Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models

Enrique Sentana
CEMFI

Gabriele Fiorentini
University of Alicante

Working Paper No. 9709
July 1997

This paper is an extensively revised version of Sentana (1992), which concentrated on identification issues. Financial support for that version from the Spanish Ministry of Education and Science and the LSE Financial Markets Group as part of the ESRC project "The Efficiency and Regulation of Financial Markets" is gratefully acknowledged. The authors are also grateful to Manuel Arellano, Antonis Demos, Lars Hansen, Andrew Harvey, Mervyn King, Jan Magnus, Danny Quah, Esther Ruiz, Mushtaq Shah, Neil Shephard and Sushil Wadhwani, as well as seminar participants at CEMFI, the European Meeting of the Econometric Society (Brussels, 1992) and the ESRC Econometric Study Group Conference (Bristol, July 1997) for very helpful comments and suggestions. Two anonymous referees have also helped us greatly improve the paper. Of course the usual caveat applies. (E-mail addresses: gafi@merlin.fae.ua.es, sentana@cemfi.es).

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain.
Tel: 341 4290551, fax: 341 4291056, http://www.cemfi.es.

Abstract

We investigate several important inference issues for factor models with dynamic heteroskedasticity in the common factors. First, we show that such models are identified if we take into account the time-variation in the variances of the factors. Our results also apply to dynamic versions of the APT, dynamic factor models, and vector autoregressions. Secondly, we propose a consistent two-step estimation procedure which does not rely on knowledge of any factor estimates, and explain how to compute correct standard errors. Thirdly, we develop a simple preliminary LM test for the presence of ARCH effects in the common factors. Finally, we conduct a Monte Carlo analysis of the finite sample properties of the proposed estimators and hypothesis tests.

1 Introduction

In recent years, increasing attention has been paid to modelling the observed changes in the volatility of many economic and financial time series. By and large, though, most theoretical and applied research in this area has concentrated on univariate series. However, many issues in finance, such as tests of asset pricing restrictions, asset allocation, performance evaluation or risk management, can only be fully addressed within a multivariate framework. Unfortunately, the application of dynamic heteroskedasticity in a multivariate context has been hampered by the sheer number of parameters involved.

Given that there are many similarities between this problem and that of modelling the unconditional covariance matrix of a large number of asset returns, it is perhaps not surprising that one of the most popular approaches to multivariate dynamic heteroskedasticity is based on the same idea as traditional factor analysis. That is, in order to obtain a parsimonious representation of conditional second moments, it is assumed that each of several observed variables is a linear combination of a smaller number of common factors plus an idiosyncratic noise term, but allowing for dynamic heteroskedasticity-type effects in the underlying factors. The factor GARCH model of Engle (1987) and the conditionally heteroskedastic latent factor model introduced by Diebold and Nerlove (1989) and extended by King, Sentana and Wadhwani (1994) are the best known examples. Such models also have the advantage of being compatible with standard factor analysis based on unconditional covariance matrices. Furthermore, they are particularly appealing in finance, where there is a long tradition of factor or multi-index models (see e.g. the Arbitrage Pricing Theory of Ross (1976)).

Although many properties of these models have already been studied in detail, either for the general class or for some of its members (see e.g. Bollerslev and Engle (1993), Engle, Ng and Rothschild (1990), Gourieroux, Monfort and Renault

(1991), Harvey, Ruiz and Sentana (1992), Kroner (1987), Lin (1992), or Nijman and Sentana (1996)), some very important inference issues have not been fully investigated yet. The purpose of the paper is to address four such remaining issues

The first issue is in what sense, if any, the identification problems of traditional factor models are altered by the presence of dynamic heteroskedasticity in the factors. This has important implications for empirical work related to the Arbitrage Pricing Theory (APT), as in static factor models individual risk premia components are only identifiable up to an orthogonal transformation. Furthermore, it also has some bearing upon the interpretation of common trend and dynamic factor models, and on the identification of fundamental disturbances and their dynamic impact in vector autoregressions.

Another important aspect is the development of alternative estimation methods. Traditionally, the preferred method of estimation for such models has been full information maximum likelihood. Unfortunately, this involves a very time consuming procedure, which is disproportionately more so as the number of series considered increases. Although using the EM algorithm combined with derivative based methods significantly reduces the computational burden (see Demos and Sentana (1996b)), it would be interesting to have simpler estimation procedures, which are nevertheless based on firm statistical grounds.

It is also of some interest to have a simple preliminary test for the presence of ARCH effects in the common factors. Moreover, since the way in which standard errors are usually computed in static factor models is only valid under conditional homoskedasticity, it is convenient to have a model diagnostic to assess the validity of such a maintained assumption.

Finally, given that the justification of such estimators and hypothesis tests is asymptotic in nature, it is useful to investigate their finite sample properties by means of simulation methods.

The rest of the paper is organized as follows. We formally introduce the model in section 2, and relate it to the most common conditional variance parametrisations. Identification issues are discussed in detail in section 3. Then, in section 4.1, we propose a simple two-step consistent estimator. We also derive an LM test for ARCH in the common factors in section 4.2. Finally, we carry out a Monte Carlo analysis in section 5. Proofs and auxiliary results are gathered in appendices.

2 Conditionally Heteroskedastic Factor Models

Consider the following multivariate model:

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

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

where $\mathbf{x}_t$ is a $N \times 1$ vector of observable random variables, $\mathbf{f}_t$ is a $k \times 1$ vector of unobserved common factors, $\mathbf{C}$ is the $N \times k$ matrix of factor loadings, with $N \geq k$ and $\operatorname{rank}(\mathbf{C}) = k$, $\mathbf{w}_t$ is a $N \times 1$ vector of idiosyncratic noises, which are conditionally orthogonal to $\mathbf{f}_t$, $\mathbf{\Gamma}$ is a $N \times N$ positive semidefinite (p.s.d.) matrix of constant idiosyncratic variances, $\mathbf{\Lambda}_t$ is a $k \times k$ diagonal positive definite (p.d.) matrix of (possibly) time-varying factor variances, which generally involve some extra parameters, $\boldsymbol{\psi}$, and $\mathbf{X}_{t-1}$ is an information set that contains the values of $\mathbf{x}_t$ up to, and including 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 shall refer to the data generation process specified by (1-2) as a multivariate conditionally heteroskedastic factor model. Note that the diagonality of $\mathbf{\Lambda}_t$ implies that the factors are conditionally orthogonal.

Such a formulation nests several models widely used in the empirical literature. In particular, it nests the conditionally heteroskedastic latent factor model introduced by Diebold and Nerlove (1989) and extended by King, Sentana and Wadhwani (1994), and the factor ARCH model of Engle (1987). These models typically assume that the unobserved factors follow univariate dynamic heteroskedastic processes, but differ in the exact parametrisation of $\mathbf{\Lambda}_t$ and $\mathbf{\Gamma}$.

For instance, in the conditionally heteroskedastic latent factor model, the idiosyncratic covariance matrix is assumed diagonal, and the variances of the factors are parametrised as univariate ARCH models, but taking into account that the values of the factors are unobserved. In particular, for the GQARCH(1,1) formulation of Sentana (1995),

$$\lambda_{jj,t} = \varphi_{j0} + \varphi_{j1} f_{jt-1|t-1} + \alpha_{j1}(f^2_{jt-1|t-1} + \lambda_{jj,t-1|t-1}) + \beta_{j1}\lambda_{jj,t-1} \tag{3}$$

where $\mathbf{f}_{t|t} = E(\mathbf{f}_t|\mathbf{X}_t)$ and $\mathbf{\Lambda}_{t|t} = V(\mathbf{f}_t|\mathbf{X}_t)$, which can be easily evaluated via the Kalman filter (see Harvey, Ruiz and Sentana (1992)). Note that the measurability of $\lambda_{jj,t}$ with respect to $\mathbf{X}_{t-1}$ is achieved in this model by replacing the unobserved factors by their best (in the conditional mean square error sense) estimates, and including a correction in the standard ARCH terms which reflects the uncertainty in the factor estimates.

Similarly, the factor GARCH(p,q) model can also be written as a particular case of (1-2), with $\mathbf{\Gamma}$ non-diagonal, and the conditional variances of the factors given by:

$$\lambda_{jj,t} = \sum_{s=1}^{q} \alpha_{js}\dot{x}^2_{jt-s} + \sum_{r=1}^{p} \beta_{jr}\lambda_{jj,t-r} \tag{4}$$

where $\dot{\mathbf{x}}_t = \mathbf{D}'\mathbf{x}_t$ and $\mathbf{D} = (\mathbf{d}_1|\dots|\mathbf{d}_k)$ is a $N \times k$ matrix of full column rank satisfying $\mathbf{D}'\mathbf{C} = \mathbf{I}_k$ (see Sentana (1997a)). Note that the measurability of $\lambda_{jj,t}$ with respect to $\mathbf{X}_{t-1}$ is achieved here by making the time-variation in second moments a function of k linear combinations of $\mathbf{x}_t$.

Finally, if $\mathbf{f}_t$ is conditionally homoskedastic, which usually corresponds to $\boldsymbol{\psi} = \mathbf{0}$, (1-2) reduces to the static orthogonal factor model (see e.g. Johnson and Wichern (1992)). But even if $\mathbf{f}_t$ is conditionally heteroskedastic, provided that it is covariance stationary, the assumption of constant factor loadings implies an unconditionally orthogonal k factor structure for $\mathbf{x}_t$. That is, the unconditional covariance matrix of $\mathbf{x}_t$, $\boldsymbol{\Sigma} = E(\boldsymbol{\Sigma}_t)$, can be written as:

$$\boldsymbol{\Sigma} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' + \boldsymbol{\Gamma} \tag{5}$$

where $\boldsymbol{\Lambda} = V(\mathbf{f}_t) = E(\boldsymbol{\Lambda}_t)$. This property makes the model considered here compatible with traditional factor analysis.

