Working Paper n° 08-02

Fourth Order Pseudo Maximum Likelihood Methods

Alberto Holly, Alain Monfort and Michael Rockinger

UNIL | Université de Lausanne

IEMS – Institute of Health Economics and Management

Fourth Order Pseudo Maximum Likelihood Methods

Alberto Holly[1], Alain Monfort[2] and Michael Rockinger[3]

[1] Institute of Health Economics and Management(IEMS), University of Lausanne, Faculty of Business and Economics, Extranef Building, CH-1015 Lausanne, Switzerland.
E-mail: alberto.holly@unil.ch
[2] Corresponding author. CNAM and CREST, 15 Boulevard Péri, 92245 Malakoff Cédex, France.
E-mail: alain.monfort@ensae.fr
[3] Swiss Finance Institute and CEPR, University of Lausanne, Faculty of Business and Economics, Extranef Building, CH-1015 Lausanne, Switzerland.
E-mail: Michael.rockinger@uni.ch

Working Paper n° 08-02

August 2008

Institute of Health Economics and Management (IEMS)
UNIL Dorigny
Extranef
1015 Lausanne
Switzerland
Phone +41 (0)21 692 33 20
Fax +41 (0)21 692 36 55

www.hec.unil.ch/iems

Fourth Order Pseudo Maximum Likelihood Methods

by Alberto Holly[a], Alain Monfort[b],

and Michael Rockinger[c]

August 2008

[a] Institute of Health Economics and Management (IEMS), University of Lausanne, Faculty of Business and Economics, Extranef Building, CH-1015 Lausanne, Switzerland. E-mail: alberto.holly@unil.ch.

[b] Corresponding author. CNAM and CREST, 15 Boulevard Péri, 92245 Malakoff Cédex, France, E-mail: alain.monfort@ensae.fr.

[c] Swiss Finance Institute, University of Lausanne, and CEPR. University of Lausanne. Faculty of Business and Economics, Extranef Building, CH-1015 Lausanne, Switzerland. E-mail: michael.rockinger@unil.ch.

Acknowledgement: The third author acknowledges support from the Swiss National Science Foundation through NCCR FINRISK (Financial Valuation and Risk Management). We are grateful to Prof. W. Gautschi for his advice concerning numerical integration and to Prof. G. V. Milovanović for providing a very useful sequence of parameters used for numerical integrations.

Keywords: Quartic Exponential Family, Pseudo Maximum Likelihood, Skewness, Kurtosis.

JEL classification: C01, C13, C16, C22

Abstract

The objective of this paper is to extend the results on Pseudo Maximum Likelihood (PML) theory derived in Gourieroux, Monfort, and Trognon (GMT) (1984) to a situation where the first four conditional moments are specified. Such an extension is relevant in light of pervasive evidence that conditional distributions are non-Gaussian in many economic situations. The key statistical tool here is the quartic exponential family, which allows us to generalize the PML2 and QGPML1 methods proposed in GMT(1984) to PML4 and QGPML2 methods, respectively. An asymptotic theory is developed which shows, in particular, that the QGPML2 method reaches the semi-parametric bound. The key numerical tool that we use is the Gauss-Freud integration scheme which solves a computational problem that has previously been raised in several econometric fields. Simulation exercises show the feasibility and robustness of the methods.

1. Introduction

It is well known that the Maximum Likelihood estimator may not only be inefficient but also inconsistent under misspecification. Specifically, this occurs when the parametric model providing the likelihood function does not contain the true distribution. The study of the relations between Maximum Likelihood Theory and misspecification has now a long history. Hood and Koopmans (1953) demonstrated that the conditionally Gaussian ML estimator is consistent and asymptotically Gaussian, even if the true distribution is not conditionally Gaussian, as soon as the first two conditional moments are well specified. They coined the label "quasi ML estimator" for this kind of estimator. White (1982) showed that, under misspecification, the ML estimator is in fact a CAN (consistent asymptotically normal) estimator of the pseudo-true value (as defined for instance in Sawa (1978)). Gourieroux, Monfort, and Trognon (1984) characterized the parametric families leading to CAN estimators of the parameters appearing in the first two conditional moments, even if the true distribution does not belong to this parametric family. These families are the linear exponential families (when only the first conditional moment is specified) and the quadratic exponential families (when the first two conditional moments are specified). The estimators thus obtained were called PML1 and PML2 estimators, respectively. Bollerslev and Wooldrige (1992) generalized the properties of the quasi generalized estimator, i.e. the Gaussian PML estimator, to the dynamic case.

The PML theories described above only consider the robustness of the estimator of the parameters appearing in the first two conditional moments. More recently, however, many econometric fields are paying greater attention to higher order conditional moments. This is particularly the case, in Financial Econometrics and in Health Econometrics. Very often, the approach used to account for higher moments is the ML method based on a choice of a parametric family, which allows for asymmetry and fat tails. A few such examples, occurring in finance are: Generalized Hyperbolic distribution [Eberlein and Keller (1995); Barndorff-Nielsen (1997)], the noncentral Student t distribution [Harvey and Siddique (1999)], and the Skewed-t distribution [Hansen (1994); Jondeau and

Rockinger (2003)]. In Health Econometrics, some examples are: The Generalized Gamma distribution proposed by Stacy (1962) and Stacy and Mihram (1965), [Manning, Basu, and Mullahy (2005)], and the Pearson Type IV distribution, which may be considered a skewed-t distribution [Holly and Pentsak (2004)]. More generally, examples can be found in various fields, including physics, astronomy, image processing, and in the biomedical sciences [see Genton (2004) and Arellano-Valle and Genton (2005)].

Current ML approaches have two types of drawbacks. First, some families may not be flexible enough to span the whole set of possible skewness (s) and kurtosis (k), namely the domain $k \geq s^2+1$. Second, as mentioned above, the risk of misspecification may lead to inconsistent estimators. If we are interested in the first four conditional moments, a natural method is the Generalized Method of Moments (GMM). There is now a large body of studies, however, suggesting that GMM estimators can have poor finite sample properties [see e.g. Tauchen (1986), Andersen and Sorenson (1996), Altonji and Segal (1996), Ziliak (1997), Doran and Schmidt (2006)]. These difficulties led Kitamura and Stutzer (1997), relatively early on, to propose an alternative estimation procedure based on the Kullback-Leibler Information Criterion.

The objective of this paper is to propose another alternative to GMM. It is an extension of the PML method developed by Gourieroux, Monfort, and Trognon (1984), henceforth referred to as GMT. This work extends GMT to a situation where the first four moments (centered or not) are known functions depending on unknown parameters. Specifically, we show that the PML estimator is consistent for any specification of the first four conditional moments, any true conditional distribution of the endogenous variable, and any marginal distribution of the exogenous variables, if and only if the PML is based on a quartic exponential family. We shall refer to this extension as PML4. We also propose an extension of the Quasi-Generalized Pseudo Maximum Likelihood (QGPML) estimator proposed by GMT (1984) based on the quartic exponential family. We show that this estimator, which we refer to as QGPML2, reaches the semiparametric bound. Beyond the robustness and nice asymptotic properties of the estimates resulting from the quartic exponential families, two additional features should be noted. First, the quartic exponential family spans the whole domain of the pairs (s, k), except for a set of measure zero. This is not necessarily the case for other parametric families of distributions such as those mentioned earlier. Second, we show how the parameters of the quartic exponential family may be obtained from a given set of moments. Thereby, we solve a numerical problem which had been already encountered, both in the econometric literature [Zellner and Highfield (1988), Ormoneit and White (1999)] and other fields [Agmon *et al.* (1979) or Mead and Papanicolaou (1984)], where the exponential family arises as an Entropy Maximizing density and which were considered as difficult. The key issue, from the computational point of view, is the use of the Gauss-Freud quadrature scheme which seems very promising for computing the numerical integrations needed in this framework.

The rest of this paper is organized as follows. Preliminary results are given in Section 2, where some properties of exponential families are briefly reviewed and the notion of a quartic exponential family is defined. This section also contains a brief presentation of the properties of M-Estimators. These preliminary results are then used to derive important properties of the exponential quartic family in Section 3. The PML4 method is defined in Section 4, and the asymptotic properties of the PML4 estimators are derived. In Section 5, we perform a similar analysis as in Section 4 but for the QGPML2 method. In Section 6, we discuss the numerical issues and describe numerical algorithms for implementing of the PML4 and QGPML2 methods. Several Monte-Carlo exercises demonstrating the usefulness of the methods proposed in our paper are presented in Section 7. This section contains a discussion on computational issues linked with the quartic exponential distribution, and it also presents four Monte-Carlo experiments, each of which numerically demonstrates a different property of PML4 or QGPML2. Conclusions are presented in Section 8. Finally, to not interrupt our discussion of the essential ideas of this paper, some proofs and other technical details are presented in the Appendix.

2. Preliminaries

2.1. **Exponential Families.**

Let us consider a measure space $(\mathcal{Y}, \mathcal{A}, \nu)$ where $\mathcal{A}$ is a $\sigma-$field and ν a $\sigma-$finite measure. An exponential family is a family of probability distributions on $(\mathcal{Y}, \mathcal{A})$ which are equivalent to ν and with pdfs of the form:

$$\ell(y, \lambda) = \exp\left[\lambda' T(y) - \psi(\lambda)\right], \qquad \lambda \in \Lambda \subset \mathbb{R}^p,$$

where $T(y)$ is a $p-$dimensional vector defined on $\mathcal{Y}$, and $\psi(\lambda)$ is a normalizing constant, equal to the Log-Laplace transform of ν^T, equivalent to the image of ν by T.

Such families have many well-known properties. Some of them will be useful in the rest of the paper, and they are summarized below [the proofs can be found for instance in Barndorff-Nielsen (1978), Monfort (1982), or Brown (1986)].

(1) Λ can be taken as the convex set where the Laplace transform of ν^T is defined.

(2) For any $\lambda \in \mathring{\Lambda}$, interior of Λ, all the moments of the statistic T exist, and in particular, we have:

$$E_\lambda(T) = \frac{\partial}{\partial \lambda}\psi(\lambda), \qquad V_\lambda(T) = \frac{\partial^2}{\partial \lambda \partial \lambda'}\psi(\lambda),$$

which implies:

$$\frac{\partial E_\lambda(T)}{\partial \lambda'} = \frac{\partial^2}{\partial \lambda \partial \lambda'}\psi(\lambda) = V_\lambda(T).$$

(3) The Fisher information matrix $I_F(\lambda)$ is equal to $V_\lambda(T) = \partial^2\psi(\lambda)/\partial\lambda\partial\lambda'$.

(4) The model is identifiable if, and only if, $I_F(\lambda)$ is invertible for any $\lambda \in \Lambda$.

(5) If the model is identifiable, then the mapping $\lambda \to E_\lambda(T)$ is injective.

2.2. **Quartic Exponential Family.**

We consider the particular case where $\mathcal{Y} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}_{\mathbb{R}}$ (the Borelian $\sigma-$field of $\mathbb{R}$), ν is the Lebesgue measure on $\mathbb{R}$, and $T(y) = (y, y^2, y^3, y^4)'$.

In other words, we consider the pdfs on $\mathbb{R}$ defined by:

$$\ell(y,\lambda) = \exp\left[\sum_{i=1}^{4} \lambda_i y^i - \psi(\lambda)\right], \qquad \text{with } \lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)'. \tag{2.1}$$

We will also use the notation:

$$\ell(y,\lambda) = \exp\left[\lambda_0 + \sum_{i=1}^{4} \lambda_i y^i\right], \qquad \text{with } \lambda_0 = -\psi(\lambda). \tag{2.2}$$

This type of density has been extensively used in the entropy literature, e.g. Golan *et al.* (1996), since it is obtained by maximizing, with respect to f, the entropy $-\int_{\mathbb{R}} f(y)\log f(y)\,dy$, under a set of data moment-consistency constraints $\int_{\mathbb{R}} y^i f(y)\ dy = m_j$, for $j = 1, \cdots, 4$, where the m_j are given, as well as a normalization constraint $\int_{\mathbb{R}} f\ (y)\,dy = 1$.

The set Λ where $\ell(y,\lambda)$ is defined is easily obtained. If $\lambda_4 < 0$, $\ell(y,\lambda)$ is always integrable. If $\lambda_4 > 0$, $\ell(y,\lambda)$ is never integrable. Finally, if $\lambda_4 = 0$, $\ell(y,\lambda)$ is integrable if $\lambda_3 = 0$ and $\lambda_2 < 0$. In other words, Λ is defined by:

$$\Lambda = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{-*} + \mathbb{R} \times \mathbb{R}^{-*} \times \{0\} \times \{0\},$$

where $\mathbb{R}^{-*}$ is the set of strictly negative numbers.

This family will be called the quartic exponential family and denoted by $\{Q(\lambda), \lambda \in \Lambda\}$. The case $\lambda_4 = \lambda_3 = 0, \lambda_2 < 0$ corresponds to the Gaussian family. Therefore, since on this set, $T(y)$ is integrable, the quartic family is not steep [see Brown (1986), p. 72]. Consequently, we cannot use the standard results developed by Barndorff-Nielsen (1978) about the range of the mapping $\lambda \to E_\lambda[T(y)]$ for steep families. For more details, see Barndorff-Nielsen (1978), Brown (1986), Letac (1992), and Appendix A.

One should note that the variance-covariance matrix of $T(Y) = (Y, Y^2, Y^3, Y^4)'$ is invertible everywhere, since otherwise there would exist a linear relation between Y, Y^2, Y^3, Y^4, i.e. the support of the distribution of Y would be made of at most four points, which is impossible since this distribution is absolutely continuous with respect to the Lebesgue measure. Therefore, using the general properties 3) and 4) we see that the model is identifiable. Moreover, using 5) we conclude that the mapping $\lambda \to [m_i(\lambda), i = 1, \ldots, 4]$, where $m_i(\lambda) = E_\lambda(Y^i)$, is injective.

Denoting by $s(\lambda)$ and $k(\lambda)$ the skewness and kurtosis $[s(\lambda) = E_\lambda[Y - E(Y)]^3 / [V_\lambda(Y)]^{3/2}$, $k(\lambda) = E_\lambda[Y - E(Y)]^4 / [V_\lambda(Y)]^2]$, respectively, it is also clear that the mapping:

$$\lambda \to [m_1(\lambda), m_2(\lambda), s(\lambda), k(\lambda)],$$

is injective. The same is true for the mapping

$$\lambda \to [m(\lambda), \sigma^2(\lambda), s(\lambda), k(\lambda)],$$

where $m(\lambda) = m_1(\lambda)$, and $\sigma^2(\lambda) = m_2(\lambda) - m_1^2(\lambda)$.

