Quadratic M-estimators for ARCH-Type Processes*

Nour Meddahi† and Eric Renault‡

First Draft: January 1995
This Draft: December 1997

*This is a revised version of Meddahi-Renault (1995) "Linear Statistical Inference for ARCH-Type Processes". We would like to thank for useful comments and discussions Ali Alami, Bruno Biais, Marine Carrasco, Ramdan Dridi, Jean-Marie Dufour, Jean-Pierre Florens, Eric Ghysels, Christian Gouriéroux, Lars Hansen, Paul Johnson, G.S. Maddala, Huston McCulloch, Theo Nijman, Bas Werker and in particular René Garcia, as well as seminar participants at Université de Toulouse, the VII World Congress of the Econometric Society, August 1995, Tokyo, Université de Montréal and Ohio State University. All errors remain ours. The first author acknowledges the financial support of the Fonds FCAR du Québec.

†Département de sciences économiques, Université de Montréal, CRDE, CIRANO and CEPR. Address: C.P. 6128, succursale Centre-ville, Montréal (Québec), H3C 3J7, Canada. E-mail: nour.meddahi@umontreal.ca

‡CREST-INSEE, 15 Bd Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: renault@ensae.fr

1 Introduction

Since the introduction of the ARCH (Autoregressive Conditional Heteroskedasticity), the GARCH (Generalized ARCH) and EGARCH (Exponential GARCH) models by Engle (1982), Bollerslev (1986) and Nelson (1991) respectively, there has been widespread interest in semi-parametric dynamic models that jointly parameterize the conditional mean and conditional variance of financial series.[1] The trade-off between predictable returns (conditional[2] mean) and risk (conditional variance) of asset returns in financial time series appears as an essential motivation for the study of these models. However, in most financial series, there are strong evidence that the conditional probability distribution of returns has asymmetries and heavy tails compared to the gaussian distribution.

This becomes all the more an issue when one realizes that GARCH regression models are usually estimated and test statistics computed based on the Quasi-Maximum Likelihood Estimator (QMLE) under the **nominal** assumption of a conditional normal log-likelihood. It is well known that this QMLE [3] is consistent in the general framework of a dynamic model under correct specification of both the conditional mean and the conditional variance.[4] Bollerslev and Wooldridge (1992) focus on the QMLE due to its simplicity, but they make the three following points: first, rather than employing QMLE, it is straightforward to construct GMM estimators; second, the results of Chamberlain (1982), Hansen (1982), White (1982b) and Cragg (1983) can be extended to produce an instrumental variables estimator asymptotically more efficient than the QMLE under non-normality; third, under enough regularity conditions, it is almost certainly possible to obtain an estimator with a variance that achieves the semiparametric lower bound (Chamberlain (1987)).

The main reason why QMLE is credited of simplicity is the regression-type interpretation of associated inference procedures allowed by the nominal normality assumption. More precisely, it is usual to interpret QML estimation and procedures of tests through the estimators and associated diagnostic tools of two regression equations: one for the conditional mean and the other one for the conditional variance. We propose here to systematize this argument and to develop a general inference theory through these two regression equations that takes into account skewness (the third moment) and kurtosis (the fourth moment). The intuition is as follows: on the one hand, since we consider a regression of the variance, we need, in order to increase the efficiency, the variance of the variance, namely the kurtosis; on the other hand, we have to perform the two regressions **jointly**. Hence, we need for the efficiency reasons to consider the covariance between the two regressions, that is the covariance between the mean and the variance, namely the skewness.

[1]See Bollerslev-Chou-Kroner (1992) and Bollerslev-Engle-Nelson (1994) for a review.

[2]The precise conditioning information is defined in the sequel.

[3]See White (1982-a, 1994), Gouriéroux-Monfort-Trognon (1984), Gouriéroux-Monfort (1993) for the consistency of the QMLE under the nominal assumption of an exponential distribution and see Broze-Gouriéroux (1995) and Newey-Steigerwald (1997) for a general QMLE theory. See also the recent book by Heyde (1997) and the surveys by Newey-McFadden (1994) and Wooldridge (1994).

[4]See Weiss (1986) for consistency of the QMLE for ARCH models, Bollerslev-Wooldridge (1992) for GARCH ones, Lee-Hansen (1994) and Lumsdaine for the IGARCH of Nelson (1990).

In this paper, we focus on the efficient estimation[5] in the case of regression equations defined by **conditional expectations** (for the first and the second moments, at least) without giving up the simplicity of the QMLE.[6] The paper has three main results.

First, we consider a general quadratic class of M-estimators (Huber (1967)) and characterize the optimal quadratic estimator which involves the conditional skewness and the conditional kurtosis. We show that the standard QMLE is asymptotically equivalent to a specific quadratic estimator which is in general suboptimal. However, the optimal quadratic estimator can be interpreted as a bivariate QMLE, with respect to the vector (y, y^2) instead of y alone.

Secondly, we state a general equivalence result between (quadratic) M-estimation and GMM (Hansen (1982)) which holds for any set of conditional moment restrictions given an information set I_{t-1}

$$E[f(y_t, \theta) \mid I_{t-1}] = 0, \qquad \theta \in \Theta \subset I\!\!R^p$$

as soon as

$$\frac{\partial f}{\partial \theta'}(y_t, \theta) \in I_{t-1}, \ \forall \theta \in \Theta,$$

that is a regression type model.

In the framework of GARCH models, this result implies that the optimal quadratic M-estimator is asymptotically equivalent to the efficient GMM (with optimal instruments), even though the class of quadratic M-estimators is generally strictly included in the GMM class. In other words, the semiparametric efficiency bound (see Chamberlain (1987)) may be reached by a quadratic estimator which features the same simplicity advantage as the QMLE. As far as inference is concerned in models defined by conditional moment restrictions, one can rely on robust QMLE inference as developed in Wooldridge (1990, 1991a-b). Of course, the QMLE paradigm applies in this case in a multivariate version, involving (y, y^2) since conditional heteroskedasticity is to be accounted for.[7]

The GMM point of view stresses the informational paradox. Efficient semiparametric estimators generally use, for feasibility, some additional information which should have been incorporated in the set of conditional moment restrictions involved in efficient GMM. This pitfall is not new (see for instance Bates and White (1990)). However, with respect to the initial set of moment restrictions, the efficient semiparametric estimator reaches the semiparametric efficiency bound (see e.g. Chamberlain (1987)).

Thirdly, our estimating procedure offers the advantage of taking into account non-gaussian skewness and kurtosis. In general the conditional skewness and the conditional kurtosis are not specified, except in the so-called semiparametric GARCH models introduced by Engle and González-Rivera (1991).[8] In this framework, the standardized residuals are i.i.d which implies

[5]Testing tools are developed in Alami-Meddahi-Renault (1998).

[6]The previous version of this paper, Meddahi and Renault (1995) stresses that, even if the regression equations are defined by linear projections (in the spirit of Drost and Nijman (1993) weak GARCH) instead of conditional expectations, regression-based quadratic M-estimators may also be still consistent. See also Franck and Zakoian (1997).

[7]For higher moments equations, conditional skewness or conditional kurtosis for example, QMLE will be defined in terms of (y, y^2, y^3) and (y, y^2, y^3, y^4).

[8]For the asymptotic properties of the semiparametric GARCH models, see Linton (1993) and Drost-Klassen (1997).

that the conditional skewness and kurtosis are constant. Hence, they coincide with the unconditional skewness and kurtosis, which can be estimated. Thus, our estimation procedure is less demanding than the nonparametric one of Engle and González-Rivera (1991). Indeed, this procedure can be applied in a more general setting than the semiparametric one, in particular when we are able to consider a sufficiently narrow information set I_{t-1} to ensure that conditional skewness and kurtosis are constant. The narrowest information set that one is allowed to consider is the σ-field I^*_{t-1} spanned by the family of measurable functions $m_t(\theta)$ and $h_t(\theta)$, indexed by $\theta \in \Theta$ which represent respectively the conditional mean and the conditional variance functions of interest. We stress this point not only to show that there are many cases where we are able to reach the efficiency bound by using only parametric techniques but also to notice that nonparametric tools can often be used as soon as the σ-field I^*_{t-1} is spanned by a finite set of random variables.

The paper is organized as follows. We first build our class of quadratic M-estimators in section 2. In this class, we show that a particular estimator is asymptotically equivalent to the QMLE. Then, we exhibit an estimator with minimum asymptotic covariance matrix in this class by a Gauss-Markov type argument. This optimal instrument takes into account the conditional skewness and the conditional kurtosis. Section 3 reconsiders the same issue through the GMM approach. The links between GMM, QMLE and M-estimation are clearly established. Finally, in section 4 we address several issues related to the feasibility and the empirical relevance of our general approach. In particular, we consider in detail the semiparametric GARCH models through a Monte Carlo study and we describe several circumstances where our methodology remains friendly even though the assumptions of semiparametric GARCH are dramatically weakened. We conclude in section 5.

2 Efficiency bound for M-estimators

In this section, we first introduce the set of dynamic models of interest. Since these models are specified by their conditional mean and their conditional variance, that is by two regression equations, it is natural to consider least-squares based estimation procedures. Therefore we introduce a large quadratic class of "generalized" M-estimators. We further characterize an efficiency bound for this class of estimators following the Bates and White (1993) concept of determination of estimators with minimum asymptotic covariance matrices.

2.1 Notation and setup[9]

Let $(y_t, z_t), t = 1, 2, .., T$ be a sequence of observable random variables with y_t a scalar and z_t of dimension K. The variable y_t is the endogenous variable of interest which has to be explained in terms of K explanatory variables z_t and past values of y_t and z_t[10]. Thus, let $I_{t-1} = (z'_t, y_{t-1}, z'_{t-1}, ..., z'_1, y_1)'$ denote the information provided by the predetermined variables, which will be called the information available at time (t-1) in the rest of the paper. We consider here the joint inference about $E(y_t \mid I_{t-1})$ and $Var(y_t \mid I_{t-1})$. These conditional mean and

[9]This first subsection is to a large extent borrowed from Wooldridge (1991).

[10]Many concepts and results of the paper could be extended easily to a multivariate vector y_t of endogenous variables. These extensions are omitted here for the sake of notational simplicity.

variance functions are jointly parameterized by a vector θ of size p:

Assumption 1: For some $\theta^o \in \Theta \subset I\!R^p$, $E(y_t \mid I_{t-1}) = m_t(\theta^0)$ and $Var(y_t \mid I_{t-1}) = h_t(\theta^0)$.

Assumption 1 provides a regression model of order 2 for which usual identifiability conditions are assumed.

Assumption 2: For every $\theta \in \Theta$, $m_t(\theta) \in I_{t-1}$, $h_t(\theta) \in I_{t-1}$ and $\left.\begin{array}{l} m_t(\theta) = m_t(\theta^0) \\ h_t(\theta) = h_t(\theta^0) \end{array}\right\} \Rightarrow \theta = \theta^0$

Typically, we have in mind GARCH-regression models where $\theta = (\alpha', \beta')'$ and $m_t(\theta)$ depends only on α ($m_t(\theta) = m_t(\alpha)$ with a slight change in notations) and $h_t(\theta)$ depends on α only through past mean values $m_\tau(\alpha), \tau < t$.

In this setting, Assumption 2 is generally replaced by a slightly stronger one:

Assumption 2'a: $\Theta = \mathcal{A} \times \mathcal{B}$, $\theta^0 = (\alpha^{0'}, \beta^{0'})'$.

For every $\alpha \in \mathcal{A}$, $\quad m_t(\alpha) = m_t(\alpha^0) \Rightarrow \alpha = \alpha^0$.

For every $\beta \in \mathcal{B}$, $\quad h_t(\alpha^0, \beta) = h_t(\alpha^0, \beta^0) \Rightarrow \beta = \beta^0$.

A local version of Assumption 2'a which is usual for least-squares based estimators of α and β is:

Assumption 2'b: $E\dfrac{\partial m_t}{\partial \alpha}(\alpha^0)\dfrac{\partial m_t}{\partial \alpha'}(\alpha^0)$ and $E\dfrac{\partial h_t}{\partial \beta}(\theta^0)\dfrac{\partial h_t}{\partial \beta'}(\theta^0)$ are positive definite.

However, the only maintained assumptions hereafter will be Assumptions 1 and 2 since the additional restrictions which characterize Assumption 2' with respect to Assumption 2 may be binding for at least two reasons. First, they exclude ARCH-M type models (Engle-Lilien-Robins (1987)), where the whole conditional variance $h_t(\theta)$ should appear in the conditional mean function $m_t(\theta)$. Second,they exclude some unidentifiable representations of GARCH type models. Let us consider for instance a GARCH-regression model which, for a given value α^0 and $\varepsilon_t = y_t - m_t(\alpha^0)$, is characterized by a GARCH(p,q) representation of ε_t^2:

$$h_t(\theta^0) = \beta_0^0 + \sum_{i=1}^{q} \beta_i^0 \varepsilon_{t-i}^2(\alpha^0) + \sum_{j=1}^{p} \beta_{q+j}^0 h_{t-j}(\theta^0) \tag{2.1}$$

or equivalently, by the following ARMA (Max(p,q),p) model for ε_t^2:

$$\varepsilon_t^2(\alpha^0) - \sum_{i=1}^{q} \beta_i^0 \varepsilon_{t-i}^2(\alpha^0) - \sum_{j=1}^{p} \beta_{q+j}^0 \varepsilon_{t-j}^2(\theta^0) = \beta_0^0 + \nu_t - \sum_{j=1}^{p} \beta_{q+j}^0 \nu_{t-j} \tag{2.2}$$

where $\nu_t = \varepsilon_t^2 - h_t(\theta)$. Therefore, the vector of parameters $\beta^0 = (\beta_i^0)_{0 \le i \le p+q}$ is identifiable (in the sense of Assumption 2'a) if and only if the ARMA representation (2.2) is minimal in the sense that there is no common factor involved in both the AR and the MA lag polynomials[11]. This excludes for instance the case: $\beta_i^0 = 0 \ \forall \ i = 1, ..p$ with nonzero β_{q+j}^0 for some $j = 1, .., q$. In other words, GARCH(p,0) models, $p = 1, 2, ...$ are excluded by Assumption 2'.

[11]Of course, the positivity requirement for the conditional variance $h_t(\beta^0)$ defined by (2.1) implies some inequality restrictions on β^0 (see Nelson and Cao (1992)) but they do not modify the identification issue as presented here.

A benchmark estimator for θ^0 is the Quasi-Maximum Likelihood Estimator (QMLE) under the nominal assumption that y_t given I_{t-1} is normally distributed. For observation t, the quasi-conditional log-likelihood apart from a constant is:

$$l_t(y_t \mid I_{t-1}, \theta) = -\frac{1}{2} log h_t(\theta) - \frac{1}{2h_t(\theta)}(y_t - m_t(\theta))^2 \tag{2.3}$$

The QMLE $\hat{\theta}_T^Q$ is obtained by maximizing the normal quasi-log-likelihood function $L_T(\theta) = \sum_{t=1}^T l_t(\theta)$. The consistency and asymptotic probability distribution of $\hat{\theta}_T^Q$ have been extensively studied by Bollerslev and Wooldridge (1992). In the framework of their assumptions, we know that the asymptotic covariance matrix of $\sqrt{T}(\hat{\theta}_T^Q - \theta^0)$ is $A^{0^{-1}}\ B^0\ A^{0^{-1}}$, which is consistently estimated by $A_T^{0^{-1}}\ B_T^0\ A_T^{0^{-1}}$ where:

$$A_T^0 = -\frac{1}{T}\sum_{t=1}^T E[\frac{\partial s_t}{\partial \theta'}(\theta^0)], \quad B_T^0 = \frac{1}{T}\sum_{t=1}^T E[s_t(\theta^0) s_t(\theta^0)'], \text{ where } \ s_t(\theta) = \frac{\partial l_t}{\partial \theta}(y_t \mid I_{t-1}, \theta).$$

More precisely, differentiation of (2.3) yields the p x 1 score function:

$$\begin{aligned} s_t(\theta) &= -\frac{1}{2h_t(\theta)}\frac{\partial h_t}{\partial \theta}(\theta) + \frac{1}{2h_t^2(\theta)}(y_t - m_t(\theta))^2 \frac{\partial h_t}{\partial \theta}(\theta) + \frac{1}{h_t(\theta)}(y_t - m_t(\theta))\frac{\partial m_t}{\partial \theta}(\theta) \\ &= \frac{1}{h_t(\theta)}\frac{\partial m_t}{\partial \theta}(\theta)\varepsilon_t(\theta) + \frac{1}{2h_t^2(\theta)}\frac{\partial h_t}{\partial \theta}(\theta)\nu_t(\theta) \end{aligned} \tag{2.4}$$

where:

$$\varepsilon_t(\theta) = y_t - m_t(\theta), \tag{2.5.a}$$

$$\nu_t(\theta) = \varepsilon_t(\theta)^2 - h_t(\theta). \tag{2.5.b}$$

Note that by Assumption 1, $\varepsilon_t(\theta^0)$ and $\nu_t(\theta^0)$ are martingale difference sequences with respect to the filtration I_{t-1}. This allows Bollerslev and Wooldridge (1992) to apply a martingale central limit theorem for the proof of asymptotic normality of the QMLE.