3 The Effects of Modelling Conditional Heteroskedasticity on Identification

3.1 Identification of Idiosyncratic Factors

The most distinctive feature of factor models is that they provide a parsimonious specification of the (dynamic) cross-sectional dependence of a vector of observable random variables. In our case, the factor structure allows us to decompose the conditional covariance matrix $\boldsymbol{\Sigma}_t$ into two parts: one which is common but of reduced rank k, $\boldsymbol{\Sigma}_{ct} = \mathbf{C}\boldsymbol{\Lambda}_t\mathbf{C}'$, and one which is specific, $\boldsymbol{\Sigma}_{st} = \boldsymbol{\Gamma}$. Unfortunately, without further restrictions on $\boldsymbol{\Gamma}$, or on the constant part of $\boldsymbol{\Lambda}_t$, we cannot separately identify one from the other. The reason is twofold. On the one hand, we are not able to differentiate the contribution to the conditional variance of conditionally homoskedastic common factors (see Engle, Ng and Rothschild (1990)). On the other, we may be able to transfer unconditional variance from the idiosyncratic terms to the common factors. For instance, if $\boldsymbol{\Gamma}$ is non-singular, we can take $\boldsymbol{\Sigma}_{ct}(-) = \mathbf{C}(\boldsymbol{\Lambda}_t + -)\mathbf{C}'$, and $\boldsymbol{\Sigma}_{st}(-) = \boldsymbol{\Gamma} - \mathbf{C} - \mathbf{C}'$, where - is any

$k \times k$ p.s.d. diagonal matrix such that the eigenvalues of - $\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C}$ are less than or equal to 1 (see Sentana (1997a)).

The most common assumption made to differentiate common from idiosyncratic effects is that $\mathbf{\Gamma}$ is diagonal (see e.g. Diebold and Nerlove (1989) or King, Sentana and Wadhwani (1994)). In this case, we say that the conditional factor structure is exact. However, in some applications, diagonality of $\mathbf{\Gamma}$ may be thought to be too restrictive. For that reason, Chamberlain and Rothschild (1983) introduced the concept of approximate factor structures, in which the idiosyncratic terms may be mildly correlated. Their definition is asymptotic in N, and amounts to the largest eigenvalue of $V\left[(w_{1t}, w_{2t}, \ldots, w_{Nt})'\right]$ remaining bounded as N increases (as in band-diagonal matrices).[1] In practice, the eigenvalues of $\mathbf{\Gamma}$ are always bounded as N is finite, and it is difficult to come up with realistic models that ensure such an asymptotic restriction.

An alternative way to differentiate common from idiosyncratic effects is to assume that $\mathbf{\Gamma}$ has reduced rank.[2] In some cases, in fact, it may be necessary to assume that $\mathbf{\Gamma}$ is both diagonal and of reduced rank. As a trivial example, consider an exact conditionally homoskedastic single factor model with $N = 2$ and $\lambda_{11} = 1$. Its covariance matrix can be written as

$$\begin{pmatrix} c_{11}^{*2} + \gamma_{11}^{*} & c_{11}^{*}c_{21}^{*} \\ & c_{21}^{*2} + \gamma_{22}^{*} \end{pmatrix}$$

with $c_{11}^{*} = \sqrt{c_{11}^{2} + \gamma_{11} - \gamma_{11}^{*}}, c_{21}^{*} = c_{21}c_{11}/c_{11}^{*}$ and $\gamma_{22}^{*} = \gamma_{22} + c_{21}^{2}\left[1 - (c_{11}/c_{11}^{*})^{2}\right]$ for any $\gamma_{11}^{*} \in [0, \gamma_{11} + c_{11}^{2}\gamma_{22}/(c_{21}^{2} + \gamma_{22})]$. Note that the extreme values of this range correspond to the two possible Heywood (i.e. singular) cases.

[1] This suggests an intuitive interpretation by analogy with univariate time series: if y_t is a covariance stationary and ergodic process (e.g. an MA model), then all the eigenvalues of the intertemporal covariance matrix $V\left[(y_1, y_2, \ldots y_T)'\right]$ remain bounded as $T \to \infty$. Unlike in a time series framework, though, there is generally no natural ordering for the variables in $\mathbf{x}_t$.

[2] The rank of $\mathbf{\Gamma}$ is related to the observability of the factors. If rank$(\mathbf{\Gamma}) = N - k$ the factors would be fully revealed by the $\mathbf{x}_t$ variables; otherwise they are only partially revealed (see King, Sentana and Wadhwani (1994))

3.2 Identification of Common Factors

But the most fundamental identification issue in factor models relates to the decomposition of $\mathbf{\Sigma}_{ct}$ into $\mathbf{C}$ and $\mathbf{\Lambda}_t$. Since the scaling of the factors is usually irrelevant, then in the case of constant variances, it is conventional to impose the assumption that the variance of each factor is unity, that is $\mathbf{\Lambda}_t = \mathbf{I}$, $\forall t$. By analogy, we may impose here the same scaling assumption on the factors unconditional variances.[3]

Suppose that we were to ignore the time-variation in the conditional variances and base our estimation in the unconditional covariance matrix of $\mathbf{x}_t$ in (5). As is well known from standard factor analysis theory, it would then be possible to generate an observationally equivalent (o.e.) model up to unconditional second moments as $\mathbf{x}_t = \mathbf{C}^*\mathbf{f}_t^* + \mathbf{w}_t$, where $\mathbf{C}^* = \mathbf{C}\mathbf{Q}'$, $\mathbf{f}_t^* = \mathbf{Q}\mathbf{f}_t$, and $\mathbf{Q}$ is an arbitrary orthogonal $k \times k$ matrix, since the unconditional covariance matrix, $\mathbf{\Sigma} = \mathbf{C}^*\mathbf{C}^{*\prime} + \mathbf{\Gamma} = \mathbf{C}\mathbf{C}' + \mathbf{\Gamma}$, remains unchanged.

Hence, some restrictions would be needed on $\mathbf{C}$. One way to impose them would be to use Dunn's (1973) set of sufficiency identification conditions for the homoskedastic factor model with orthogonal factors. These conditions are zero-type restrictions that guarantee that $\mathbf{C}$ is locally identifiable up to column sign changes. For instance, when $\mathbf{C}$ is otherwise unrestricted, imposing $c_{ij} = 0$ for $j > i, i = 1, 2, \ldots k$ (i.e. $\mathbf{C}$ lower trapezoidal) ensures identification.[4] Although such restrictions are often arbitrary, the factors can be orthogonally rotated to simplify their interpretation once the model has been estimated. In some other

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

[4] Other alternative sets of sufficient local identifiability restrictions have been suggested. For example, Jennrich (1978) proves that when $\mathbf{C}$ is otherwise unrestricted, fixing not necessarily to zero the $k(k-1)/2$ supra-diagonal coefficients of (a permutation of) $\mathbf{C}$ also guarantees identifiability. From a computational point of view, though, the most convenient uniqueness condition in the unrestricted case is $\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C}$ diagonal (see e.g. Johnson and Wichern (1992)).

cases, identifiability can be achieved by imposing plausible a priori restrictions. For example, if in a two factor model it is believed that the second factor only affects a subset of the variables (say the first N_1, with $N_1 < N$, so that $c_{i2} = 0$ for $i = N_1 + 1, \ldots, N$) the non-zero elements of $\mathbf{C}$ will always be identifiable.

However, when time variation in $\mathbf{\Lambda}_t$ is explicitly recognized in estimation, the set of admissible $\mathbf{Q}$ matrices is substantially reduced, as the conditional covariance matrix of the transformed factors $\mathbf{f}_t^* = \mathbf{Qf}_t$ has to remain diagonal $\forall t$. In this context, the following result can be stated:

Proposition 1 *Let $\boldsymbol{\lambda}_t = vecd(\mathbf{\Lambda}_t)$ denote the $k \times 1$ vector containing the diagonal of $\mathbf{\Lambda}_t$. If the stochastic processes in $\boldsymbol{\lambda}_t$ are linearly independent, in the sense that there is no vector $\boldsymbol{\alpha} \in \mathcal{R}^k, \boldsymbol{\alpha} \neq \mathbf{0}$, such that $\boldsymbol{\alpha}'\boldsymbol{\lambda}_t = 0, \forall t$, $\mathbf{C}$ is unique under orthogonal transformations other than column permutations and sign changes.*

Notice the generality of Proposition 1 since it has been obtained without assuming any particular parametrisation for the dynamic heteroskedasticity; it relies only on the conditional orthogonality of the factors, the linearly independent time-variation of their variances, and the constancy of $\mathbf{C}$. One possible way to gain some intuition on this result is to recall that parameter identifiability can be obtained in many econometric models by looking at higher order moments. Since conditional normality with changing variances is incompatible with unconditional normality, but at the same time implies autocorrelation in $vech(\mathbf{x}_t\mathbf{x}_t')$, Proposition 1 provides an example in which identifiability comes from considering dynamic fourth-order, as opposed to second order, moments.

If the processes in $\boldsymbol{\lambda}_t$ were linearly dependent, though, identification problems would re-appear. Given the parametrisations used in empirical work (see section 2.1), it is difficult to envisage situations in which this will be the case, unless two or more 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 characterized by a scalar covariance matrix $\lambda_{kk,t}\mathbf{I}_{k_2}$, while the others have an unrestricted diagonal covariance matrix $\mathbf{\Lambda}_{1t}$. 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 2 *Let $\boldsymbol{\lambda}_{1t} = vecd(\mathbf{\Lambda}_{1t})$. If the stochastic processes in $(\boldsymbol{\lambda}'_{1t}, \lambda_{kk,t})$ are linearly independent, $\mathbf{C}_1$ is unique under orthogonal transformations other than column permutations and sign changes.*

For practical purposes, Proposition 2 could be re-stated so that it would refer only to the empirically relevant case in which $\lambda_{kk,t} = 1, \forall t$. However, in its present form it makes it clear that the lack of identifiability comes from the factors having common, rather than constant, variances.

Finally, note that the imposition of unnecessary restrictions on $\mathbf{C}$ by analogy with standard factor models may produce misleading results. An important implication of our results is that if such restrictions were nevertheless made, at least they could then be tested. However, the accuracy that can be achieved in estimating $\mathbf{C}$ depends on how much linearly independent variability there is in $\mathbf{\Lambda}_t$, for if the elements of this matrix are essentially constant, identifiability problems will reappear.