It is important to check whether the previous mapping is also surjective, which is to say that it can reach any admissible value of (m_1, m_2, s, k). It is well known that the set $\mathcal{D}$ of admissible values of (m, σ^2, s, k) is defined by:

$$m \in \mathbb{R}, \quad \sigma^2 \geq 0, \quad s \in \mathbb{R}, \quad k \geq s^2 + 1.$$

The latter inequality is obtained, for instance, by noting that the variance-covariance matrix of (Y, Y^2) where $E(Y) = 0, V(Y) = 1$, is given by

$$\begin{pmatrix} 1 & s \\ s & k-1 \end{pmatrix},$$

and therefore that $k - 1 - s^2 \geq 0$. Moreover, the boundary $k = s^2 + 1$ is reached if Y and Y^2 are linked linearly, that is, if the support is made of at most two points. Therefore, this boundary clearly cannot be reached by the quartic family, and the boundary point $\sigma^2 = 0$ cannot be reached either (for the same reason). Therefore, the natural question is now the following: is the range D of the mapping $\lambda \to [m(\lambda), \sigma^2(\lambda), s(\lambda), k(\lambda)]$ defined by:

$$\left\{m \in \mathbb{R}, \quad \sigma^2 > 0, \quad s \in \mathbb{R}, \quad k > s^2 + 1\right\} \ ?$$

The answer is no, but it can be shown [see Junk (2000)] that the range is almost equal to this set, in the sense that all the admissible values of (m, σ^2, s, k) can be reached except for those corresponding to the set of measure zero, defined by $s = 0$, $k > 3$. In particular, we can approach all the points of the half line $s = 0$, $k > 3$, in the plane (s, k), as closely as we wish.

2.3. **M-estimators and Quasi-Generalized M-estimators.**

Let us consider an endogenous variable Y_i and a vector of exogenous variables X_i. For simplicity, we assume that (Y_i, X_i) for $i = 1, \cdots, n$ are i.i.d. Standard extensions can be found in Gallant (1987), Holly (1993), or White (1994). To each possible conditional distribution of the Y_i's given the X_i's, we associate a parameter $\theta \in \Theta \subset \mathbb{R}^K$. In particular, the value of the parameter corresponding to the true conditional distribution of the Y_i's given the X_i's is called the true value of the parameter, and it is denoted by θ_0. The true distribution of the sequence $(Y_i,\ X_i,\ i \in \mathbb{N})$ is denoted by P_0. Throughout this paper, we adopt the notation corresponding to a conditional static model, but the results could be extended to a stationary conditional dynamic model by replacing Y_i by Y_t and X_i by $(Y_{t-1}, \ldots, Y_1, X_t, \ldots, X_1)$.

An M-estimator of θ_0 is an estimator $\hat{\theta}_n$ obtained by maximizing, with respect to θ, an objective function of the form:

$$\sum_{i=1}^{n} \varphi(Y_i, X_i,\ \theta). \tag{2.3}$$

Under standard regularity conditions [see e.g. Chamberlain (1987), Newey (1990), White (1994), Gourieroux and Monfort (1995a)], it can be shown that $\hat{\theta}_n$ is a consistent estimator of θ_0, for any θ_0, if the limit function $\varphi_\infty(\theta, P_0) = P_0 \lim \left[\frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i, X_i,\ \theta)\right]$ has a unique maximum at $\theta = \theta_0$. Moreover, the limit function can be written:

$$\varphi_\infty(\theta, P_0) = E_X E_0 \varphi(Y, X,\ \theta), \tag{2.4}$$

where E_0 is the conditional expectation operator associated to the true conditional distribution of Y_i, given that $X_i = x$ (independent of i) and E_X is the expectation with respect to the distribution P_X, of any X_i.

Let us now consider two subvectors θ^* and θ^{**} of θ. These subvectors are not necessarily disjoint, and in particular, we can have $\theta^* = \theta^{**} = \theta$.

We assume that θ_0^*, the true value of θ^*, can be consistently estimated by $\hat{\theta}_n^*$ defined by:

$$\hat{\theta}_n^* = \underset{\theta^*}{\text{Argmax}} \sum_{i=1}^{n} \varphi\left[Y_i, X_i, \theta^*, a\left(X_i, \hat{\theta}_n^{**}\right)\right], \tag{2.5}$$

where a is some function, and $\hat{\theta}_n^{**}$ is a consistent estimator of θ_0^{**}, the true value of θ^{**}. In other words, θ_0^* gives the unique maximum in θ^* of:

$$P_0 \lim \left[\frac{1}{n} \sum_{i=1}^{n} \varphi\left[Y_i, X_i, \theta^*, a\left(X_i, \hat{\theta}_n^{**}\right)\right]\right] \tag{2.6}$$

$$= E_X E_0 \varphi\left[Y, X, \theta^*, a\left(X, \theta_0^{**}\right)\right]. \tag{2.7}$$

Such an estimator is called a Quasi-Generalized M-estimator of θ_0^* (QGM estimator). The corresponding unfeasible M-estimator $\hat{\theta}_{0n}^*$ is defined by:

$$\hat{\theta}_{0n}^* = \underset{\theta^*}{\text{Argmax}} \sum_{i=1}^{n} \varphi\left[Y_i, X_i, \theta^*, a\left(X_i, \theta_0^{**}\right)\right], \tag{2.8}$$

and is also consistent.

As far as the asymptotic behavior of the M and QGM estimators is concerned, we have the following properties.

The asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is $\mathcal{N}\left[0, J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)\right]$ where

$$J(\theta_0) = -E_X E_0 \left[\frac{\partial^2 \varphi(Y, X, \theta_0)}{\partial\theta\partial\theta'}\right], \tag{2.9}$$

$$I(\theta_0) = E_X E_0 \left[\frac{\partial \varphi(Y, X, \theta_0)}{\partial\theta} \frac{\partial \varphi(Y, X, \theta_0)}{\partial\theta'}\right]. \tag{2.10}$$

A nice property of the QGM-estimator $\hat{\theta}_n^*$ of θ_0^* is the following. If

$$E_0 \left[\frac{\partial^2 \varphi(Y, X, \theta_0^*, a\left(X, \theta_0^{**}\right))}{\partial\theta^*\partial a'} \mid X\right] = 0, \tag{2.11}$$

then $\sqrt{n}(\hat{\theta}_n^* - \theta_0^*)$ has the same asymptotic distribution as $\sqrt{n}(\hat{\theta}_{0n}^* - \theta_0^*)$. Namely, $\mathcal{N}\left[0, \tilde{J}^{-1}(\theta_0)\tilde{I}(\theta_0)\tilde{J}^{-1}(\theta_0)\right]$, where $\tilde{J}(\theta_0)$ and $\tilde{I}(\theta_0)$ are obtained from (2.9) and (2.10), once θ has been replaced by θ^* and $\varphi(Y, X, \theta)$ by $\varphi(Y, X, \theta^*, a\left(X, \theta_0^{**}\right))$.

3. Properties of the Exponential Quartic Family

Let us denote by

$$\ell(y, \lambda) = \exp(\lambda_0 + \lambda_1 y + \lambda_2 y^2 + \lambda_3 y^3 + \lambda_4 y^4),$$

the pdf of the exponential family, and where $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)'$.

We know that the range Λ of λ is: $\Lambda = \mathbb{R}^3 \times \mathbb{R}^{-*} + \mathbb{R} \times \mathbb{R}^{-*} \times \{0\} \times \{0\}$, and let us denote by D the set of the values of (m_1, σ^2, s, k) which can be reached by this family. We have seen in Subsection 2.2 that:

$$D = \left\{ m_1 \in \mathbb{R}, \quad \sigma^2 > 0, \quad s \in \mathbb{R}, \quad k > s^2 + 1 \right\} - \left\{ s = 0, k > 3 \right\}.$$

Let us denote by M the corresponding range of $m = (m_1, m_2, m_3, m_4)'$, where $m_i = E(Y^i)$.

We know from Subsection 2.2 that the mapping $m(\lambda)$ from Λ to M is bijective, and we denote by $\lambda(m)$ the inverse function and $\lambda_0(m) = -\psi\left[\lambda(m)\right]$.

Proposition 1. *We have:*

$$\frac{\partial \lambda_0(m)}{\partial m} + \frac{\partial \lambda'(m)}{\partial m} m = 0. \tag{3.1}$$

Proof. We have:

$$\operatorname{Log} \ell\left[y, \lambda(m)\right] = \lambda_0(m) + \sum_{j=1}^{4} \lambda_j(m) y^j,$$

$$\frac{\partial \operatorname{Log} \ell\left[y, \lambda(m)\right]}{\partial m} = \frac{\partial \lambda_0(m)}{\partial m} + \sum_{j=1}^{4} \frac{\partial \lambda_j(m)}{\partial m} y^j.$$

The result follows by taking the expectation and using the fact that the score vector is of zero mean. □

Corollary 1. *We have:*

$$\frac{\partial^2 \lambda_0(m)}{\partial m \partial m'} + \sum_{j=1}^{4} \frac{\partial^2 \lambda_j(m)}{\partial m \partial m'} m_j + \frac{\partial \lambda'(m)}{\partial m} = 0. \tag{3.2}$$

Proof. The proof is straightforward by differentiating the identity (3.1) of Proposition 1 once more. □

Let us denote by Σ the variance-covariance matrix of $T(Y) = (Y, Y^2, Y^3, Y^4)'$ which is positive definite (since the support of Y is not reduced to point masses).

Proposition 2. *We have $\partial m / \partial \lambda' = \Sigma$, and therefore, $\partial \lambda / \partial m' = \Sigma^{-1}$.*

Proof. This is a direct consequence of the general property 2) of Subsection 2.1. □

Proposition 3. *For any pair $m, m_0 \in M$, we have:*

$$\lambda_0(m) + \lambda'(m) m_0 \leq \lambda_0(m_0) + \lambda'(m_0) m_0,$$

and the equality holds if and only if $m = m_0$.

Proof. From Kullback's inequality, we know that:

$$E_{m_0}\left\{\operatorname{Log} \ell\left[y, \lambda(m)\right]\right\} \leq E_{m_0}\left\{\operatorname{Log} \ell\left[y, \lambda(m_0)\right]\right\},$$

or

$$E_{m_0}\left[\lambda_0(m) + \lambda'(m)T(Y)\right] \leq E_{m_0}\left[\lambda_0(m_0) + \lambda'(m_0)T(Y)\right],$$

and the result follows. □

4. PML 4 Method

4.1. **Definition.**

We adopt a semi-parametric approach based on the specifying of the conditional moments up to fourth order. This is obviously equivalent to specifying (m_1, m_2, m_3, m_4) or (m_1, σ^2, s, k). Moreover, to satisfy the inequality $k > s^2+1$, it could be convenient to specify (m_1, σ^2, s, k^*), where $k^* = k - s^2 - 1$, which could be called the *over-kurtosis*, since (s, k^*) is only constrained to belong to $\mathbb{R}^+ \times \mathbb{R}^+$. We consider the latter parametrization, but the results could be adapted to other parametrizations in a straightforward manner.

We therefore specify the following functions: $m(x_i, \theta_1)$, $\sigma^2(x_i, \theta_2)$, $s(x_i, \theta_3)$, and $k^*(x_i, \theta_4)$. Note that θ_1, θ_2, θ_3, and θ_4 may have some components in common, and we denote by θ the union of θ_1, θ_2, θ_3, and θ_4 without repetition (in particular, we could have $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta$). We denote by Θ the range of θ.

For a given x_i and θ, we can compute the coefficients $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ of the quartic exponential distribution having the same mean, variance, skewness and kurtosis. As mentioned in Section 2, this can always be done unless the skewness is zero and the kurtosis larger than 3, but even then these values can be closely approached. Let us denote these coefficients by $\lambda_j(x_i, \theta)$, $j = 0, \ldots, 4$.

Definition 1. *The fourth order Pseudo Maximum Likelihood estimator of θ_0, called PML4 and denoted by $\hat{\theta}_n$ is defined by:*

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{Argmax}} \sum_{i=1}^{n} \sum_{j=0}^{4} \lambda_j(x_i, \theta) y_i^j .$$

Condition 1. *We assume that the semiparametric model is identifiable, i.e. that:*

(1) if $m_1(x_i, \theta_1) = m_1(x_i, \bar{\theta}_1)$, $\sigma^2(x_i, \theta_2) = \sigma^2(x_i, \bar{\theta}_2)$, $s(x_i, \theta_3) = s(x_i, \bar{\theta}_3)$, $k^*(x_i, \theta_4) = k^*(x_i, \bar{\theta}_4)$ (P_X almost surely), then we have $\theta = \bar{\theta}$.

Note that Condition 1 is equivalent to

(2) $\lambda_j(x_i, \theta) = \lambda_j(x_i, \bar{\theta})$, $j = 0, \ldots, 4$ (P_X almost surely) implies $\theta = \bar{\theta}$.

4.2. **Asymptotic Properties.**

Proposition 4. *Under standard regularity conditions, if the semi-parametric model is identifiable, the PML4 estimator $\hat{\theta}_n$ is consistent.*

Proof. From the properties of the M estimators mentioned in Subsection 2.3, we have to prove that the limit function (2.4) $\varphi_\infty(\theta, P_0) = E_X E_0 \varphi(Y, X, \theta)$ has a unique maximum at θ_0. Here we have:

$$\begin{aligned}\varphi_\infty(\theta, P_0) &= E_X E_0 \left[\sum_{j=0}^{4} \lambda_j(X, \theta) Y^j\right], \\ &= E_X \left[\lambda_0(X, \theta) + \sum_{j=1}^{4} \lambda_j(X, \theta) m_{j0}\right].\end{aligned}$$

Using Proposition 3, we know that

$$\varphi_\infty(\theta, P_0) \leq \varphi_\infty(\theta_0, P_0),$$

and that $\varphi_\infty(\theta, P_0) = \varphi_\infty(\theta_0, P_0)$ if and only if $\lambda_j(X, \theta) = \lambda_j(X, \theta_0),\ j = 0, \ldots, 4$, P_X almost surely in X, and, therefore, using the identification assumption, if and only if $\theta = \theta_0$. The result follows. □

Proposition 5. *Under standard regularity conditions, if the semi-parametric model is identifiable, $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ is asymptotically distributed as*

$$\mathcal{N}\left[0, J^{-1}(\theta_0)\, I(\theta_0)\, J^{-1}(\theta_0)\right],$$

where

$$\begin{aligned}J(\theta_0) &= E_X \left[\frac{\partial m'(X, \theta_0)}{\partial \theta} \Sigma^{-1}(X, \theta_0) \frac{\partial m(X, \theta_0)}{\partial \theta'}\right], \\ I(\theta_0) &= E_X \left[\frac{\partial m'(X, \theta_0)}{\partial \theta} \Sigma^{-1}(X, \theta_0) \Omega(X) \Sigma^{-1}(X, \theta_0) \frac{\partial m(X, \theta_0)}{\partial \theta'}\right],\end{aligned}$$

and where $\Sigma(X, \theta_0)$ is the conditional variance-covariance matrix of $T(Y) = (Y, Y^2, Y^3, Y^4)'$ given X in the quartic conditional distribution associated with $\lambda_j(X, \theta_0),\ j = 0, \ldots, 4$. Meanwhile, $\Omega(X)$ is the true conditional variance-covariance matrix of T given X.