Since we are concerned by "**quadratic** statistical inference", the form of the score function (2.4) in relation with error terms $\varepsilon_t(\theta)$ and $\nu_t(\theta)$ of "regression models" (2.5.a) and (2.5.b) suggests a quadratic interpretation of the QMLE. More precisely, we consider a modified score function:

$$\tilde{s}_t(\theta) = \frac{1}{h_t(\theta^0)}\frac{\partial m_t}{\partial \theta}(\theta)\varepsilon_t(\theta) + \frac{1}{2h_t^2(\theta^0)}\frac{\partial h_t}{\partial \theta}(\theta)(\varepsilon_t(\theta^0)^2 - h_t(\theta)), \tag{2.6}$$

which is the negative of the gradient vector with respect to θ of the quadratic form:

$$\frac{\varepsilon_t^2(\theta)}{2h_t(\theta^0)} + \frac{(\varepsilon_t^2(\theta^0) - h_t(\theta))^2}{4h_t^2(\theta^0)}. \tag{2.7}$$

The idea to base our search for linear procedures of inference on this quadratic form appears natural since (see Appendix A1):

$$\tilde{s}_t(\theta^0) = s_t(\theta^0) \text{ and } E[\frac{\partial \tilde{s}_t}{\partial \theta'}(\theta^0)] = E[\frac{\partial s_t}{\partial \theta'}(\theta^0)] \tag{2.8}$$

so that the replacement of s by $\tilde{s}$ does not modify the matrices A_T and B_T that characterize the asymptotic probability distribution of the "estimator" obtained by solving the first-order conditions: $\sum_{t=1}^{T} s_t(\theta) = 0$. Therefore, we may hope to build, through this modified score function, a regression-based estimator asymptotically equivalent to the QMLE. We are going to introduce such an estimator in the following subsection as a particular element of a large class of quadratic generalized M-estimators.

2.2 A quadratic class of generalized M-estimators

As usual, a regression-based estimation of GARCH-type regression models raises two main difficulties. First, we have to take into account simultaneously the two dynamic regressions:

$$y_t = m_t(\theta) + \varepsilon_t(\theta), \quad E[\varepsilon_t(\theta^0) \mid I_{t-1}] = 0, \tag{2.9.a}$$

$$\varepsilon_t^2(\theta) = h_t(\theta) + \nu_t(\theta), \quad E[\nu_t(\theta^0) \mid I_{t-1}] = 0. \tag{2.9.b}$$

Second, the dependent variable of regression equation (2.9.b) depends on the unknown parameter θ so that we must have at our disposal a first stage consistent estimator $\tilde{\theta_T}$ of θ^0. However, such an estimator is generally easy to obtain. For instance, in the framework of Assumption 2'a, $\tilde{\theta}_T = (\tilde{\alpha}'_T, \tilde{\beta}'_T)'$ where we can choose in a first stage $\tilde{\alpha}_T$ as a (non linear) least squares estimator of α^0 in the regression equation (2.9.a):

$$\tilde{\alpha_T} = Arg\, \underset{\alpha}{Min} \sum_{t=1}^{T} (y_t - m_t(\alpha))^2 \tag{2.10.a}$$

and, in a second stage, $\tilde{\beta}_T$ as a (non linear) least squares estimator of β^0 in the regression equation (2.9.b) after replacement of α^0 by $\tilde{\alpha}_T$:

$$\tilde{\beta}_T = Arg\, \underset{\beta}{Min} \sum_{t=1}^{T} (\varepsilon_t(\tilde{\alpha}_T)^2 - h_t(\tilde{\alpha}_T, \beta))^2 \tag{2.10.b}$$

After obtaining such a preliminary consistent estimation $\tilde{\theta_T}$ of θ^0, it is then natural to try to improve it by considering more general weighting schemes of the two regression equations, that is to say general M-estimators of the type:

$$\hat{\theta}_T(\tilde{\theta}_T, \gamma_T) = Arg\, \underset{\theta}{Min} \sum_{t=1}^{T} q_t(\theta, \tilde{\theta}_T, \gamma_T) \tag{2.11.a}$$

where $\Lambda_{t,T}$ is a symmetric positive matrix, $\gamma_T = (\Lambda_{t,T})_{T \geq t \geq 1}$ and:

$$q_t(\theta, \tilde{\theta}_T, \gamma_T) = \frac{1}{2}(\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))\Lambda_{t,T}(\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))', \tag{2.11.b}$$

Indeed, since we have only parametric methodologies in mind[12], we shall always consider weighting matrices $\Lambda_{t,T}$ of the following form: $\Lambda_{t,T} = \Lambda_t(\omega_T)$, where ω_t is I_t-measurable and $\Lambda_t(\omega)$ is a symmetric positive matrix for every ω in a parametric space $\mathcal{V} \subset I\!R^n$. To derive weak consistency of the resulting estimator $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma)$, $\gamma = (\Lambda_t)_{t \geq 1}$ (with a slight change of notation) we shall maintain the following assumption (see Wooldridge (1994) for notations and terminology):

[12]However, many results of this paper could be extended to the case of nonparametric consistent estimator $\Lambda_{t,T}$ of weighting matrices Λ_t. See Linton (1994) for a review of this type of approach.

Assumption 3: Let $\mathcal{V} \subset I\!R^n$, let Λ_t be a sequence of random matricial functions defined on $\mathcal{V}$. For every $\omega \in \mathcal{V}$, $\Lambda_t(\omega)$ is a symmetric 2 x 2 matrix. We assume that:

(A.3.1) Θ and $\mathcal{V}$ are compact.

(A.3.2) $\tilde{\theta}_T \xrightarrow{P} \theta^0 \quad \in \Theta$ and $\omega_T \xrightarrow{P} \omega^* \quad \in \mathcal{V}$.

(A.3.3) m_t, h_t and Λ_t satisfy the standard measurability and continuity requirements. In particular $m_t(\theta)$, $h_t(\theta)$ and $\Lambda_t(\omega)$ are I_{t-1} measurable for every $(\theta, \omega) \in \Theta$ x $\mathcal{V}$.

(A.3.4) $q_t^{\gamma}(\theta, \tilde{\theta}_T, \omega) = \frac{1}{2}(\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))\Lambda_t(\omega)(\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))'$

satisfies the Uniform Weak Law of Large Numbers (UWLLN) on Θ x Θ x $\mathcal{V}$.

(A.3.5) $\Lambda_t(\omega^*)$ is positive definite.

We are then able (see Appendix B) to derive the consistency result based on the usual analogy principle argument.

Proposition 2.1 *Under Assumptions 1, 2, 3, the estimator* $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma)$ *defined by:*

$$\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma) = \underset{\theta \in \Theta}{Arg\ Min} \sum_{t=1}^{T} (\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))\Lambda_t(\omega_T)(\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))' \tag{2.12}$$

where $\gamma = (\Lambda_t)_{t \geq 1}$, *is weakly consistent towards* θ^0.

Note that the quadratic M-estimator that we have suggested in the previous subsection (see the objective function (2.7)) by analogy with the QMLE belongs to the general class considered here when $\omega^* = \theta^0$ and

$$\Lambda_t(\theta) = \begin{bmatrix} \dfrac{1}{h_t(\theta)} & 0 \\ 0 & \dfrac{1}{2h_t^2(\theta)} \end{bmatrix} \tag{2.13}$$

By extending to a dynamic setting the quadratic principle of estimation first introduced by Crowder (1987) for transversal data, we may be led to consider more general weighting matrices. Indeed, we may guess that the weighting matrix (2.13) is optimal in the gaussian case where, by the well-known kurtosis characterization of the gaussian probability distribution:

$$h_t(\theta^0) = Var[\varepsilon_t(\theta^0) \mid I_{t-1}] \Longrightarrow 2h_t^2(\theta^0) = Var[\varepsilon_t^2(\theta^0) \mid I_{t-1}] = Var[\nu_t(\theta^0) \mid I_{t-1}].$$

On the other hand, a leptokurtic conditional probability distribution function (which is a widespread finding for financial time series) may lead to a different weight of $2h_t^2(\theta^0)$ for $\nu_t^2(\theta^0)$ while skewness may lead to a non-diagonal weighting matrix Λ_t. Of course, the relevant criterion for the choice of a sequence $\gamma = (\Lambda_t)_{t \geq 1}$ of weighting matrices is the asymptotic covariance matrix of the corresponding estimator $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma)$.

As far as the asymptotic probability distribution is concerned, the following assumptions are usual (see for instance Bollerslev and Wooldridge (1992)).

Assumption 4: In the framework of Assumption 3, we assume that:

(A.4.1.) $\theta^0 \in int\Theta$, $\omega^* \in int\mathcal{V}$, interiors of the corresponding parameter spaces Θ and

$\mathcal{V}$, and $\sqrt{T}(\tilde{\theta}_T - \theta^0) = Op(1)$ $\sqrt{T}(\omega_T - \omega^*) = Op(1)$.

(A.4.2.) $m_t(.)$ and $h_t(.)$ are twice continuously differentiable on int Θ for all I_{t-1}.

(A.4.3.) Denote by: $s_t^\gamma(\theta,\lambda,\omega) = \dfrac{\partial q_t^\gamma}{\partial \theta}(\theta,\lambda,\omega)$, which is assumed squared-integrable, and $[s_t^\gamma(\theta,\lambda,\omega)(s_t^\gamma(\theta,\lambda,\omega))']$ satisfies the UWLLN on $\Theta \times \Theta \times \mathcal{V}$ with:

$B_\gamma^0 = \lim_{T\to\infty} \dfrac{1}{T}\sum_{t=1}^{T} E[s_t^\gamma(\theta^0,\theta^0,\omega^*)s_t^\gamma(\theta^0,\theta^0,\omega^*)']$ is positive definite; $s_t^\gamma(\theta^0,\theta^0,\omega^*)$ satisfies the central limit theorem: $\dfrac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t^\gamma(\theta^0,\theta^0,\omega^*) \xrightarrow{d} \mathcal{N}[0, B_\gamma^0]$.

(A.4.4.) $\dfrac{\partial s_t^\gamma}{\partial \theta'}(\theta,\lambda,\omega)$ and $\dfrac{\partial s_t^\gamma}{\partial \lambda'}(\theta,\lambda,\omega)$ satisfy the UWLLN on $\Theta \times \Theta \times \mathcal{V}$ with

$A_\gamma^0 = \lim_{T\to\infty} \dfrac{1}{T}\sum_{t=1}^{T} E\left[\dfrac{\partial s_t^\gamma}{\partial \theta'}(\theta^0,\theta^0,\omega^*)\right]$ positive definite.

Note that:

$$q_t^\gamma(\theta,\lambda,\omega) = \frac{1}{2}[\varepsilon_t(\theta), \varepsilon_t^2(\lambda) - h_t(\theta)]\Lambda_t(\omega)\begin{bmatrix} \varepsilon_t(\theta) \\ \varepsilon_t^2(\lambda) - h_t(\theta) \end{bmatrix},$$

$$s_t^\gamma(\theta,\lambda,\omega) = -\left[\frac{\partial m_t}{\partial \theta}(\theta), \frac{\partial h_t}{\partial \theta}(\theta)\right]\Lambda_t(\omega)\begin{bmatrix} \varepsilon_t(\theta) \\ \varepsilon_t^2(\lambda) - h_t(\theta) \end{bmatrix}, \text{ and}$$

$$\frac{\partial s_t^\gamma}{\partial \theta'}(\theta,\lambda,\omega) = \left[\frac{\partial m_t}{\partial \theta}(\theta), \frac{\partial h_t}{\partial \theta}(\theta)\right]\Lambda_t(\omega)\begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta) \\ \frac{\partial h_t}{\partial \theta'}(\theta) \end{bmatrix} + c_t(\theta,\lambda,\omega),$$

with $E[c_t(\theta^0,\theta^0,\omega^*) \mid I_{t-1}] = 0$. Therefore,

$$E\left[\frac{\partial s_t^\gamma}{\partial \theta'}(\theta^0,\theta^0,\omega^*)\right] = E\left[\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\Lambda_t(\omega)\begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right] \text{ and}$$

$$E\left[s_t^\gamma(\theta^0,\theta^0,\omega^*)s_t^\gamma(\theta^0,\theta^0,\omega^*)'\right] = E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\Lambda_t(\omega)\Sigma_t\Lambda_t(\omega)\begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right\},$$

where $\Sigma_t = Var\left[\begin{bmatrix} \varepsilon_t(\theta^0) \\ \nu_t(\theta^0) \end{bmatrix} \mid I_{t-1}\right]$.

Therefore, if Assumption 4 is maintained in particular for the canonical weighting matrix $\Lambda_t = Id_2$, the positive definiteness of A_0 and B_0 corresponds[13] to the following assumption:

[13]In general, the non-singularity of an expectation matrix $E[x\Lambda x']$ where x is a $p \times K$ random matrix and Λ is a $K \times K$ random symmetric positive matrix depends on Λ. But, intuitively, the non singularity of $E(xx')$ is not only necessary (for $\Lambda = Id_K$) but often sufficient.

Assumption 4': **(i)** $\Sigma_t = Var \begin{bmatrix} \varepsilon_t(\theta^0) & \\ & \mid I_{t-1} \\ \nu_t(\theta^0) & \end{bmatrix}$ is positive definite.

$$\textbf{(ii)}\ E\left[\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right] \text{ is positive definite.}$$

The first item of Assumption 4' is indeed very natural since we are interested in the asymptotic probability distribution of least squares based estimators of θ from the two dynamic regression equations (2.9.a) and (2.9.b). If the error terms $\varepsilon_t(\theta^0)$ and $\nu_t(\theta^0)$ were conditionally (given I_{t-1}) perfectly correlated, this should introduce a restriction on θ, changing dramatically the estimation issue. The second item is directly related to the statement of Assumption 2'b in the case of a GARCH-regression model $\theta = (\alpha', \beta')'$ conformable to Assumption 2'a. In this case, $\dfrac{\partial m_t}{\partial \beta} = 0$ so that:

$$E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right\} = E\left\{\begin{bmatrix} \dfrac{\partial m_t}{\partial \alpha}(\alpha^0)\dfrac{\partial m_t}{\partial \alpha'}(\alpha^0) + \dfrac{\partial h_t}{\partial \alpha}(\theta^0)\dfrac{\partial h_t}{\partial \alpha'}(\theta^0) & \dfrac{\partial h_t}{\partial \alpha}(\theta^0)\dfrac{\partial h_t}{\partial \beta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \beta}(\theta^0)\dfrac{\partial h_t}{\partial \alpha'}(\theta^0) & \dfrac{\partial h_t}{\partial \beta}(\theta^0)\dfrac{\partial h_t}{\partial \beta'}(\theta^0) \end{bmatrix}\right\}$$

is automatically positive definite when Assumption 2'b is fulfilled (see Appendix A2).

It is worth noticing however that the framework of Assumptions 3 and 4 is fairly general and does not exclude for instance ARCH-M type models (where the whole vector θ of parameters appears in the conditional expectation $m_t(\theta)$) since a first step consistent estimator $\tilde{\theta}_T$ such as $\sqrt{T}(\hat{\theta}_T - \theta^0) = Op(1)$ is always available, for instance a QMLE conformable to (2.3).