3.3 Extensions

Proposition 1 can also be applied to other closely related models, and in particular to the model in Harvey, Ruiz and Sentana (1992). Theirs is a general state space formulation for $\mathbf{x}_t$, with unrestricted mean dynamics, in which some unobservable components show dynamic conditional heteroskedasticity. In this section, we shall explicitly consider the application of Proposition 1 to some well-known special cases which are empirically relevant.

3.3.1 Conditionally Heteroskedastic in Mean Factor Models

Several recent studies based on dynamic versions of the APT have estimated conditionally heteroskedastic factor models in which the variances of the common factors affect the mean of $\mathbf{x}_t$ (see e.g. Engle, Ng and Rothschild (1990), King, Sentana and Wadhwani (1994), or Ng, Engle and Rothschild (1992)). The models typically considered in those studies can be expressed as:

$$\mathbf{x}_t = \mathbf{C}\boldsymbol{\Lambda}_t\boldsymbol{\tau} + \mathbf{C}\mathbf{f}_t + \mathbf{w}_t$$

where $\boldsymbol{\tau}$ is a $k \times 1$ vector of "price of risk" coefficients. Notice that if $\boldsymbol{\tau} = \mathbf{0}$, we return to the previous case. Since the proof of Proposition 1 is based on the diagonality of the conditional variance of $\mathbf{f}_t$, it is straightforward to show that the columns of $\mathbf{C}$ and $\boldsymbol{\tau}'$ corresponding to factors with linearly independent time-varying variances are identifiable (up to sign changes and permutations).

3.3.2 Conditionally Heteroskedastic Dynamic Factor Models

The formulation considered in section 2 is also a special case of the so-called dynamic factor model, which constitutes a popular specification for multivariate time series applications because of its plausibility and parsimony (see e.g. Engle and Watson (1982) or Peña and Box (1987)). For simplicity, we shall just consider here the case in which the factor dynamics can be captured by a VAR(1) process. Specifically,

$$\mathbf{x}_t = \mathbf{C}\mathbf{y}_t + \mathbf{w}_t; \qquad \mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{f}_t$$

where $\mathbf{y}_t$ is a $k \times 1$ vector of dynamic factors, $\mathbf{A}$ is the matrix of VAR coefficients and $\mathbf{f}_t$ and $\mathbf{w}_t$ are defined as in (1-2). If $\mathbf{A} = \mathbf{0}$, we go back to the traditional (i.e. static) factor model. On the other hand, when $\mathbf{A} = \mathbf{I}$ we have the *common trends* model (see e.g. Harvey (1989) or Stock and Watson (1988)). If $\mathbf{f}_t$ is conditionally homoskedastic, it is well known that an o.e. model (up to unconditional second moments) can be obtained by orthogonally rotating $\mathbf{y}_t$. That is, for any orthogonal matrix $\mathbf{Q}$, the model $\mathbf{x}_t = \mathbf{C}^*\mathbf{y}_t^* + \mathbf{w}_t$, $\mathbf{y}_t^* = \mathbf{A}^*\mathbf{y}_{t-1}^* + \mathbf{f}_t^*$ where $\mathbf{y}_t^* = \mathbf{Q}\mathbf{y}_t, \mathbf{f}_t^* = \mathbf{Q}\mathbf{f}_t, \mathbf{C}^* = \mathbf{C}\mathbf{Q}'$ and $\mathbf{A}^* = \mathbf{Q}\mathbf{A}\mathbf{Q}'$, is o.e. Again, Proposition 1 implies that linearly independent time-variability in the conditional variances of $\mathbf{f}_t$ will eliminate the nonidentifiability of the matrix $\mathbf{C}$.

3.3.3 Vector Autoregressive Moving Average Models

Our results also apply to models with N common factors, no idiosyncratic noise and linear mean dynamics, such as VARMA(r, s) models. Again, for simplicity consider the following VAR(1):

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{u}_t; \qquad \mathbf{u}_t = \mathbf{C}\mathbf{f}_t$$

where $\mathbf{f}_t$, a $N \times 1$ vector defined as in (1-2), could perhaps be better understood in this context as conditionally orthogonal "fundamental" shocks affecting the process $\mathbf{x}_t$. Given that $\mathbf{f}_t$ is white noise, we can estimate this model without taking into account the time-variation in conditional variances. But then $\mathbf{C}$ is not identifiable without extra restrictions. This problem is well known and has received substantial attention in macroeconometrics. To solve it, some authors impose short run restrictions such as $\mathbf{C}$ lower triangular (cf. the discussion in section 3.2). More recently, Blanchard and Quah (1989) have achieved identifiability by means of restrictions on some elements of the long run multipliers $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{C}$. But suppose that some elements of $\mathbf{f}_t$ have time-varying conditional variances and this is explicitly recognized in estimation. Then Proposition 2 implies that the columns of $\mathbf{C}$ associated with those disturbances are identifiable.

In this context, we can perhaps shed more light on Proposition 1 by reinterpreting it as a uniqueness result for the disturbances, $\mathbf{f}_t$. Given the way in which the model is defined, we know that there is a set of disturbances, conditionally uncorrelated with each other, that can be written as a (time-invariant) linear combination of the innovations in $\mathbf{x}_t$, namely, $\mathbf{f}_t = \mathbf{C}^{-1}\mathbf{u}_t$. If $k_2 \leq 1$, Proposition 2 then says that there is only one such set.[5]

3.3.4 Oblique Factor Models with Constant Conditional Covariances

So far we have assumed that the factors are conditionally orthogonal, since this has been a maintained assumption in all existing empirical applications. However, as the following proposition shows, it turns out that most of the identifiability is coming from the fact the conditional covariances of conditionally orthogonal factors are (trivially) constant over time

Proposition 3 *Let $\mathbf{\Lambda}_t$ be a $k\times\ k$ positive definite matrix of (possibly) time-varying factor variances but constant conditional covariances, and let $\boldsymbol{\lambda}_t = vecd(\mathbf{\Lambda}_t)$. If the stochastic processes in $(\boldsymbol{\lambda}'_t, 1)$ are linearly independent, $\mathbf{C}$ is unique under orthogonal transformations other than column permutations and sign changes*

Notice that the main difference with Proposition 1 is that identification problems reappear in oblique factor models when a single factor has constant conditional variance. The reason is that we can transfer unconditional variance from the conditionally homoskedastic factor to the others. This is not possible if the factors have to remain conditionally orthogonal.

Factor models with constant conditional covariances arise more commonly than it may appear. For instance, the factor ARCH model of Engle (1987) is o.e. to a whole family of oblique factor models with constant conditional covariances, whose limiting cases are the conditionally orthogonal factor model in (4), and a model with a singular idiosyncratic covariance matrix (see Sentana (1997a) for details). In fact, we can always express any conditionally heteroskedastic factor model as an oblique factor model with constant conditional covariances and a

[5] However, it is important to emphasize that Proposition 1 is not an existence result, in that it does not say whether or not such disturbances exist to begin with. Rather, it takes them as given.

singular idiosyncratic covariance matrix, since:

$$\mathbf{x}_t = \mathbf{C}\mathbf{f}_t^G + \mathbf{w}_t^G$$

$$\begin{pmatrix} \mathbf{f}_t^G \\ \mathbf{w}_t^G \end{pmatrix} \mid \mathbf{X}_{t-1} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Lambda}_t + (\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} - \mathbf{C}\left(\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}' \end{pmatrix} \right]$$

where $\mathbf{f}_t^G = (\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{x}_t$ are the Generalized Least Squares (GLS) estimates of the common factors (see Gourieroux, Monfort and Renault (1991)). These factor scores are different from the minimum (conditional) mean square error estimates, but closely related as $\mathbf{f}_t^G = \left[\mathbf{I} + (\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})^{-1}\mathbf{\Lambda}_t^{-1}\right]\mathbf{f}_{t|t}$.

4 Estimation and Testing

In model (1-2), the parameters of interest, $\boldsymbol{\phi}'=(\mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\psi}')$, where $\mathbf{c} = vec(\mathbf{C})$ and $\boldsymbol{\gamma} = vech(\mathbf{\Gamma})$ or $vecd(\mathbf{\Gamma})$, are usually estimated jointly from the log-likelihood function of the observed variables, $\mathbf{x}_t$. Ignoring initial conditions, the log-likelihood function of a sample of size T takes the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, where:

$$l_t(\boldsymbol{\phi}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma}| - \frac{1}{2}\mathbf{x}_t'\left(\mathbf{C}\mathbf{\Lambda}_t\mathbf{C}' + \mathbf{\Gamma}\right)^{-1}\mathbf{x}_t \tag{6}$$

and $\mathbf{\Lambda}_t = diag\left[\boldsymbol{\lambda}_t(\boldsymbol{\phi})\right]$, which allows the conditional variances of the factors to depend not only on $\boldsymbol{\psi}$, but also on the static factor model parameters $\mathbf{c}$ and $\boldsymbol{\gamma}$.

Since the first order conditions are particularly complicated in this case (see appendix B), a numerical approach is usually required. Unfortunately, the application of standard quasi-Newton optimisation routines results in a very time consuming procedure, which is disproportionately more so as the number of series considered increases. In this respect, Demos and Sentana (1996b) show that using the EM algorithm combined with derivative-based methods significantly reduces the computational burden. Nevertheless, it is still of some interest to have simpler alternative estimation procedures.

4.1 Two-step consistent estimation procedures

Most empirical applications of the factor GARCH model have been carried out using a two-step univariate GARCH method under the assumption that the matrix $\mathbf{D}$ is *known*. First, univariate models are fitted to $\dot{x}_{jt} = \mathbf{d}_j'\mathbf{x}_t, j = 1, 2, \ldots, k$. Then, the estimated conditional variances are taken as data in the estimation of N univariate models for each $x_{it}, i = 1, 2, \ldots, N$. However, such a procedure ignores cross-sectional correlations and parameter restrictions, and thus sacrifices efficiency. For that reason, Demos and Sentana (1996b) proposed an EM-based *restricted* maximum likelihood estimator which exploits those restrictions but maintains the assumption of known $\mathbf{D}$. In the general case, an equivalent assumption would be that the matrix $\mathbf{D}' = (\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Gamma}^{-1}$ is known, which is tantamount to $\mathbf{f}_t^G$ being observed. Under such a maintained assumption, it is possible to prove that consistent estimates of $\mathbf{C}, \mathbf{\Gamma}$ and $\boldsymbol{\psi}$ can be obtained by combining the estimates of the marginal model for $\mathbf{f}_t^G$ with the estimates from the OLS regression of each x_{it} on $\mathbf{f}_t^G$ (see Sentana (1997b) for details). Unfortunately, the consistency of such restricted ML estimators crucially depends on the correct specification of the factor scores (see Lin (1992) for the factor GARCH case).