Proof. See Appendix B. □

In the formulas giving $J(\theta_0)$ and $I(\theta_0)$ we have the Jacobian matrices $\partial m(X, \theta_0)/\ \partial \theta'$. If the parametrization used is not $m = (m_1, m_2, m_3, m_4)'$ but instead $\mu = (m_1, \sigma^2, s, k^*)'$, we must compute $\partial m(X, \theta_0)/\partial \theta'$ as a function of $\mu = [m_1(X, \theta_1), \sigma^2(X, \theta_2), s(X, \theta_3)$, and $k^*(X, \theta_4)]'$, and we get:

$$\frac{\partial m(X, \theta_0)}{\partial \theta'} = \frac{\partial m}{\partial \mu'} \frac{\partial \mu(X, \theta_0)}{\partial \theta'}.$$

Therefore:

Corollary 2. *We have,*

$$J(\theta_0) = E_X \left[\frac{\partial \mu'(X,\theta_0)}{\partial \theta} \frac{\partial m'}{\partial \mu} \Sigma^{-1}(X,\theta_0) \frac{\partial m}{\partial \mu'} \frac{\partial \mu(X,\theta_0)}{\partial \theta'} \right], \tag{4.1}$$

$$\begin{aligned} I(\theta_0) = E_X \Bigg[& \frac{\partial \mu'(X,\theta_0)}{\partial \theta} \frac{\partial m'}{\partial \mu} \Sigma^{-1}(X,\theta_0) \Omega(X) \\ & \times \Sigma^{-1}(X,\theta_0) \frac{\partial m}{\partial \mu'} \frac{\partial \mu(X,\theta_0)}{\partial \theta'} \Bigg]. \end{aligned} \tag{4.2}$$

Given this, Propositions 4 and 5 show that the PML4 method based on the quartic exponential family provides consistent and asymptotically normal estimators of certain parameters which specify the conditional moments of order one to four. It is also important to note that the unique family with these properties is the generalized quartic family

$$\exp \left[\lambda_0(m) + \sum_{i=1}^{4} \lambda_i(m) y^i + a(y) \right],$$

where $m = (m_1, m_2, m_3, m_4)$.

Proposition 6. *Let $f(m)$ be a family of pdfs on $\mathbb{R}$ indexed by their moments $m = (m_1, m_2, m_3, m_4)'$. If the PML method based on the maximization of*

$$\sum_{i=1}^{n} \operatorname{Log} f\left[y_i, m_1(x_i,\theta_1), m_2(x_i,\theta_2), m_3(x_i,\theta_3), m_4(x_i,\theta_4)\right]$$

is consistent for any specification of the conditional moments, any true conditional distribution satisfying the moment specification for some value θ_0 of $\theta = (\theta_1, \dots, \theta_4)'$, and any distribution P_X of X, then we have:

$$f(y,m) = \exp \left[\lambda_0(m) + \sum_{i=1}^{4} \lambda_i(m) y^i + a(y) \right].$$

Proof. Under the assumptions of Proposition 6, we must have, in particular, the consistency property in a model without exogenous variables. Furthermore, this model must possess the parametrization $\theta_i = E(Y^i), i = 1, \dots, 4$, where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$ belongs to the interior of the domain defined by:

$$\begin{vmatrix} 1 & \theta_1 \\ \theta_1 & \theta_2 \end{vmatrix} \geq 0; \qquad \begin{vmatrix} 1 & \theta_1 & \theta_2 \\ \theta_1 & \theta_2 & \theta_3 \\ \theta_2 & \theta_3 & \theta_4 \end{vmatrix} \geq 0.$$

In other words, if for all θ_i $(i = 1, \dots, 4)$ we have $E(Y^i - \theta_i) = 0$, then we must have:

$$E \left[\frac{\partial \operatorname{Log} f(Y,\theta)}{\partial \theta} \right] = 0.$$

Using a version of the Farkas Lemma [see Lemma 8.1 in Gourieroux-Monfort (1995a), p. 252], we conclude that:

$$\frac{\partial \operatorname{Log} f(y,\theta)}{\partial \theta} = \sum_{i=1}^{4} \lambda_i(\theta)(y^i - \theta_i).$$

Integrating the latter equation gives the result. □

Note that the generalized quartic family can be seen as a quartic family with respect to the modified measure $d\nu^*(y) = \exp(a(y))d\nu(y)$, ν being the Lebesgue measure on $\mathbb{R}$.

5. QGPML2 Method

5.1. **Alternative Parametrization.**

We have seen that the quartic exponential family can be equivalently parametrized by $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)'$, by $m = (m_1, m_2, m_3, m_4)'$, or by $\mu = (m_1, \sigma^2, s, k^*)'$. There is a fourth parametrization that will be of great interest, namely $(m_1, m_2, \lambda_3, \lambda_4)'$ or equivalently $\nu = (m_1, \sigma^2, \lambda_3, \lambda_4)'$. First, we have to show that this is indeed a genuine parametrization.

Proposition 7. *There is a one-to-one relationship between* $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ *and* $(m_1, m_2, \lambda_3, \lambda_4)$.

Proof. We have to prove that for any (λ_3, λ_4) the relationship

$$(\lambda_1, \lambda_2) \to [m_1(\lambda_1, \lambda_2, \lambda_3, \lambda_4), m_2(\lambda_1, \lambda_2, \lambda_3, \lambda_4)]$$

is one-to-one. We have seen that the Jacobian matrix $\partial m/\partial \lambda' = \Sigma$ is symmetric and positive definite $\forall m$, so the same is true for the upper (2×2) block-diagonal submatrix. Moreover, for any given (λ_3, λ_4) fixed, the section of Λ is convex, and therefore, using Theorem 6 in Gale-Nikaido (1965), we obtain the required result. □

The previous result means that, starting from the quartic family $Q(\Lambda)$, we can reparametrize it as $Q(m_1, \sigma^2, \lambda_3, \lambda_4)$, and therefore, fixing (λ_3, λ_4) at any admissible value $(\lambda_3^0, \lambda_4^0)$, we get a quadratic exponential family $Q(m_1, \sigma^2, \lambda_3^0, \lambda_4^0)$ in the sense of GMT (1984). We denote by $\lambda_1^*(m_1, \sigma^2, \lambda_3, \lambda_4)$, $\lambda_2^*(m_1, \sigma^2, \lambda_3, \lambda_4)$ and $\lambda_0^*(m_1, \sigma^2, \lambda_3, \lambda_4)$ the functions giving $\lambda_1, \lambda_2, \lambda_0$ in terms of $m_1, \sigma^2, \lambda_3, \lambda_4$.

5.2. **Efficient Semiparametric Estimator of Conditional Means and Variances.**

Let us consider a parametric specification of the conditional mean and variance: $m_1(X_i, \theta_1)$ and $\sigma^2(X_i, \theta_2)$, where θ, made of a (non redundant) union of θ_1 and θ_2 belongs to $\Theta \subset \mathbb{R}^K$. We know that the semiparametric efficiency bound for the asymptotic variance-covariance matrix of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$, where $\tilde{\theta}_n$

is a consistent estimator of θ_0, is given by [see e.g. Gourieroux-Monfort (1995b), Chapter 23]:

$$B = \left\{ E_X \left[E_0 \left(\frac{\partial r'(Y,X,\theta_0)}{\partial \theta} \right) V_0^{-1} \left(r(Y,X,\theta_0) \right) E_0 \left(\frac{\partial r(Y,X,\theta_0)}{\partial \theta'} \right) \right] \right\}^{-1}, \tag{5.1}$$

where,

$$r(Y,X,\theta) = \begin{bmatrix} Y - m_1(X,\theta_1) \\ Y^2 - m_1^2(X,\theta_1) - \sigma^2(X,\theta_2) \end{bmatrix}.$$

Therefore, we have:

$$\begin{aligned} E_0 \left(\frac{\partial r(Y,X,\theta_0)}{\partial \theta'} \right) &= \begin{bmatrix} -\frac{\partial m_1(X,\theta_1)}{\partial \theta_1} & 0 \\ -2\frac{\partial m_1(X,\theta_1)}{\partial \theta_1} m_1(X,\theta_1) & -\frac{\partial \sigma^2(X,\theta_2)}{\partial \theta_2} \end{bmatrix} \\ &= - \begin{bmatrix} 1 & 0 \\ 2m_1(X,\theta_1) & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial m_1(X,\theta_1)}{\partial \theta_1} & 0 \\ 0 & \frac{\partial \sigma^2(X,\theta_2)}{\partial \theta_2} \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} V_0 \left(r(Y,X,\theta_0) \right) &= V_0 \begin{bmatrix} Y & \\ Y^2 & \mid X \end{bmatrix} \\ &= \Omega_1(X,\theta_0), \text{ say,} \end{aligned}$$

and

$$\begin{aligned} B(\theta_0) = \Bigg\{ E_X \Bigg[&\begin{pmatrix} \frac{\partial m_1(X,\theta_1)}{\partial \theta_1} & 2\frac{\partial m_1(X,\theta_1)}{\partial \theta_1} m_1(X,\theta_1) \\ 0 & \frac{\partial \sigma^2(X,\theta_2)}{\partial \theta_2} \end{pmatrix} \Omega_1^{-1}(X,\theta_0) \\ &\times \begin{pmatrix} \frac{\partial m_1(X,\theta_1)}{\partial \theta_1} & 0 \\ 2\frac{\partial m_1(X,\theta_1)}{\partial \theta_1'} m_1(X,\theta_1) & \frac{\partial \sigma^2(X,\theta_2)}{\partial \theta_2} \end{pmatrix} \Bigg] \Bigg\}^{-1}. \end{aligned} \tag{5.2}$$

5.3. **QGPML2 Method.**

We assume that the conditional mean and variance are specified as $m_1(X_i,\theta_1)$ and $\sigma^2(X_i,\theta_2)$, and the conditional skewness and over-kurtosis are specified as $s(X_i,\theta_3)$ and $k^*(X_i,\theta_4)$.

We can first estimate (θ_1,θ_2) by the PML2 method based on the Gaussian family, i.e. by solving the problem:

$$(\tilde{\theta}_{1n}, \tilde{\theta}_{2n}) = \underset{\theta_1,\theta_2}{\text{Argmin}} \sum_{i=1}^{n} \text{Log}\, \sigma^2(X_i,\theta_2) + \frac{[Y_i - m_1(X_i,\theta_1)]^2}{\sigma^2(X_i,\theta_2)}. \tag{5.3}$$

Next, we compute

$$\hat{u}_i = \frac{Y_i - m_1(X_i,\tilde{\theta}_{1n})}{\sigma(X_i,\tilde{\theta}_{2n})},$$

and obtain consistent estimators of $\tilde{\theta}_{3n}$ and $\tilde{\theta}_{4n}$ of θ_3 and θ_4 from the nonlinear regressions of $\hat{u}_i^3$ on $s(X_i, \theta_3)$ and $\hat{u}_i^4 - s(X_i, \tilde{\theta}_{3n})^2 - 1$ on $k^*(X_i, \theta_4)$. Explicitly, this corresponds to obtaining the pair $(\tilde{\theta}_{3n}, \tilde{\theta}_{4n})$, verifying:

$$\begin{aligned}
\tilde{\theta}_{3n} &= \underset{\theta_3}{\operatorname{argmin}} \sum_{i=1}^{N} (\hat{u}_i^3 - s(X_i, \theta_3))^2, \\
\tilde{\theta}_{4n} &= \underset{\theta_4}{\operatorname{argmin}} \sum_{i=1}^{N} (\hat{u}_i^4 - s(X_i, \tilde{\theta}_{3n})^2 - 1 - k^*(X_i, \theta_4))^2.
\end{aligned}$$

Then, noting $\tilde{\theta}_n = (\tilde{\theta}_{1n}, \tilde{\theta}_{2n}, \tilde{\theta}_{3n}, \tilde{\theta}_{4n})'$, we define:

$$\begin{aligned}
\tilde{m}_{1i} &= m_1(x_i, \tilde{\theta}_{1n}); \qquad \tilde{\sigma}_i^2 = \sigma^2(x_i, \tilde{\theta}_{2n}); \\
\tilde{s}_i &= s(x_i, \tilde{\theta}_{3n}); \qquad \tilde{k}_i^* = k^*(x_i, \tilde{\theta}_{4n}); \\
\tilde{\lambda}_{ji} &= \lambda_j(\tilde{m}_{1i}, \tilde{\sigma}_i^2, \tilde{s}_i, \tilde{k}_i^*), \quad j = 3, 4.
\end{aligned}$$

Definition 2. *The Quasi Generalized PML2 (QGPML2) estimator* $(\hat{\theta}_{1n}, \hat{\theta}_{2n})$ *of* $(\theta_{01}, \theta_{02})$ *is defined by maximizing with respect to* (θ_1, θ_2) :

$$\begin{aligned}
L_n^{(2)}(\theta_1, \theta_2) = \sum_{i=1}^{n} \{ & \lambda_0^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \tilde{\lambda}_{3i}, \tilde{\lambda}_{4i} \right] \\
& + \lambda_1^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \tilde{\lambda}_{3i}, \tilde{\lambda}_{4i} \right] y_i \\
& + \lambda_2^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \tilde{\lambda}_{3i}, \tilde{\lambda}_{4i} \right] y_i^2 \}
\end{aligned}$$

Note that, using the parametrization $(m_1, \sigma^2, \lambda_3, \lambda_4)$ the quartic family of pdf can be written:

$$\begin{aligned}
f(y_i | m_1, \sigma^2, \lambda_3, \lambda_4) = \exp & \left[\lambda_0^* \left(m_1, \sigma^2, \lambda_3, \lambda_4 \right) + \lambda_1^* \left(m_1, \sigma^2, \lambda_3, \lambda_4 \right) y_i \right. \\
& \left. + \lambda_2^* \left(m_1, \sigma^2, \lambda_3, \lambda_4 \right) y_i^2 + \lambda_3 y_i^3 + \lambda_4 y_i^4 \right],
\end{aligned}$$

and, therefore, the objective function of Definition 2 is equivalent to:

$$\sum_{i=1}^{n} \operatorname{Log} f(y_i \mid m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \lambda_3(x_i, \tilde{\theta}_n), \lambda_4(x_i, \tilde{\theta}_n)),$$

since the terms $\lambda_3(x_i, \tilde{\theta}_n) y_i^3 + \lambda_4(x_i, \tilde{\theta}_n) y_i^4$ do not depend on (θ_1, θ_2). The method is called QGPML2 because only y_i and y_i^2 are involved, and it is clearly an example of a Quasi-Generalized M-estimator.

