvey et al. (1992)). If $\boldsymbol{\psi} = \mathbf{0}$, so that $\boldsymbol{\Lambda}_t(\mathbf{c},\boldsymbol{\gamma},\mathbf{0}) = \mathbf{I}\ \forall t$, we obtain $\partial\boldsymbol{\lambda}_t'(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})/\partial\mathbf{c} = \mathbf{0}$, $\partial\boldsymbol{\lambda}_t'(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})/\partial\boldsymbol{\gamma} = \mathbf{0}$ and $\partial\boldsymbol{\lambda}_t'(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})/\partial\boldsymbol{\psi} = dg\left[\mathbf{f}_{t-1|t-1}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0})\mathbf{f}_{t-1|t-1}'(\mathbf{c},\boldsymbol{\gamma},\mathbf{0}) + \boldsymbol{\Omega}_{t-1|t-1}(\mathbf{c},\boldsymbol{\gamma},\mathbf{0}) - \mathbf{I}\right]$.

Therefore, since $E\left(\mathbf{C}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0 - \mathbf{C}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0\right) = 0$, the orthogonality conditions implicit in the last k elements of the score are simply $cov[f_{jt|t}^2(\mathbf{c}_0,\boldsymbol{\gamma}_0,\mathbf{0}), f_{jt-1|t-1}^2(\mathbf{c}_0,\boldsymbol{\gamma}_0,\mathbf{0})] = 0$. Moreover, since $\mathbf{f}_{t|t}(vec(\mathbf{C}_0\mathbf{Q}_0),\boldsymbol{\gamma}_0,\mathbf{0}) = \mathbf{Q}_0'\mathbf{f}_{t|t}(\mathbf{c}_0,\boldsymbol{\gamma}_0,\mathbf{0})$, the same interpretation holds if we replace $\mathbf{C}_0$ by $\mathbf{C}_0\mathbf{Q}_0$.

Bollerslev and Wooldridge (1992) also prove that $\mathbf{H}_t(\boldsymbol{\phi}) = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ satisfies $E\left[\mathbf{H}_t(\boldsymbol{\phi}_0)|\mathbf{X}_{t-1}\right] = -\frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t\right]/\partial\boldsymbol{\phi}(\boldsymbol{\Sigma}_t^{-1}\otimes\boldsymbol{\Sigma}_t^{-1})\partial vec\left[\boldsymbol{\Sigma}_t\right]/\partial\boldsymbol{\phi}'$. When

$\boldsymbol{\psi}_0 = \mathbf{0}$,

$$-E\left[\frac{\partial^2 l_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\psi} \partial \mathbf{c}'}|\mathbf{X}_{t-1}\right] = \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\psi}} \mathbf{E}_k'(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C} \otimes \mathbf{C}'\boldsymbol{\Sigma}^{-1})$$

$$-E\left[\frac{\partial^2 l_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\gamma}'}|\mathbf{X}_{t-1}\right] = \frac{1}{2}\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\psi}}(\mathbf{C}'\boldsymbol{\Sigma}^{-1} \odot \mathbf{C}'\boldsymbol{\Sigma}^{-1})$$

Therefore, if we apply the law of iterated expectations, it is clear that the information matrix is block diagonal between static and dynamic variance parameters under the null of conditional homoskedasticity. Alternatively, this result can be derived directly from the form of the score by noticing that when $\boldsymbol{\psi}_0 = \mathbf{0}$, $vech(\mathbf{x}_t\mathbf{x}_t' - \boldsymbol{\Sigma}_0)$ is serially independent over time. Therefore, it remains valid if we replace $\mathbf{C}_0$ by $\mathbf{C}_0\mathbf{Q}_0$. Finally, it is also worth noting that under conditional homoskedasticity

$$-E\left[\frac{\partial^2 l_t(\boldsymbol{\phi}_0)}{\partial \mathbf{c} \partial \mathbf{c}'}|\mathbf{X}_{t-1}\right] = (\mathbf{C}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) + (\mathbf{C}_0'\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0)\mathbf{K}_{Nk}$$

$$-E\left[\frac{\partial^2 l_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\gamma} \partial \mathbf{c}'}|\mathbf{X}_{t-1}\right] = \mathbf{E}_N'(\boldsymbol{\Sigma}_0^{-1}\mathbf{C}_0 \otimes \boldsymbol{\Sigma}_0^{-1})$$

$$-E\left[\frac{\partial^2 l_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}|\mathbf{X}_{t-1}\right] = \frac{1}{2}(\boldsymbol{\Sigma}_0^{-1} \odot \boldsymbol{\Sigma}_0^{-1})$$

where $\mathbf{K}_{Nk}$ is the commutation matrix of orders N and k (see Magnus and Neudecker (1988)).