Here, we shall develop a two-step consistent estimation procedure which does not rely on knowledge of $\mathbf{f}_t^G$ for those cases in which the idiosyncratic covariance matrix is diagonal. For clarity of exposition, we initially assume that the matrix $\mathbf{C}$ is identifiable even if we ignore the time-variation in $\mathbf{\Lambda}_t$.

The rationale for our proposed two-step estimator is as follows. We saw in section 2 that if $\mathbf{f}_t$ and $\mathbf{w}_t$ are covariance stationary, the unconditional covariance matrix, $\mathbf{\Sigma}$, inherits the factor structure (cf. (5)). As our first step, therefore, we can estimate the unconditional variance parameters $\mathbf{c}$ and $\boldsymbol{\gamma}$ by pseudo-maximum likelihood using a standard factor analytic routine. Note that such estimators satisfy $(\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}}) = \arg\max_{\mathbf{c},\boldsymbol{\gamma}} L_T(\mathbf{c}, \boldsymbol{\gamma}, \mathbf{0})$. It is easy to see that $(\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}})$ are root-T

consistent, as the expected value of the score of the estimated model evaluated at the true parameter values is 0 under our assumptions. However, since the first derivatives are proportional to $vech(\mathbf{x}_t\mathbf{x}_t')$ (see appendix B), the score does not preserve the martingale difference property when there are ARCH effects in the common factors, and it is necessary to compute robust standard errors which take into account its serial correlation.

Having obtained consistent estimates of $\mathbf{c}$ and $\boldsymbol{\gamma}$, we can then estimate the conditional variance parameters by maximizing (6) with respect to $\boldsymbol{\psi}$ keeping $\mathbf{c}$ and $\boldsymbol{\gamma}$ fixed at their pseudo-maximum likelihood estimates. That is, our second step estimator is $\hat{\boldsymbol{\psi}} = \arg\max_{\psi} L_T(\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\psi})$. On the basis of well-known results from Durbin (1970), it is clear that $\hat{\boldsymbol{\psi}}$ is also root-T consistent. However, since the asymptotic covariance matrix is not generally block-diagonal between static and dynamic variance parameters (see appendix B), 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 (6) evaluated at the two-step estimators $\hat{\mathbf{c}}, \hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\psi}}$ (see Lin (1992) for an analogous correction in the factor GARCH case).

When $\mathbf{C}$ is not identifiable from the unconditional covariance matrix, $\hat{\boldsymbol{\gamma}}$ remains consistent, but $\hat{\mathbf{c}}$ is only consistent up to an orthogonal transformation. As discussed in section 3.2, the reason is that by assuming unconditional normality in estimation, we are neglecting very valuable information in dynamic fourth order moments. One possibility would be to replace the Gaussian quasi-likelihood in the first-step by an alternative objective function which 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. In any case, note that if we were to iterate our proposed two step procedure and achieved convergence, we would obtain fully efficient maximum likelihood estimates of all model parameters. 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 maximization of (6) with respect to $\boldsymbol{\psi}$ holding $\mathbf{c}$ and $\boldsymbol{\gamma}$ fixed.

4.2 A simple LM test for ARCH in the common factors

Despite the simplicity of the two-step procedure, the numerical maximization of (6) with respect to $\boldsymbol{\psi}$ in models such as (3) still involves the use of the Kalman filter to produce estimates of $f_{jt-1|t-1}$ and $\lambda_{jj,t-1|t-1}$ once per parameter per iteration. Therefore, it is of some interest to have a simple preliminary test for the presence of ARCH effects in the common factors. Moreover, since the way in which standard errors are usually computed in static factor models is only valid under conditional homoskedasticity, it is convenient to have a model diagnostic to assess the validity of such a maintained assumption.

If the factors were observable, we could easily carry out standard LM tests for ARCH on each of them. For the ARCH(1) case, for instance, that would entail regressing 1 on $(f_{jt}^2-1)(f_{jt-1}^2-1)$, or equivalently f_{jt}^2-1 on f_{jt-1}^2-1. Unfortunately, the factors are generally unobserved. Nevertheless, we can derive similar tests using some factor estimates instead. Under conditional normality, $\mathbf{f}_{t|t}$, the Kalman filter based estimates of the underlying factors, satisfy:

$$\mathbf{f}_{t|t}|\mathbf{X}_{t-1} \sim N\left[\mathbf{0}, \boldsymbol{\Lambda}_{t|t}\right] \tag{7}$$

where $\boldsymbol{\Lambda}_{t|t} = \left[\boldsymbol{\Lambda}_t^{-1} + (\mathbf{C}'\boldsymbol{\Gamma}^{-1}\mathbf{C})^{-1}\right]^{-1}$. As a result, $\mathbf{f}_{t|t}$ will be conditionally homoskedastic if and only if $\boldsymbol{\Lambda}_t$ is constant over time. Hence, had we data on $\mathbf{f}_{t|t}$, we could test whether or not the moment condition $cov\left[f_{jt|t}^2, f_{jt-1|t-1}^2\right] = 0$ holds for $j = 1, \ldots, k$. Importantly, the aggregation results in Nijman and Sentana (1996) imply that linear combinations of multivariate factor models like (1-2), whose weights are not orthogonal to $\mathbf{C}$, will follow weak GARCH processes. Therefore, such moment tests will have non-trivial power since under the alternative $f_{jt|t}$ will show serial correlation in the squares.

In practice, we must base the tests on $\mathbf{f}_{t|t}$ evaluated at the parameter estimates under the null. In particular, we will use

$$\hat{\mathbf{f}}_{t|t} = \hat{\mathbf{\Lambda}}_{t|t}\hat{\mathbf{C}}'\hat{\mathbf{\Gamma}}^{-1}\mathbf{x}_t$$

where

$$\hat{\mathbf{\Lambda}}_{t|t} = \left[\mathbf{I} + \left(\hat{\mathbf{C}}'\hat{\mathbf{\Gamma}}^{-1}\hat{\mathbf{C}}\right)^{-1}\right]^{-1} \quad \forall t$$

It turns out that the presence of parameter estimates does not affect the asymptotic distribution of such tests, as the information matrix is block diagonal between $\boldsymbol{\psi}$ and $(\mathbf{c}, \boldsymbol{\gamma})$ under the null (see appendix B). Furthermore, we also prove in appendix B that our proposed moment test is precisely the standard LM test for conditional homoskedasticity in the common factors based on the score of (6) evaluated under H_0. Therefore, we can compute a two-sided χ^2_1 test against ARCH(1) in each common factor as T times the uncentred R^2 from the regression of either 1 on $(\hat{f}^2_{jt|t} + \hat{\lambda}_{jj,t|t} - 1)$ times $(\hat{f}^2_{jt-1|t-1} + \hat{\lambda}_{jj,t-1|t-1} - 1)$ (outer-product version), or $(\hat{f}^2_{jt|t} + \hat{\lambda}_{jj,t|t} - 1)$ on $(\hat{f}^2_{jt|t} + \hat{\lambda}_{jj,t|t} - 1)$ (Hessian-based version). In fact, more powerful variants of these tests can be obtained by taking the one-sided nature of the alternative hypothesis into account through the sign of the relevant regression coefficient (see Demos and Sentana (1996a)).

5 Monte Carlo Evidence

In a recent paper, Lin (1992) analyzes different estimation methods for the factor GARCH model of Engle (1987) by means of a detailed Monte Carlo analysis. In this section, we shall conduct a similar exercise for the conditionally heteroskedastic latent factor model in (3). 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. Nevertheless, we select the parameter values, and in particular the signal-to-noise ratio, so as to reflect empirically relevant situations.

5.1 A single factor model

We first generated 8000 samples of 240 observations each (plus another 100 for initialization) of a trivariate single factor model using the NAG library G05DDF routine. Such a sample size corresponds roughly to twenty years of monthly data, five years of weekly data or one year of daily data. Since the performance of the different estimators depends on $\mathbf{C}$ and $\mathbf{\Gamma}$ mostly through the scalar quantity $(\mathbf{C}'\mathbf{\Gamma}^{-1}\mathbf{C})$, the model considered is:

$$x_{it} = c_i f_t + w_{it} \qquad (i = 1, 2, 3)$$

with $\mathbf{c} = (1, 1, 1)'$, $\lambda_t = (1-\alpha-\beta)+\alpha(f^2_{t-1|t-1}+\lambda_{t-1|t-1})+\beta\lambda_{t-1}$ and $\mathbf{\Gamma} =\gamma\mathbf{I}$. Two values of γ have been selected, namely 2 or 1/2, corresponding to low and high signal to noise ratios, and three pairs of values for α and β, namely $(0, 0), (.2, .6)$ and $(.4, .4)$, which represent constant variances, persistent but smooth GARCH behaviour, and persistent but volatile conditional variances respectively. It is worth mentioning that the pair $\alpha = .2, \beta = .6$ matches roughly what we tend to see in the empirical literature. In order to minimize experimental error, we use the same set of underlying random numbers in all designs. Maximization of the log-likelihood (6) with respect to $\mathbf{c}, \boldsymbol{\gamma}, \alpha$ and β was carried out using the NAG library E04JBF routine. Initial values of the parameters were obtained by means of the EM algorithm in Demos and Sentana (1996b).

For scaling purposes, we use $c_1^2 + c_2^2 + c_3^2 = 1$, and leave the constant part of the conditional variance free. In order to guarantee the positivity and stationarity restrictions $0 \leq \beta \leq 1 - \alpha \leq 1$, we use the re-parametrisation $\alpha = \sin^2(\theta_1)$

and $\beta = \sin^2(\theta_2)(1-\alpha)$. Similarly, we used $\gamma_i = (\gamma_i^*)^2$. We also set λ_1 to the unconditional variance of the common factor to start up the recursions. But since this implies that β is not identified if $\alpha = 0$, we set $\beta = 0$ whenever $\alpha = 0$.