Moreover, we obtain the following important property:

Proposition 8. *The QGPML2 estimator* $(\hat{\theta}_{1n}, \hat{\theta}_{2n})$ *is asymptotically equivalent to the unfeasible estimator based on the maximization of*

$$\begin{aligned} L_{n0}^{(2)}(\theta_1, \theta_2) = \sum_{i=1}^{n} \{ & \lambda_0^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \lambda_3(x_i, \theta_0), \lambda_4(x_i, \theta_0) \right] \\ & + \lambda_1^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \lambda_3(x_i, \theta_0), \lambda_4(x_i, \theta_0) \right] y_i \\ & + \lambda_2^* \left[m_1(x_i, \theta_1), \sigma^2(x_i, \theta_2), \lambda_3(x_i, \theta_0), \lambda_4(x_i, \theta_0) \right] y_i^2 \}. \end{aligned}$$

Proof. According to the result given in Equation (2.11), we have to check that:

$$E_0 \left[\frac{\partial^2}{\partial \begin{pmatrix} m_1 \\ \sigma^2 \end{pmatrix} \partial \begin{pmatrix} \lambda_3 \\ \lambda_4 \end{pmatrix}'} \left(\lambda_0^* + \lambda_1^* Y_i + \lambda_2^* Y_i^2 \right) \mid X \right] = 0.$$

Differentiating Log $f(y_i \mid x_i; m_1, \sigma^2, \lambda_3, \lambda_4)$ with respect to m_1 and σ^2 and then taking the expectation we get:

$$\frac{\partial \lambda_0^*}{\partial \begin{pmatrix} m_1 \\ \sigma^2 \end{pmatrix}} + \frac{\partial \lambda_1^*}{\partial \begin{pmatrix} m_1 \\ \sigma^2 \end{pmatrix}} m_1 + \frac{\partial \lambda_2^*}{\partial \begin{pmatrix} m_1 \\ \sigma^2 \end{pmatrix}} m_2 = 0,$$

for any $(m_1, \sigma^2, \lambda_3, \lambda_4)$. Therefore, differentiating further with respect to λ_3 and λ_4, we still get zero. □

Proposition 9. *The QGPML2 estimator* $(\hat{\theta}_{1n}, \hat{\theta}_{2n})$ *reaches the semiparametric bound.*

Proof. See Appendix C. □

From consistent estimators of the asymptotic variance-covariance matrix $B(\theta_0)$, we can deduce Wald and score tests, as well as asymptotic confidence regions.

We also note that, from the proof of Proposition 9, we know that the matrices $\tilde{I}$ and $\tilde{J}$ are equal, and therefore we can also use Likelihood-ratio type tests [see Gourieroux and Monfort 1995b, chapter 18]. More precisely, denoting by $\hat{\theta}_{1n}^0$ and $\hat{\theta}_{2n}^0$ the constrained QGMPL2 estimators obtained by maximizing $L_n^{(2)}(\theta_1, \theta_2)$ under the null, we can use the test statistic:

$$\xi_n^R = 2 \left[L_n^{(2)}(\hat{\theta}_{1n}, \hat{\theta}_{2n}) - L_n^{(2)}(\hat{\theta}_{1n}^0, \hat{\theta}_{2n}^0) \right]. \tag{5.4}$$

If the null is $g(\theta_1, \theta_2) = 0$ where g is an $r-$dimensional vector, then, under the null, ξ_n^R is asymptotically distributed as a $\chi^2(r)$. If $(\theta_1', \theta_2')'$ is a $p-$dimensional vector and the null is of the form $\theta_1 = h_1(\gamma)$ and $\theta_2 = h_2(\gamma)$, where γ is a $q-$dimensional vector, then under the null, ξ_n^R is asymptotically distributed as a $\chi^2(p-q)$.

6. Numerical Implementation

The implementation of the PML4 and QGPML2 methods necessitates the numerical algorithms that will be described in this section. To this end, let us first introduce the useful notion of a canonical quartic family.

6.1. **Canonical Quartic Family.**

Let us consider the quartic family $Q(\lambda)$ with pdfs:

$$\exp\left(\sum_{i=0}^{4} \lambda_i y^i\right),$$

with $\lambda_0 = -\operatorname{Log} \int \exp\left(\sum_{i=1}^{4} \lambda_i y^i\right) dy$, and $\lambda_4 \leq 0$.

If $\lambda_4 = 0$, we have $\lambda_3 = 0$ and $\lambda_2 < 0$, and we get the normal family. If $\lambda_4 < 0$ and if we consider a linear transformation, $Y = a + bZ$, $b > 0$, then the distribution of Z has a pdf equal to:

$$\exp\left[\operatorname{Log} b + \lambda_0 + \lambda_1(a+bz) + \lambda_2(a+bz)^2 + \lambda_3(a+bz)^3 + \lambda_4(a+bz)^4\right].$$

In particular, the coefficients of z^4 and z^3 are respectively $\lambda_4 b^4$ and $\lambda_3 b^3 + 4\lambda_4 a b^3$. Therefore, choosing $b = (-\lambda_4)^{-1/4}$ and $a = -\lambda_3/(4\lambda_4)$, we obtain a pdf of the form:

$$\exp\left[\alpha_0(\alpha,\beta) + \alpha z + \beta z^2 - z^4\right], \tag{6.1}$$

which will be called a *canonical form* and denoted by $Q^*(\alpha,\beta)$. Note that, since the skewness and the kurtosis are invariant under linear mapping, the range of the pair (s,k) reached by the canonical forms is the same as that for the entire $Q(\lambda)$ family. Namely, it is the universal set $k \geq s^2+1$, except $\{s=0, k>3\}$. Furthermore

$$\begin{aligned}
\alpha_0(\alpha,\beta) &= \operatorname{Log} b + \lambda_0 + \lambda_1 a + \lambda_2 a^2 + \lambda_3 a^3 + \lambda_4 a^4, \\
\alpha &= (\lambda_1 + 2\lambda_2 a + 3\lambda_3 a^2 + 4\lambda_4 a^3) b, \\
\beta &= (\lambda_2 + 3\lambda_3 a + 6\lambda_4 a^2) b^2.
\end{aligned}$$

Also, for given values of $\alpha_0(\alpha,\beta), \alpha, \beta$ and λ_3, λ_4, one may obtain λ_0, λ_1, and λ_2 using:

$$\lambda_2 = \beta/b^2 - 3a\lambda_3 - 6\lambda_4 a^2, \tag{6.2}$$

$$\lambda_1 = \alpha/b - 2\lambda_2 a - 3\lambda_3 a^2 - 4\lambda_4 a^3, \tag{6.3}$$

$$\lambda_0 = \alpha_0(\alpha,\beta) - \operatorname{Log} b - (\lambda_1 a + \lambda_2 a^2 + \lambda_3 a^3 + \lambda_4 a^4). \tag{6.4}$$

6.2. **Implementation of the PML4 Method.**

To compute the values of $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ corresponding to a given $m = (m_1, m_2, m_3, m_4)'$, we can proceed as follows.

(1) Compute $s(m), k(m)$.[1]

(2) Find a pair of (α, β) corresponding to $s(m), k(m)$.

(3) Compute the mean $m_1^*(\alpha, \beta)$ and the variance $\sigma^{2*}(\alpha, \beta)$ corresponding to the canonical pdf $Q^*(\alpha, \beta)$.

(4) The linear transform $Y = a + bZ$, $b > 0$, gives $m_1 = a + b\, m_1^*(\alpha, \beta)$, and $\sigma = b\,\sigma^*(\alpha, \beta)$, where $\sigma^2 = m_2 - m_1^2$. That is, $b = \sigma/\sigma^*(\alpha, \beta)$ and $a = m_1 - b\, m_1^*(\alpha, \beta)$.

(5) Compute $\lambda_4 = -(b)^{-4}$, $\lambda_3 = -4\lambda_4 a$, and use equations (6.2), (6.3), and (6.4) to get $\lambda_0, \lambda_1, \lambda_2$.

6.3. **Computation of the Functions $\alpha(s,k)$, $\beta(s,k)$ Using the Gauss-Freud Method.**

In subsection 6.2 we have seen that the key step of the computation is step (2). For a given pair (s, k), we must find (α, β) such that the pdf (6.1) has a skewness equal to s and a kurtosis equal to k.

Using the notation:

$$q(z; \alpha, \beta) = \exp(\alpha z + \beta z^2 - z^4),$$

we have to compute the integrals:

$$I_j(\alpha, \beta) = \int_{-\infty}^{\infty} z^j\, q(z; \alpha, \beta)\, dz, \qquad j = 0, \cdots, 4. \tag{6.5}$$

Once these integrals are known, we easily get the moments:

$$m_j(\alpha, \beta) = I_j(\alpha, \beta)/I_0(\alpha, \beta), \qquad j = 1, \cdots, 4,$$

Therefore $s(\alpha, \beta)$, $k(\alpha, \beta)$. Finally, we have to minimize in (α, β) the distance:

$$[s - s(\alpha, \beta)]^2 + [k - k(\alpha, \beta)]^2, \tag{6.6}$$

It is well known that the computation of the parameters for such a problem may be difficult. The basic reason for this is that the function $q(z; \alpha, \beta)$ may have two maxima for very different values of z, and it may take very small values in a large area between these two values of z. Furthermore, one of the

[1] Skewness and kurtosis are given, respectively, by

$$s(m) = \frac{m_3 - 3m_2 m_1 + 2m_1^3}{(m_2 - m_1^2)^{3/2}}, \quad k(m) = \frac{m_4 - 4m_3 m_1 + 6m_2 m_1^2 - 3m_1^4}{(m_2 - m_1^2)^2}.$$

maxima may be far out in the tails and yet contribute a relatively important probability mass. This implies that integration methods of the Newton type, based on an equidistant grid, are inadequate. More precisely, approximations of the form:

$$\hat{I}_j(\alpha, \beta) = \delta \sum_{i=0}^{N} z_i^j \, \exp(\alpha z_i + \beta z_i^2 - z_i^4),$$

with $z_i - z_{i-1} = \delta$ and $i = 1, \cdots, N$, or even improvements thereof, such as Simpson's scheme, may necessitate very large values for $N, -z_0$, and z_N to achieve acceptable precision. Typically, we would have to take values like $N = 20'000$, and $-z_0 = z_N = 80$, which makes the optimization of (6.6) very difficult. In addition, "smarter" integration techniques, based on the Gauss-Lagrange scheme, may be problematic since such a scheme requires first a transform of $\mathbb{R}$ into $(-1, 1)$, by the logistic map. This transform essentially varies in a neighborhood of the origin, and it tends to lose information contained in the tails. As a consequence, such schemes even when performed with a large number of abscissas, tend to be inaccurate even for relatively low values of kurtosis. For this reason, we adopte the Gauss-Freud method [see Freud (1986) for the seminal work and Levin and Lubinsky (2000) for a recent research-monograph], which is designed to accurately approximate integrals of the kind:

$$\int_{-\infty}^{\infty} f(z) \exp(-z^4) \, dz.$$

This method leads to approximations of the form:

$$I_j^*(\alpha, \beta) = \sum_{i=0}^{N} z_i^j \, \exp(\alpha z_i + \beta z_i^2) w_i, \tag{6.7}$$

where the abscissa z_i and the weights w_i are very precisely adapted to the shape of the function to be integrated. Further details on how the z_i and the w_i may be computed may be found in Gautschi (2004, Part 1).[2] Thus, the proposed algorithm has two advantages over the other numerical methods: it uses results on numerical integration specifically related to the integration problem and, moreover, the calculation of parameters involves computing only

[2] Essentially, the weights w_j and abscissa x_j can be obtained as eigenvectors and eigenvalues of a Jacobi matrix (see Golub and Welch 1969). This matrix, in turn, requires a sequence of parameters for which a stable estimation algorithm has been proposed by Noschese and Pasquini (1999) for the $\exp(-z^4)$ weight function. Prof. Milovanović implemented this algorithm and made the resulting parameters available to the public via the website of Prof. Gautschi. On this website, one may find the file: coefffreud4.txt under `www.cs.purdue.edu/archives/2001/wxg/tables`. To obtain the x_j and w_j, we use his routine `Gauss.m` to be found under: `www.cs.purdue.edu/archives/2002/wxg/codes`.

two parameters, resulting in a significant gain in time. We performed all of the numerical integrations using $N = 100$.

6.4. **Implementation of the QGPML2 Method.**

The numerical problem is the following: given any admissible value of $(m_1, \sigma^2, \lambda_3, \lambda_4)$, compute $\lambda_i^*(m_1, \sigma^2, \lambda_3, \lambda_4)$, for $i = 0, 1, 2$.