Table 1: One Factor Model
Mean biases and standard deviations
for unconditional variance parameters

		γ_0=0.5				γ_0=2.0			
		c		γ		c		γ	
		CML	UML	CML	UML	CML	UML	CML	UML
α_0=0.0	bias	-.0006	-.0006	-.0036	-.0036	-.0054	-.0051	-.0297	-.0287
β_0=0.0	std.dev.	.0265	.0265	.0729	.0730	.0789	.0771	.3093	.3025
α_0=0.2	bias	-.0006	-.0006	-.0036	-.0036	-.0055	-.0054	-.0290	-.0292
β_0=0.6	std.dev.	.0269	.0270	.0720	.0729	.0795	.0786	.3045	.3034
α_0=0.4	bias	-.0006	-.0006	-.0035	-.0037	-.0055	-.0058	-.0282	-.0300
β_0=0.4	std.dev.	.0277	.0282	.0700	.0729	.0795	.0818	.2913	.3047

$c = (c_1 + c_2 + c_3)/3 \qquad \gamma = (\gamma_1 + \gamma_2 + \gamma_3)/3$

Table 2: One Factor Model

Proportion of estimates at the boundary of the parameter space

	γ_0=0.5				γ_0=2.0			
	$\hat{\alpha}=0, \hat{\beta}=0$		$\hat{\alpha}\neq 0, \hat{\beta}=0$		$\hat{\alpha}=0, \hat{\beta}=0$		$\hat{\alpha}\neq 0, \hat{\beta}=0$	
	CML	2S	CML	2S	CML	2S	CML	2S
α_0=0.0,β_0=0.0	.556	.557	.265	.264	.552	.552	.286	.282
α_0=0.2,β_0=0.6	.022	.027	.091	086	.118	.137	.198	.167
α_0=0.4,β_0=0.4	.003	.005	.074	.070	.049	.059	.218	.185

Table 3: One Factor Model
Mean biases and standard deviations
for conditional variance parameters

		γ_0=0.5				γ_0=2.0			
		α		β		α		β	
		CML	2S	CML	2S	CML	2S	CML	2S
α_0=0.2	bias	.007	-.002	-.106	-.103	.019	-.007	-.183	-.162
β_0=0.6	std.dev.	.112	.104	.253	.250	.172	.149	302	.299
α_0=0.4	bias	-.004	-.030	-.043	-.039	-.015	-.065	-.081	-.058
β_0=0.4	std.dev.	.151	.134	.196	.195	.222	.190	.257	.257

Table 4: Two Factor Model
Mean biases and standard deviations
for unconditional variance parameters

$\mathbf{c}_0 = (0,0,0,1,1,1;1,1,1,0,0,0)'$

		$\alpha_0=0.2$ $\beta_0=0.6$						$\alpha_0=0.4$ $\beta_0=0.4$					
		$\gamma_0=0.5$			$\gamma_0=2.0$			$\gamma_0=0.5$			$\gamma_0=2.0$		
		CML	RML	UML	CML	RML	UML	CML	RML	UML	CML	RML	UML
c_{a1}	bias	.0014	-.0013	-.0012	.0215	-.0001	-.0004	.0026	-.0014	-.0014	.0136	-.0003	-.0004
	s.d.	.1349	.0554	.0556	.1912	.1006	.1003	.1018	.0570	.0577	.1603	.1005	.1034
c_{b1}	bias	-.0120	-.0033	-.0034	-.0504	-.0147	-.0149	-.0011	-.0035	-.0037	-.0357	-.0145	-.0159
	s.d.	.0600	.0279	.0282	.1365	.0814	.0829	.0578	.0286	.0295	.1187	.0803	.0858
c_{a2}	bias	-.0178	-.0016	-.0016	-.0549	-.0117	-.0117	-.0069	-.0015	-.0016	-.0347	-.0113	-.0117
	s.d.	.0611	.0269	.0269	.1513	.0810	.0809	.0306	.0271	.0271	.1208	.0808	.0807
c_{b2}	bias	.0033	-.0005	-.0006	.0098	-.0026	-.0020	-.0004	-.0003	-.0005	.0036	-.0018	-.0012
	s.d.	.1285	.0405	.0407	.1934	.1010	.1011	.0835	.0396	.0408	.1556	.0974	.0974
γ	bias	-.0058	-.0057	-.0058	-.0437	-.0428	-.0441	-.0059	-.0058	-.0059	-.0429	-.0417	-.0444
	s.d.	.0724	.0723	.0730	.3141	.3100	.3136	.0712	.0712	.0730	.3048	.3018	.3142