Moreover, Assumptions 3 and 4 are stated in a framework sufficiently general to allow for non-stationary *score processes*, for which $E\left[s_t^\gamma(\theta^0, \theta^0, \omega^*) s_t^\gamma(\theta^0, \theta^0, \omega^*)'\right]$ and $E\left[\dfrac{\partial s_t^\gamma}{\partial \theta'}(\theta^0, \theta^0, \omega^*)\right]$ could depend on t. This case is important since it occurs as soon as non-markovian (for instance MA) components are allowed either in the conditional mean (AR**MA** processes) or in the conditional variance (**G**ARCH processes). In any case, the following result holds:

Proposition 2.2 *Under Assumptions 1, 2, 3 and 4, the estimator $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma)$ defined by:*

$$\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma) = \underset{\theta \in \Theta}{Arg\ Min} \sum_{t=1}^{T} (\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta)) \Lambda_t(\omega_T) (\varepsilon_t(\theta), \varepsilon_t^2(\tilde{\theta}_T) - h_t(\theta))'$$

is asymptotically normal, with asymptotic covariance matrix $A_\gamma^{0^{-1}} B_\gamma^0 A_\gamma^{0^{-1}}$ where:

$$A_\gamma^0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t(\omega^*) \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right\},$$

$$B^0_\gamma = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial\theta}(\theta^0), \frac{\partial h_t}{\partial\theta}(\theta^0)\right]\Lambda_t(\omega^*)\Sigma_t(\theta^0)\Lambda_t(\omega^*)\left[\begin{array}{c}\dfrac{\partial m_t}{\partial\theta'}(\theta^0)\\ \dfrac{\partial h_t}{\partial\theta'}(\theta^0)\end{array}\right]\right\},$$

$$\Sigma_t(\theta^0) = Var\left[\left(\begin{array}{c}\varepsilon_t(\theta^0)\\ \nu_t(\theta^0)\end{array}\right)\middle| I_{t-1}\right] \textit{ and } w^* = \text{Plim} w_T.$$

We are now able to be more precise about our regression based interpretation of the QMLE:
Proposition 2.3 *If Assumptions 1, 2, 3, 4 are fulfilled for*

$$\gamma^Q = (\Lambda^Q_t)_{t\geq 1},\ \Lambda^Q_t(\omega^*) = \left[\begin{array}{cc}\dfrac{1}{h_t(\theta^0)} & 0\\ 0 & \dfrac{1}{2h^2_t(\theta^0)}\end{array}\right]$$

then $\hat\theta_T(\tilde\theta_T, \omega_T, \gamma^Q)$ *is asymptotically equivalent to the QMLE* $\hat\theta^Q_T$.

But, as announced in the introduction, Proposition 2.2 suggests the possibility to build regression based consistent estimators $\hat\theta_T(\tilde\theta_T, \omega_T, \gamma)$ which, for a convenient choice of γ and $\omega^* = \text{Plim }\omega_T$, could be (asymptotically) strictly more accurate than QMLE. This will be the main purpose of the next subsection 2.3. Let us only notice at this stage that, according to Proposition 2.2, the asymptotic accuracy of $\hat\theta_T(\tilde\theta_T, \omega_T, \gamma)$ depends on $(\tilde\theta_T, \omega_T, \gamma = (\Lambda_t)_{t\geq 1})$ only through: $\Lambda_t(\omega^*), t \geq 1$, whatever the consistent estimators $\tilde\theta_T$ and ω_T of θ^0 and ω^* may be.

2.3 Determination of estimators with minimum asymptotic covariance matrices

Our purpose in this section is to address an efficiency issue as in Bates and White (1993), that is to find an optimal estimator in the class defined by Assumptions 3 and 4. Our main result is then the following:

Theorem 2.1 *If the GARCH regression model:*

$$\left\{\begin{array}{l} y_t = m_t(\theta) + \varepsilon_t(\theta), E(\varepsilon_t(\theta^0) \mid I_{t-1}) = 0\\ \varepsilon^2_t(\theta) = h_t(\theta) + \nu_t(\theta), E(\nu_t(\theta^0) \mid I_{t-1}) = 0\end{array}\right.$$

fulfills Assumptions 1 and 2 and: $\Sigma_t(\theta^0) = Var\left[\begin{array}{c}\varepsilon_t(\theta^0)\\ \nu_t(\theta^0)\end{array}\middle| I_{t-1}\right]$ *is positive definite, a sufficient condition for an estimator of the class defined by Assumptions 3 and 4 being of minimum asymptotic covariance matrix in that class is that, for all* t *and all* I_{t-1}*:*

$$\Lambda_t(\omega^*) = \Sigma_t(\theta^0)^{-1}.$$

The corresponding asymptotic covariance matrix is $(A^0)^{-1}$ *with:*

$$A^0 = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[\left[\frac{\partial m_t}{\partial\theta}(\theta^0), \frac{\partial h_t}{\partial\theta}(\theta^0)\right]\Sigma^{-1}_t(\theta^0)\left[\begin{array}{c}\dfrac{\partial m_t}{\partial\theta'}(\theta^0)\\ \dfrac{\partial h_t}{\partial\theta'}(\theta^0)\end{array}\right]\right]$$

Of course, this theorem leaves unsolved the general issue of estimating $\Sigma_t(\theta^0)$ to get a feasible estimator in practice.[14] This issue will be addressed in more details in Section 3. At this stage we only stress the statistical interpretation of the optimal weighting matrix:

$$\Lambda_t(\omega^*) = \Sigma_t(\theta^0)^{-1} = \begin{bmatrix} h_t(\theta^0) & M_{3t}(\theta^0)h_t(\theta^0)^{\frac{3}{2}} \\ M_{3t}(\theta^0)h_t(\theta^0)^{\frac{3}{2}} & (3K_t(\theta^0)-1)h_t^2(\theta^0) \end{bmatrix}^{-1}. \tag{2.14}$$

with

$$M_{3t}(\theta^0) = E[u_t^3(\theta^0) \mid I_{t-1}], \tag{2.15.a}$$

$$K_t(\theta^0) = \frac{1}{3}E[u_t^4(\theta^0) \mid I_{t-1}]. \tag{2.15.b}$$

When one derives the first-order conditions associated with this optimal M-estimator, one obtains equations similar to some previously proposed in the literature for some particular cases: the i.i.d setting of Crowder (1987) and the stationary markovian setting of Wefelmeyer (1996). In other words, Theorem 2.1 suggests to improve the usual QMLE by taking into account non-gaussian conditional skewness and kurtosis while, by Proposition 2.3, the QMLE $\hat{\theta}_T^Q$ should be inefficient if: $M_{3t}(\theta^0) \neq 0$ or $K_t(\theta^0) \neq 1$.

Let us first consider the simplest case of symmetric innovations ($M_{3t}(\theta^0) = 0$). In this case, the role of $K_t(\theta^0)$ is to provide the optimal relative weights for the two regression equations (2.9.a) and (2.9.b). In case of asymmetry ($M_{3t}(\theta^0) \neq 0$), Theorem 2.1 stresses the importance of taking into account the conditional correlation between these two equations through a suitably weighted cross-product of the two errors. Indeed, Meddahi and Renault (1996) documents the role of this correlation as a form of leverage effect, according to Black (1976).

In order to highlight the role of conditional skewness and kurtosis to build efficient M-estimators, we shall use the following reparametrization $\gamma = (a_t, b_t, c_t)_{t\geq 1}$ of the sequence $\gamma = (\Lambda_t)_{t\geq 1}$ such as

$$\Lambda_t = 2\begin{bmatrix} \dfrac{a_t}{h_t(\theta^0)} & \dfrac{c_t}{h_t(\theta^0)^{3/2}} \\ \dfrac{c_t}{h_t(\theta^0)^{3/2}} & \dfrac{b_t}{h_t^2(\theta^0)} \end{bmatrix}. \tag{2.16}$$

In other words, the class of M-estimators defined by assumption 3 consists of the following:

$$\underset{\theta}{Arg\ Min} \sum_{t=1}^{T} a_t \frac{\varepsilon_t(\theta)^2}{h_t(\theta^0)} + b_t \frac{(\varepsilon_t(\tilde{\theta_T})^2 - h_t(\theta))^2}{h_t(\theta^0)^2} + 2c_t \frac{\varepsilon_t(\theta)(\varepsilon_t(\tilde{\theta_T})^2 - h_t(\theta))}{h_t(\theta^0)^{\frac{3}{2}}}, \tag{2.17}$$

for various choices of the weights $(a_t, b_t, c_t) \in I_{t-1}$ ensuring that Λ_t is positive definite ($a_t > 0$, $b_t > 0$ and $a_t b_t > c_t^2$). Of course, the M-estimator (2.17) is unfeasible and its practical

[14] On the other hand, when a consistent estimator $\hat{\Sigma}_{t,T}$ of $\Sigma_t(\theta^0)$ is available, Theorem 2.1 directly provides a consistent estimator of the asymptotic covariance matrix of the optimal estimator $\hat{\theta}_T$ by:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial m_t}{\partial \theta}(\hat{\theta}_T), \frac{\partial h_t}{\partial \theta}(\hat{\theta}_T)\right]\hat{\Sigma}_{t,T}^{-1}\begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\hat{\theta}_{t,T}) \\ \dfrac{\partial h_t}{\partial \theta'}(\hat{\theta}_{t,T}) \end{bmatrix}$$

.

implementation should lead to replace θ^0 by the consistent preliminary estimator $\tilde{\theta}_t$. But Theorem 3.1 above implies that the optimal choice of a_t, b_t, c_t should be:

$$a_t^* = (3K_t(\theta^0) - 1) \ \times \ b_t^*, \quad b_t^* = \frac{1}{2}\frac{1}{3K_t(\theta^0) - 1 - M_{3t}(\theta^0)^2}, \quad c_t^* = -M_{3t} \ \times \ b_t^*. \tag{2.18}$$

For feasibility we need a preliminary estimation of the optimal weights a_t^*, b_t^*, c_t^* as detailed in section 3 below. Moreover, by Proposition 2.3, we have a M-estimator asymptotically equivalent to the QMLE $\hat{\theta}_T^Q$ by choosing the following constant weights:

$$(a_t, b_t, c_t) = (\frac{1}{2}, \frac{1}{4}, 0). \tag{2.19}$$

One of the issues addressed in section 3 below is the estimation of weights (a_t^*, b_t^*, c_t^*) which allows one to improve the choice (2.19), that is to obtain a M-estimator which is more accurate than the QMLE. Indeed, it is important to keep in mind that the usual QMLE is inefficient since it does not fully take into account the information included in the two regression equations (2.9). On the other hand, if one considers these two equations as a SUR system:

$$\begin{cases} y_t = m_t(\theta) + \varepsilon_t, & E[\varepsilon_t \mid I_{t-1}] = 0, \\ \varepsilon_t^2(\theta^0) = h_t(\theta) + \nu_t, & E[\nu_t \mid I_{t-1}] = 0, \end{cases} \tag{2.20}$$

it is clear that the QMLE written from the **joint** probability distribution of $(y_t, \varepsilon_t^2(\theta^0))$ (and not only from y_t as the usual QMLE) considered as a gaussian vector with conditional variance $\Sigma_t(\theta^0)$ coincides with the optimal M-estimator characterized by Theorem 2.1, when $\varepsilon_t^2(\theta^0)$ has been replaced by a first stage estimator $\varepsilon_t^2(\tilde{\theta}_T)$. Another way to interpret such an estimator is to compute the QMLE with gaussian pseudo-likelihood from the following SUR system (equivalent to (2.20))

$$\begin{cases} y_t = m_t(\theta) + \varepsilon_t, & E[\varepsilon_t \mid I_{t-1}] = 0, \\ y_t^2 = m_t^2 + h_t(\theta) + \eta_t, & E[\eta_t \mid I_{t-1}] = 0. \end{cases} \tag{2.21}$$

Of course, both the QMLE and the optimal M-estimator (as previously defined) are unfeasible. Their practical implementation would need (see section 3) a first stage estimation of the conditional variance matrix $\Sigma_t(\theta^0)$. But we stress here that a quasi-generalized PML1 as in Gourieroux-Monfort-Trognon (1984) is optimal since it takes into account the informational content of the parametric model for the two first moments (with a parametric specification of the third and fourth ones) as soon as it is written in a **multivariate way** about (y_t, y_t^2).

3 Instrumental Variable Interpretations

3.1 An equivalence result

Let us consider the general conditional moment restrictions:

$$E[f(y_t, \theta) \mid I_{t-1}] = 0, \quad \theta \in \Theta \subset I\!R^p \tag{3.1}$$

which uniquely define the true unknown value θ^0 of the vector θ of unknown parameters. For any sequence $(\Lambda_t)_{t\geq 1}$ of positive definite matrices of size H (same size that f), one may define a M-estimator $\hat{\theta}_T$ of θ^0 as:

$$\hat{\theta}_T = \underset{\theta\in\Theta}{Arg\ Min} \sum_{t=1}^{T} f(y_t, \theta)' \Lambda_t f(y_t, \theta). \tag{3.2}$$

Under general regularity conditions, this estimator will be characterized by the first order conditions:

$$\sum_{t=1}^{T} \frac{\partial f'}{\partial \theta}(y_t, \hat{\theta}_T)\Lambda_t f(y_t, \hat{\theta}_T) = 0. \tag{3.3}$$

By a straightforward generalization of the proof of Proposition 2.1, the consistency of such an estimator is ensured by the following assumptions:

$$f(y_t, \theta) - f(y_t, \theta^0) \in I_{t-1}, \forall \theta \in \Theta \tag{3.4.a}$$

and

$$\Lambda_t \in I_{t-1.}$$

But, in such a case:

$$\frac{\partial f'}{\partial \theta}(y_t, \theta^0) \in I_{t-1}, \forall \theta \in \Theta \tag{3.4.b}$$

and the M-estimator $\hat{\theta}_T$ can be reinterpreted as the GMM estimator associated with the following unconditional moment restrictions (implied by (3.1)):

$$E[\frac{\partial f'}{\partial \theta}(y_t, \theta)\Lambda_t f(y_t, \theta)] = 0. \tag{3.5}$$

We have proved that any M-estimator of our quadratic class (by extending the terminology of previous sections) is a GMM estimator based on (3.1) and corresponding to a particular choice of instruments.[15]

Conversely, we would like to know if the efficiency bound of GMM (corresponding to optimal instruments) may be reached by M-estimators.

Three types of results are available concerning efficient GMM based on conditional moment restrictions (3.1).

i) First, it has been known since Hansen (1982) that the optimal choice of instruments is given by $D_t(\theta^0)\Sigma_t(\theta^0)^{-1}$ where:

$$D_t(\theta^0) = E[\frac{\partial f'}{\partial \theta}(y_t, \theta^0) \mid I_{t-1}] \text{ and } \Sigma_t(\theta^0) = Var[f(y_t, \theta^0) \mid I_{t-1}].$$

In other words, the GMM efficiency bound associated with (3.1) is characterized by the just identified unconditional moment restrictions:

$$E[D_t(\theta^0)\Sigma_t^{-1}(\theta^0)f(y_t, \theta)] = 0. \tag{3.6}$$

ii) In practise, we cannot use the moments conditions (3.6) since the parameter θ^0 as well as the functions $D_t(.)$ and $\Sigma_t(.)$ are unknown. θ^0 could be replaced by a first stage consistent estimator $\tilde{\theta}_T$ without modifying the asymptotic probability distribution of the resulting GMM estimator (see e.g. Wooldridge (1994)). In our case, that is a regression type model, the function $D_t(.)$ is known (assumption (3.4)): $D_t(\theta) = \frac{\partial f}{\partial \theta'}(y_t, \theta)$. Hence, the main issue is the estimation of the conditional variance $\Sigma_t(\theta^0)$. Either we have a parametric form of the conditional variance (section 4) and we can compute the optimal instrument, without however taking into account the information included in the conditional variance matrix $\Sigma_t(\theta^0)$.[16] Or

[15]Note that this result is different from the well known one where we reinterpret a score function as a moment condition.

[16]In other words, our "efficient" GMM estimation with optimal instruments (with respect to the initial set of restrictions) is only a second best one.

this conditional variance could be nonparametrically estimated at fast enough rates to obtain an asymptotically efficient GMM estimator (see e.g. Newey (1990) and Robinson (1991) for the cross-section case). But in the dynamic case, nonparametric estimation is difficult. In particular the fast enough consistency cannot generally be obtained in non Markovian settings where the dimension of conditioning information is growing with the sample size T.[17] In this latter case, as summarized by Wooldridge (1994) "*little is known about the efficiency bounds for the GMM estimator. Some work is available in the linear case; see Hansen (1985) and Hansen, Heaton and Ogaki (1988).*"[18]

In what follows we assume that the efficient GMM estimator $\hat{\theta}_T$ with optimal instruments is obtained by solving the moment conditions:

$$\frac{1}{T}\sum_{t=1}^{T} D_t(\tilde{\theta}_T)\Sigma_t^{-1}(\tilde{\theta}_T)f(y_t,\hat{\theta}_T) = 0, \tag{3.7}$$

where $\tilde{\theta}_T$ is a first-stage consistent estimator such that $\sqrt{T}(\tilde{\theta}_T - \theta^0) = O_p(1)$. This assumption will be maintained throughout all this section.

iii) In a context of homoskedastic "errors" $f(y_t,\theta^0)$, $t = 1, 2, ..T$, Rilstone (1992) noticed that an obvious alternative is the estimator that solves the moment conditions simultaneously over both the residuals and the instruments, that is the solution of θ:

$$\sum_{t=1}^{T} D_t(\theta)f(y_t,\theta) = 0. \tag{3.8}$$

Rilstone (1992) suggests to refer to $\hat{\theta}_T$ as the "two-step" and $\hat{\theta}_T^*$ (solution of (3.8)) as the "extremum" estimator.