In this respect, it is important to mention that when α and/or β are 0, the parameter values lie on the boundary of the admissible range. The distribution of the ML estimator and associate tests in those situations has been studied by Self and Liang (1987) and Wolak (1989). When $\alpha = 0$, for instance, we could use the result in case 2, theorem 2 of Self and Liang (1987), to show that the asymptotic distribution of the ML estimators of $(\beta, \alpha, \mathbf{c}', \boldsymbol{\gamma}')$ should be a $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ mixture of a) the usual asymptotic distribution, b) the asymptotic distribution of a restricted ML estimator which sets $\alpha = \beta = 0$, and c) the asymptotic distribution of a restricted ML estimator which only sets $\beta = 0$. We take into account these results in order to compute standard errors.

It is also important to mention that joint estimates are always at least as efficient as two-step estimates in this context, since the information matrix is block-diagonal between unconditional and conditional variance parameters under the null of no ARCH.

Table 1 presents mean biases and standard deviations across replications for joint and two-step maximum likelihood estimates of the static factor model parameters $\mathbf{c}$ and $\boldsymbol{\gamma}$. For simplicity of exposition, only averages across equations are included (in particular, $c = (c_1 + c_2 + c_3)/3$ and $\gamma = (\gamma_1 + \gamma_2 + \gamma_3)/3$). Note that all estimates are very mildly downward biased. At the same time, it seems that the more variability there is in conditional variances, the better joint estimates are relative to two-step estimates. Nevertheless, the differences are minor, at least for the sample sized used.

Given the large number of parameters involved, we summarize the performance of the estimates of the asymptotic covariance matrix of these estimators by computing the experimental distribution of some simple test statistics. In particular, we test $c_1 = c_2 = c_3$, and $\gamma_1 = \gamma_2 = \gamma_3$. Both tests should have asymptotic χ^2_2 distributions under the null. Standard errors for joint ML estimates are computed from the Hessian. On the other hand, the usual sandwich estimator with a 4-lag triangular window is employed for two-step estimates of the static factor parameters. The results, which are not reported for conciseness, suggest that the size distortions are not very large.

Our experimental design also allows us to analyze the performance of the different LM test for ARCH under the null, and under two alternatives. In order to evaluate their size properties, we employ the **p-value discrepancy plots** proposed by Davidson and MacKinnon (1996), which are plots of the difference between actual and nominal test size versus nominal test size for all possible test sizes. If the asymptotic distribution is correct, p-value discrepancy plots should be close to the x axis. Figure 1 shows such plots for the one-sided and two-sided versions of the outer-product and Hessian-based forms of the LM test. As expected, the outer-product versions have much larger distortions than the Hessian-based ones, whose sizes are fairly accurate.

In order to display the simulation evidence on the power of the different tests, we employ the **size-power curves** of Davidson and MacKinnon (1996), which are plots of test power versus actual test size for all possible test sizes. The main advantage of size-power plots is that they allow us to see immediately the effect on power of different parameter values, as well as to compare the relative powers of test statistics that have different null distributions. Figure 2 presents such plots for the Hessian-based one-sided and two-sided tests. 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 (1996a)).

Table 2 presents the proportion of estimates of α and β which are at the boundary of the parameter space. Asymptotically, the proportions of $\alpha = \beta = 0$ and $\alpha \neq 0, \beta = 0$ should be $(\frac{1}{2}, \frac{1}{4})$ under the null of no ARCH, and $(0, 0)$ under the alternative. However, the results show that $\alpha = 0$, and especially $\beta = 0$ occur more frequently than what the asymptotic distribution would suggest. This is particularly true 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 joint and two-step maximum likelihood estimates of α and β. In this respect, it is important to mention that since β is not identified when $\alpha = 0$, the reported values for β correspond to those cases in which α is not estimated as 0. Note that the $\alpha's$ obtained are rather more accurate than the $\beta's$. Also note that the biases for the joint estimates of α are smaller than for the two-step ones, although the latter have smaller Monte Carlo variability. In contrast, the downward biases in β are larger for joint ML estimates. To some extent, these biases reflect the larger proportion of zero $\beta's$ in Table 2.

5.2 A two factor model

We have also simulated the following six-variate model with two factors:

$$x_{it} = c_{i1} f_{1t} + c_{i2} f_{2t} + w_{it}$$

with $\lambda_{11,t} = (1 - \alpha - \beta) + \alpha(f^2_{1t-1|t-1} + \lambda_{11,t-1|t-1}) + \beta\lambda_{11,t-1}$, $\lambda_{22,t} = 1$ and $\mathbf{\Gamma} = \gamma\mathbf{I}$. Please note that according to Proposition 1, the parameters in $\mathbf{C}$ are identified without further restrictions, provided that $\alpha \neq 0$ and we take into account the time-variation in conditional second moments.

Two sets of values for $\mathbf{C}$ have been selected, $\mathbf{c}'=(0, 0, 0, 1, 1, 1; 1, 1, 1, 0, 0, 0)$ and $\mathbf{c}'= (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, 1, 1; 1, 1, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. The first design corresponds to two trivariate single factor models like the one considered in the previous subsection put together, while the second design introduces "correlation" in the columns of $\mathbf{C}$.

For each value of $\mathbf{c}$, two values of γ have been selected, namely 2 and 1/2, corresponding to low and high signal to noise ratios. Then for each of the four combinations, we consider two pairs of values for α and β, namely $(.4, .4)$ and $(.2, .6)$, in order to obtain persistent but volatile conditional variances, and the more realistic persistent but smooth GARCH behaviour. 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 estimation details are the same as in section 5.1.[6]

Table 4 presents mean biases and standard deviations across replications for joint and two step maximum likelihood estimates, as well as a restricted ML estimator which imposes the same identifying restriction as the two step estimator, namely $c_{62} = 0$. Such an estimator is efficient when the overidentifying restriction is true, but becomes inconsistent when it is false. More precisely, if $\mathbf{C}$ is not unconditionally identifiable, restricted and two-step ML estimators of $\mathbf{c}$ are consistent for the orthogonal transformation of the true parameter values which zeroes c_{62}. For simplicity of exposition, only certain averages across equations are included (in particular, $c_{a1} = (c_{11}+c_{21}+c_{31})/3$, $c_{b1} = (c_{41}+c_{51}+c_{61})/3$, $c_{a2} = (c_{12}+c_{22}+c_{32})/3$, $c_{b2} = (c_{42} + c_{52})/2$, and $\gamma = (\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + \gamma_6)/6$).

The first panel of Table 4 contains the results for those designs in which $\mathbf{c}'=(0, 0, 0, 1, 1, 1; 1, 1, 1, 0, 0, 0)$. Not surprisingly, the restricted ML estimator is clearly the best as far as estimates of the factor loadings are concerned. However, it turns out that the two-step estimator performs very similarly, except when there is significant variability in conditional variances, which is in line with the results for the single factor model. On the other hand, the joint ML estimator is the

[6] One additional issue that arose during the simulations with two factor models was that, occasionally, some idiosyncratic variances were estimated as 0. The incidence of these so-called Heywood cases increased with the value of γ, and especially c_{62}. Nevertheless, since at worst only 35 out of 2000 replications had this problem, we discarded them, and replaced them by new ones.

worst performer when the signal to noise ratio and the variability in $\lambda_{11,t}$ are low, but comes very close to the restricted ML in the opposite case.[7] This behaviour is not unexpected, given that the identifiability of the joint ML estimator comes from the fact that $\lambda_{11,t}$ changes over time, while the identifiability of the other two estimators is obtained from the restriction $c_{62} = 0$. Nevertheless, it seems that the latter identifiability condition is more informative than the former, which should be borne in mind in empirical work.

In contrast, there are only minor differences between the different estimates of the idiosyncratic variance parameters, which are always identified. Obviously, their Monte Carlo standard deviations increase when γ changes from 1/2 to 2, but the coefficients of variation remain approximately the same.

The second panel of Table 4 contains the results for those designs in which $\mathbf{c}'=(\frac{1}{4},\frac{1}{4},\frac{1}{4},1,1,1;1,1,1,\frac{1}{4},\frac{1}{4},\frac{1}{4})$. Note that the different estimates of γ_j are hardly affected. As expected, though, the behaviour of both restricted and two-step factor loading estimators radically changes, as they clearly become inconsistent. In contrast, the performance of the joint estimates of $\mathbf{c}$ is basically the same as in the first panel.

In order to summarize the performance of the estimates of the asymptotic covariance matrix of these estimators, we computed the experimental distribution of some simple test statistics. 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 our choice of parameter values, the plims of all the estimators satisfy these restrictions even when the assumption $c_{62} = 0$ is false. Therefore, all five tests should have asymptotic χ^2_2 distributions. The results, not reported for conciseness, suggest that the size distortions associated with the two-step estimator, for which the usual sandwich expression with a 4-lag triangular window is employed, are small, but larger than

[7] Since the joint estimates of $\mathbf{c}$ are not identified when $\alpha = 0$, the reported values correspond to those cases in which α is not estimated as 0.

those for joint and restricted ML estimators.

Our design also allows us to consider the finite sample distribution of the likelihood ratio test for the restriction $c_{62} = 0$, both under the null and under the alternative. The p-value discrepancy plot presented in Figure 3 shows that nominal test sizes are fairly accurate at the 5% level, although less so when γ is small. For very large significance levels, however, the size distortions are higher, because the LR test takes the value 0 when α is estimated as 0. The distribution of this test under the alternative, though, is far more interesting, as it provides a summary indicator of the determinants of the information content in our identifiability restrictions. Figure 4 present the size-power curves for the four experimental designs in which $c_{62} \neq 0$. Although null and alternative experimental designs differ in more than one parameter, we have done the required implicit size-corrections in these plots using the closest match (cf. Davidson and MacKinnon (1996)). Not surprisingly, the absolute power of the test is small, as 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, it is clear that 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, as stated in Proposition 1.