$\mathbf{c}_0 = (\frac{1}{4},\frac{1}{4},\frac{1}{4},1,1,1;1,1,1,\frac{1}{4},\frac{1}{4},\frac{1}{4})'$

		$\alpha_0=0.2$ $\beta_0=0.6$						$\alpha_0=0.4$ $\beta_0=0.4$					
		$\gamma_0=0.5$			$\gamma_0=2.0$			$\gamma_0=0.5$			$\gamma_0=2.0$		
		CML	RML	UML	CML	RML	UML	CML	RML	UML	CML	RML	UML
c_{a1}	bias	-.0121	.0955	.1040	-.0038	.0920	.0987	-.0072	.0955	.1076	-.0045	.0841	.1007
	s.d.	.1296	.0424	.0417	.1820	.0815	.0809	.0939	.0473	.0455	.1526	.0841	.0838
c_{b1}	bias	-.0158	-.0373	-.0394	-.0457	-.0437	-.0468	-.0079	-.0360	-.0415	-.0305	-.0406	-.0486
	s.d.	.0622	.0296	.0297	.1305	.0787	.0785	.0439	.0315	.0317	.1057	.0778	.0813
c_{a2}	bias	-.0151	.0151	.0150	-.0562	-.0021	-.0035	-.0048	.0149	.0150	-.0337	-.0009	-.0038
	s.d.	.0611	.0309	.0312	.1655	.0997	.1012	.0324	.0306	.0313	.1326	.0977	.1018
c_{b2}	bias	-.0142	-.1339	-.1418	-.0164	-.1433	-.1561	-.0075	-.1210	-.1417	-.0210	-.1260	-.1566
	s.d.	.1293	.0480	.0485	.1916	.1344	.1399	.0796	.0473	.0488	.1564	.1283	.1413
γ	bias	-.0060	-.0063	-.0060	-.0517	-.0519	-.0524	-.0060	-.0067	-.0061	-.0505	-.0525	-.0537
	s.d.	.0725	.0726	.0732	.3269	.3267	.3210	.0711	.0715	.0732	.3117	.3266	.3114

$c_{a1} = (c_{11}+c_{21}+c_{31})/3$, $c_{b1} = (c_{41}+c_{51}+c_{61})/3$, $c_{a2} = (c_{12}+c_{22}+c_{32})/3$,
$c_{b2} = (c_{42}+c_{52})/2$, $\gamma = \frac{1}{6}\sum_{i=1}^{6}\gamma_i$

Table 5: Two Factor Model

Proportion of estimates at the boundary of the parameter space

$\mathbf{c}_0 = (0, 0, 0, 1, 1, 1; 1, 1, 1, 0, 0, 0)'$

	γ_0=0.5						γ_0=2.0					
	$\hat{\alpha} = 0, \hat{\beta} = 0$			$\hat{\alpha} \neq 0, \hat{\beta} = 0$			$\hat{\alpha} = 0, \hat{\beta} = 0$			$\hat{\alpha} \neq 0, \hat{\beta} = 0$		
	CML	RML	2S	CML	RML	2S	CML	RML	2S	CML	RML	2S
α_0=0.2,β_0=0.6	.034	.034	.038	.114	.088	.084	.146	.145	.153	.226	.190	.166
α_0=0.4,β_0=0.4	.004	.004	.004	.097	.077	.072	.064	.064	.072	.260	.222	.188

$\mathbf{c}_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, 1, 1; 1, 1, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})'$

	γ_0=0.5						γ_0=2.0					
	$\hat{\alpha} = 0, \hat{\beta} = 0$			$\hat{\alpha} \neq 0, \hat{\beta} = 0$			$\hat{\alpha} = 0, \hat{\beta} = 0$			$\hat{\alpha} \neq 0, \hat{\beta} = 0$		
	CML	RML	2S	CML	RML	2S	CML	RML	2S	CML	RML	2S
α_0=0.2,β_0=0.6	.034	.048	.054	.095	.124	.117	.156	.167	.195	.201	.225	.183
α_0=0.4,β_0=0.4	.004	.011	.015	.095	.099	.093	.062	.095	.109	.297	.227	.186

Table 6: Two Factor Model
Mean biases and standard deviations
for conditional variance parameters