The natural generalization to heteroskedastic errors of the extremum estimator suggested by Rilstone (1992) is now $\hat{\theta}_T^*$ defined as solution of the following system of equations:

$$\frac{1}{T}\sum_{t=1}^{T} D_t(\hat{\theta}_T^*)\Sigma_t^{-1}(\tilde{\theta}_T)f(y_t,\hat{\theta}_T^*) = 0. \tag{3.9}$$

By identification with (3.3), one observes that $\hat{\theta}_T^*$ is nothing but that our efficient quadratic M-estimator. Thus, by extending the equivalence argument of Rilstone (1992), one gets an equivalence result between GMM and M-estimation which was never (to the best of our knowledge) clearly stated until now:[19]

Theorem 3.1 *If for conditional moment restricitions (3.1) conformable to (3.4), one considers the efficient GMM $\hat{\theta}_T$ associated with optimal instruments (defined by (3.7)) and the efficient quadratic M-estimator $\hat{\theta}_T^*$ (defined by (3.9)), under standard regularity conditions (Assumptions 1,2,3,4, adapted to the setting of section 3), $\hat{\theta}_T$ and $\hat{\theta}_T^*$ are consistent, asymptotically normal and have the same asymptotic probability distribution.*

[17]We recall that an ARCH model can be markovian in the opposite of the GARCH one.

[18]See also Kuersteiner (1997) and Guo-Phillips (1997).

[19]The proof is similar to the proof of Proposition 2.2.

Note that a key difference between our setting and Rilstone's is that we assume by (3.4) that: $\frac{\partial f'}{\partial \theta}(y_t, \theta) \in I_{t-1}$ and therefore $D_t(\theta^0) = \frac{\partial f'}{\partial \theta}(y_t, \theta^0)$. Thus we are able to interpret Rilstone's suggestion as a **quadratic M-estimator**. In other words, we give support, a posteriori, to Rilstone's terminology of "extremum" estimator to refer to $\hat{\theta}_T^*$.

3.2 Application to ARCH-type processes

The general equivalence result of section 3.1 can be applied to our ARCH-type setting defined by Assumptions 1 to 4 by considering:[20]

$$f(y_t, \theta) = [y_t - m_t(\theta), (y_t - m_t(\theta^0))^2 - h_t(\theta)]', \tag{3.10}$$

or, given $\tilde{\theta}_T$ as a first-stage estimator of θ^0,

$$\tilde{f}(y_t, \theta) = [y_t - m_t(\theta), (y_t - m_t(\tilde{\theta}_T))^2 - h_t(\theta)]'.$$

With such a convention, the "error term" $\tilde{f}(y_t, \theta)$ fulfills the crucial assumption (3.4) which allows us to apply the equivalence Theorem 3.1. Since we know from Chamberlain (1987) that the GMM efficiency bound is indeed the semiparametric efficiency bound, we conclude that *the efficient way to use the information provided by the parametric specification $m_t(.)$ and $h_t(.)$ of conditional mean and variance is the optimal quadratic M-estimation principle defined by Theorem 2.1.*

In other words, besides its intuitive appeal, the equivalence result is important in two respects. The QMLE and its natural improvements in terms of quadratic M-estimation is considered as a simpler method than GMM (see Bollerslev and Wooldridge (1992) as mentioned in the introduction above and previous work by Crowder (1987) and Wefelmeyer (1996)). Also, the GMM theory provides the benchmark for optimal use of available information in terms of semiparametric efficiency bounds.

Since GMM with optimal instruments as well as optimal quadratic M-estimators are generally unfeasible without preliminary adaptive estimation of higher order conditional moments, one is often led to use parametric specifications of these moments. Typically, parametric specifications of conditional skewness and kurtosis (see section 4) will allow one to compute both optimal quadratic M-estimator and optimal instruments. But, as already explained, such an approach is flawed by a logical internal inconsistency since, if one knows the parametric specification $M_{3t}(\theta)$ and $K_t(\theta)$ of conditional skewness and kurtosis, for inference one should use the set of conditional moments restrictions associated to the following "augmented" f:

$$f(y_t, \theta) = [y_t - m_t(\theta), (y_t - m_t(\theta^0))^2 - h_t(\theta), (y_t - m_t(\theta^0))^3 - M_{3t}(\theta)h_t^{3/2}(\theta), (y_t - m_t(\theta^0))^4 - 3K_t(\theta)h_t^2(\theta)]'. \tag{3.11}$$

With respect to (3.11), the optimal GMM associated with (3.10) will generally be inefficient. Note that the augmented f, as defined by (3.11) under the assumption (3.4) allows one to apply our equivalence result. In other words, the new efficiency bound associated with (3.11)

[20]Note that we can also consider the instrumental variable estimation based on $E[(y_t - m_t(\theta), (y_t - m_t(\theta))^2 - h_t(\theta))' \mid I_{t-1}] = 0$. Given an instrument z_t, the corresponding estimator is consistent and asymptotically equivalent to the estimator based on $E[(y_t - m_t(\theta), (y_t - m_t(\theta^0))^2 - h_t(\theta))' \mid I_{t-1}] = 0$ with the same instrument.

(which is generally smaller than the one associated to (3.10)) can generate estimation strategies conformable to our section 4 (see below). Furthermore, the efficiency bound will be reached by multivariate QMLE which would consider $f(y_t, \theta)$ as a gaussian vector.

Indeed, the main lesson of the above results is perhaps that, for a given number of moments involved (order 1,2,3,...), multivariate QMLE and the associated battery of inference tools (see Gouriéroux, Monfort and Trognon (1984), Wooldridge (1990, 1991a, 1991b)) allow one to reach the semiparametric efficiency bound. Moreover, the reduction of information methodology emphasized in section 4 (see below) will often simplify the feasibility of an "**optimal**" QMLE by providing a principle of reduction of the set of admissible strategies. The search for such a principle is not new in statistics (see unbiasedness, invariance, ... principles) and is fruitful if it does not rule out the most natural strategies. This is clearly the case for interesting examples that we have listed in section 4.

4 Information adjusted M-estimators and linear interpretations

4.1 The semiparametric ARCH-type model

To obtain a feasible estimator of which asymptotic variance achieves the efficiency bound of Theorem 2.1, we generally require a nonparametric estimation of dynamic conditional third and fourth moments. These issues will be discussed in more detail in section 4.2 below.

Engle and González-Rivera (1991) have introduced the so-called "semiparametric ARCH model" to simplify the nonparametric estimation. By assuming that the standardized errors $u_t(\theta^0) = \varepsilon(\theta^0)/\sqrt{h_t(\theta^0)}$ are i.i.d, they are led to perform a nonparametric probability density estimation in a static setting which provides a semi-nonparametric inference technique about θ^0. Our purpose in this section is to show that this semiparametric model allows us to compute easily an optimal semiparametric estimator.

Surprisingly, Engle and González-Rivera (1991) stress the role of conditional skewness and kurtosis but their i.i.d assumption imposes some restrictions on the whole probability distribution of the error process. Alternatively, we consider in this section an "independence" assumption which is only defined through third and fourth moments:

Assumption 5: The standardized errors $u_t(\theta^0)$ have constant conditional skewness $M_{3t}(\theta^0)$ and conditional kurtosis $K_t(\theta^0)$.

In other words, $M_{3t}(\theta^0)$ and $K_t(\theta^0)$ are assumed to coincide with unconditional skewness and kurtosis coefficients of the u_t process:

$$M_3(\theta^0) = E(u_t^3(\theta^0)) \tag{4.1.a}$$

$$K(\theta^0) = \frac{1}{3}E(u_t^4(\theta^0)) \tag{4.1.b}$$

An advantage of Assumption 5 (with respect to the more restrictive Engle and González-Rivera (1991) semiparametric setting) is that it is fully characterized by a set of conditional moment restrictions:

$$E(u_t^3(\theta^0) - M_3(\theta^0) \mid I_{t-1}) = 0 \tag{4.2.a}$$

$$E(u_t^4(\theta^0) - 3K(\theta^0) \mid I_{t-1}) = 0 \tag{4.2.b}$$

which are testable by GMM overidentification tests.

Moreover, let us assume that we have at our disposal a first-step consistent estimator $\tilde{\theta_T}$ of θ^0 (it could be the QMLE). Thanks to Assumption 5, we are then able to compute consistent estimators of skewness and kurtosis coefficients of $u_t(\theta^0)$:

$$\hat{M}_{3,T}(\tilde{\theta_T}) = \frac{1}{T}\sum_{t=1}^{T} u_t^3(\tilde{\theta_T}) \tag{4.3.a}$$

$$\hat{K}_T(\tilde{\theta_T}) = \frac{1}{3T}\sum_{t=1}^{T} u_t^4(\tilde{\theta_T}) \tag{4.3.b}$$

Note that under Assumption 5, $\hat{M}_{3,T}(\tilde{\theta_T})$ (resp $\hat{K}_T(\tilde{\theta_T})$) is a consistent estimator of both $M_{3t}(\theta^0)$ and $M_3(\theta^0)$ (resp $K_t(\theta^0)$ and $K(\theta^0)$). Therefore, we obtain a feasible M-estimator of θ^0 by considering $\hat{\theta_T}^* = \hat{\theta}_T(\tilde{\theta}_T, \hat{\omega}_T^*, \hat{\gamma}^*)$.

Theorem 4.1 *Let us consider the estimator $\hat{\theta}_T^*$ defined by:*

$$\hat{\theta}_T^* = \underset{\theta}{ArgMin} \sum_{t=1}^{T} \hat{a}_T^* \frac{\varepsilon_t(\theta)^2}{h_t(\tilde{\theta}_T)} + \hat{b}_T^* \frac{(\varepsilon_t(\tilde{\theta}_T)^2 - h_t(\theta))^2}{h_t(\tilde{\theta}_T)^2} + 2\hat{c}_T^* \frac{\varepsilon_t(\theta)(\varepsilon_t(\tilde{\theta}_T)^2 - h_t(\theta))}{h_t(\tilde{\theta}_T)^{\frac{3}{2}}}$$

where: $\hat{a}_T^* = (3\hat{K}_T(\tilde{\theta}_T) - 1) \times \hat{b}_T^*,\ \hat{b}_T^* = \dfrac{1}{2}\dfrac{1}{3\hat{K}_T(\tilde{\theta}_T) - 1 - \hat{M}_{3,T}(\tilde{\theta}_T)^2},\ \hat{c}_T^* = -\hat{M}_{3,T}(\tilde{\theta}_T) \times \hat{b}_T^*,$

and where $\tilde{\theta}_T$ is a weakly consistent estimator of θ^0 such that $\sqrt{T}(\tilde{\theta}_T - \theta^0) = O_P(1)$ (e.g. a consistent asymptotically normal estimator). Then under Assumptions 1, 2, 3, 4 and 5, $\hat{\theta}_T^*$ is a weakly consistent estimator of θ^0, asymptotically normal, of which asymptotic covariance matrix coincides with the efficiency bound Σ^0 defined by Theorem 2.1.

We then have in a sense constructed an optimal M-estimator of θ^0. Of course, this optimality is defined relatively to a given set of estimating restrictions, namely Assumption 1. In particular, the informational content of Assumption 5 is not take into account (see section 3). However, for normal errors u_t, our estimator is asymptotically equivalent to $\hat{\theta}_T^Q$, which in this case is the **Maximum Likelihood Estimator** (MLE). This is a direct consequence of Proposition 2.3, Theorems 2.1 and 4.1.[21] On the other hand, in the semiparametric setting proposed by Engle and González-Rivera (1991) (and more generally in our framework defined by Assumptions 1 to 5), Theorem 2.1 provides the best choice of weights Λ_t to take into account non-normal skewness and kurtosis coefficients. In particular, in this latter case, our estimator **strictly dominates** (without a genuine additional computational difficulty) the usual QMLE based on nominal normality. The QMLE appears to be a judicious way to estimate only if we are sure that conditional skewness and kurtosis are respectively equal to 0 and 1.

[21]Proposition 2.3, Theorems 2.1 and 4.1 prove respectively that: first, $\hat{\theta}_T^Q$ is asymptotically equivalent to the estimator $\hat{\theta}_T(\tilde{\theta}_T, \omega^*, \gamma^Q)$ of our class; second, γ^Q is an optimal choice of γ in the normal case; third, $\hat{\theta}_T(\tilde{\theta}_T, \omega^*, \gamma^Q)$ may be replaced by a feasible estimator without loss of efficiency.

4.2 Relaxing the assumption of semiparametric ARCH

Our semiparametric ARCH type setting has allowed us to consistently estimate (conditional) skewness and kurtosis by their empirical counterparts. If we are not ready to maintain Assumption 5, we know that the empirical skewness and kurtosis coefficients (4.3) are only consistent estimates of marginal skewness and kurtosis. Therefore, Theorem 4.1 does not provide in general an efficient estimator as characterized by Theorem 2.1. We propose in this section a general methodology to construct "efficient" estimators, where the efficiency concept is possibly weakened by restricting ourselves to more specific models and estimators. The basic tool for doing this is the following remark which is a straightforward corollary of Proposition 2.2:

Let us consider a sequence of σ-fields J_t, $t = 0, 1, 2, ..$, such that, for any $\theta \in \Theta$:

$$m_t(\theta),\ h_t(\theta)\ \in\ J_{t-1} \subset I_{t-1}. \tag{4.4}$$

Under assumptions 1, 2, 3, 4 and the notations of proposition 2.2, we consider the class $\mathcal{C}^J$ of M-estimators $\hat{\theta}(\tilde{\theta}_T, \omega_T, \gamma)$ such that:

$$\Lambda_t(\omega) \in J_{t-1} \tag{4.5}$$

for any $\omega \in \mathcal{V}$ and $t = 1, 2, ..T$.

Since $m_t(\theta)$ and $h_t(\theta)$ are assumed to be J_{t-1} measurable for any θ, the class $\mathcal{C}^J$ is large and contains in particular every M-estimator (2.17) associated to constant weights a_t, b_t, c_t. Therefore, by looking for a M-estimator optimal in the class $\mathcal{C}^J$, we are in particular improving the QMLE which corresponds (in terms of asymptotic equivalence) to the constant weights $(\frac{1}{2}, \frac{1}{4}, 0)$.

For such an estimator, the asymptotic covariance matrix $A_\gamma^{0^{-1}} B_\gamma^0 A_\gamma^{0^{-1}}$ admits a slightly modified expression deduced from Proposition 2.2 by replacing $\Sigma_t(\theta^0)$ by:

$$\Sigma_t^J(\theta^0) = Var[\begin{pmatrix} \varepsilon_t(\theta^0) \\ \nu_t(\theta^0) \end{pmatrix} \mid J_{t-1}] = E[\Sigma_t(\theta^0) \mid J_{t-1}].$$

This suggests the following generalization of Theorem 2.1:

Theorem 4.2 *Under the assumptions of Theorem 2.1, a sufficient condition for an estimator of the class $\mathcal{C}^J$ (according to (4.4)/(4.5)) to have the minimum asymptotic covariance matrix in this class is that, for all t:*

$$\Lambda_t(\omega^*) = (\Sigma_t^J(\theta^0))^{-1}.$$

Notice that Theorem 4.2 is not identical to Theorem 2.1 since it can be applied to sub-σ fields $J_{t-1} \subset I_{t-1} = \sigma(z_t, y_\tau, z_\tau, \tau < t)$ without even assuming that $(J_t), t = 0, 1, 2..$ is an increasing filtration. If for instance we consider a linear regression model with ARCH disturbances:

$$m_t(\theta) = a + x_t' b, \qquad h_t(\theta) = \omega + \sum_{i=1}^{q} \alpha_i (y_{t-i} - m_{t-i}(\theta))^2, \tag{4.6}$$

where $x_t = (x_t^1, x_t^2, ..x_t^H)'$ and x_t^h, $h = 1, ..H$, is a given variable in I_{t-1}, we can consider:

$$J_{t-1} = \sigma(x_t, y_{t-i}, x_{t-i}, i = 1, 2.., q).$$

Thus, Theorem 4.2 suggests a large set of applications which were not previously considered in the literature. The basic idea of these applications is that one could try to find a reduction J_{t-1} of the information set such that conditional skewness and kurtosis with respect to this new information set admit simpler forms which can be consistently estimated. Below, we consider three types of "simplified" conditional skewness and kurtosis.