Table 5 presents the proportion of estimates of α and β which are at the boundary of the parameter space. In all cases, the proportions of $\alpha = \beta = 0$ and $\alpha \neq 0, \beta = 0$ should be $(0, 0)$ asymptotically. But as in the single factor model, the results show that $\alpha = 0$, and especially $\beta = 0$ occur more frequently than what the asymptotic distribution would suggest. This is particularly true 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 joint, restricted and two-step maximum likelihood estimators of α and β. Once more, the $\alpha's$ are estimated rather more accurately than the $\beta's$, which reflects the larger proportion of zero $\beta's$ in Table 5. As in Table 4, though, there are significant differences between the first and second panel. While the performance of joint ML estimator is by and large independent of whether or not $c_{62} = 0$, the behaviour of the restricted and two-step estimators radically changes, and they clearly become inconsistent.

6 Conclusions

In this paper we investigate some important issues related to the identification, estimation and testing of multivariate conditionally heteroskedastic factor models. We begin by re-examining the identification problems of traditional factor analysis. It turns out that the model considered here only suffers from lack of identification in as much as the variances of some of the common factors are constant. Thus, there is a non-trivial advantage in explicitly recognizing the existence of dynamic heteroskedasticity when estimating factor analytic models. Our results also apply to other popular time series models, and in particular, to dynamic versions of the APT in which the variances of the common factors affect the mean of $\mathbf{x}_t$. Importantly, our result could also be useful in the interpretation of common trend-dynamic factor models, and in the identification of fundamental disturbances from vector autoregressions.

Secondly, we propose a root-T consistent two-step estimation procedure for these models which does not rely on knowledge of (some consistent estimates of) the factors. For those cases in which the idiosyncratic covariance matrix is diagonal, and the factor loadings are identified even if we ignore the time-variation in the factor variances, our procedure involves estimating the factor loadings and idiosyncratic variances by pseudo-maximum likelihood based on the unconditional covariance matrix. Then, the conditional variance parameters are estimated by maximizing the log-likelihood function of the observed variables keeping the static factor model parameters fixed at their pseudo-maximum likelihood estimates. In this respect, we also explain how to compute correct standard errors.

Thirdly, we develop a simple preliminary moment test for the presence of ARCH effects in the common factors, which can also be employed as a model diagnostic. This is particularly relevant because the way in which standard errors are usually computed in static factor models is only valid under conditional homoskedasticity. Importantly, we prove that our proposed test is precisely the standard LM test for conditional homoskedasticity in the common factors based on the score of the joint model evaluated under the null. Not surprisingly, it can be computed as T times the uncentred R^2 from an auxiliary regression involving squares of the best estimates of the factors and their lags. In fact, more powerful versions of these tests can be obtained by taking the one-sided nature of the alternative hypothesis into account.

Finally, we investigate the finite sample properties of our proposed estimators and hypothesis tests by simulation methods in order to assess the reliability of their asymptotic distributions in practice. Our results suggest that: (i) the efficiency of joint ML estimates of $\mathbf{c}$ and $\boldsymbol{\gamma}$ relative to two-step estimates 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, which results in significant downward biases; and (vi) although time-variation in factor variances ensures identification in practice, traditional conditions on $\mathbf{C}$ are more informative, as long as they are correct.

The conditionally heteroskedastic factor model in (1-2) is a special case of the general approximate conditional factor representation $\mathbf{\Sigma}_t = \mathbf{C}_t\mathbf{C}_t + \mathbf{\Gamma}_t$, where $\mathbf{C}_t$ is a $N \times k$ matrix of measurable functions of the information set and $\mathbf{\Gamma}_t$ is such that its eigenvalues remain bounded as N increases. In this framework, our model can be-written as $\mathbf{x}_t = \mathbf{C}_t\mathbf{f}_t^{\#} + \mathbf{w}_t$, where $V(\mathbf{w}_t|\mathbf{X}_{t-1}) = \mathbf{\Gamma}, V(\mathbf{f}_t^{\#}|\mathbf{X}_{t-1}) = \mathbf{I}$ and $\mathbf{C}_t = \mathbf{C}\mathbf{\Lambda}_t^{1/2}$, so that the loadings of different variables on each conditionally homoskedastic factor change proportionately over time (see Engle, Ng and Rothschild (1990)). The motivation for such an assumption is twofold. First, it provides a parsimonious and plausible specification of the time variation in $\mathbf{\Sigma}_t$, and for that reason has been the only one adopted so far in empirical applications. Second, it implies that the unconditional factor representation of $\mathbf{x}_t$ is well defined (provided unconditional variances are bounded), which makes it compatible with the standard approach based on $\mathbf{\Sigma}$, and therefore empirically relevant. Notice that even if $\mathbf{\Gamma}_t$ is diagonal, the unconditional variance of a process characterized by a conditional factor representation may very well lack an unconditional factor structure for any $k < N$ (see Hansen and Richard (1987)). Although the model is not identifiable if $\mathbf{C}_t$ is unspecified, this paper shows that the statistical properties of alternative plausible formulations of the general conditional factor model would certainly merit a close look.

References

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

Bollerslev, T. and Engle, R.F. (1993): "Common persistence in conditional variances", *Econometrica* 61, 166-187.

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.

Chamberlain, G. and Rothschild, M. (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica* 51, 1281-1304.

Davidson, R. and MacKinnon, J.G. (1996): "Graphical methods for investigating the size and power of hypothesis tests", mimeo, GREQAM.

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

Demos, A. and Sentana, E. (1996a): "Testing for GARCH effects: A one-sided approach", CEMFI Working Paper 9611.

Demos, A. and Sentana, E. (1996b): "An EM algorithm for conditionally heteroskedastic factor models", CEMFI Working Paper 9615, forthcoming in *Journal of Business and Economic Statistics.*

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.

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. (1987): "Multivariate ARCH with factor structures - cointegration in variance", mimeo, University of California at San Diego.

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 Watson, M. (1981): "A one-factor multivariate time series model of metropolitan wage rates", *Journal of the American Statistical Association* 76, 774-781.

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

Hansen, L.P. and Richard, S.F. (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models", *Econometrica* 55, 587-613.

Harvey, A.C. (1989): *Forecasting, Structural Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A.C., Ruiz, E. and Sentana, E. (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.

Johnson, R.A. and Wichern, D.W. (1992): *Applied Multivariate Statistical Analysis*, 3rd edition, Prentice-Hall.

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

Kroner, K.F. (1987): "Estimating and testing for factor GARCH", mimeo, University of California at San Diego.

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

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.

Ng, V.M.; Engle, R.F. and Rothschild, M. (1992): "A multi-dynamic factor model for stock returns", *Journal of Econometrics* 52, 245-266.

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

Peña, D. and Box, G.E.P. (1987): "Identifying a simplifying structure in time series", *Journal of the American Statistical Association* 82, 836-843.

Ross, S. (1976): "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, 13, 341-360.

Self, S.G. and Liang, K.Y. (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. (1995): "Quadratic ARCH models", *Review of Economic Studies* 62, 639-661.

Sentana, E. (1997a): "The relation between conditionally heteroskedastic factor models and factor GARCH models", mimeo, CEMFI.

Sentana, E. (1997b): "Risk and return in the Spanish stock market: Some evidence from individual assets", CEMFI Working Paper 9702, forthcoming in *Investigaciones Económicas.*

Stock, J.H. and Watson, M.W. (1988): "Testing for common trends", *Journal of the American Statistical Association* 83, 1097-1107.

Wolak, F.A. (1989): "Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models", *Econometric Theory* 5, 1-35.

Appendices

A Proofs

A.1 Proposition 1

Let $\mathbf{Q}$ be an arbitrary $k \times k$ orthogonal matrix with typical element $[\mathbf{Q}]_{ij} = q_{ij}$ such that $\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}_k$. Since the covariance matrix of the transformed factors $\mathbf{f}_t^* = \mathbf{Q}\mathbf{f}_t$, is $\mathbf{\Lambda}_t^* = \mathbf{Q}\mathbf{\Lambda}_t\mathbf{Q}'$, with typical element $[\mathbf{\Lambda}_t^*]_{ij} = \sum_{l=1}^k \lambda_{ll,t} q_{il} q_{jl}$, conditional orthogonality requires $\sum_{l=1}^k \lambda_{ll,t} q_{il} q_{jl} = 0$ for $j > i, i = 1, 2, \ldots k$ and $t = 1, 2, \ldots T$. For a given i, j $(j > i)$, these restrictions can be expressed in matrix notation as:

$$\tilde{\mathbf{\Lambda}}_T \mathbf{q}_{ij} = 0 \cdot \boldsymbol{\iota}_T \tag{A1}$$

where $\tilde{\mathbf{\Lambda}}_T = \begin{pmatrix} \lambda_{11,1} & \lambda_{22,1} & \cdots & \lambda_{kk,1} \\ \lambda_{11,2} & \lambda_{22,2} & \cdots & \lambda_{kk,2} \\ \vdots & \vdots & & \vdots \\ \lambda_{11,T} & \lambda_{22,T} & \cdots & \lambda_{kk,T} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}_1' \\ \boldsymbol{\lambda}_2' \\ \vdots \\ \boldsymbol{\lambda}_T' \end{pmatrix}$ is a $T \times k$ matrix, $\boldsymbol{\iota}_T$ a $T \times 1$ vector of ones and $\mathbf{q}_{ij} = \begin{pmatrix} q_{i1}q_{j1} & q_{i2}q_{j2} & \ldots & q_{ik}q_{jk} \end{pmatrix}'$ a $k \times 1$ vector. We can regard (A1) as a set of T homogenous linear equations in k unknowns, $\mathbf{q}_{ij}$. Given that rank $\left(\tilde{\mathbf{\Lambda}}_T\right) = k$ when the stochastic processes in $\boldsymbol{\lambda}_t$ are linearly independent, the only solution to the above system of equations is $\tilde{\mathbf{\Lambda}}_T \mathbf{q}_{ij} = 0 \cdot \boldsymbol{\iota}_k$. irrespectively of i and j. That is, we must have that for all $j > i, i = 1, 2, \ldots k$, $q_{il}q_{jl} = 0$ for $l = 1, 2, \ldots, k$, which in turn requires $q_{il} = 0$ and/or $q_{jl} = 0$. Therefore, there cannot be two elements in any column of $\mathbf{Q}$ which are different from 0. Given that $\mathbf{Q}$ is an orthogonal matrix, the only admissible transformations are permutations of Cholesky square roots of the unit matrix, $\mathbf{I}_k^{1/2}$, where $\left\{\mathbf{I}_k^{1/2}\right\}_{ij} = \pm 1$ for $i = j$ and 0 otherwise. $\square$