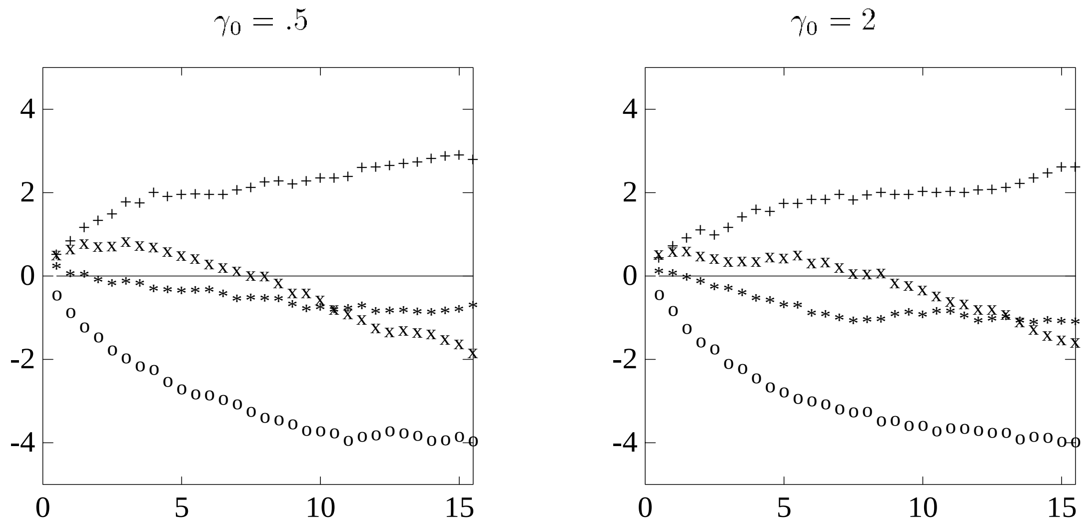
$\mathbf{c}_0 = (0, 0, 0, 1, 1, 1; 1, 1, 1, 0, 0, 0)'$

		γ_0=0.5						γ_0=2.0					
		α			β			α			β		
		CML	RML	2S	CML	RML	2S	CML	RML	2S	CML	RML	2S
α_0=0.2	bias	.025	.007	-.003	-.128	-.109	-.106	.062	.025	-.014	-.219	-.192	-.166
β_0=0.6	std.dev.	.115	.112	.103	.259	.248	.246	.192	.181	.146	.305	.301	.301
α_0=0.4	bias	.010	-.003	-.032	-.057	-.047	-.041	.017	-.012	-.081	-.108	-.089	-.055
β_0=0.4	std.dev.	.150	.150	.131	.197	.193	.193	.224	.226	.186	.258	.256	.258

$\mathbf{c}_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, 1, 1; 1, 1, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})'$

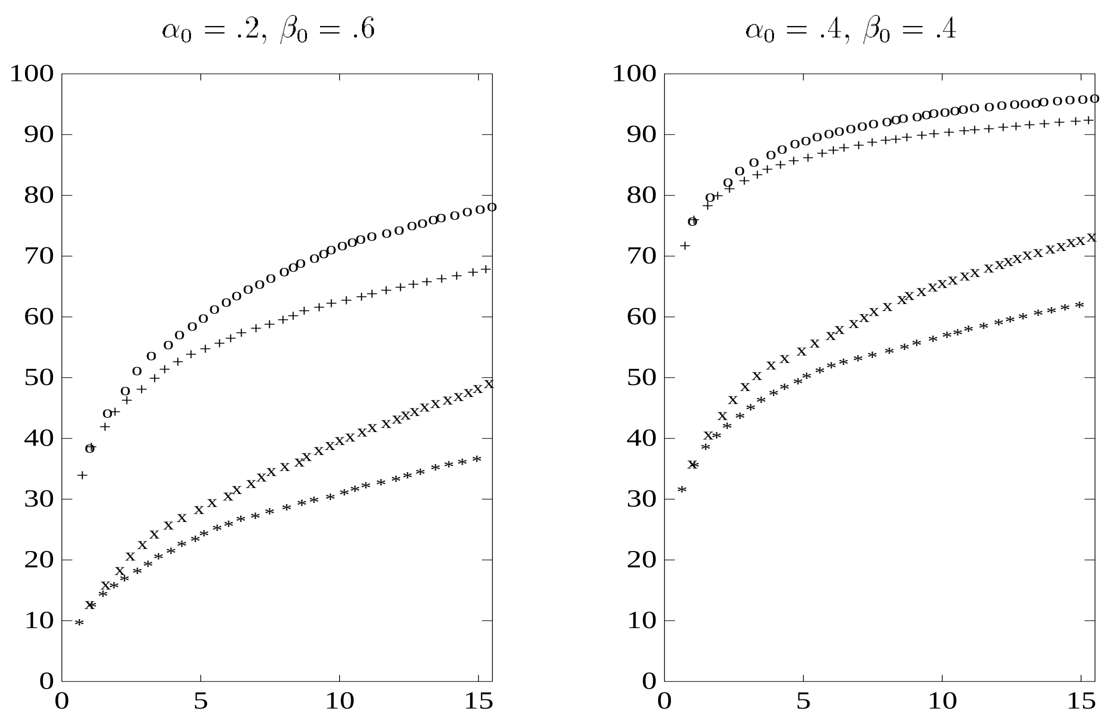
		γ_0=0.5						γ_0=2.0					
		α			β			α			β		
		CML	RML	2S	CML	RML	2S	CML	RML	2S	CML	RML	2S
α_0=0.2	bias	.027	-.021	-.030	-.116	-.120	-.115	.064	-.013	-.047	-.201	-.207	-.175
β_0=0.6	std.dev.	.120	.108	.100	.256	.273	.269	.208	.164	.134	.306	.312	.311
α_0=0.4	bias	.012	-.061	-.088	-.055	-.027	-.019	.017	-.079	-.144	-.095	-.074	-.036
β_0=0.4	std.dev.	.156	.150	.133	.202	.214	.215	.234	.217	.176	.266	.271	.275