Application 1: Constant conditional skewness and kurtosis.

Let us first imagine that a reduction J_{t-1} of the information set I_{t-1} (conformable to (4.4)) allows one to obtain constant conditional skewness and kurtosis:

$$M_3(\theta^0) = E[u_t(\theta^0)^3 \mid J_{t-1}] = E[M_{3t}(\theta^0) \mid J_{t-1}], \tag{4.7.a}$$

$$K(\theta^0) = \frac{1}{3}E[u_t(\theta^0)^4 \mid J_{t-1}] = E[K_t(\theta^0) \mid J_{t-1}]. \tag{4.7.b}$$

If this is the case, it is true in particular for the minimal information set:

$$J_{t-1} = I^*_{t-1} = \sigma(m_t(\theta), h_t(\theta), \theta \in \Theta).$$

For notational simplicity, we will focus on this case. Therefore, the hypothesis (4.7) may be tested by considering the moment conditions:

$$E[u_t(\theta^0)^3 - M_3 \mid I^*_{t-1}] = 0 \quad \text{and} \quad E[u_t(\theta^0)^4 - 3K \mid I^*_{t-1}] = 0.$$

More precisely, one can perform an overidentification Hansen's test on the following set of conditional moment restrictions associated with the vector $(\theta', M_3, K)'$ of unknown parameters:

$$\begin{cases} E[y_t - m_t(\theta) \mid I^*_{t-1}] = 0, & E[(y_t - m_t(\theta))^2 - h_t(\theta) \mid I^*_{t-1}] = 0, \\ E[u_t(\theta^0)^3 - M_3 \mid I^*_{t-1}] = 0, & E[u_t(\theta^0)^4 - 3K \mid I^*_{t-1}] = 0. \end{cases}$$

Let us notice that if we consider example (4.6), we are led to test orthogonality conditions like:

$$Cov[u_t^3(\theta^0), f(x_t, x_\tau, y_\tau, \tau < t)] = 0 \;\text{ and }\; Cov[u_t^4(\theta^0), f(x_t, x_\tau, y_\tau, \tau < t)] = 0 \tag{4.8}$$

for any real valued function f. Taking into account the parametric specification (4.6), it is quite natural to consider, as particular testing functions f, the polynomials of degree 1 and 2 with respect to the variables components of $(x_t, x_{t-i}, y_{t-i}, i = 1, 2, ..q)$. In any case, if one trusts assumption (4.7), one can use the following result:

Theorem 4.3 *Under assumptions (4.7) with the assumptions of Theorem 2.1, the estimators $\hat{\theta}^*_T$ defined by Theorem 4.1 is of minimum asymptotic covariance matrix in the minimal class $\mathcal{C}^{I^*}$.*

In other words, thanks to a reduction $\mathcal{C}^{I^*}$ of the class of M-estimators we consider, assumption (4.7) is a sufficient condition (much more general than the semiparametric ARCH setting) to ensure that the M-estimator $\hat{\theta}^*_T$ computed from empirical skewness and kurtosis is optimal in a second-best sense and particularly, more accurate than the QMLE.

Indeed, to ensure that $\hat{\theta}^*_T$ is better than the usual QMLE, it is sufficient to know that $\hat{\theta}^*_T$ is optimal in the subclass $\mathcal{C}^0$ of $\mathcal{C}^{I^*}$ of M-estimators associated to constant weights: $(a_t, b_t, c_t) = (a, b, c)$. This optimality is ensured by a weaker assumption than (4.7) as shown by the following:

Proposition 4.1 *If the following orthogonality conditions are fulfilled:*

$$Cov[(\frac{M_{3t}(\theta^0)}{K_t(\theta^0)}), \frac{1}{h_t(\theta^0)}\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0)] = 0, \quad Cov[(\frac{M_{3t}(\theta^0)}{K_t(\theta^0)}), \frac{1}{h_t(\theta^0)^2}\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)] = 0,$$

$$Cov[(\frac{M_{3t}(\theta^0)}{K_t(\theta^0)}), \frac{1}{h_t(\theta^0)^{3/2}}\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)] = 0,$$

then the estimator $\hat{\theta}_T^*$ defined in Theorem 4.1 is of minimum asymptotic covariance matrix in the class $\mathcal{C}^0$ of M-estimators defined by constant weights (a, b, c).

The orthogonality assumptions of proposition 4.1 are minimal in the sense that they are a weakening of (4.7) which involves only the functions of $J_{t-1} = I_{t-1}^*$ which do appear in the variance calculations.

Application 2: "Linear models" of the conditional skewness and kurtosis.

It turns out that there are situations where, while the assumption (4.7) of constant conditional skewness and kurtosis could not be maintained, one may trust a more general parametric model (associated with a reduction J_{t-1} of the information set):

$$\begin{cases} M_{3t}^J(\theta^0) = E[M_{3t}(\theta^0) \mid J_{t-1}] = M_3^C(m_t(\theta^0), h_t(\theta^0), \lambda) \\ K_t^J(\theta^0) = E[K_3(\theta^0) \mid J_{t-1}] = K^C(m_t(\theta^0), h_t(\theta^0), \lambda) \end{cases} \tag{4.9}$$

where λ is a vector of nuisance parameters and $M_3^C(.)$ and $K^C(.)$ are known functions.

An example of such a situation is provided by Drost and Nijman (1993) in the context of temporal aggregation of a symmetric semiparametric ARCH(1) process. Indeed, one of the weaknesses of the semiparametric GARCH framework considered in subsection 4.1 is its lack of robustness with respect to temporal aggregation (see Drost and Nijman (1993) and Meddahi and Renault (1996)). Thus it is important to be able to relax the assumption of semiparametric GARCH if we are not sure of the relevant frequency of sampling (which should allow us to maintain the semiparametric assumption). Following Drost and Nijman (1993), Example 3 page 918, let us consider the following semiparametric symmetric ARCH(1) process:

$$\begin{cases} y_t = \sqrt{h_t(\theta^0)}u_t, \quad h_t(\theta^0) = \psi^0 + \alpha^0 y_{t-1}^2, \\ u_t \text{ i.i.d}, \quad E[u_t] = 0, \quad Var(u_t) = 1, \quad E[u_t^3] = 0. \end{cases} \tag{4.10}$$

If one now imagines that the sampling frequency is divided by 2, one observes $y_{2t},\ t \in \mathbf{Z}$, which defines a reduced information filtration:

$$I_{2t}^{(2)} = \sigma(y_{2\tau}, \tau \leq t).$$

Due to this reduction of past information, we now have to redefine the conditional variance process:

$$h_{2t}^{(2)}(\theta^0) = Var[y_{2t} \mid I_{2t-2}^{(2)}].$$

The parametric form of $h_{2t}^{(2)}(\theta^0)$ can be deduced from (4.10) by elementary algebra: $h_{2t}^{(2)} = E[h_{2t} \mid I_{2t-2}^{(2)}]$

with: $h_{2t} = \psi + \alpha y_{2t-1}^2 = \psi + \alpha u_{2t-1}^2(\psi + \alpha y_{2t-2}^2) = \psi + \alpha(\psi + \alpha y_{2t-2}^2) + \alpha(\psi + \alpha y_{2t-2}^2)(u_{2t-1}^2 - 1).$

Therefore:

$$h_{2t}^{(2)} = \psi(1 + \alpha) + \alpha^2 y_{2t-2}^2 \tag{4.11}$$

and
$$u_{2t}^{(2)} = \frac{y_{2t}}{\sqrt{h_{2t}^{(2)}}} = u_{2t}\sqrt{\frac{h_{2t}}{h_{2t}^{(2)}}} = u_{2t}\sqrt{\lambda_{2t} + u_{2t-1}^2(1-\lambda_{2t})} \tag{4.12}$$
with
$$\lambda_{2t} = \frac{\psi}{h_{2t}^{(2)}}. \tag{4.13}$$

By a simple development of $E[(u_{2t}^{(2)})^4 \mid I_{2t-2}^{(2)}]$ from (4.11), one gets:
$$E[(u_{2t}^{(2)})^4 \mid I_{2t-2}^{(2)}] = 3K[\lambda_{2t}^2(3K-1) - 2\lambda_{2t}(3K-1) + 3K] \tag{4.14}$$
where
$$K = \frac{1}{3}E[u_t^4 \mid I_{t-1}] = \frac{1}{3}E[u_t^4].$$
In other words, while conditional kurtosis was constant with a given frequency, it is now time-varying and stochastic (through the process λ_{2t}) when the sampling frequency is divided by 2. On the other hand, the symmetry assumption is maintained:
$$E[(u_{2t}^{(2)})^3 \mid I_{2t-2}^{(2)}] = 0.$$

This example suggests a class of models where, for a reduced information J_t, one has the following relaxation of (4.7):
$$K_t^J(\theta^0) = \frac{1}{3}E[u_t(\theta^0)^4 \mid J_{t-1}] = \lambda_0 + \frac{\lambda_1}{h_t(\theta^0)} + \frac{\lambda_2}{(h_t(\theta^0))^2} \tag{4.15}$$

and, in this case $M_{3t}^J(\theta^0) = 0$.

Such a parametric form of conditional kurtosis has been suggested by temporal aggregation arguments.[22] Moreover, it corresponds to some empirical evidence already documented for instance by Bossaerts, Hafner and Hardle (1995) who notice that while higher conditional volatility is associated with large changes in exchange rate quotes, conditional kurtosis is higher for small quote changes.

In any case, whatever the parametric model (4.9) we have in mind, it can be used to compute an estimator asymptotically equivalent to the efficient one in the class $\mathcal{C}^{\mathbf{J}}$ (defined by Theorem 4.2). The procedure may be the following. First, compute standardized residuals $\tilde{u}_t(\tilde{\theta}_T)$ associated with a first-stage consistent estimator $\tilde{\theta}_T$. Then, compute a consistent estimator $\tilde{\lambda}_T$ of λ from (4.9), for instance by minimizing the sum of squared deviations:
$$\sum_{t=1}^{T}[\tilde{u}_t^3(\tilde{\theta}_T) - M_3^c(m_t(\tilde{\theta}_T), h_t(\tilde{\theta}_T), \lambda)]^2 + [\frac{1}{3}\tilde{u}_t^4(\tilde{\theta}_T) - K^c(m_t(\tilde{\theta}_T), h_t(\tilde{\theta}_T), \lambda)]^2.$$

For the example (4.15) we only have to perform linear OLS of $\frac{1}{3}\tilde{u}_t^4(\tilde{\theta}_T)$ with respect to 1, $\dfrac{1}{h_t(\tilde{\theta}_T)}$ and $\dfrac{1}{(h_t(\tilde{\theta}_T))^2}$. Finally, use the adjusted conditional skewness $M_3^c(m_t(\tilde{\theta}_T), h_t(\tilde{\theta}_T), \tilde{\lambda})$ and kurtosis $K^c(m_t(\tilde{\theta}_T), h_t(\tilde{\theta}_T), \tilde{\lambda})$ to compute a weighting matrix $\tilde{\Lambda}_{t,T} = [\tilde{\Sigma}_{t,T}^J(\tilde{\theta}_T)]^{-1}$. By Proposition 2.2, the estimator $\hat{\theta}_T$ deduced from $\tilde{\theta}_T$ and the weighting matrices $\tilde{\Lambda}_{t,T}$, $t = 1, 2.., T$, will be of minimal asymptotic covariance matrix in the class $\mathcal{C}^J$.

[22]See also Hansen (1994), DeJong, Drost and Werker (1996), El-Babsiri and Zakoian (1997), for examples of heteroskewness and heterokurtosis models.

Application 3: Nonparametric regression models of the conditional skewness and kurtosis.

The two applications above always assume a fully specified parametric model for conditional skewness and kurtosis (with respect to a reduced filtration J). In this respect, they suffer from the usual drawback: In order to compute an "efficient" M-estimator, we need additional information which could theoretically be used for defining a better estimator (see section 3 for some insights on this paradox). A way to avoid this problem is to look for weighting matrices Λ_t, $t = 1, 2.., T$, which are deduced from a nonparametric estimation of the conditional variance $\Sigma_t(\theta^0)$. But for such a semiparametric strategy, the usual disclaimer applies: if the process is not markovian in such a way that $\Sigma_t(\theta^0)$ depends on I_{t-1} through an infinite number of lagged values $y_\tau, \tau < t$, the nonparametric estimation cannot be performed in general. Moreover, non Markovian dynamics of conditional higher order moments is a common situation since, for instance in a GARCH framework, dynamics (4.15) of conditional kurtosis are not markovian. Of course, one may always imagine limiting a priori the number of lags taken into account in the nonparametric estimation (see e.g. Masry and Tjostheim (1995)), but there is then a trade off between the misspecification bias and the curse of dimensionality problem.

Thus a reduction of the information set may be very useful. Indeed, when $\Sigma_t(\theta^0)$ cannot be consistently estimated, it may be the case that a reduction J of the information filtration provides a new covariance matrix $\Sigma_t^J(\theta^0)$ which depends only on a finite number of given functions. For instance with the minimal information set:

$$J_{t-1} = I_{t-1}^* = \sigma(m_t(\theta), h_t(\theta), \theta \in \Theta)$$

we may hope that $M_{3t}^J(\theta^0)$ and $K_t^J(\theta^0)$ depend only on a finite number of functions of lagged values of $(m_t(\theta^0), h_t(\theta^0))$. By extending the main idea of Application 2, one may imagine for instance that $K_t^J(\theta^0)$ is an unknown function of the q variables $h_{t-i_j}(\theta^0)$, $i_j \in \mathbf{N}^*$, $j = 1, 2.., q$. In such a case, the estimation procedure described in Application 2 can be generalized by replacing the second stage nonlinear regression by a nonparametric kernel estimation of the regression function of $\tilde{u}_t^3(\tilde{\theta}_T)$ and $\tilde{u}_t^4(\tilde{\theta}_T)$ on relevant variables.

4.3 Multistage linear least squares procedures

In this section we show that all the estimators considered above (except the ones which involve nonparametric kernel estimation) admit asymptotically equivalent versions which can be computed by using only linear regression packages.

We have already stressed (see (2.10)) that in standard settings, a first-stage consistent estimator $\tilde{\theta}_T$ can be obtained with nonlinear regression packages. Of course, with Newton regression (see e.g. Davidson and MacKinnon (1993)) these nonlinear regressions can be replaced with linear ones. It remains to be explained how we are able to compute an efficient M-estimator (that is an estimator asymptotically equivalent to the efficient one defined by Theorem 4.1, Theorem 4.2 or Application 2) by using only linear tools. Indeed, this is a general property of our quadratic M-estimators as it is stated in the following theorem:

Theorem 4.4 *Consider, in the context of Assumptions 1, 2, 3, 4, a M-estimator $\hat{\theta}_T^1$ defined* *by:*

$$\hat{\theta}_T^1 = \underset{\theta}{ArgMin} \sum_{t=1}^{T} \phi_t'(\theta, \tilde{\theta}_T) \Lambda_{t,T}(\tilde{\theta}_T) \phi_t(\theta, \tilde{\theta}_T)$$

where, for $t = 1, \; 2...,$ ϕ_t *is a known function of class* C^2 *on* $(int\Theta)^2$ *such that* $E[\phi_t(\theta^0, \theta^0) \mid I_{t-1}] = 0$. *Then* $\hat{\theta}_T^1$ *is asymptotically equivalent to* $\hat{\theta}_T^2$ *defined by*

$$\hat{\theta}_T^2 = \underset{\theta}{ArgMin} \sum_{t=1}^{T} [\phi_t(\tilde{\theta}_T, \tilde{\theta}_T) + \frac{\partial \phi_t}{\partial \theta'}(\tilde{\theta}_T, \tilde{\theta}_T)(\theta - \tilde{\theta}_T)]' \Lambda_{t,T}(\tilde{\theta}_T) [\phi_t(\tilde{\theta}_T, \tilde{\theta}_T) + \frac{\partial \phi_t}{\partial \theta'}(\tilde{\theta}_T, \tilde{\theta}_T)(\theta - \tilde{\theta}_T)]$$

where $\frac{\partial \phi_t}{\partial \theta'}$ *denotes the jacobian matrix of* ϕ_t *with respect to its first occurence.*

This theorem implicitly assumes that ϕ_t verifies the standard measurability, continuity and differentiability conditions which ensure consistency and asymptotic normality of the associated estimators. This is typically the case under Assumptions 1, 2, 3, 4, if:

$$\phi_t(\theta, \lambda) = (\varepsilon_t(\theta), \varepsilon_t^2(\lambda) - h_t(\theta)).$$

The basic idea of Theorem 4.4, namely a Newton-based modification of the initial objective function to produce a two-step estimation method without loss of efficiency is not new in econometrics. From the seminal paper by Hartley (1961) and its application to dynamic models by Hatanaka (1974), Trognon and Gouriéroux (1990) have developed a general theory (see also Pagan (1986)). Indeed, the proof of Theorem 4.4 shows that we are confronted with a case where there is no efficiency loss produced by a direct two-stage procedure and thus, we do not need to build an "approximate objective function" as in Trognon and Gouriéroux (1990). By application of the same methodology, all the procedures described above can be performed by linear regressions, including the preliminary estimation of conditional skewness and kurtosis functions.