According to this approach, it is necessary to allow for densities of any given mean and variance. However, the numerical integration scheme uses the kernel $\exp(-z^4)$, a symmetric kernel that weights those observations in a neighborhood of 0. We expect that this may create numerical difficulties for random variables whose mean is distant from 0. For this reason, we consider a computation strategy of the λ_i^* where, in a preliminary step, observations are studentized.[3]

Thus, instead of considering the pdf,

$$\exp\left[\lambda_0^* + \lambda_1^* y + \lambda_2^* y^2 + \lambda_3 y^3 + \lambda_4 y^4\right],$$

it is useful to characterize the associated density, which has a mean of zero and a variance of 1. This density is related to the previous one by the linear transformation $Y = m_1 + \sigma Z$, where Z represents a random variable with mean 0 and variance 1. The corresponding density, which will be called the *Studentized exponential quartic*, is written as:

$$\exp\left[\delta_0 + \delta_1 z + \delta_2 z^2 + \delta_3 z^3 + \delta_4 z^4\right].$$

We have the relations $\delta_4 = \sigma^4 \lambda_4 < 0$ and $\delta_3 = \sigma^3 \lambda_3 + \frac{4m_1}{\sigma}\delta_4$. Since, in the QGPML2 approach, m_1, σ^2, λ_3, λ_4 are given, the parameters δ_3, δ_4 are also given. Once $\delta_0, \delta_1, \delta_2$ corresponding to a zero mean and unit variance have been obtained using the method described below, one may revert to the initial parameters using:

$$\lambda_0^* = \left[\sigma^4 \delta_0 - m_1 \sigma^3 \delta_1 + m_1^2 \sigma^2 \delta_2 - m_1^3 \sigma \delta_3 + m_1^4 \delta_4\right] / \sigma^4 - \mathrm{Log}(\sigma),$$
$$\lambda_1^* = \left[\sigma^3 \delta_1 - 2m_1 \sigma^2 \delta_2 + 3m_1^2 \sigma \delta_3 - 4m_1^3 \delta_4\right] / \sigma^4,$$
$$\lambda_2^* = \left[\sigma^2 \delta_2 - 3m_1 \sigma \delta_3 + 6m_1^2 \delta_4\right] / \sigma^4.$$

Using the notation,

$$q(z; \delta_1, \delta_2) = \exp\left[\delta_1 z + \delta_2 z^2 + \delta_3 z^3 + \delta_4 z^4\right],$$

[3] This studentization is not required for PML4 where the mean m_1^* (and variances $(\sigma^*)^2$) turn out to be close to 0 (1).

we have to compute the integrals:

$$I_j(\delta_1,\delta_2) = \int_{-\infty}^{\infty} z^j\, q(z;\delta_1,\delta_2)\, dz, \quad j = 0,\cdots,2. \tag{6.8}$$

Using the change of variable $u = (-\delta_4)^{1/4} z$, these integrals become:

$$I_j(\delta_1,\delta_2) = \int_{u=-\infty}^{+\infty} \left((-\delta_4)^{-1/4}u\right)^j \exp\Big[\delta_0 + \delta_1(-\delta_4)^{-1/4}u + \delta_2\left((-\delta_4)^{-1/4}u\right)^2 + \delta_3\left((-\delta_4)^{-1/4}u\right)^3 - u^4\Big](-\delta_4)^{-1/4}du \quad j = 0,\cdots,2. \tag{6.9}$$

Now, the kernel $\exp(-u^4)$ appears again, and we may use the Gauss-Freud method outlined in Section 6.3. Once these integrals are efficiently evaluated, we may compute the moments:

$$m_j(\delta_1,\delta_2) = I_j(\delta_1,\delta_2)/I_0(\delta_1,\delta_2), \quad j = 1,2.$$

The parameters δ_1 and δ_2 are obtained by minimizing the distance:

$$[m_1(\delta_1,\delta_2)]^2 + [\sigma^2(\delta_1,\delta_2) - 1]^2, \tag{6.10}$$

where $\sigma^2(\delta_1,\delta_2) = m_2(\delta_1,\delta_2) - m_1^2(\delta_1,\delta_2)$. Eventually, $\delta_0 = -\operatorname{Log} I_0(\delta_1,\delta_2)$.

7. Numerical Examples

In this section, after discussing the computation of the parameters of the quartic exponential for given moments of order 1 to 4, we will discuss several Monte-Carlo exercises demonstrating the usefulness of the methods at hand. The examples we want to discuss are 1) a comparison between various estimation techniques for small samples, 2) a study of the performance of PML4 in the case of misspecification, and 3) an illustration of QGPML2.

7.1. **Computation of the Quartic Exponential.**

As discussed in Section 6, feasibility of the PML4 estimation hinges on the ability to efficiently compute the parameters λ_0, $\cdots$, λ_4 of the quartic exponential for given m_1, m_2, m_3, m_4. Ormoneit and White (1999) attribute to Agmon *et al.* (1981) the first attempts to compute the "correct" λ^*. Agmon *et al.* (1981) considered the maximization of entropy under moment constraints, that is the so-called primal problem. They computed the integrations using the Gauss-Lagrange scheme after mapping the domain of integration $(-\infty,\infty)$ into $(-1,1)$. Zellner and Highfield (1988) propose computation of the λ^* by seeking the zeros of the first order conditions that result from the entropy maximization (i.e. the dual approach). Maasoumi (1983) reported difficulties with this method that Ormoneit and White (1999) corroborate. Ormoneit and White

(1999) also map the domain $(-\infty, \infty)$ into $(-1, 1)$, and they use the Gauss-Lagrange scheme. Moreover, they feed intelligent starting values into their optimization and stabilize the computation of the exponential to avoid numerical overflow. Last but not least, to our knowledge, they are the first ones in the literature to have acknowledged numerical difficulties in the λ parameter computation along the segment $s = 0$, $k > 3$. Indeed, we know that no density exists for this segment, based on theoretical grounds as discussed earlier.

Our method hinges instead on obtaining the parameters α and β of the canonical form for given skewness and kurtosis. In this section, we wish to discuss the precision of these computations.[4] In order to evaluate the algorithm which gives the parameters α and β of the canonical form, we considered a grid covering the range of values of kurtosis from 1.5 to 20. For each value of kurtosis, k, we considered a grid for skewness ranging from 0.1 to $\sqrt{k-1} - 0.1$. For each point of this grid, say s, k, we computed α and β as described in Section 6.3 and recomputed the associated skewness and kurtosis, say $\tilde{s}, \tilde{k}$.[5] In **Figure 1**, circles represent the points of the skewness-kurtosis grid for which we evaluated the parameters and, + symbols represent those points for which the distance $D = (s - \tilde{s})^2 + (k - \tilde{k})^2 > 10^{-5}$.

This figure demonstrates several interesting phenomena 1) Even though, based on theoretical grounds, no density can exist on the segment $(s = 0, k \geq 3)$, it is still possible to obtain a density for parameters close to the excluded segment. 2) Even for very large values of kurtosis (limited in the figure to 20), we obtain a large range of values of skewness for which a highly accurate density may be obtained. We also constructed a similar graph where kurtosis was allowed to took values up to 150. We find that, even for a kurtosis of 150, the range of skewness, where $D < 10^{-5}$, ranges from 2.5 to 12.1, still a very respectable domain.

Many of the difficulties encountered in earlier attempts at such calculations disappear in our approach. The linear transform $Y = a + bZ$ allows us to write the exponential quartic in terms of $\exp(-z^4)$. Then, by replacing $\exp(-z^4)$ by well-behaved discrete weights (this leads to formula (6.7)), we integrate directly over the range $(-\infty, \infty)$, thus obviating the use of the logistic map. Next, we

[4] All of the programming was performed in the MATLAB environment. We implemented the code on both Mac OS X and Windows Vista machines. All of the simulations were performed on a PC with an Intel quadricore processor running four MATLAB clones in parallel. To increase the speed of the computations, we transcribed the central part of the programs into the C language and called it via a MEX interface.

[5] We perform the required minimization using the Nelder-Meade approach.

only optimize over two parameters, rather than four. We also feed optimized starting values for α and β into the optimizer. These starting values are the α and β corresponding to those points in the domain represented in Figure 1 that are closest to some given values of skewness and kurtosis. The skewness and kurtosis, and their associated α and β, are stored once and for all in some file that is read into memory as the program is initialized.

To further understand some of the difficulties encountered in earlier studies, we obtained the parameters λ_j, for $j = 0, \cdots, 4$ for extremely skewed cases, and evaluated the resulting densities at points far out in the tail (say $z = 50$ for a centered and reduced density) and still found a small, yet significant probability mass. The logistic map, which transforms $(-\infty, \infty)$ into $(-1, 1)$ used in the earlier work, may therefore have 'fudged' the behavior of the density for relatively large values of skewness and kurtosis. We also note that earlier work required many integration points (each evaluation costs time) whereas using abscissas and weights that are made specifically for the $\exp(-z^4)$ weighting function reduces the number of points for which the integrand needs to be evaluated.[6]

As the numerical exercises that follow will demonstrate, the time necessary to compute the required α and β parameters is of an order that is suitable for applying the exponential quartic in many econometric problems. The computation of one set of α and β requires about 0.015 seconds, allowing for about 66 density constructions per second. It is clear that the method proposed here is not confined to econometrics and may prove useful in other fields as well.

Similarly to the protocol described above, in the context of the parameter estimations related to QGPML2, we verified the precision of the computation of δ_i, for $i = 0, \cdots, 2$ for given δ_3, δ_4, yielding a density with mean 0 and variance 1.

7.2. **A First Experiment.**

The objective of this first experiment was to demonstrate that the PML4 estimation may provide estimates which are superior to either the PML2 or the GMM estimators in an unconditional setting. We first discuss the choice of a data generating process, and then we focus on the estimation techniques. A priori, many distributions could be used for this experiment (Student-t, distributions in the Pearson family, Gamma, etc.) Preliminary work made it clear

[6] We presently do not incorporate a selection rule on the number of abscissas, N, which may further decrease the speed of the computation of the α and β.

that a distribution should be chosen from which draws could be obtained in a very rapid manner. For this reason we settled on the family of skewed Laplace distributions, denoted sLD. These distributions have been used to price options in the context of extreme return realizations, for example, by Gourieroux and Monfort (2006). This family of distributions has three parameters, $b_0 > 0, b_1 > 0$, and c, and its pdf is defined by:

$$f(z; b_0, b_1, c) = \begin{cases} \frac{b_0 b_1}{b_0+b_1} \exp\left[b_0(z-c)\right], & \text{if } z \leq c, \\ \frac{b_0 b_1}{b_0+b_1} \exp\left[-b_1(z-c)\right], & \text{if } z > c. \end{cases} \tag{7.1}$$

The mean, variance, skewness and kurtosis of this density are given by:

$$\begin{aligned} m_1(c, b_0, b_1) &= E[Y] = c + \frac{1}{b_1} - \frac{1}{b_0}, &(7.2)\\ \sigma^2(c, b_0, b_1) &= Var[Y] = \frac{1}{b_0^2} + \frac{1}{b_1^2}, &(7.3)\\ s(c, b_0, b_1) &= \frac{2}{\sigma^3}\left[\frac{1}{b_1^3} - \frac{1}{b_0^3}\right], &(7.4)\\ k(c, b_0, b_1) &= \frac{9}{\sigma^4}\left[\frac{1}{b_1^4} + \frac{1}{b_0^4}\right] + \frac{6}{\sigma^4 b_0^2 b_1^2}. &(7.5) \end{aligned}$$

Since this density may be viewed as describing a mixture of exponentials, we use the inverse c.d.f. technique to simulate random draws from it.

In this first experiment, we focus on the situation where $c = 0$. Indeed, without an additional assumption on one of the parameters of the sLD, we would not be able, in the following, to obtain parameter estimates based on the PML2 principle.

We simulated 10,000 samples, each of a length of either $T = 25, 50$ or 100 i.i.d. observations. The estimation techniques used were ML, PML2, PML4 and GMM. Let us describe the way in which we implemented these estimations.

For ML, we maximized for each sample, the log-likelihood obtained from (7.1):

$$L^{ML} = \sum_{i=1}^{T} \operatorname{Log} f(z_i; b_0, b_1).$$

For PML2, we considered the objective function:

$$L^{PML2} = -T \operatorname{Log} \sigma(b_0, b_1) - \frac{1}{2}\sum_{i=1}^{T}\left(\frac{z_i - m_1(b_0, b_1)}{\sigma(b_0, b_1)}\right)^2.$$

For PML4, we formed the objective function:

$$L^{PML4} = \sum_{i=1}^{T} \lambda_0 + \lambda_1 y_i + \lambda_2 y_i^2 + \lambda_3 y_i^3 + \lambda_4 y_i^4,$$

where the parameters $\lambda_0, \cdots, \lambda_4$ were computed as described in section 6.2 for $(m_1(b_0, b_1), \sigma(b_0, b_1)^2, s(b_0, b_1), k(b_0, b_1))$.

For GMM, we defined, (see Hansen (1982)), the 4×1 vector,

$$X_i(b_0, b_1) = \begin{bmatrix} z_i - m_1(b_0, b_1) \\ z_i^2 - m_1^2(b_0, b_1) - \sigma^2(b_0, b_1) \\ \left(\frac{z_i - m_1(b_0,b_1)}{\sigma(b_0,b_1)}\right)^3 - s(b_0, b_1) \\ \left(\frac{z_i - m_1(b_0,b_1)}{\sigma(b_0,b_1)}\right)^4 - k(b_0, b_1) \end{bmatrix},$$

and considered the distance:

$$J = g_T(b_0, b_1)' S^{-1} g_T(b_0, b_1), \text{ where } g_T(b_0, b_1) = \frac{1}{T} \sum_{i=1}^{T} X_i(b_0, b_1).$$

The GMM estimates were obtained as those parameters minimizing the distance J. The matrix S that appears in the distance was obtained by using, as a first step, the identity matrix, and as a second step, the asymptotic variance-covariance matrix. Thus, the GMM estimates are asymptotically optimal.

We performed the simulation using $(b_0, b_1) = (2.41, 1.30)$.[7] This point corresponds to the moments $(m_1, m_2, s, k) = (1.30, 0.35, 1.15, 6.91)$.

Table 1 displays several statistics for 10,000 simulations and for various sample sizes. However, the main result that this Table conveys is that, for all sample sizes considered, the MSE of the ML estimation dominates, as expected, over all other methods. However, we find that the PML4 technique yields estimates with an MSE up to less than half of the one for PML2. Concerning GMM, the parameters often hit the boundaries that we imposed during the estimation. The MSE of the parameters obtained by this method is the largest. This suggests that for situations where the econometrician has no prior information on the skewed distribution to use for the estimation, the PML4 technique may be a most useful one.