A.2 Proposition 2

In this case (A1) also applies, but since $\boldsymbol{\lambda}_t' = (\boldsymbol{\lambda}_{1t}', \lambda_{kk,t}\boldsymbol{\iota}_{k_2}')$, we can re-write it as

$$\bar{\boldsymbol{\Lambda}}_T \bar{\mathbf{q}}_{ij} = 0 \cdot \boldsymbol{\iota}_T \tag{A2}$$

where $\bar{\boldsymbol{\Lambda}}_T = \begin{pmatrix} \lambda_{11,1} & \lambda_{22,1} & \cdots & \lambda_{k_1k_1,1} & \lambda_{kk,1} \\ \lambda_{11,2} & \lambda_{22,2} & \cdots & \lambda_{k_1k_1,2} & \lambda_{kk,2} \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda_{11,T} & \lambda_{22,T} & \cdots & \lambda_{k_1k_1,T} & \lambda_{kk,T} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}_{11}', & \lambda_{kk,1} \\ \boldsymbol{\lambda}_{12}' & \lambda_{kk,2} \\ \vdots & \vdots \\ \boldsymbol{\lambda}_{1T}' & \lambda_{kk,T} \end{pmatrix}$ is a $T \times (k_1+1)$ matrix, and $\bar{\mathbf{q}}_{ij} = \begin{pmatrix} q_{i1}q_{j1} & q_{i2}q_{j2} & \cdots & q_{ik_1}q_{jk_1} & \sum_{l=k_1+1}^{k} q_{il}q_{jl} \end{pmatrix}'$ a $(k+1) \times 1$ vector.

Since rank $\left(\bar{\boldsymbol{\Lambda}}_T\right) = k_1 + 1$ by assumption, $\bar{\mathbf{q}}_{ij} = 0 \cdot \boldsymbol{\iota}_{k_1+1}$ irrespectively of i and j. That is, for all $j > i, i = 1, 2, \ldots k$ we must have $q_{il}q_{jl} = 0$ for $l = 1, 2, \ldots, k_1$ 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. Let's partition $\mathbf{Q}$ comformably as:

$$\begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}$$

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. $\square$

A.3 Proposition 3

First of all, note that $\boldsymbol{\Lambda}_t^* = \mathbf{Q}\boldsymbol{\Lambda}_t\mathbf{Q}' = \mathbf{Q}dg(\boldsymbol{\Lambda}_t)\mathbf{Q}' + \mathbf{Q}[\boldsymbol{\Lambda}_t - dg(\boldsymbol{\Lambda}_t)]\mathbf{Q}'$. But since $\boldsymbol{\Lambda}_t - dg(\boldsymbol{\Lambda}_t)$ is time-invariant by assumption, constant conditional covariances for $\boldsymbol{\Lambda}_t^*$ simply requires that $\mathbf{Q}dg(\boldsymbol{\Lambda}_t)\mathbf{Q}'$ is also time-invariant. Given that the ij^{th} element of $\mathbf{Q}dg(\boldsymbol{\Lambda}_t)\mathbf{Q}'$ is $\sum_{l=1}^{k} \lambda_{ll,t}q_{il}q_{jl}$, this requires $\sum_{l=1}^{k} \lambda_{ll,t}q_{il}q_{jl} = \phi_{ij}$ for $j > i, i = 1, 2, \ldots k$ and $t = 1, 2, \ldots T$. For a given i, j $(j > i)$, these restrictions can be expressed in matrix notation as:

$$\tilde{\mathbf{\Lambda}}_T \mathbf{q}_{ij} = \phi_{ij} \cdot \boldsymbol{\iota}_T \tag{A3}$$

We can regard (A3) as a set of T non-homogenous linear equations in k unknowns, $\mathbf{q}_{ij}$. Given that rank $\left(\tilde{\mathbf{\Lambda}}_T | \boldsymbol{\iota}_T\right) = k+1$ when the stochastic processes in $(\boldsymbol{\lambda}_t', 1)$ are linearly independent, the only way the above system of equations can have a solution is if $\phi_{ij} = 0$ for all $j > i, i = 1, 2, \ldots k$. That is, if $\mathbf{Q}dg(\mathbf{\Lambda}_t)\mathbf{Q}'$ remains diagonal for $t = 1, 2, \ldots T$. In that case, the proof of Proposition 1 applies. $\square$

B The score and information matrix of a conditionally heteroskedastic factor model

Let $\boldsymbol{\phi}' = (\mathbf{c}', \boldsymbol{\gamma}', \boldsymbol{\psi}')$ denote the vector of parameters of interest, with $\mathbf{c} = vec(\mathbf{C})$ and $\boldsymbol{\gamma} = vecd(\mathbf{\Gamma})$. Bollerslev and Wooldridge (1992) and Kroner (1987) show that the score function $\mathbf{s}_t(\boldsymbol{\phi}) = \partial l_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi}$ of any conditionally heteroskedastic multivariate model with zero conditional mean is given by the following expression:

$$\mathbf{s}_t(\boldsymbol{\phi}) = \frac{1}{2}\frac{\partial vec'\left[\mathbf{\Sigma}_t\right]}{\partial \boldsymbol{\phi}}\left[\mathbf{\Sigma}_t^{-1} \otimes \mathbf{\Sigma}_t^{-1}\right] vec\left[\mathbf{x}_t\mathbf{x}_t' - \mathbf{\Sigma}_t\right]$$

Then, since the differential of $\mathbf{\Sigma}_t$ is

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

(cf. Magnus and Neudecker (1988)), we have that the three terms of the Jacobian corresponding to $\mathbf{c}, \boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ will be:

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

$$\frac{\partial vec\left[\mathbf{\Sigma}_t\right]}{\partial \boldsymbol{\gamma}'} = \mathbf{E}_N + (\mathbf{C} \otimes \mathbf{C})\mathbf{E}_k \frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\phi})}{\partial \boldsymbol{\gamma}'}$$

$$\frac{\partial vec\left[\mathbf{\Sigma}_t\right]}{\partial \boldsymbol{\psi}'} = (\mathbf{C} \otimes \mathbf{C})\mathbf{E}_k \frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\phi})}{\partial \boldsymbol{\psi}'}$$

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

After some straightforward algebraic manipulations, we get

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

$$\frac{1}{2}\begin{bmatrix} \partial\boldsymbol{\lambda}_t'(\boldsymbol{\phi})/\partial\mathbf{c} \\ \partial\boldsymbol{\lambda}_t'(\boldsymbol{\phi})/\partial\boldsymbol{\gamma} \\ \partial\boldsymbol{\lambda}_t'(\boldsymbol{\phi})/\partial\boldsymbol{\psi} \end{bmatrix} vecd\left[\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{C}\right]$$

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

$$\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C}\boldsymbol{\Lambda}_t - \boldsymbol{\Sigma}_t^{-1}\mathbf{C}\boldsymbol{\Lambda}_t = \boldsymbol{\Gamma}^{-1}E\left[(\mathbf{x}_t - \mathbf{C}\mathbf{f}_t)\mathbf{f}_t'|\mathbf{X}_T, \boldsymbol{\phi}\right]$$

$$\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Gamma}^{-1}E\left[(\mathbf{x}_t - \mathbf{C}\mathbf{f}_t)(\mathbf{x}_t - \mathbf{C}\mathbf{f}_t)'|\mathbf{X}_T, \boldsymbol{\phi}\right]\boldsymbol{\Gamma}^{-1}$$

$$\mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} - \mathbf{C}'\boldsymbol{\Sigma}_t^{-1}\mathbf{C} = \boldsymbol{\Lambda}_t^{-1}E\left[\mathbf{f}_t\mathbf{f}_t' - \boldsymbol{\Lambda}_t|\mathbf{X}_T, \boldsymbol{\phi}\right]\boldsymbol{\Lambda}_t^{-1}$$

where $E\left[\cdot|\mathbf{X}_T, \boldsymbol{\phi}\right]$ refers to expectations conditional on all observed $\mathbf{x}_t's$ and the parameter values $\boldsymbol{\phi}$. Therefore, we can interpret the score of the log-likelihood function for $\mathbf{x}_t$ as the expected value given $\mathbf{X}_T$ of the sum of the (unobservable) scores corresponding to the conditional log-likelihood function of $\mathbf{x}_t$ given $\mathbf{f}_t$, and the marginal log-likelihood function of $\mathbf{f}_t$ (cf. Demos and Sentana (1996b)). Note that these expressions only involve $\mathbf{f}_{t|T} = E\left[\mathbf{f}_t|\mathbf{X}_T, \boldsymbol{\phi}\right] = \mathbf{f}_{t|t}$ and $\boldsymbol{\Lambda}_{t|T} = E\left[\mathbf{f}_t\mathbf{f}_t'|\mathbf{X}_T, \boldsymbol{\phi}\right] = \boldsymbol{\Lambda}_{t|t}$.

As a simple yet important example, consider the following ARCH(1)-type conditional variance specification

$$\lambda_{jj,t} = (1 - \alpha_{j1}) + \alpha_{j1}(f^2_{jt-1|t-1} + \lambda_{jj,t-1|t-1})$$

so that $\boldsymbol{\psi}' = (\alpha_{11}, \alpha_{21}, ..., \alpha_{k1})$. If the true parameter configuration corresponds to the case of conditional homoskedasticity, i.e. $\boldsymbol{\psi}_0 = \mathbf{0}$, so that $\boldsymbol{\Lambda}_t = \mathbf{I}, \forall t$, then $\partial\lambda_{jj,t}(\boldsymbol{\phi}_0)/\partial\mathbf{c} = \mathbf{0}$, $\partial\lambda_{jj,t}(\boldsymbol{\phi}_0)/\partial\boldsymbol{\gamma} = \mathbf{0}$ and $\partial\lambda_{jj,t}(\boldsymbol{\phi}_0)/\partial\alpha_{j1} = (f^2_{jt-1|t-1} + \lambda_{jj,t-1|t-1} - 1)$. Since $\lambda_{jj,t} = 1$ under the null, and

$$E\left[f^2_{jt|t} + \lambda_{jj,t|t} - 1\right] = E\left[E\left(f^2_{jt} - 1|\mathbf{X}_t\right)\right] = E\left[f^2_{jt} - 1\right] = 0$$

then the orthogonality conditions implicit in the last k elements of the score are simply $cov(f^2_{jt|t}, f^2_{jt-1|t-1}) = 0$.