4.4 Monte Carlo evidence

Until now we have only presented theoretical asymptotic properties of our various estimators. In the following, we present a Monte Carlo study which compare the asymptotic variances is several cases. Thus we consider a large sample size (1000). We want to give a flavor of the importance of taking into account conditional skewness and kurtosis. A complete discussion of the small-sample is done in Alami, Meddahi and Renault (1998) (AMR hereafter). We consider the following DGP:

$$y_t = c + \rho y_{t-1} + \varepsilon_t \tag{4.16.a}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 \tag{4.16.b}$$

$\theta = (c, \rho, \omega, \alpha)'$ with $\theta^0 = (1, 0.7, 0.5, 0.5)'$, with three possible probability distributions for the i.i.d standardized residuals $u_t = \frac{\varepsilon_t}{\sqrt{h_t}}$: standard Normal, standardized Student T(5) and standardized Gamma $\Gamma(1)$.

For each experiment, we have performed 400 replications. The main goal of these experiments is to compare, for the three probability distributions above, three natural estimators:[23]

1) Two-stage OLS, that is OLS on (4.16.a) to compute residuals $\hat{\varepsilon}_t$ and OLS on the approximated regression equation associated with (4.16.b): $\hat{\varepsilon}_t^2 \simeq \omega + \alpha \hat{\varepsilon}_{t-1}^2 + \nu_t$.

2) QMLE.

3) Our efficient M-estimator from Theorem 4.1.

Since our efficient M-estimator is a two-stage one (based on a first stage consistent estimator $\tilde{\theta}_T$), the finite sample properties might depend heavily on the choice of $\tilde{\theta}_T$. Therefore, we consider below four versions of our efficient M-estimator:

- **Version C1**: $\tilde{\theta}_T =$ OLS,
- **Version C2**: $\tilde{\theta}_T =$ QMLE,
- **Version C3**: "Iterated OLS",
- **Version C4**: "Iterated QMLE",

where "Iterated OLS" (resp QMLE) means that $\tilde{\theta}_T^{(5)}$ is defined from the following algorithm: $\tilde{\theta}_T^{(1)}$ is the "version C1" (resp C2) efficient estimator, and for $p = 2, 3, 4, 5$, $\tilde{\theta}_T^{(p)}$ is the efficient estimator computed with $\tilde{\theta}_T^{(p-1)}$ as a first-stage estimator $\tilde{\theta}_T$. For these small-scale experiments, we have simplified this theoretical procedure by using, at each stage, only one step of the numerical routine of optimization.[24]

The results of our Monte Carlo experiments are presented in tables 1, 2, 3 which correspond respectively to cases 1, 2 and 3. We provide the mean over our 400 replications, and between brackets, the Monte Carlo standard error.[25]

The Monte Carlo results lead to four preliminary conclusions:

i) **The ARCH parameters (ω and α) are very badly estimated by OLS**. This inefficiency is more and more striking when one goes from Table 1 to Table 3. While the heteroskedasticity parameter is underestimated by OLS by almost 20 percent in the gaussian case, it is underestimated by almost 50 percent in the gamma case, that is when both leptokurtosis and skewness are present.

ii) **Despite the inefficiency of OLS, it can be used as a first-stage estimator for efficient estimation without a dramatic loss of efficiency with respect to the use of QMLE as a first-stage estimator**. In other words, C1 (resp C3) is not very different from C2 (resp C4). In particular, the difference is negligible in the iterated case: C3 and C4

[23] A large variety of estimators should be considered. For example, OLS could be iterated to perform QGLS. In any case, we know that the asymptotic accuracy of QGLS is worse than QMLE in case 1 (for the estimation of c and ρ, see Engle (1982)). Thus QGLS is not studied here, to focus on our main issue of improving QMLE.

[24] We provide in AMR (1998) additional experiments to show that such a simplification has almost no impact on the value of $\tilde{\theta}_T^{(5)}$.

[25] Mean and Monte carlo standard errors are obtained without any procedure of variance reduction. See AMR (1998) for a comparison with theoretical standard errors.

provide almost identical results (large sample size). However, we will now focus on C2 and C4 (efficient estimator with initial estimator QML) that we want to compare to b, that is QMLE.

iii) **As far as one is concerned by the estimation of the first-order dynamics (c end ρ), the use of an efficient procedure (C2 and C4) provides important efficiency gains for non-gaussian distributions, particularly when skewness is present**. The most striking result is that the efficient estimator of ρ is almost twice more accurate than QML in the case of gamma errors. On the other hand, iteration does not appear very fruitful (C4 almost identical to C2) due to the large sample size.

iv) **The efficient estimator of the heteroskedasticity parameters α is more accurate than QMLE.** The efficiency gain reached almost 50 percent in case of gamma errors. However, one has to be cautious when interpreting this conclusion for two reasons. First, it is important to use the iterated version of the efficient estimator, since, otherwise, α could be severely underestimated. Second, the efficiency gain in the case of a symmetric distribution (Student case) is only due (see the expression of the score) to the finite sample gain in estimation of c and ρ.

In any case, we conclude that, for accurate estimation of both first-order and second-order dynamics (ρ and α), the efficient estimation method provides a genuine efficiency gain in the case of skewed innovations. As already noticed by Engle and González-Rivera (1991), fat tails without skewness matter less. On the other hand, there is no loss implied by efficient estimation with respect to QML, at least for sample sizes 1000 with an iterated version of the estimator. Moreover, since one can use OLS as a first-stage estimator, efficient estimation does not imply dramatic numerical complexity with respect to QML. In other words, we conclude that for estimation, QML is strictly dominated by efficient procedures in all respects.

5 Conclusion

In this paper, we consider the estimation of time series models defined by their conditional mean and variance. We introduce a large class of quadratic M-estimators and characterize the optimal estimator which involves conditional skewness and kurtosis. We show that this optimal estimator is more efficient than the QMLE under non-normality. Furthermore, it is as efficient as the optimal GMM as well as the bivariate QMLE based on the dependent variable and its square. We also extend this study to higher order moments.

We apply our methodology to the so-called semiparametric GARCH models of Engle and González-Rivera (1991). A monte Carlo analysis confirms the relevance of our approach, in particular the importance of skewness. The recent work by Guo and Phillips (1997) also stress the skewness effect. We also present several cases where we can apply our methodology while the semiparametric setting (standardized residuals are i.i.d) is violated. A Monte Carlo analysis in such cases is considered in AMR (1998). Moreover, such cases, typically heteroskewness and heterokurtosis, introduce specific problems in testing for heteroskedasticity as detailed in AMR (1998).

References

Alami, A., N. Meddahi and E. Renault (1998), "Finite-Sample Properties of Some Alternative ARCH-Regression Estimators and Tests," mimeo, CREST.

Bates C.E. and H. White (1990), "Efficient Instrumental Variables Estimation for Systems of Implicit Heterogeneous Nonlinear Dynamic Equation with Nonspherical Errors," *Dynamic Econometric Modeling*, edited by W. Barnett, E. Berndt and H. White, Cambridge University Press.

Bates C.E. and H. White (1993), "Determination of Estimators with Minimum Asymptotic Variance," *Econometric Theory*, 9, 633-648.

Black, F. (1976), "Studies in Stock Price Volatility Changes," *Proceedings of the 1976 Business Meeting of the Business and Statistics Section, American Statistical Association*, 177-181.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T, R.Y. Chou et K.F. Kroner (1992), "ARCH Modeling in Finance : A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., R. Engle and D.B. Nelson (1994), "ARCH Models," in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol IV, 2959-3038, Elsevier, North-Holland, Amsterdam.

Bollerslev, T. and J.F. Wooldridge (1992), "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143-172.

Bossaerts, P., C. Hafner and W. Hardle (1995), "Foreign Exchange Rates Have Surprising Volatility," unpublished manuscript, Caltech.

Broze, L. and C. Gouriéroux (1995), "Pseuddo Maximum Likelihood Method, Adjusted Pseudo-Maximum Likelihood Method and Covariance Estimators," *Journal of Econometrics*, forthcoming.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5-46.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Cragg, J.G. (1983), "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 51, 751-764.

Crowder, M.J. (1987), "On linear and Quadratic Estimating Functions," *Biometrika*, 74, 591-597.

Davidson, R. and J.G. Mackinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

De Jong, F., F. Drost and B.J.M. Werker (1996), "Exchange Rate Target Zones: A New Approach," Manuscript, Tilburg University.

Drost, F. and C.A.J. Klassen (1997), "Efficient Estimation in Semiparametric GARCH Models," *Journal of Econometrics*, 81, 193-221.

Drost, F. and TH.E. Nijman (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909-927.

El Babsiri, M. and J.M. Zakoian (1997), "Contemporaneous Asymmetry in GARCH Processes," CREST DP 9703.

Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Engle, R. and G. González-Rivera (1991), "Semiparametric GARCH models," *Journal of Business and Economic Statistics*, 9, 345-359.

Engle, R., D.M. Lilien and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Models," *Econometrica*, 55, 391-407.

Franck, C. and J.M. Zakoian (1997), "Estimating Weak GARCH Representations," unpublished manuscript, CREST, Paris.

Gouriéroux, C. and A. Monfort (1993), "Pseudo-Likelihood Methods," in *The Handbook Of Statistics*, vol 11, 335-362, edited by G.S. Maddala, C.R. Rao and H.D. Vinod, North-Holland.

Gouriéroux, C., A. Monfort and A Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681-700.

Guo, B. and P.C.B. Phillips (1997), "Efficient Estimation of Second Moment Parameters in the ARCH Model," unpublished manuscript, Yale University.

Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation," *Internation Economic Review*, 35, 705-730.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hansen, L.P. (1985), "A Method for Calculating Bounds in the Asymptotic Covariance Matrix of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203-238.

Hansen, L.P, J.C. Heaton and M. Ogaki (1988), "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863-871.

Hartley, H.O. (1961), "The Modified Gauss-newton Method for the Fitting of Non-Linear Regressions Function by Least Squares," *Technometrics*, 3.

Hatanaka, M. (1974), "An Efficient Two-Step Estimator for the Dynamic Adjusted Model with Autoregressive Errors," *Journal of Econometrics*, 2, 199-220.

Heyde, C.C. (1997), *Quasi-Likelihood And Its Appalicatio: A general Approach to Optimal Parameter Estimation*, Springer Series in Statistics, Springer-Verlag, New York.

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions," *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley, University of California Press.

Kuersteiner, G.M. (1997), "Efficient IV estimation for Autoregressive Models with Conditional Heterogeneity," unpublished manuscript, MIT.

Lee, S.W. and B.E. Hansen (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29-52.

Linton, O. (1993), "Adaptive Estimation in ARCH Models," Econometric Theory, 9, 539-569.

Linton, O. (1994), "Estimation in Semiparametric Models: a Review," in *Advances in Econometrics and Quantitative Economics*, edited by G.S. Maddala, P.C. Phillips and T.N. Srinivasan.

Lumsdaine, R.L. (1996), "Consitency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1)," *Econometrica*, 64, 575-596.

Masry, E. and D. Tjostheim (1995), "Nonparametric Estimation and Identification of Nonlinear ARCH Time Series," *Econometric Theory*, 11, 258-289.

Meddahi, N. and E. Renault (1995), "Linear Statistical Inference for ARCH-Type Processes," mimeo, GREMAQ, Toulouse.

Meddahi, N. and E. Renault (1996), "Aggregations and Marginalization of GARCH and Stochastic Volatility Models," mimeo, GREMAQ, Toulouse.

Nelson, D.B. (1990), "Stationarity and Persistence in GARCH(1,1) Models," *Econometric Theory*, 6, 318-334.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

Nelson, D.B. and C.Q. Cao (1992), "Inequality Constraints in the Univariate GARCH Model," *Journal of Business and Economic Statistics*, 10, 229-235.

Newey, W.K. (1990), "Efficient Instrumental Variables Estimation of Nonlinear Econometric Models," *Econometrica*, 58, 809-837.

Newey, W.K. (1993), "Efficient Estimation of Models with Conditional Moment Restrictions," in *The Handbook Of Statistics*, vol 11, 419-454, edited by G.S. Maddala, C.R. Rao and H.D. Vinod, North-Holland.

Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol IV, 2111-2245, Elsevier, North-Holland, Amsterdam.

Newey, W.K. and D.G. Steigerwald (1997), "Asymptotic Bias for Quasi-Maximum Likelihood in Conditional Heteroskedasticity Models," *Econometrica*, 65, 587-599.

Pagan, A. (1986), "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies*, LIII, 517-538.

Rilstone, P. (1992), "Semiparametric IV Estimation with Parameter Dependent Instruments," *Econometric Theory*, 8, 403-406.

Robinson, P.M. (1991), "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica*, 59, 755-786.

Trognon, A. and C. Gouriéroux (1990), "A Note on the Efficiency of Two-Step Estimation Methods," in *Essays in Honor of Edmond Malinvaud*, Volume 3, 232-248, The MIT Press.

Wefelmeyer, W. (1996), "Quasi-Likelihood Models and Optimal Inference," *Annals of Statistics*, 24, 405-422.

Weiss, A.A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107-131.

White, H. (1982a), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-26.

White, H. (1982b), "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483-499.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press.

Wooldridge, J.F. (1990), "A Unified Approach to Robust Regression Based Specification Tests," *Econometric Theory*, 6, 17-43.

Wooldridge, J.F. (1991a), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47, 5-46.

Wooldridge, J.F. (1991b), "Specification Testing and Quasi-Maximum Likelihood Estimation," *Journal of Econometrics*, 48, 29-55.

Wooldridge, J.F. (1994), "Estimation and Inference for Dependent Processes," in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol IV, 2639-2738, Elsevier, North-Holland, Amsterdam.