7.3. **A Second Experiment.**

In this experiment, we illustrate the robustness of the PML4 method with respect to the shape of the distribution as compared to that of the ML method. In order to be fair with the ML, we consider a situation in which the family of pdfs used to build the likelihood is allowed to reach the true moments up to fourth order, although it does not contain the true pdf. More precisely, the misspecified likelihood will be based on the family of skewed Laplace distributions (sLD), and the true pdf will be that of a mixture of two normals whose

[7] We also used other points, but the results are very similar to the ones reported here.

moments, up to the fourth order, are reachable by the sLD family. The skewed Laplace density family defined in (7.2) is conveniently reparametrized by the mean, μ, the standard error, σ, and a parameter, $\theta \in [0, \pi/2]$, such that the standardized variable $(z-\mu)/\sigma$ follows the zero-mean, unit-variance skewed Laplace density corresponding to $b_0 = 1/\sin\theta$, $b_1 = 1/\cos\theta$, $c = \sin\theta - \cos\theta$. The pairs (s, k) that are reachable by the skewed Laplace density family are given by (7.4) and (7.5), where b_0, b_1 are replaced respectively by $1/\sin\theta$ and $1/\cos\theta$, respectively, and σ is replaced by 1. Thus, the misspecified likelihood function is,

$$\prod_{i=1}^{T} \frac{1}{\sigma} f\left(\frac{z_i-\mu}{\sigma}; \frac{1}{\sin\theta}, \frac{1}{\cos\theta}, \sin\theta - \cos\theta\right), \tag{7.6}$$

where f is given in (7.1).

The PML4 objective function is obtained by specifying the mean, variance, skewness, and kurtosis as μ, σ^2, $2(\cos^3\theta - \sin^3\theta)/(\sigma^3)$, and $9(\cos^4\theta + \sin^4\theta)/(\sigma^4) + 6\cos^2\theta\sin^2\theta/(\sigma^4)$, respectively.

The data generating processes are mixtures of two normals such that $\mu = 0$, $\sigma^2 = 1$, with the skewness and kurtosis being obtained from the previous formulae with $\theta = 10$ (in degrees), namely $s = 1.90$ and $k = 8.65$. The family of two normals can reach any set of mean, variance, skewness, kurtosis, and this, moreover, can be reached in a non-unique way. Here, we choose two data generating processes.[8] First, we consider the unique mixture satisfying $\mu = 0, \sigma^2 = 1$, $s = 1.90$, and $k = 8.65$ with the same variance in the two normal components. We find the mixture,

$$0.0451\,\mathcal{N}(3.483, 0.654) + (1 - 0.0451)\,\mathcal{N}(-0.164, 0.654),$$

where 0.654 is the standard deviation. Second, we choose the mixture maximizing the entropy under the four constraints on the moments and find

$$0.102\,\mathcal{N}(2.036, 1.426) + (1 - 0.102)\,\mathcal{N}(-0.232, 0.597).$$

For these two DGPs, we use 10,000 simulations of $T = 100$ observations. For each simulation, we compute the misspecified ML estimates and the PML4 estimates of μ, σ, θ, where the true values are (0,1,10).

Table 2 displays the results from this simulation. Inspection of this Table reveals, for both DGPs, qualitatively similar results for the estimates of the parameters and the MSE. We find rather good estimates of both the mean and

[8] More details on the construction of these processes are available upon request. The formulae on how to obtain these mixtures may be found in Titterington *et al.* (1985).

the variance. The MSE of $\hat{\sigma}$ tends to be better for the PML4 estimation than for the ML one.[9] The deterioration of the estimate of the parameter θ, which commands higher moments is, however, dramatic if one uses a misspecified model. For instance, focusing on the ML estimation, whereas the MSE of $\hat{\mu}$ is 0.0004, it becomes 327.49 for $\hat{\theta}$. Inspecting of the average estimate of $\hat{\theta}$ and its 5- and 95-percentiles reveals that the estimate is nearly three times the estimate of its true value. The 10% confidence intervals do not even contain the true value. On the other hand, if one considers the PML4 estimations, the deterioration of the MSE of $\hat{\theta}$ is much less with an MSE=30.08. Also, the confidence intervals now contain the true value.

As this experiment suggests, in the case of skewed and kurtic data, great care needs to be exercised if one uses ML estimation with a risk of misspecification. The PML4 method appears to be much more robust.

7.4. **A Third Experiment.**

In the previous examples, we demonstrated the usefulness of the PML4 technique in an unconditional setting. In this section, we illustrate the usefulness of the PML4 technique in a conditional setting where the numerical complexity is much greater. There are situations where the modeling of the variation in higher moments may be of importance *per se*. For instance, Hansen (1994) considered a model in the spirit of Garch with time varying skewness and kurtosis.

A generic model that captures variation in the higher moments is given by:

$$\begin{aligned} y_i &= \mu + \sigma\varepsilon_i, \\ \text{where } \varepsilon_i &\sim D(0,1,s(x_i),k^*(x_i)), \\ s(x_i) &= 0.5 + a x_i, \\ k^*(x_i) &= 2 + b x_i, b > 0, \\ x_i &\sim U[1/2,3/2], \text{ i.i.d.} \end{aligned} \tag{7.7}$$

In the second line, D stands for some distribution where skewness and over-kurtosis depends on some exogenous variable, x_i. The next two lines specify how skewness and over-kurtosis are parametrized. We recall that the kurtosis, k, is related to the over-kurtosis, k^*, by $k = 1 + s^2 + k^*$. For the simulations, we take the lower moments to be $\mu = 0$ and $\sigma = 1$. Furthermore, we take $a = 1$ and $b = 2$. As long as the condition $b > 0$ is imposed in the numerical computation, the model will be well defined. The intercepts 0.5 and 2 in the

[9] This result is not due to statistical variation as we could verify by running the simulation several times.

specification of $s(x_i)$ and $k^*(x_i)$ guarantee that the distribution will be skewed (here $s > 1$) and fat-tailed (here $k > 4$).

The D distribution that we choose is the mixture of two normal distributions with identical variances as discussed in the previous subsection. In the Monte-Carlo exercise, we simulate 1,500 samples, each with $T = 100$ observations. The estimations require between about 60 and 340 sec, with an average time of about 130 sec.

Table 3 contains the statistics associated with the various estimations. As this Table demonstrates, even though the numerical complexity behind the PML4 computation is significant, this method may be actually implemented even in a Monte Carlo framework with many replications (here 1,500). With the feasibility of the method already demonstrated, we may now turn to the interpretation of the statistics. We first note that the parameters tend to be estimated rather well if one uses the average of the estimates. We find that the parameter estimates are skewed and kurtic, and we note that the MSE of the parameter estimates increases with the order of the moment that a given parameter describes. The MSE of the mean μ is 0.04. The MSE of the parameter b that describes the kurtosis of the distribution takes a much higher value of 4.490. We observe that parameters describing higher moments have higher associated MSE.

We conclude this section by noticing that our method may obviously be used in real applications, for models which may have more parameters, since it then has to be estimated only once.

7.5. **A Fourth Experiment: QGPML2.**

To validate the QGPML2 approach, we consider as DGP the observation (y_i, x_i) generated by:

$$\begin{aligned}
&y_i = a x_i + \exp(b x_i)\varepsilon_i, \\
&u_i \sim U(0,1), \quad i.i.d. \\
&x_i = (1 + 29\, u_i)/10, \\
&\theta_i = (1 + 29\, u_i)\frac{\pi}{180}, \\
&\varepsilon_i \sim sLD\left(\frac{1}{\sin\theta_i}, \frac{1}{\cos\theta_i}, \sin\theta_i - \cos\theta_i\right).
\end{aligned} \tag{7.8}$$

The first line specifies the mean and the variance of the model as depending on some exogenous variable x_i. The second line defines the u_i as uniform draws. From this basic source of exogenous randomness, we construct x_i as uniform random numbers $U(1/10, 3)$. The fourth line specifies θ_i as an angle that varies between 1 and 30 degrees. The ratio $\pi/180$ converts this angle into radians. The last equation specifies that ε_i is distributed according to the skewed Laplace distribution with mean 0, variance 1, and known skewness and kurtosis. A similar parametrization was chosen in Subsection 7.3. We select as parameters $a = 1$ and $b = 1$.

A preliminary simulation revealed that, for this parametrization, the skewness (kurtosis) of the ε_i ranges between 0 and 2 (respectively, 6 and 9).

The QGPML2 estimation is based on the following steps:

(1) Estimate a and b via PML2. This is tantamount to obtaining a, b by maximizing the function:

$$\sum_{i=1}^{T} -bx_i - \frac{1}{2}\left(\frac{y_i - ax_i}{\exp(bx_i)}\right)^2.$$

(2) Compute the first step estimators $\tilde{\lambda}_{j,i}$ for $j = 1, \cdots, 4$ by using the information contained in $s(x_i)$ and $k(x_i)$. This computation is described in section 5.3. Notice that only $\tilde{\lambda}_{3i}$ and $\tilde{\lambda}_{4i}$ are used in the next step.

(3) Maximize, with respect to a and b, the objective function:

$$\sum_{i=1}^{T} \lambda^*_{0,i} + \lambda^*_{1,i} y_i + \lambda^*_{2,i} y_i^2,$$

where, $\lambda^*_{j,i} = \lambda(ax_i, \exp(bx_i), \tilde{\lambda}_{3i}, \tilde{\lambda}_{4i})$. The computation of the $\lambda^*_{i,j}$ may be found in section 6.4. The resulting a and b estimates are the QGPML2 estimates.

Table 4 reports some statistics for the simulations. Each time we use $N = 30,000$, a rather large number of simulations to ascertain that the findings are not spurious. We consider samples of size $T = 25$ and $T = 100$.[10]

Inspecting the table reveals that, as expected, the dispersion of the estimates obtained in the larger sample tends to be better. Comparing the dispersion of the parameters between the PML2 estimates and the QGPML2 estimates reveals an improvement when using QGPML2.[11] For instance, for $T = 25$, the improvement of the RMSE of the parameter intervening in the mean, a, is 2.6%.

[10]Here, we focus on a large number of simulations each involving a relatively small sample. The time required for this simulation could alternatively be devoted to the estimation of a model with either a larger sample or a more complex model structure involving several parameters.

[11]We did not pursue a search for settings where the efficiency gain may be more important, as we simply wished to demonstrate the feasibility of the method here. We leave this pursuit for future research.

8. Conclusion

In this paper, we generalize the PML2 and QGPML1 methods proposed in Gourieroux, Monfort, and Trognon (1984). The main objective of these methods was to propose consistent and asymptotically normal estimators of the parameters which appear in the specification of the first two conditional moments, based on the optimization of possibly misspecified likelihood functions.

Here, we extend this approach by considering the first four conditional moments. A key tool is the quartic exponential family. This family allows us to introduce PML4 and QGPML2 estimators, respectively, generalizing PML2 and QGPML1. A complete asymptotic theory is proposed. In particular, it is shown that the QGPML2 estimator reaches the semi-parametric bound based on the first two moments.

Another key issue is the numerical computation of the exponential quartic density parameters for given values of the first four moments. The solution adapted in this paper, which is based on an approach proposed by G. Freud (1986), appears to be very quick and stable, and it solves technical problems, which had been stressed in different strands of the literature, e.g. Maasoumi (1993), Ormoneit and White (1999).

In numerical studies, we not only demonstrate the feasibility of the proposed estimation methods, but also show that PML4 may provide more efficient estimates, in particular for small samples where GMM based estimates may have encountered difficulties. We also consider an example where an econometrician uses either a misspecified ML model or the PML4 model. In that case, the PML4 model demonstrates superior results. Lastly, we show the feasibility of the QGPML2 estimation, and in that context, we again prove gains in efficiency.

Our estimation method may prove useful in many econometric applications that involve non-Gaussianity of some random variable. Beyond this, the proposed numerical techniques may be of relevance in Bayesian analysis, independent component analysis, and possibly physics, i.e. in situations where non-Gaussian distributions may be of relevance.

Appendix A. Properties of the Exponential Families

Let us consider the exponential family $\ell(y, \lambda) = \exp[\lambda' T(y) - \psi(\lambda)]$ defined in Section 2.1. An important issue is to characterize the range of the mapping $\lambda \to E_\lambda[T(y)]$, for $\lambda \in \mathring{\Lambda}$. Let us denote this mapping by ξ, and we denote $M = \xi(\mathring{\Lambda})$, and C the closed convex hull of the support of ν^T.

We recall below a necessary and sufficient condition for which $M = \mathring{C}$ due to Barndorff-Nielsen (1978). To this end, we need to recall the definition of a *steep function* [or alternatively an *essentially smooth convex function* (Rockafellar 1970)]. Introduce the following definition. Let $\varphi : \mathbb{R}^k \to (-\infty, \infty)$ be a proper convex function and $D = \{x \in \mathbb{R}^k : \varphi(x) < \infty\}$. Assume that $\mathring{D}$ is not empty and φ is differentiable throughout $\mathring{D}$. Such a function will be defined to be *steep* at x, where $x \in D \setminus \mathring{D}$, if $\lim_{i\to\infty} \mid \frac{\partial \varphi}{\partial x}(x_i) \mid = \infty$ whenever $x_1, x_2, \ldots,$ is a sequence of points in $\mathring{D}$ converging to x. Furthermore, φ will be called *steep* if it is steep at all $x \in D \setminus \mathring{D}$.

An exponential family is called steep if its cumulant generating function, ψ, is steep. It follows from this definition that a natural exponential family is called steep if $\lim_{i\to\infty} \mid E_{\lambda_i}(T) \mid = +\infty$ whenever $\lambda_1, \lambda_2, \ldots,$ is a sequence of points in $\mathring{\Lambda}$ converging to a point $\lambda \in \Lambda \setminus \mathring{\Lambda}$.

A result due to Brown (1986, p. 72) provides a convenient, necessary, and sufficient condition for steepness. According to this result, the exponential family is steep if and only if $E_\lambda(\| T \|) = \infty$ for all $\lambda \in \Lambda \setminus \mathring{\Lambda}$. Unfortunately, this condition is not satisfied by the exponential quartic family since T is integrable in the Gaussian case, corresponding to values of λ satisfying $\lambda_4 = \lambda_3 = 0$, $\lambda_2 < 0$, and which are in $\Lambda \setminus \mathring{\Lambda}$.