Let $\mathbf{H}_t(\boldsymbol{\phi}) = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ denote Hessian matrix of $l_t(\boldsymbol{\phi})$. Bollerslev and Wooldridge (1992) also prove that

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

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

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

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

where we use the fact that the Hadamard (or element by element) product of two $m \times n$ matrices, $\mathbf{R}$ and $\mathbf{S}$, can be written as $\mathbf{R} \odot \mathbf{S} = \mathbf{E}'_m(\mathbf{R} \otimes \mathbf{S})\mathbf{E}_n$ (see Magnus (1988)).

Since $E\left[\partial\lambda_{jj,t}(\boldsymbol{\phi}_0)/\partial\alpha_{j1}\right] = E\left[f^2_{jt-1|t-1} + \lambda_{jj,t-1|t-1} - 1\right] = 0$, it is clear that the information matrix is block diagonal between static and dynamic variance parameters under the null of conditional homoskedasticity.

Finally, it is also worth noting that under conditional homoskedasticity

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

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

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

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	
		ML	2S	ML	2S	ML	2S	ML	2S
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
α_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

Table 2: One Factor Model
Proportion of estimates at the boundary of the parameter space

	γ_0=0.5				γ_0=2.0			
	$\alpha = 0, \beta = 0$		$\alpha \neq 0, \beta = 0$		$\alpha = 0, \beta = 0$		$\alpha \neq 0, \beta = 0$	
	ML	2S	ML	2S	ML	2S	ML	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	
		ML	2S	ML	2S	ML	2S	ML	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

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

		$\alpha_0 = 0.2$ $\beta_0 = 0.6$						$\alpha_0 = 0.4$ $\beta_0 = 0.4$					
		$\gamma_0 = 0.5$			$\gamma_0 = 2.0$			$\gamma_0 = 0.5$			$\gamma_0 = 2.0$		
		ML	R	2S	ML	R	2S	ML	R	2S	ML	R	2S
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

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

		$\alpha_0 = 0.2$ $\beta_0 = 0.6$						$\alpha_0 = 0.4$ $\beta_0 = 0.4$					
		$\gamma_0 = 0.5$			$\gamma_0 = 2.0$			$\gamma_0 = 0.5$			$\gamma_0 = 2.0$		
		ML	R	2S	ML	R	2S	ML	R	2S	ML	R	2S
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

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					
	$\alpha=0,\beta=0$			$\alpha\neq 0,\beta=0$			$\alpha=0,\beta=0$			$\alpha\neq 0,\beta=0$		
	ML	R	2S	ML	R	2S	ML	R	2S	ML	R	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					
	$\alpha=0,\beta=0$			$\alpha\neq 0,\beta=0$			$\alpha=0,\beta=0$			$\alpha\neq 0,\beta=0$		
	ML	R	2S	ML	R	2S	ML	R	2S	ML	R	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		
		ML	R	2S	ML	R	2S	ML	R	2S	ML	R	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		
		ML	R	2S	ML	R	2S	ML	R	2S	ML	R	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

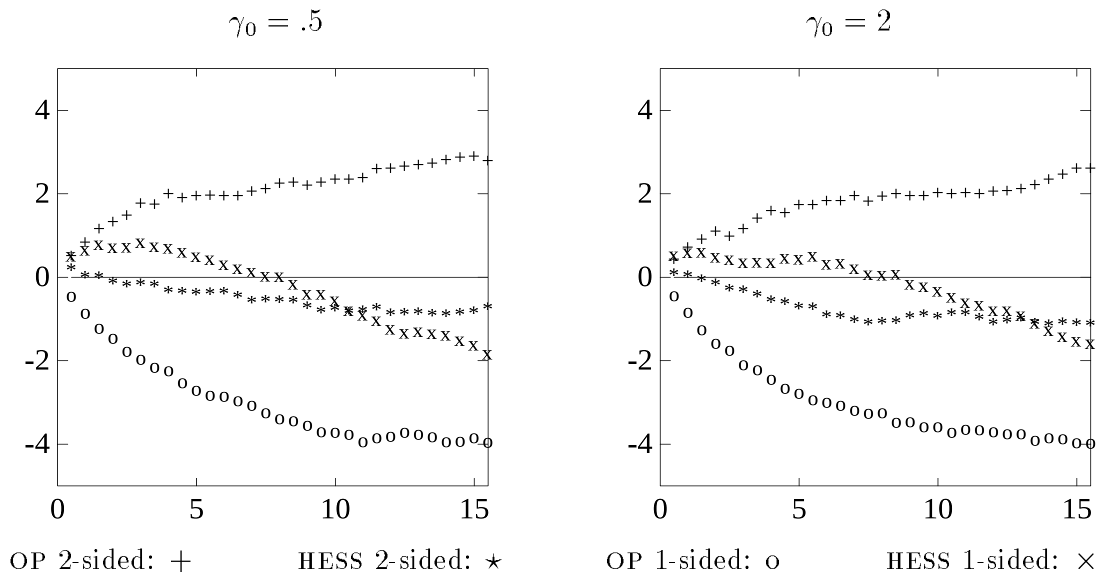


Figure 2: Test for ARCH in common factor
Size–Power curves

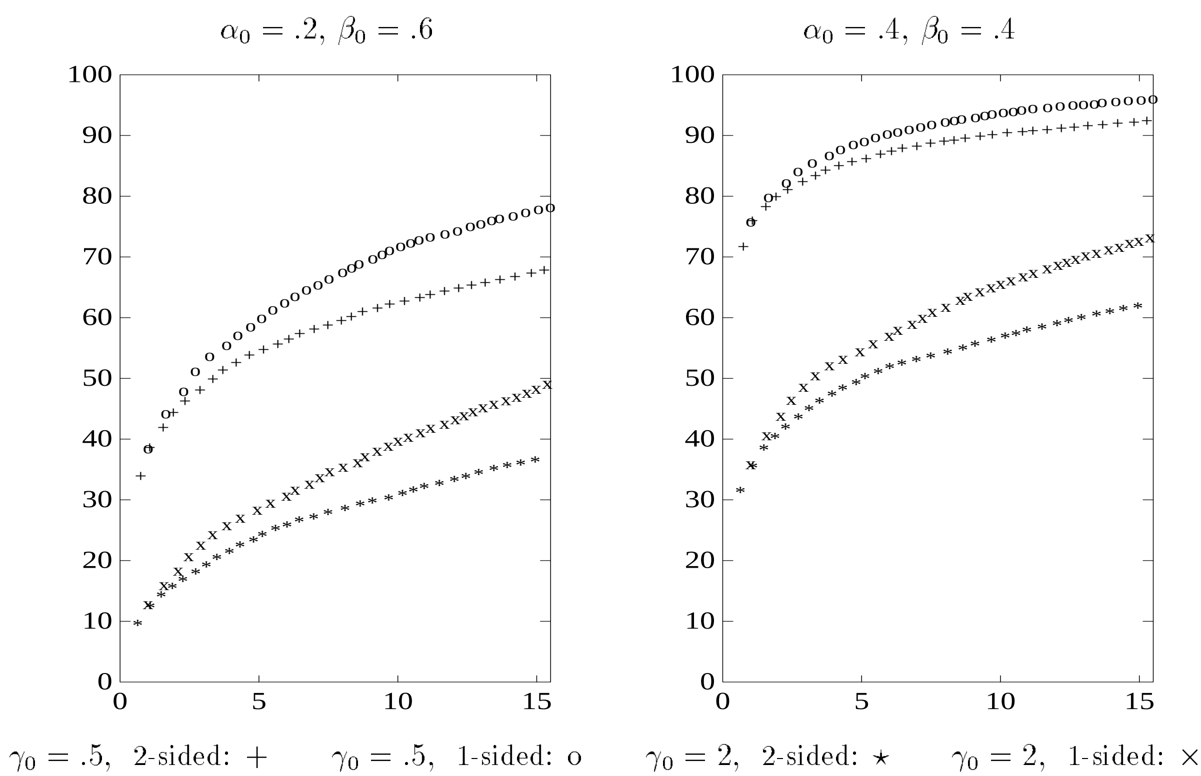
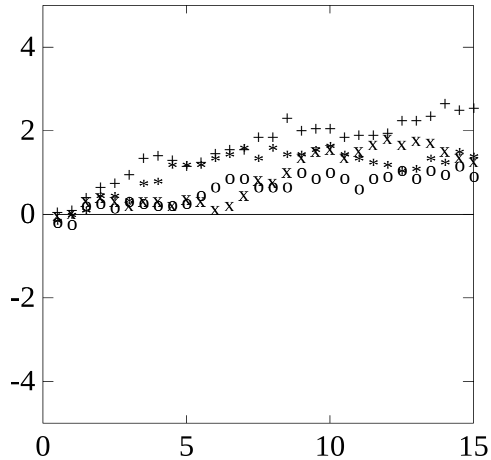


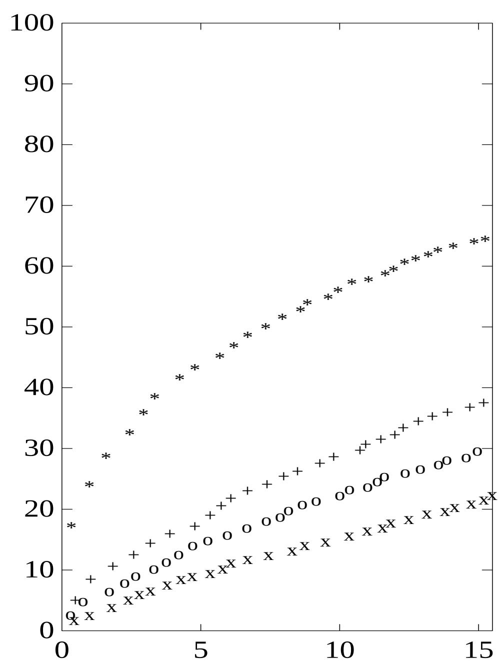
Figure 3: Likelihood Ratio Test for overidentifying restriction
P-value discrepancy plots



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



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

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