Appendix A1

We have $\tilde{s}_t(\theta^0) = \dfrac{1}{h_t(\theta^0)}\dfrac{\partial m_t}{\partial\theta}(\theta^0)\varepsilon_t(\theta^0) + \dfrac{1}{2h_t^2(\theta^0)}\dfrac{\partial h_t}{\partial\theta}(\theta^0)\nu_t(\theta^0)$

which is, by (2.4), equal to $s_t(\theta^0)$. We have

$$\frac{\partial s_t}{\partial\theta'}(\theta) = -\frac{1}{h_t^2(\theta)}\frac{\partial m_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta)\varepsilon_t(\theta) + \frac{1}{h_t(\theta)}\frac{\partial^2 m_t}{\partial\theta\partial\theta'}(\theta)\varepsilon_t(\theta) - \frac{1}{h_t(\theta)}\frac{\partial m_t}{\partial\theta}(\theta)\frac{\partial m_t}{\partial\theta'}(\theta)$$

$$-\frac{1}{h_t^3(\theta)}\frac{\partial h_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta)\nu_t(\theta) + \frac{1}{2h_t^2(\theta)}\frac{\partial^2 h_t}{\partial\theta\partial\theta'}(\theta)\nu_t(\theta) - \frac{1}{2h_t^2(\theta)}\frac{\partial h_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta), \text{ and}$$

$$\frac{\partial\tilde{s}_t}{\partial\theta'}(\theta) = \frac{1}{h_t(\theta^0)}\frac{\partial^2 m_t}{\partial\theta\partial\theta'}(\theta)\varepsilon_t(\theta) - \frac{1}{h_t(\theta^0)}\frac{\partial m_t}{\partial\theta}(\theta)\frac{\partial m_t}{\partial\theta'}(\theta)$$

$$+\frac{1}{2h_t^2(\theta^0)}\frac{\partial^2 h_t}{\partial\theta\partial\theta'}(\theta)(\varepsilon_t^2(\theta^0) - h_t(\theta)) - \frac{1}{2h_t^2(\theta^0)}\frac{\partial h_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta).$$

Due to Assumptions 1 and 2, we have

$$E[\frac{\partial s_t}{\partial\theta'}(\theta^0)] = -E[\frac{1}{h_t(\theta^0)}\frac{\partial m_t}{\partial\theta}(\theta)\frac{\partial m_t}{\partial\theta'}(\theta)] - E[\frac{1}{2h_t^2(\theta^0)}\frac{\partial h_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta)], \text{ and}$$

$$E[\frac{\partial\tilde{s}_t}{\partial\theta'}(\theta^0)] = -E[\frac{1}{h_t(\theta^0)}\frac{\partial m_t}{\partial\theta}(\theta)\frac{\partial m_t}{\partial\theta'}(\theta)] - E[\frac{1}{2h_t^2(\theta^0)}\frac{\partial h_t}{\partial\theta}(\theta)\frac{\partial h_t}{\partial\theta'}(\theta)], \text{ that is}$$

$$E[\frac{\partial s_t}{\partial\theta'}(\theta^0)] = E[\frac{\partial\tilde{s}_t}{\partial\theta'}(\theta^0)].$$

Appendix A2

To prove the assertion, we must show that the matrix is invertible (since it is positive). We have

$$E\left\{\begin{bmatrix} \frac{\partial m_t}{\partial\alpha}(\alpha^0)\frac{\partial m_t}{\partial\alpha'}(\alpha^0) + \frac{\partial h_t}{\partial\alpha}(\theta^0)\frac{\partial h_t}{\partial\alpha'}(\theta^0) & \frac{\partial h_t}{\partial\alpha}(\theta^0)\frac{\partial h_t}{\partial\beta'}(\theta^0) \\ \frac{\partial h_t}{\partial\beta}(\theta^0)\frac{\partial h_t}{\partial\alpha'}(\theta^0) & \frac{\partial h_t}{\partial\beta}(\theta^0)\frac{\partial h_t}{\partial\beta'}(\theta^0) \end{bmatrix}\right\}$$

$$= E\begin{bmatrix} \frac{\partial m_t}{\partial\alpha}(\alpha^0)\frac{\partial m_t}{\partial\alpha'}(\alpha^0) & 0 \\ 0 & 0 \end{bmatrix} + E\left[\frac{\partial h_t}{\partial\theta}(\theta^0)\frac{\partial h_t}{\partial\theta'}(\theta^0)\right].$$

Let us consider a vector $Z = (Z_1', Z_2')'$ such that

$$E\left\{\begin{bmatrix} \frac{\partial m_t}{\partial\alpha}(\alpha^0)\frac{\partial m_t}{\partial\alpha'}(\alpha^0) + \frac{\partial h_t}{\partial\alpha}(\theta^0)\frac{\partial h_t}{\partial\alpha'}(\theta^0) & \frac{\partial h_t}{\partial\alpha}(\theta^0)\frac{\partial h_t}{\partial\beta'}(\theta^0) \\ \frac{\partial h_t}{\partial\beta}(\theta^0)\frac{\partial h_t}{\partial\alpha'}(\theta^0) & \frac{\partial h_t}{\partial\beta}(\theta^0)\frac{\partial h_t}{\partial\beta'}(\theta^0) \end{bmatrix}\right\} Z = 0. \text{ Hence}$$

$$Z' \ E \left\{ \begin{bmatrix} \frac{\partial m_t}{\partial \alpha}(\alpha^0)\frac{\partial m_t}{\partial \alpha'}(\alpha^0) + \frac{\partial h_t}{\partial \alpha}(\theta^0)\frac{\partial h_t}{\partial \alpha'}(\theta^0) & \frac{\partial h_t}{\partial \alpha}(\theta^0)\frac{\partial h_t}{\partial \beta'}(\theta^0) \\ \frac{\partial h_t}{\partial \beta}(\theta^0)\frac{\partial h_t}{\partial \alpha'}(\theta^0) & \frac{\partial h_t}{\partial \beta}(\theta^0)\frac{\partial h_t}{\partial \beta'}(\theta^0) \end{bmatrix} \right\} Z = 0, \text{ that is}$$

$Z_1' E\left[\frac{\partial m_t}{\partial \alpha}(\alpha^0)\right] \ Z_1 + Z' \ E\left[\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)\right] \ Z = 0$. The two terms are nonnegatives. Hence,

$Z_1' E\left[\frac{\partial m_t}{\partial \alpha}(\alpha^0)\right] \ Z_1 = 0$ and $Z' \ E\left[\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)\right] \ Z = 0$. By the first part of Assumption 2'b, we conclude that $Z_1 = 0$. Then, by the second part of this assumption, we conclude that $Z_2 = 0$.□

Appendix B

Proof of Proposition 2.1: By the usual Jennrich (1969) argument, it is sufficient to check that, when T goes to infinity, (2.12) defines an asymptotic minimization program of which only solution is θ^0. The objective limit function is 2 $E[q_t^\gamma(\theta, \theta^0, \omega^*)]$. We have:

$E[q_t^\gamma(\theta, \theta^0, \omega^*)] = E[(\varepsilon_t(\theta), \varepsilon_t^2(\theta^0) - h_t(\theta))\Lambda_t(\omega^*)(\varepsilon_t(\theta), \varepsilon_t^2(\theta^0) - h_t(\theta))']$

Let us define $X_t(\theta) = (\varepsilon_t(\theta), \varepsilon_t^2(\theta^0) - h_t(\theta))'$. Straightforward calculus show that:

$E[q_t^\gamma(\theta, \theta^0, \omega^*)] - E[q_t^\gamma(\theta^0, \theta^0, \omega^*)] = E[(X_t(\theta) - X_t(\theta^0))'\Lambda_t(\omega^*)(X_t(\theta) + X_t(\theta^0))]$

$= E[(X_t(\theta) - X_t(\theta^0))'\Lambda_t(\omega^*)(X_t(\theta) - X_t(\theta^0))] + 2E[(X_t(\theta) - X_t(\theta^0))'\Lambda_t(\omega^*)X_t(\theta^0)]$

We have $X_t(\theta) - X_t(\theta^0) = (m_t(\theta^0) - m_t(\theta), h_t(\theta^0) - h_t(\theta))'$

which is, by Assumption 2, I_{t-1}-adapted. This is also the case for $\Lambda_t(\omega^*)$ (by A.3.3). We have also, by Assumption 1, $E[X_t(\theta^0) \mid I_{t-1}] = 0$. Hence:

$E[(X_t(\theta) - X_t(\theta^0))'\Lambda_t(\omega^*)X_t(\theta^0)] = 0$, and then

$E[q_t^\gamma(\theta, \theta^0, \omega^*)] - E[q_t^\gamma(\theta^0, \theta^0, \omega^*)] = E[(X_t(\theta) - X_t(\theta^0))'\Lambda_t(\omega^*)(X_t(\theta) - X_t(\theta^0))] \geq 0.$

In other words, θ^0 is an argminimum of the function $E[q_t^\gamma(\theta, \theta^0, \omega^*)]$. To complete the proof, we need to prove that θ^0 is the unique minimum. Let us consider another minimum θ^*. We have:

$E[q_t^\gamma(\theta^*, \theta^0, \omega^*)] - E[q_t^\gamma(\theta^0, \theta^0, \omega^*)] = 0 = E[(X_t(\theta^*) - X_t(\theta^0))'\Lambda_t(\omega^*)(X_t(\theta^*) - X_t(\theta^0))].$

Hence $\quad (X_t(\theta^*) - X_t(\theta^0))'\Lambda_t(\omega^*)(X_t(\theta^*) - X_t(\theta^0)) = 0.$

By A.3.5, $\Lambda_t(\omega^*)$ is definite positive; Hence $X_t(\theta^*) = X_t(\theta^0)$ and by Assumption 2, $\theta^* = \theta^0$.□

Proof of Proposition 2.2: The estimator $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma)$ ($\hat{\theta}_T$ hereafter) is defined by (see Assumption 4): $\sum_{t=1}^T s_t^\gamma(\hat{\theta}_T, \tilde{\theta}_T, \omega_T) = 0$. By a Taylor expansion around $(\theta^0, \theta^0, \omega^*)$, we get:

$$\sum_{t=1}^T s_t^\gamma(\theta^0, \theta^0, \omega^*) + \sum_{t=1}^T \frac{\partial s_t^\gamma}{\partial \theta'}(\theta^0, \theta^0, \omega^*)(\hat{\theta}_T - \theta^0)$$

$$+ \sum_{t=1}^T \frac{\partial s_t^\gamma}{\partial \lambda'}(\theta^0, \theta^0, \omega^*)(\tilde{\theta}_T - \theta^0) + \sum_{t=1}^T \frac{\partial s_t^\gamma}{\partial \omega'}(\theta^0, \theta^0, \omega^*)(\omega_T - \omega^*) = o_P(1)$$

We have $\quad s_t^{\gamma}(\theta,\lambda,\omega) = -\left[\frac{\partial m_t}{\partial \theta}(\theta), \frac{\partial h_t}{\partial \theta}(\theta)\right] \Lambda_t(\omega) \begin{bmatrix} \varepsilon_t(\theta) \\ \varepsilon_t^2(\lambda) - h_t(\theta) \end{bmatrix}.$

Hence $\quad \frac{\partial s_t^{\gamma}}{\partial \lambda'}(\theta,\lambda,\omega) = -\left[\frac{\partial m_t}{\partial \theta}(\theta), \frac{\partial h_t}{\partial \theta}(\theta)\right] \Lambda_t(\omega) \begin{bmatrix} 0 \\ -2\frac{\partial m_t}{\partial \lambda'}(\lambda)\varepsilon_t(\lambda) \end{bmatrix},$

then $E[\frac{\partial s_t^{\gamma}}{\partial \lambda'}(\theta^0,\theta^0,\omega^*)] = E\left[-\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t(\omega^*) \begin{bmatrix} 0 \\ -2\frac{\partial m_t}{\partial \lambda'}(\theta^0)\varepsilon_t(\theta^0) \end{bmatrix}\right] = 0.$

Let ω_i a component of ω. We have $\frac{\partial s_t^{\gamma}}{\partial \omega_i}(\theta,\lambda,\omega) = -\left[\frac{\partial m_t}{\partial \theta}(\theta), \frac{\partial h_t}{\partial \theta}(\theta)\right] \frac{\partial \Lambda_t}{\partial w_i}(\omega) \begin{bmatrix} \varepsilon_t(\theta) \\ \varepsilon_t^2(\lambda) - h_t(\theta) \end{bmatrix}.$

Then $E[\frac{\partial s_t^{\gamma}}{\partial \omega'}(\theta^0,\theta^0,\omega^*)] = E\left[-\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \frac{\partial \Lambda_t}{\partial w_i}(\omega^*) \begin{bmatrix} \varepsilon_t(\theta^0) \\ \varepsilon_t^2(\theta^0) - h_t(\theta^0) \end{bmatrix}\right] = 0.$

Hence: $\quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t^{\gamma}(\theta^0,\theta^0,\omega^*) + [\frac{1}{T}\sum_{t=1}^{T} \frac{\partial s_t^{\gamma}}{\partial \theta'}(\theta^0,\theta^0,\omega^*)]\ \sqrt{T}(\hat{\theta}_T - \theta^0) = o_P(1)$

and $\quad \sqrt{T}(\hat{\theta}_T - \theta^0) = -A_{\gamma}^{0^{-1}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t^{\gamma}(\theta^0,\theta^0,\omega^*) + o_P(1)$. By Assumption 4, we conclude that $\sqrt{T}(\hat{\theta}_T - \theta^0)$ is asymptotically normal, with asymptotic covariance matrix equal to $A_{\gamma}^{0^{-1}} B_{\gamma}^{0} A_{\gamma}^{0^{-1}}$.□

Proof of Proposition 2.3: From Proposition 2.2, we have

$\sqrt{T}(\hat{\theta}_T(\tilde{\theta},\omega_T,\gamma^Q) - \theta^0) = -A_{\gamma^Q}^{0^{-1}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t^{\gamma^Q}(\theta^0,\theta^0,\omega^*) + o_P(1)$, with

$$s_t^{\gamma^Q}(\theta^0,\theta^0,\omega^*) = -\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^Q(\omega^*) \begin{bmatrix} \varepsilon_t(\theta^0) \\ \varepsilon_t^2(\theta^0) - h_t(\theta^0) \end{bmatrix}$$

$= \frac{1}{h_t(\theta^0)} \frac{\partial m_t}{\partial \theta}(\theta^0)\varepsilon_t(\theta^0) + \frac{1}{2h_t^2(\theta^0)} \frac{\partial h_t}{\partial \theta}(\theta^0)\nu_t(\theta^0)$, and

$A_{\gamma^Q}^{0} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial s_t^{\gamma^Q}}{\partial \theta'}(\theta^0,\theta^0,\omega^*)\right]$ where

$$E\left[\frac{\partial s_t^{\gamma^Q}}{\partial \theta'}(\theta^0,\theta^0,\omega^*)\right] = E\left[\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^Q(\omega^*) \begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right]$$

$$= E\left[\frac{1}{h_t(\theta^0)} \frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0) + \frac{1}{2h_t^2(\theta^0)} \frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)\right].$$

The estimator $\hat{\theta}_T^Q$ is defined by $0 = \sum_{t=1}^{T} s_t(\hat{\theta}_T^Q)$ where $s_t(\theta)$ is defined by (2.4). By Taylor expansion around θ^0, we get: $\sqrt{T}(\hat{\theta}_T^Q - \theta^0) = -A^{0^{-1}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t(\theta^0) + o_P(1)$,

where A^0 is defined by $A^0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial s_t}{\partial \theta'}(\theta^0)\right]$ and

$$E\left[\frac{\partial s_t}{\partial \theta'}(\theta^0)\right] = E\left[\frac{1}{h_t(\theta^0)} \frac{\partial m_t}{\partial \theta}(\theta^0) \frac{\partial m_t}{\partial \theta'}(\theta^0) + \frac{1}{2h_t^2(\theta^0)} \frac{\partial h_t}{\partial \theta}(\theta^0) \frac{\partial h_t}{\partial \theta'}(\theta^0)\right].$$

In other words, we have: $s_t^{\gamma^Q}(\theta^0) = s_t(\theta^0)$ and $E[\frac{\partial s_t^{\gamma^Q}}{\partial \theta'}(\theta^0)] = E[\frac{\partial s_t}{\partial \theta}(\theta^0)]$, then $A_{\gamma^Q}^0 = A^0$. Hence:

$\sqrt{T}(\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma^Q) - \hat{\theta}_T^Q) = o_P(1)$, that is $\hat{\theta}_T(\tilde{\theta}_T, \omega_T, \gamma^Q)$ and $\hat{\theta}_T^Q$ are asymptotically equivalent.□

Proof of Theorem 2.1: This proof is adapted from Newey (1993, page 423). Let

$$z_t^{\gamma} = \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t(\omega^*) \begin{bmatrix} \varepsilon_t(\theta^0) \\ \\ \nu_t(\theta^0) \end{bmatrix} \text{ and } z_t^{\gamma^*} = \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Sigma_t(\theta^0)^{-1} \begin{bmatrix} \varepsilon_t(\theta^0) \\ \\ \nu_t(\theta^0) \end{bmatrix}.$$

We have $E[z_t^{\gamma} z_t^{\gamma^{*\prime}}] = E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t(\omega^*) \begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right\}$, and

$$E[z_t^{\gamma} z_t^{\gamma'}] = E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t(\omega^*) \Sigma_t(\theta^0)^{-1} \Lambda_t(\omega^*) \begin{bmatrix} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix}\right\}.$$

Of course, we have $A_{\gamma^*}^0 = B_{\gamma^*}^0$. Hence:

$A_{\gamma}^{0^{-1}} B_{\gamma}^0 A_{\gamma}^{0^{-1}} - A_{\gamma^*}^{0^{-1}} B_{\gamma^*}^0 A_{\gamma^*}^{0^{-1}}$

$$= (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma^{*\prime}}])^{-1} (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma'}]) (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma'}])^{-1} - (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma^{*\prime}}])^{-1}$$

$$= (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma^{*\prime}}])^{-1} \left\{(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma'}])\right.$$

$$\left. -(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma^{*\prime}}]) (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma^{*\prime}}])^{-1} (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma'}])\right\} (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma'}])$$

$= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[R_t R_t']$, with

$$R_t = (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma^{*\prime}}])^{-1} \left\{z_t^{\gamma} - (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma} z_t^{\gamma^{*\prime}}]) (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[z_t^{\gamma^*} z_t^{\gamma^{*\prime}}])^{-1} z_t^{\gamma^*}\right\}. \square$$