Appendix B. Computation of $J(\theta_0)$ and $I(\theta_0)$

$$
\begin{aligned}
J(\theta_0) &= -\frac{\partial^2 \varphi_\infty(\theta, P_0)}{\partial\theta\partial\theta'} \\
&= -E_X\left[\frac{\partial^2 \lambda_0(X,\theta_0)}{\partial\theta\partial\theta'} + \sum_{j=1}^{4} \frac{\partial^2 \lambda_j(X,\theta_0)}{\partial\theta\partial\theta'} m_{j0}\right],
\end{aligned}
$$

We have (omitting the variables X and θ_0):

$$
\begin{aligned}
\frac{\partial \lambda_j}{\partial\theta} &= \frac{\partial m'}{\partial\theta}\frac{\partial \lambda_j}{\partial m}, \quad j = 0, \ldots, 4, \\
\frac{\partial^2 \lambda_j}{\partial\theta\partial\theta'} &= \frac{\partial m'}{\partial\theta}\frac{\partial^2 \lambda_j}{\partial m \partial m'}\frac{\partial m}{\partial\theta'} + \sum_{k=1}^{4} \frac{\partial \lambda_j}{\partial m_k}\frac{\partial^2 m_k}{\partial\theta\partial\theta'},
\end{aligned}
$$

$$J(\theta_0) = -E_X \left\{ \frac{\partial m'}{\partial \theta} \left[\frac{\partial^2 \lambda_0}{\partial m \partial m'} + \sum_{j=1}^{4} \frac{\partial^2 \lambda_j}{\partial m \partial m'} m_{j0} \right] \frac{\partial m}{\partial \theta'} \right\}$$
$$- \left\{ \sum_{k=1}^{4} \left[\frac{\partial \lambda_0}{\partial m_k} + \frac{\partial \lambda_1}{\partial m_k} m_{10} + \cdots + \frac{\partial \lambda_4}{\partial m_k} m_{40} \right] \frac{\partial^2 m_k}{\partial \theta \partial \theta'} \right\},$$

and using Proposition 1, Corollary 1, and Proposition 2:

$$\begin{aligned} J(\theta_0) &= E_X \left[\frac{\partial m'}{\partial \theta} \frac{\partial \lambda'}{\partial m} \frac{\partial m}{\partial \theta'} \right] \\ &= E_X \left[\frac{\partial m'}{\partial \theta} \Sigma^{-1} \frac{\partial m}{\partial \theta'} \right]. \end{aligned}$$

Similarly,

$$\begin{aligned} I(\theta_0) &= E_X E_0 \left[\frac{\partial \varphi(Y, X, \theta_0)}{\partial \theta} \frac{\partial \varphi(Y, X, \theta_0)}{\partial \theta'} \right] \\ &= E_X E_0 \left[\left(\frac{\partial \lambda_0}{\partial \theta} + \frac{\partial \lambda'}{\partial \theta} T \right) \left(\frac{\partial \lambda_0}{\partial \theta'} + T' \frac{\partial \lambda}{\partial \theta'} \right) \right], \end{aligned}$$

where $\varphi(Y, X, \theta) = \sum_{j=0}^{4} \lambda_j(X, \theta) Y^j$, and $T' = (Y, Y^2, Y^3, Y^4)$.

Using Proposition 1, we have (omitting the variables X and θ_0):

$$\begin{aligned} I(\theta_0) &= E_X E_0 \left[\frac{\partial \lambda'}{\partial \theta} (T - m)(T - m)' \frac{\partial \lambda}{\partial \theta'} \right] \\ &= E_X \left(\frac{\partial \lambda'}{\partial \theta} \Omega \frac{\partial \lambda}{\partial \theta'} \right) \\ &= E_X \left(\frac{\partial m'}{\partial \theta} \frac{\partial \lambda'}{\partial m} \Omega \frac{\partial \lambda}{\partial m'} \frac{\partial m}{\partial \theta'} \right) \\ &= E_X \left(\frac{\partial m'}{\partial \theta} \Sigma^{-1} \Omega \Sigma^{-1} \frac{\partial m}{\partial \theta'} \right). \end{aligned}$$

Appendix C. Asymptotic Behavior of the QGPML2

C.1. **Preliminaries.**

Let us consider the quartic family parametrized by $\xi' = (m_1, \sigma^2, \lambda_3, \lambda_4)$. If we fix λ_3, λ_4 to a given value of λ_3^*, λ_4^*, then we obtain a family, indexed by (m_1, σ^2), with log-density:

$$\sum_{j=0}^{2} \lambda_j^*(\xi_{12}) y^j + \lambda_3^* y^3 + \lambda_4^* y^4,$$

where $\xi_{12} = (m_1, \sigma^2)'$ and $\lambda_j^*(\xi_{12})$ is a notation which stands for $\lambda_j^*(m_1, \sigma^2, \lambda_3^*, \lambda_4^*)$. It is the log-density of a quadratic exponential family (see GMT (1984)).

Differentiating with respect to ξ_{12}, we get

$$\sum_{j=0}^{2} \frac{\partial \lambda_j^*}{\partial \xi_{12}} m_j = 0, \qquad \text{(with } m_0 = 1\text{)} \tag{C.1}$$

and

$$\sum_{j=0}^{2} \frac{\partial^2 \lambda_j^*}{\partial \xi_{12} \partial \xi_{12}'} m_j + \frac{\partial \lambda^{*\prime}}{\partial \xi_{12}} \frac{\partial m_{12}}{\partial \xi_{12}'} = 0, \tag{C.2}$$

with $\lambda^* = (\lambda_1^*, \lambda_2^*)'$ and $m_{12} = (m_1, m_2)'$.

Moreover,

$$\frac{\partial m_{12}}{\partial \xi_{12}'} = \begin{pmatrix} 1 & 0 \\ 2m_1 & 1 \end{pmatrix}. \tag{C.3}$$

The variance-covariance matrix Σ_1 of $(Y, Y^2)'$ in this family is found easily. For instance, we note that the pdf with respect to the measure μ^* defined by $d\mu^*(y) = \exp(\lambda_3^* y^3 + \lambda_4^* y^4) dy$ is $\exp\left[\lambda_0^*(\xi_{12}) + \lambda_1^*(\xi_{12}) y + \lambda_2^*(\xi_{12}) y^2\right]$. Using the parametrization $(\lambda_1^*, \lambda_2^*)$ and the general property of the exponential family given in 2.1 (2):

$$\Sigma_1 = \frac{\partial m_{12}}{\partial \lambda^{*\prime}},$$

which, written as a function of ξ_{12}, leads to:

$$\Sigma_1 = \frac{\partial m_{12}}{\partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \lambda^{*\prime}}.$$

We therefore have:

$$\frac{\partial \xi_{12}}{\partial \lambda^{*\prime}} = \left(\frac{\partial m_{12}}{\partial \xi_{12}'} \right)^{-1} \Sigma_1,$$

hence:

$$\frac{\partial \lambda^*}{\partial \xi_{12}'} = \Sigma_1^{-1} \frac{\partial m_{12}}{\partial \xi_{12}'}. \tag{C.4}$$

C.2. **Computation of $\tilde{J}(\theta_0)$.**

In the estimation based on the unfeasible equivalent estimator of the QGPML2, we have, noting $\theta_{12} = (\theta_1, \theta_2)$:

$$\tilde{J}(\theta_0) = -E_X \left\{ \sum_{j=0}^{2} \frac{\partial^2 \lambda_j^*}{\partial \theta_{12} \partial \theta_{12}'} \left[m_1(X, \theta_{10}), \sigma^2(X, \theta_{20}), \right.\right.$$
$$\left.\left. \lambda_3(X, \theta_0), \lambda_4(X, \theta_0) \right] m_j(X, \theta_0) \right\}.$$

But we have also (omitting X and the θ's):

$$\frac{\partial \lambda_j^*}{\partial \theta_{12}} = \frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial \lambda_j^*}{\partial \xi_{12}},$$

$$\frac{\partial^2 \lambda_j^*}{\partial \theta_{12} \partial \theta_{12}'} = \frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial^2 \lambda_j^*}{\partial \xi_{12} \partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \theta_{12}'} + \sum_{k=1}^{2} \frac{\partial \lambda_j^*}{\partial \xi_k} \frac{\partial^2 \xi_k}{\partial \theta_{12} \partial \theta_{12}'}.$$

Therefore,

$$\tilde{J}(\theta_0) = -E_X \left[\frac{\partial \xi_{12}'}{\partial \theta_{12}} \sum_{j=0}^{2} \frac{\partial^2 \lambda_j^*}{\partial \xi_{12} \partial \xi_{12}'} m_j \frac{\partial \xi_{12}}{\partial \theta_{12}'} \right] - E_X \sum_{k=1}^{2} \left[\sum_{j=0}^{2} \frac{\partial \lambda_j^*}{\partial \xi_k} m_j \right] \frac{\partial^2 \xi_k}{\partial \theta_{12} \partial \theta_{12}'},$$

or, using (C.1) and (C.2),

$$\tilde{J}(\theta_0) = E_X \left[\frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial \lambda^{*'}}{\partial \xi_{12}} \frac{\partial m_{12}}{\partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \theta_{12}'} \right],$$

or, using (C.4),

$$\tilde{J}(\theta_0) = E_X \left[\frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial m_{12}'}{\partial \xi_{12}} \Sigma_1^{-1} \frac{\partial m_{12}}{\partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \theta_{12}'} \right]. \tag{C.5}$$

C.3. **Computation of $\tilde{I}(\theta_0)$.**

Letting $T_1 = (Y, Y^2)'$, we have,

$$\tilde{I}(\theta_0) = E_X E_0 \left[\left(\frac{\partial \lambda_0^*}{\partial \theta_{12}} + \frac{\partial \lambda^{*\prime}}{\partial \theta_{12}} T_1 \right) \left(\frac{\partial \lambda_0^*}{\partial \theta_{12}'} + T_1' \frac{\partial \lambda^*}{\partial \theta_{12}'} \right) \right],$$

or, using (C.1),

$$\tilde{I}(\theta_0) = E_X E_0 \left[\frac{\partial \lambda^{*\prime}}{\partial \theta_{12}} (T_1 - m_{12}) (T_1 - m_{12})' \frac{\partial \lambda^*}{\partial \theta_{12}'} \right] = E_X \left[\frac{\partial \lambda^{*\prime}}{\partial \theta_{12}} \Omega_1 \frac{\partial \lambda^*}{\partial \theta_{12}'} \right].$$

We also have

$$\frac{\partial \lambda^{*'}}{\partial \theta_{12}} = \frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial \lambda^{*\prime}}{\partial \xi_{12}} = \frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial m_{12}'}{\partial \xi_{12}} \Sigma_1^{-1},$$

which allows us to write

$$\tilde{I}(\theta_0) = E_X \left[\frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial m_{12}'}{\partial \xi_{12}} \Sigma_1^{-1} \Omega_1 \Sigma_1^{-1} \frac{\partial m_{12}}{\partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \theta_{12}'} \right].$$

However, since the moments of order one to four are well specified in the pseudo family, the true conditional variance $\Omega_1(X, \theta_0)$ of T_1 is that corresponding to the model. Namely, they correspond to $\Sigma_1(X, \theta_0)$, and we get:

$$\tilde{I}(\theta_0) = E_X \left[\frac{\partial \xi_{12}'}{\partial \theta_{12}} \frac{\partial m_{12}'}{\partial \xi_{12}} \Omega_1^{-1} \frac{\partial m_{12}}{\partial \xi_{12}'} \frac{\partial \xi_{12}}{\partial \theta_{12}'} \right],$$

and hence, from (C.5):

$$\tilde{I}(\theta_0) = \tilde{J}(\theta_0).$$

Moreover, using (C.3),

$$\frac{\partial m_{12}}{\partial \xi_{12}'}\frac{\partial \xi_{12}}{\partial \theta_{12}'} = \begin{pmatrix} 1 & 0 \\ 2m_1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial m_1(X,\theta_1)}{\partial \theta_1} & 0 \\ 0 & \frac{\partial \sigma^2(X,\theta_2)}{\partial \theta_2} \end{pmatrix},$$

and $\tilde{I}(\theta_0)^{-1} = \tilde{J}(\theta_0)^{-1} = B(\theta_0)$, as given in (5.2).

For the sake of simplicity, we have assumed that $m_1(X, \theta_1)$ is only a function of θ_1 and that $\sigma^2(X, \theta_2)$ is only a function of θ_2. The result, however, is easily generalized to the case $m_1(X, \theta_{12}), \sigma^2(X, \theta_{12})$, with the only difference being that $\partial \xi_{12}/\partial \theta_{12}'$ is no longer diagonal.

References

[1] Agmon, N., Y. Alhassid, and R.D. Levine (1979): "An Algorithm for Finding the Distribution of Maximal Entropy," *Journal of Computational Physics,* 30, 250–259.

[2] Andersen, T.G., and B. Sorenson (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business & Economic Statistics,* 14, 328–352.

[3] Altonji, J.G. and Segal, L.M. (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics,* 14(3), 353–366.

[4] Arellano-Valle, R.B., and M.G., Genton (2005): "On Fundamental Skew Distributions," *Journal of Multivariate Analysis,* 96, 93–116.

[5] Barndorff-Nielsen, O.E. (1978): *Information and Exponential Families in Statistical Theory.* Wiley, Chichester.

[6] Barndorff-Nielsen, O.E. (1997): "Normal Inverse Gaussian Distributions and Stochastic Volatility Modeling," *Scandinavian Journal of Statistics,* 24, 1–13.

[7] Brown, L.D. (1986): "Fundamentals of Statistical Exponential Families," *Institute of Mathematical Statistics,* Hayward, California.

[8] Bollerslev, T., and J.M. Wooldridge (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews,* 11(2), 143–172.

[9] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics,* 34, 305–334.

[10] Doran, H.E., and P. Schmidt (2006): "GMM Estimators with Improved Finite Sample Properties using Principal Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model," *Journal of Econometrics,* 133(1), 387–409.

[11] Eberlein, E., and U. Keller (1995): "Hyperbolic Distributions in Finance," *Bernoulli,* 1, 281–299.

[12] Freud, G., (1986): "On the Greatest Zero of an Orthogonal Polynomial, I," *Journal of Approximation Theory,* 46, 16–24.

[13] Gale D., and H. Nikaido (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen,* 159(2), 81–93.

[14] Gallant A.R. (1987): *Nonlinear Statistical Models.* John Wiley & Sons, New York.

[15] Gautschi, W. (2004): *Orthogonal Polynomials: Computation and Approximation.* Oxford Science Publications, Oxford University Press.

[16] Genton M.G. (2004): "Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality," Edited vol., *Chapman & Hall/CRC*, Boca Raton, Florida.