Figure 1: Test for ARCH in common factor
P-value discrepancy plots



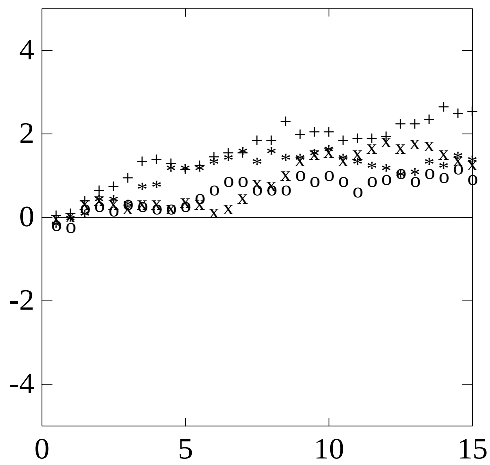
OP 2-sided: $+$ HESS 2-sided: $\star$ OP 1-sided: o HESS 1-sided: $\times$

Figure 2: Test for ARCH in common factor
Size–Power curves



$\gamma_0 = .5$, 2-sided: $+$ $\gamma_0 = .5$, 1-sided: o $\gamma_0 = 2$, 2-sided: $\star$ $\gamma_0 = 2$, 1-sided: $\times$

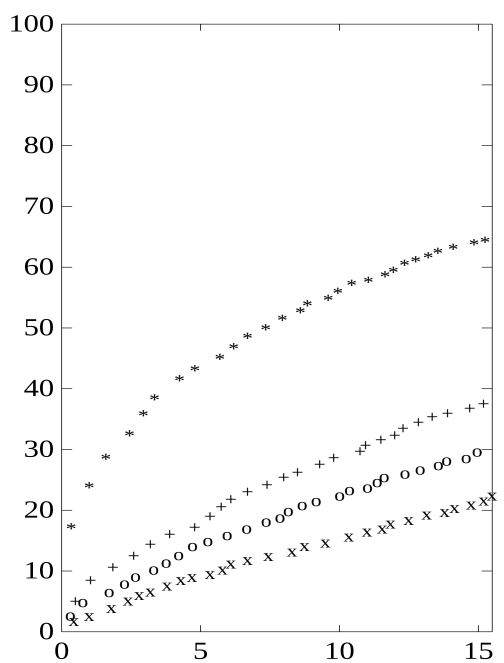
Figure 3: Likelihood Ratio Test for overidentifying restriction
P-value discrepancy plots



$\alpha_0 = .2,\ \beta_0 = .6,\ \gamma_0 = 2$: $\times$ $\alpha_0 = .4,\ \beta_0 = .4,\ \gamma_0 = 2$: o

$\alpha_0 = .2,\ \beta_0 = .6,\ \gamma_0 = .5$: $+$ $\alpha_0 = .4,\ \beta_0 = .4,\ \gamma_0 = .5$: $\star$

Figure 4: Likelihood Ratio Test for overidentifying restriction
Size–Power curves



$\alpha_0 = .2,\ \beta_0 = .6,\ \gamma_0 = 2$: $\times$ $\alpha_0 = .4,\ \beta_0 = .4,\ \gamma_0 = 2$: o

$\alpha_0 = .2,\ \beta_0 = .6,\ \gamma_0 = .5$: $+$ $\alpha_0 = .4,\ \beta_0 = .4,\ \gamma_0 = .5$: $\star$