Proof of Theorem 4.1: $\hat{\theta}_T^*$ is conformable to the large family of estimators defined by Proposition 2.2 with

$$\Lambda_t^*(\omega_T) = \begin{bmatrix} \dfrac{\hat{a}_{t,T}^*}{h_t(\tilde{\theta}_T)} & \dfrac{\hat{c}_{t,T}^*}{h_t^{3/2}(\tilde{\theta}_T)} \\ \dfrac{\hat{c}_{t,T}^*}{h_t^{3/2}(\tilde{\theta}_T)} & \dfrac{\hat{b}_{t,T}^*}{h_t^2(\tilde{\theta}_T)} \end{bmatrix} \text{ and } \omega_T = \tilde{\theta}_T.$$

Hence, by Proposition 2.1 $\hat{\theta}_T^*$ is consistent and by Proposition 2.2, it asymptotically normal with asymptotic covariance matrix equal to $(A^{*0})^{-1}B^{*0}(A^{*0})^{-1}$ with

$$A^{0*} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^*(\theta^0) \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\},$$

$$B^{0*} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^*(\theta^0)\Sigma_t(\theta^0)\Lambda_t^*(\theta^0) \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\}.$$

To complete the proof, it is sufficient to show that is asymptotic covariance matrix is equal to $A^{0^{-1}}$ defined by Theorem 2.1. We have (by (2.14) and (2.18)): $\Lambda_t^*(\theta^0) = \Sigma_t(\theta^0)$. Hence $A^{0*} = A^0$ and $B^{0*} = B^0$ and then $(A^{0*})^{-1}B^{0*}(A^{0*})^{-1} = (A^0)^{-1}$.□

Proof of Theorem 4.2 Let us denote by $\hat{\theta}_T^J$ the estimator conformable to the weighting matrix defined by $(\Sigma_t^J(\theta^0))^{-1}$. By Propositions 2.1 and 2.2, we know that this estimator is consistent and asymptotically normal with asymptotic covariance matrix equal to $(A^{J*0})^{-1}B^{J*0}(A^{J*0})^{-1}$ with

$$A^{0*J} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] (\Sigma_t^J(\theta^0))^{-1} \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\},$$

$$B^{0*J} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] (\Sigma_t^J(\theta^0))^{-1}\Sigma_t(\theta^0)(\Sigma_t^J(\theta^0))^{-1} \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\}.$$

We have:

$$B^{0*J} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] (\Sigma_t^J(\theta^0))^{-1}\Sigma_t(\theta^0)(\Sigma_t^J(\theta^0))^{-1} \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\} \mid J_{t-1} \right\}$$

$$= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] (\Sigma_t^J(\theta^0))^{-1} E\left\{\Sigma_t(\theta^0) \mid J_{t-1}\right\} (\Sigma_t^J(\theta^0))^{-1} \begin{bmatrix} \dfrac{\partial m_t}{\partial \theta'}(\theta^0) \\ \dfrac{\partial h_t}{\partial \theta'}(\theta^0) \end{bmatrix} \right\}$$

$$= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right](\Sigma_t^J(\theta^0))^{-1}(\Sigma_t^J(\theta^0))(\Sigma_t^J(\theta^0))^{-1}\left[\begin{array}{c}\frac{\partial m_t}{\partial \theta'}(\theta^0)\\ \frac{\partial h_t}{\partial \theta'}(\theta^0)\end{array}\right]\right\}$$

$$= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right](\Sigma_t^J(\theta^0))^{-1}\left[\begin{array}{c}\frac{\partial m_t}{\partial \theta'}(\theta^0)\\ \frac{\partial h_t}{\partial \theta'}(\theta^0)\end{array}\right]\right\}.$$

Define a weighting matrix $\Lambda_t(\omega) \in J_{t-1}$ and $\hat{\theta}_T$ the corresponding estimator. Its asymptotic covariance matrix is $(A_J^0)^{-1}B_J^0(A_J^0)^{-1}$ with

$$A_J^0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\Lambda_t(\omega)\left[\begin{array}{c}\frac{\partial m_t}{\partial \theta'}(\theta^0)\\ \frac{\partial h_t}{\partial \theta'}(\theta^0)\end{array}\right]\right\},$$

$$B_J^0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\Lambda_t(\omega)\Sigma_t(\theta^0)\Lambda_t(\omega)\left[\begin{array}{c}\frac{\partial m_t}{\partial \theta'}(\theta^0)\\ \frac{\partial h_t}{\partial \theta'}(\theta^0)\end{array}\right]\right\}.$$

By the same argument as for B^{0*J}, we can prove that:

$$B_J^0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right]\Lambda_t(\omega)\Sigma_t^J(\theta^0)\Lambda_t(\omega)\left[\begin{array}{c}\frac{\partial m_t}{\partial \theta'}(\theta^0)\\ \frac{\partial h_t}{\partial \theta'}(\theta^0)\end{array}\right]\right\}.$$

With these formulas, it is clear that by the same argument than in the proof of Theorem 2.1, we can prove that $(A_J^0)^{-1}B_J^0(A_J^0)^{-1} - (A^{0*J})^{-1}B^{0*J}(A^{0*J})^{-1}$ is positive, that is $\hat{\theta}_T^J$ is of minimum asymptotic covariance matrix in the class $\mathcal{C}^{\mathbf{J}}$.□

Proof of Theorem 4.3: This a direct application of the Theorem 4.2 with $\mathcal{C}^{\mathbf{J}} = \mathcal{C}^{\mathbf{I}*}$. In this case:

$$\Sigma_t^{I^*}(\theta^0) = E[\Sigma_t(\theta^0) \mid I_{t-1}^*] = \left[\begin{array}{ll} h_t(\theta^0) & h_t^{3/2}(\theta^0)E[M_{3t}(\theta^0) \mid I_{t-1}^*] \\ h_t^{3/2}(\theta^0)E[M_{3t}(\theta^0) \mid I_{t-1}^*] & h_t^2(\theta^0)(3E[K_t(\theta^0) \mid I_{t-1}^*] - 1) \end{array}\right]$$

$$= \left[\begin{array}{ll} h_t(\theta^0) & h_t^{3/2}(\theta^0)M_3(\theta^0) \\ h_t^{3/2}(\theta^0)M_3(\theta^0) & h_t^2(\theta^0)(3K(\theta^0) - 1) \end{array}\right].\square$$

Proof of Proposition 4.1: Define Λ_t^c by

$$\Lambda_t^c = \left[\begin{array}{cc} \frac{a}{h_t(\theta^0)} & \frac{c}{h_t^{3/2}(\theta^0)} \\ \frac{c}{h_t^{3/2}(\theta^0)} & \frac{b}{h_t^2(\theta^0)} \end{array}\right].$$

The corresponding estimator has an asymptotic covariance matrix equal to $(A_C^0)^{-1}B_C^0(A_C^0)^{-1}$ with

$$A_C^0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^c \left[\begin{array}{c} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{array}\right]\right\},$$

$$B_c^0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^c \Sigma_t(\theta^0)\Lambda_t^c \left[\begin{array}{c} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{array}\right]\right\}.$$

We have, by the definition of Λ_t^c and by (2.14):

$$\Lambda_t^c \Sigma_t(\theta^0)\Lambda_t^c = \left[\begin{array}{cc} \frac{a^2+2acM_{3t}(\theta^0)+c^2(3K_t(\theta^0)-1)}{h_t(\theta^0)} & \frac{ac+(c^2+ab)M_{3t}(\theta^0)+cb(3K_t(\theta^0)-1)}{h_t^{3/2}(\theta^0)} \\ \frac{ac+(c^2+ab)M_{3t}(\theta^0)+cb(3K_t(\theta^0)-1)}{h_t^{3/2}(\theta^0)} & \frac{c^2+2bcM_{3t}(\theta^0)+b^2(3K_t(\theta^0)-1)}{h_t^2(\theta^0)} \end{array}\right].$$

Hence $\left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^c \Sigma_t(\theta^0)\Lambda_t^c \left[\begin{array}{c} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{array}\right]$

$$= \left[\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0)\frac{a^2+2acM_{3t}(\theta^0)+c^2(3K_t(\theta^0)-1)}{h_t(\theta^0)}\right]$$
$$+\left[\left(\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)+\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0)\right)\frac{ac+(c^2+ab)M_{3t}(\theta^0)+cb(3K_t(\theta^0)-1)}{h_t^{3/2}(\theta^0)}\right]$$
$$+\left[\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)\frac{c^2+2bcM_{3t}(\theta^0)+b^2(3K_t(\theta^0)-1)}{h_t^2(\theta^0)}\right].$$

By the orthogonality conditions of Proposition 4.1, we conclude that:

$$E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^c \Sigma_t(\theta^0)\Lambda_t^c \left[\begin{array}{c} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{array}\right]\right\}$$
$$= E\left[\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0)\right] E\left[\frac{a^2+2acM_{3t}(\theta^0)+c^2(3K_t(\theta^0)-1)}{h_t(\theta^0)}\right]$$
$$+E\left[\left(\frac{\partial m_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)+\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial m_t}{\partial \theta'}(\theta^0)\right)\right] E\left[\frac{ac+(c^2+ab)M_{3t}(\theta^0)+cb(3K_t(\theta^0)-1)}{h_t^{3/2}(\theta^0)}\right]$$
$$+E\left[\frac{\partial h_t}{\partial \theta}(\theta^0)\frac{\partial h_t}{\partial \theta'}(\theta^0)\right] E\left[\frac{c^2+2bcM_{3t}(\theta^0)+b^2(3K_t(\theta^0)-1)}{h_t^2(\theta^0)}\right]$$
$$= E\left\{ \left[\frac{\partial m_t}{\partial \theta}(\theta^0), \frac{\partial h_t}{\partial \theta}(\theta^0)\right] \Lambda_t^c \Sigma_t^c(\theta^0)\Lambda_t^c \left[\begin{array}{c} \frac{\partial m_t}{\partial \theta'}(\theta^0) \\ \frac{\partial h_t}{\partial \theta'}(\theta^0) \end{array}\right]\right\},\ \text{where}$$

$$\Sigma_t^c(\theta^0) = \begin{bmatrix} \frac{1}{h_t(\theta^0)} & \frac{M_3(\theta^0)}{h_t^{3/2}(\theta^0)} \\ \frac{M_3(\theta^0)}{h_t^{3/2}(\theta^0)} & \frac{(3K(\theta^0)-1)}{h_t^2(\theta^0)} \end{bmatrix}.$$

With this formulas, and by an argument similar to the proof of Theorem 2.1 (or Theorem 4.3), we complete the proof.□

Proof of Theorem 4.4: The estimators $\hat{\theta}_T^1$ and $\hat{\theta}_T^2$ are respectively defined by:

$$\sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\hat{\theta}_T^1, \tilde{\theta}_T)\Lambda_{t,T}(\tilde{\theta}_T)\phi(\hat{\theta}_T^1, \tilde{\theta}_T) = 0 \text{ and}$$

$$\sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\tilde{\theta}_T, \tilde{\theta}_T)\Lambda_{t,T}(\tilde{\theta}_T)[\phi(\tilde{\theta}_T, \tilde{\theta}_T) + \frac{\partial \phi}{\partial \theta'}(\tilde{\theta}_T, \tilde{\theta}_T)(\hat{\theta}_T^2 - \tilde{\theta}_T)] = 0.$$

For sake of notational simplicity, we replace $\Lambda_{t,T}(\tilde{\theta})$ by $\Lambda_t(\theta^0)$ without changing the asymptotic probability distributions of $\hat{\theta}_T^1$ and $\hat{\theta}_T^2$ (see e.g. Proposition 2.2). Then, with a Taylor expansion around (θ^0, θ^0) (of the two functions at the points $(\hat{\theta}_T^1, \tilde{\theta}_T)$ and $(\hat{\theta}_T^2, \tilde{\theta}_T)$), and since $E[\phi_t(\theta^0, \theta^0) \mid I_{t-1}] = 0$ we get for $\hat{\theta}_T^1$:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\phi(\theta^0, \theta^0) + \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0)\ \sqrt{T}(\hat{\theta}_T^1 - \theta^0)$$

$$+\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\frac{\partial \phi}{\partial \lambda'}(\theta^0, \theta^0)\ \sqrt{T}(\tilde{\theta}_T - \theta^0) = o_P(1)$$

where $\frac{\partial \phi_t}{\partial \lambda'}$ denotes the jacobian matrix of ϕ_t with respect to its second occurrence, and for $\hat{\theta}_T^2$:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)[\phi(\theta^0, \theta^0) + \frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0)(\theta^0 - \theta^0)]$$

$$+\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0)\ \sqrt{T}(\hat{\theta}_T^2 - \theta^0)$$

$$+\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\left\{\frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0) + \frac{\partial \phi}{\partial \lambda'}(\theta^0, \theta^0) - \frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0)\right\}\ \sqrt{T}(\tilde{\theta}_T - \theta^0) = o_P(1).$$ Hence

$$\sqrt{T}(\hat{\theta}_T^1 - \theta^0) = -(A_1^0)^{-1}\left\{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\phi(\theta^0, \theta^0) + \tilde{A}^0\ \sqrt{T}(\tilde{\theta}_T - \theta^0)\right\} + o_P(1),$$

$$\sqrt{T}(\hat{\theta}_T^2 - \theta^0) = -(A_1^0)^{-1}\left\{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\phi(\theta^0, \theta^0) + \tilde{A}^0\ \sqrt{T}(\tilde{\theta}_T - \theta^0)\right\} + o_P(1),$$

with $A_1^0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\frac{\partial \phi}{\partial \theta'}(\theta^0, \theta^0)\right]$ and

$$\tilde{A}^0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial \phi'}{\partial \theta}(\theta^0, \theta^0)\Lambda_t(\theta^0)\frac{\partial \phi}{\partial \lambda'}(\theta^0, \theta^0)\right].$$

We conclude that: $\sqrt{T}(\hat{\theta}_T^2 - \hat{\theta}_T^1) = o_P(1)$, that is $\hat{\theta}_T^2$ and $\hat{\theta}_T^1$ are asymptotically equivalent.□

Table 1: Gaussian errors

	a (OLS)	b (QMLE)	C1 (Initial OLS)	C2 (Initial QMLE)	C3 (Iterated OLS)	C4 (Iterated QMLE)
c	1.022 (0.129)	1.015 (0.077)	1.014 (0.078)	1.015 (0.076)	1.014 (0.076)	1.014 (0.076)
ρ	0.694 (0.038)	0.696 (0.022)	0.696 (0.023)	0.696 (0.022)	0.696 (0.22)	0.696 (0.022)
ω	0.576 (0.090)	0.502 (0.035)	0.506 (0.037)	0.502 (0.035)	0.502 (0.035)	0.501 (0.035)
α	0.416 (0.106)	0.496 (0.063)	0.484 (0.068)	0.488 (0.067)	0.496 (0.063)	0.496 (0.063)

Table 2: Student errors

	a (OLS)	b (QMLE)	C1 (Initial OLS)	C2 (Initial QMLE)	C3 (Iterated OLS)	C4 (Iterated QMLE)
c	1.021 (0.157)	1.016 (0.096)	1.014 (0.092)	1.016 (0.086)	1.015 (0.086)	1.014 (0.086)
ρ	0.694 (0.046)	0.695 (0.027)	0.696 (0.026)	0.695 (0.025)	0.696 (0.025)	0.696 (0.025)
ω	0.654 (0.183)	0.505 (0.059)	0.522 (0.104)	0.509 (0.060)	0.513 (0.100)	0.507 (0.060)
α	0.329 (0.124)	0.498 (0.153)	0.466 (0.127)	0.470 (0.121)	0.482 (0.121)	0.492 (0.139)

Table 3: Gamma errors

	a (OLS)	b (QMLE)	C1 (Initial OLS)	C2 (Initial QMLE)	C3 (Iterated OLS)	C4 (Iterated QMLE)
c	1.056 (0.145)	1.011 (0.106)	1.013 (0.075)	1.011 (0.066)	1.001 (0.066)	1.001 (0.065)
ρ	0.682 (0.046)	0.696 (0.031)	0.696 (0.022)	0.696 (0.019)	0.697 (0.019)	0.697 (0.018)
ω	0.711 (0.443)	0.499 (0.060)	0.509 (0.062)	0.501 (0.052)	0.503 (0.052)	0.502 (0.052)
α	0.279 (0.120)	0.502 (0.145)	0.480 (0.130)	0.474 (0.108)	0.494 (0.103)	0.495 (0.102)