[17] Golan, A., G. Judge, and D. Miller (1996): *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* Wiley, Chichester.

[18] Golub, G.H., and J.H. Welch (1969): "Calculation of Gauss Quadrature Rules," *Mathematics of Computations,* 23(106), 221–230.

[19] Gourieroux C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica,* 52(1), 681–700.

[20] Gourieroux C., and A. Monfort (1995a): *Statistics and Econometric Models: Volume One.* Cambridge University Press, Cambridge.

[21] Gourieroux C., and A. Monfort (1995b): *Statistics and Econometric Models: Volume Two.* Cambridge University Press, Cambridge.

[22] Gourieroux C., and A. Monfort (2006): "Pricing with Splines," *Annales d'Economie et de Statistique,* 82, 4–33.

[23] Hansen, L.P. (1982): "Large Sample Properties of the Generalized Methods of Moments," *Econometrica,* 50, 1029–1054.

[24] Hansen, B. (1994): “Autoregressive Conditional Density Estimation,” *International Economic Review,* 35, 705–730.

[25] Harvey, C.R., and Siddique, A. (1999): “Autoregressive Conditional Skewness,” *Journal of Financial and Quantitative Analysis,* 34(4), 465–487.

[26] Holly A. (1993): “Asymptotic Theory for Nonlinear Econometric Models: Estimation,” in Aart J. de Zeeuw (ed.) *Advanced Lectures in Quantitative Economics II,* Academic Press.

[27] Holly, A., and Y. Pentsak (2004): “Maximum Likelihood Estimation of the Conditional Mean $E(y|x)$ for Skewed Dependent Variables in Four-parameter Families of Distribution,” Technical Report, Institute of Health Economics and Management (IEMS), University of Lausanne, Switzerland.

[28] Hood, W., and T. Koopmans (1953): “The Estimation of Simultaneous Linear Economic Relationships,” in *Studies in Econometric Method.* New Haven: Yale University Press.

[29] Jondeau, E., and M. Rockinger (2003): “Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements,” *Journal of Economic Dynamics and Control,* 27(10), 1699–1737.

[30] Junk, M. (2000): “Maximum Entropy for Reduced Moment Problems,” *Mathematical Models and Methods in Applied Sciences,* 10(7), 1001–1025.

[31] Kitamura, Y., and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica,* 65(4), 861–874.

[32] Levin, E., and D.S. Lubinsky (2001): *Orthogonal Polynomials for Exponential Weights.* CMS Books in Mathematics, Springer-Verlag, New-York.

[33] Letac, G. (1992): “Lectures on Natural Exponential Families and their Variance Functions.” *Monografias de matemática,* 50, IMPA, Rio de Janeiro.

[34] Manning W., A. Basu, and J. Mullahy (2005): “Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data,” *Journal of Health Economics,* 24, 465–488.

[35] Maasoumi, E. (1993): “A Compendium To Information Theory in Economics and Econometrics,” *Econometric Reviews,* 12(2), 137–181.

[36] Mead, L.R., and N. Papanicolaou (1984): “Maximum Entropy in the Problem of Moments,” *Journal of Mathematical Physics,* 25(8), 2404–2417.

[37] Monfort, A. (1982): *Cours de Statistique Mathématique.* Economica, Paris.

[38] Newey, W.K. (1990): “Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics,* 5, 99–135.

[39] Noschese S., and L. Pasquini (1999): “On the Nonnegative Solution of a Freud Three-term Recurrence,” *Journal of Approximation Theory,* 99, 54–67.

[40] Ormoneit, D., and H. White (1999): “An Efficient Algorithm to Compute Maximum Entropy Densities,” *Econometric Reviews,* 18(2), 127–140.

[41] Rockafellar, R.T. (1970): *Convex Analysis.* Princeton University Press.

[42] Sawa, T. (1978): “Information Criteria for Discriminating Among Alternative Regression Models,” *Econometrica,* 46(6), 1273-1291.

[43] Stacy, E. (1962): “A Generalization of Gamma Distribution,” *Annals of Mathematical Statistics,* 33, 1187–1192.

[44] Stacy, E., and G. Mihram (1965): “Parameter Estimation for a Generalized Gamma Distribution,” *Technometrics,* 7, 349–358.

[45] Tauchen, G. (1986): “Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data,” *Journal of Business & Economic Statistics,* 4(4), 397–416.

[46] Titterington, D.M., A.F.M. Smith, and U.E., Makov (1985): *Statistical Analysis of Finite Mixture Distributions.* John Wiley, Chichester.

[47] White, H. (1982): "Maximum Likelihood of Misspecified Models," *Econometrica,* 50(1), 1–25.

[48] White, H. (1994): *Estimation, Inference, and Specification Analysis.* Cambridge University Press, Cambridge, UK.

[49] Zellner, A., and R.A. Highfield (1988): "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions," *Journal of Econometrics,* 37(2), 195–209.

[50] Ziliak, J.P. (1997) : "Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics,* 15, 419–431.

	Mean $\hat{b}_0$	5-ptile $\hat{b}_0$	95-ptile $\hat{b}_0$	MSE $\hat{b}_0$	Mean $\hat{b}_1$	5-ptile $\hat{b}_1$	95-ptile $\hat{b}_1$	MSE $\hat{b}_1$	Total MSE
					$T = 100$				
ML	2.440	1.477	3.278	0.166	1.317	1.050	1.651	0.024	0.190
PML4	2.504	1.470	3.494	0.231	1.338	1.053	1.694	0.028	0.260
PML2	2.508	1.492	3.669	0.276	1.332	1.035	1.700	0.030	0.306
GMM	2.873	0.001	10.000	2.182	1.650	0.000	5.000	0.587	2.769
					$T = 50$				
ML	2.498	1.487	3.742	0.314	1.346	0.976	1.861	0.054	0.367
PML4	2.606	1.461	4.062	0.454	1.385	0.988	1.946	0.068	0.522
PML2	2.647	1.459	4.585	0.683	1.375	0.963	1.929	0.068	0.751
GMM	3.108	0.001	10.000	2.972	1.824	0.000	5.000	0.999	3.971
					$T = 25$				
ML	2.540	1.389	4.131	0.504	1.397	0.882	2.225	0.128	0.632
PML4	2.709	1.386	4.528	0.739	1.466	0.907	2.408	0.173	0.912
PML2	2.860	1.380	6.058	1.675	1.456	0.879	2.365	0.169	1.844
GMM	3.559	0.001	10.000	5.096	2.160	0.000	5.000	2.209	7.305

Table 1: This Table presents statistics for the estimates of a skewed Laplace distribution where the true parameters are $b_0 = 2.414$ and $b_1 = 1.301$. There are 10'000 simulations for sample sizes of $T = 100$, 50, and 25. The 5 (95) percentile of these simulations is denoted by 5$-$ptile (95$-$ptile). The mean squared error is denoted by MSE. The Total MSE is the sum of the MSE of $\hat{b}_0$ plus the one of $\hat{b}_1$.

	Mixture of normals with constant variances (True parametres are $(\mu, \sigma, \theta) = (0, 1, 10)$.												
	Mean $\hat{\mu}$	5-ptile $\hat{\mu}$	95-ptile $\hat{\mu}$	MSE $\hat{\mu}$	Mean $\hat{\sigma}$	5-ptile $\hat{\sigma}$	95-ptile $\hat{\sigma}$	MSE $\hat{\sigma}$	Mean $\hat{\theta}$	5-ptile $\hat{\theta}$	95-ptile $\hat{\theta}$	MSE $\hat{\theta}$	Total MSE
ML	0.000	-0.034	0.034	0.0004	0.950	0.915	0.987	0.0029	27.943	23.951	31.737	327.49	327.49
PML4	-0.001	-0.036	0.034	0.0004	0.998	0.966	1.030	0.0004	8.797	0.055	15.043	23.99	23.99
	Mixture of normals, 5 parameter case												
	Mean $\hat{\mu}$	5-ptile $\hat{\mu}$	95-ptile $\hat{\mu}$	MSE $\hat{\mu}$	Mean $\hat{\sigma}$	5-ptile $\hat{\sigma}$	95-ptile $\hat{\sigma}$	MSE $\hat{\sigma}$	Mean $\hat{\theta}$	5-ptile $\hat{\theta}$	95-ptile $\hat{\theta}$	MSE $\hat{\theta}$	Total MSE
ML	0.001	-0.037	0.039	0.0005	0.956	0.909	1.002	0.0027	24.936	21.194	28.673	228.24	228.25
PML4	-0.002	-0.044	0.040	0.0006	0.994	0.939	1.057	0.0013	8.881	0.050	16.321	30.08	30.08

Table 2: This Table displays the results from 10'000 simulations of samples, each containing $T = 1'000$ observations. The true DGP is either a mixture of normals with constant variances or a 5 parameter normal whose entropy is maximal. The true parameters are $\mu = 0$, $\sigma = 1$, $s = 1.90$, and $k = 8.65$. We estimate via ML the (misspecified) skewed Laplace distribution for which the skewness-kurtosis parameters translate into $\theta = 10$, see also (7.6). We also estimate θ via PML4 where we use a functional specification of skewness and kurtosis corresponding to the skewed Laplace distribution. The meaning of the heading is the same as in Table 1.

	μ	σ	a	b
True parameters	0	1	1	2
Average	0.006	1.015	1.109	2.510
Std	0.071	0.136	0.641	2.058
Median	0.000	1.011	1.057	1.794
Min	-0.277	0.695	0.000	0.000
Max	0.304	1.950	4.000	7.000
Sk	0.666	0.980	1.886	1.290
σ_{Sk}	0.063	0.063	0.063	0.063
Ku	5.607	6.869	9.500	3.398
σ_{Ku}	0.126	0.126	0.126	0.126
MSE	0.005	0.019	0.422	4.490

Table 3: This Table presents the statistics of 1'500 PML4 estimates of μ, σ, a and b as described in the model (7.7). Each sample contained $T = 100$ observations. Sk, Ku represent the skewness and kurtosis of the parameters. By σ_{Sk} and σ_{Ku} we represent the standard deviation of the skewness and kurtosis estimates.

	PML2			
	$T = 25$		$T = 100$	
True parameters	$a = 1$	$b = 1$	$a = 1$	$b = 1$
Mean	0.996	0.915	0.994	0.956
STD	0.567	0.249	0.401	0.177
min	0.001	0.001	0.001	0.331
max	4.464	2.206	2.848	1.619
RMSE	0.567	0.263	0.401	0.182
	QGPML2			
	$T = 25$		$T = 100$	
True parameters	$a = 1$	$b = 1$	$a = 1$	$b = 1$
Mean	0.997	0.917	0.998	0.957
STD	0.552	0.247	0.393	0.176
min	0.001	0.001	0.001	0.330
max	3.880	2.200	2.543	1.641
ΔRMSE (%)	2.606	0.937	2.193	0.728

Table 4: This Table reports the results of the QGPML2 simulation described in model (7.8). The true parameters are $a = 1$, and $b = 1$. The RMSE is defined as $\left(\frac{1}{M}\sum_{j=1}^{M}(\hat{\theta}^{(j)} - \theta)^2\right)^{1/2}$, where $\theta = a$ or b. Here, the superscript $j = 1, \cdots, M$ denotes a simulation. We took $M = 30'000$. By ΔRMSE (%) we denote the percentage gain in the MSE if one uses QGPML2 instead of PML2.

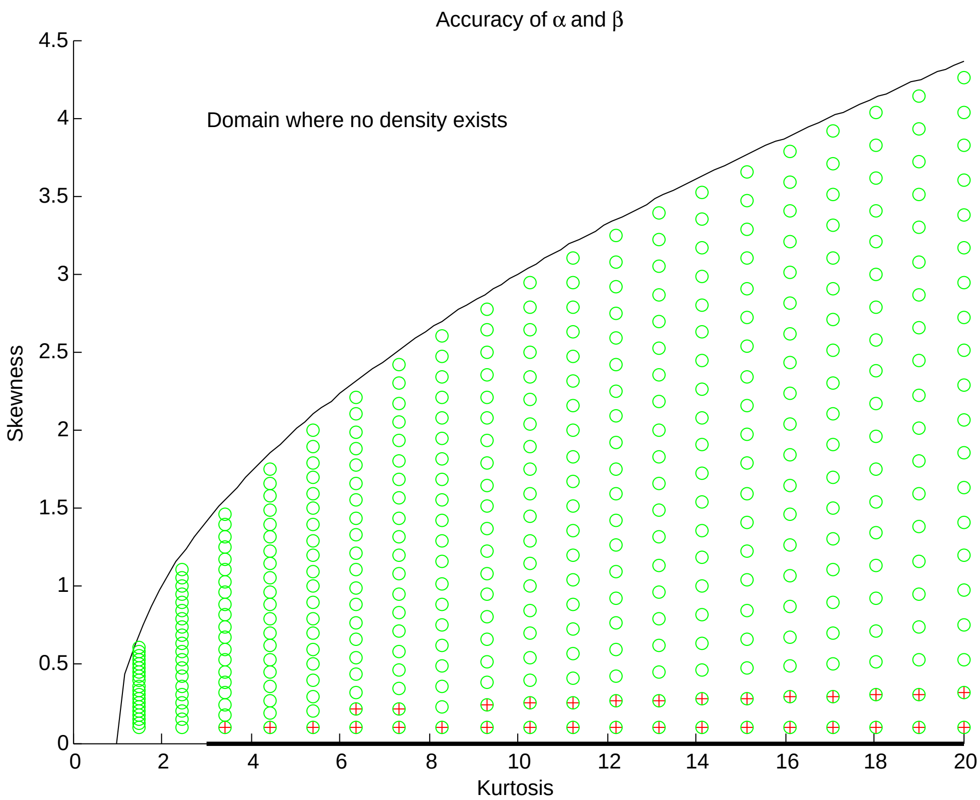

FIGURE 1. This figure represents the skewness-kurtosis domain for which a density exists (the domain is symmetric with respect to the horizontal axis). The circles represent those points for which we computed the parameters α and β. The symbol $+$ represents those points for which the distance between the original skewness and kurtosis and the recomputed skewness and kurtosis (after evaluation of the α and β) is larger than 10^{-5}.

Institute of Health Economics and Management (IEMS)
UNIL Dorigny
Extranef
1015 Lausanne
Switzerland
Phone +41 (0)21 692 33 20
Fax +41 (0)21 692 36 55

www.hec.unil.ch/iems