Efficient GMM Estimation Using the Empirical Characteristic Function*

Marine Carrasco[†]
University of Rochester
Department of Economics

Jean-Pierre Florens
GREMAQ and IDEI,
Université Toulouse I

February 2002

Abstract

This paper extends the GMM method proposed by Carrasco and Florens (2000) to handle moment conditions based on the empirical characteristic function (e.c.f.). While the common practice of selecting a finite number of grid points results in a loss of efficiency, our GMM method exploits the full continuum of moment conditions provided by the e.c.f.. We show that our estimator is asymptotically efficient. Its implementation requires a smoothing parameter that can be selected by minimizing the mean square error of the estimator. An application to a convolution of distributions is discussed. A Monte Carlo experiment shows that the finite sample properties of our estimator are good.

JEL classification: C13

Keywords: Convolution; Efficiency; GMM; Mixture; Stable distribution.

*The authors are grateful to Christia Gouriéroux, Fra cis Kramarz, Nour Meddahi, a d Josef Per told for helpful comme ts. They also tha the participa ts of the co fere ces: Eco ometrics Society World Co gress (2000) a d Paris-Berli (2000). A former versio of this paper has circulated u der the ame "Estimatio of a Co volutio via the Empirical Characteristic Fu ctio ".

†Address: U iversity of Rochester, Departme t of Eco omics, Har ess Hall 226, Rochester, NY 14627, USA. Tel: (585) 275 7307. Fax: (585) 256-2309. Email: csco@troi.cc.rochester.edu.

1. Introduction

In many circumstances, the likelihood function does not have a simple tractable expression. Examples, that will be developed later, are the convolution of distributions and the stable law. In such instances, estimation using the characteristic function offers a nice alternative to maximum likelihood method. The empirical characteristic function (e.c.f.) has been used for inference as early as 1977 in a paper by Feuerverger and Mureika. Feuerverger and McDunnough (1981b) show that an efficient estimator can be obtained by matching the e.c.f. ψ_n with the theoretical characteristic function ψ in the following manner

$$\int w_\theta(t)\left(\psi_n(t) - \psi(t)\right) dt = 0$$

where w_θ is the inverse Fourier transform of the score. This weighting function is computable for special cases like the Cauchy distribution, but is impossible to obtain such cases as a general stable distribution. Alternatively, Feuerverger and McDunnough propose to apply the Generalized Method of Moment (GMM) on a finite number of moments $\psi_n(t) - \psi(t)$, $t = t_1, t_2, ..., t_q$ obtained from a discretization of an interval of $\mathbf{R}$. This research path has been followed by many authors either for estimation (Feuerverger and McDunnough, 1981a, Feuerverger, 1990, Tran, 1998, Singleton, 2001) or for goodness of fit tests (Koutrouvelis and Kellermeier, 1981, Epps and Pulley, 1983). A difficulty with that method is that certain choices must be made before estimating the parameters: the choice of (a) the number of points q and (b) the grid $t_1, t_2, ..., t_q$. For a given q, Schmidt (1982) proposes to choose the grid that minimizes the determinant of the asymptotic covariance matrix of the estimators. The problem of the choice of q remains. One reason for the success of the GMM approach is that Feuerverger and McDunnough (1981b) show that the asymptotic variance of the GMM estimator "can be made arbitrarily close to the Cramer-Rao bound by selecting the grid sufficiently fine and extended". This drives them (and other researchers who followed) to conclude that their estimator is asymptotically efficient. This is not true. When the grid is too fine, the covariance matrix becomes singular and the GMM objective function is not bounded, hence the efficient GMM estimator can not be computed.

Our contribution is to give a tractable and relatively easy to implement method that delivers asymptotically efficient estimators. Instead of taking a finite grid, we apply GMM to the continuum of moment conditions resulting from the e.c.f., using the method developed by Carrasco and Florens (2000) (to be referred as CaFl). In particular, we introduce a penalization term, α_n, which converges to zero with the sample size and guarantees that the GMM objective function is always bounded. A close investigation shows that Carrasco and Florens' results give a rationale to Feuerverger and McDunnough's approach. Using our continuous GMM method avoids the explicit derivation of the optimal weighting function as in Feuerverger and McDunnough. We give a general method to estimate it from the data.

As models frequently include explanatory variables, we discuss the efficient estimation based on the conditional characteristic function. We also show that our estimation procedure can be used to construct goodness-of-fit tests. Since the choice of the penalization term, α_n, is potentially a problem in practice, we derive a method to choose it endogenously from the data. We propose to select the value that minimizes the mean square error of the estimator. As the GMM estimator is unbiased for any value of α_n, we derive a higher-order asymptotic bias.

In Section 2, we give the principal definitions and three examples. Section 3 reviews the results of CaFl. Section 4 explains how to obtain efficient estimators using the (unconditional) characteristic function. In Section 5, we turn our attention to the use of the conditional characteristic function. Section 6 discusses the implementation of the method and the endogenous choice of the penalization term. Section 7 presents a specification test. Section 8 develops an example and discusses the results of a limited Monte Carlo experiment. Finally, Section 9 concludes. The list of the assumptions and the proofs of the main propositions are in the Appendix.

2. Definitions and examples

2.1. Definitions

Suppose $X_1, ..., X_n$ are i.i.d. realizations of the same random variable X with density $f(x;\theta)$ and distribution function $F_\theta(x)$, X is possibly multivariate so that $X \in \mathbf{R}^p$. The parameter $\theta \in \mathbf{R}^q$ is the parameter of interest with true value θ_0. Let $\psi_\theta(t)$ denote the characteristic function of X

$$\psi_\theta(t) \equiv \int e^{it'x} dF_\theta(x) = E^\theta\left(e^{it'X}\right)$$

and $\psi_n(t)$ denote the empirical characteristic function

$$\psi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{it'X_j}.$$

The focus of the paper is on moment conditions of the form

$$h(t, X_j; \theta) = e^{it'X_j} - \psi_\theta(t). \tag{2.1}$$

Obviously h satisfies

$$E^\theta\left[h(t, X_j; \theta)\right] = 0 \text{ for all } t \text{ in } \mathbf{R}^q.$$

Our aim is to use this continuum of moment conditions to obtain an efficient estimator of θ.

Let π be a probability density function on $\mathbf{R}^p$ and $L^2(\pi)$ be the Hilbert space of complex valued functions such that

$$L^2(\pi) = \left\{ f : \mathbf{R}^p \to \mathbf{C} \mid \int |f(t)|^2 \pi(t)\, dt < \infty \right\}.$$

Note that while $e^{it'X}$ and $\psi_\theta(t)$ are not necessarily integrable with respect to Lebesgue measure on $\mathbf{R}^p$, $e^{it'X}$ and $\psi_\theta(t)$ belong to $L^2(\pi)$ for any probability measure π because $\left|e^{it'X}\right| = 1$ and $|\psi_\theta(t)| \leq \psi_\theta(0) = 1$. A candidate for π is the standard Normal distribution. The inner product on $L^2(\pi)$ is defined as

$$\langle f, g \rangle = \int f(t)\, \overline{g(t)} \pi(t)\, dt \tag{2.2}$$

where $\overline{g(t)}$ denotes the complex conjugate of $g(t)$. Let $\|.\|$ denote the norm in $L^2(\pi)$. In the following, whenever f and g are $l \times 1-$vectors of elements of $L^2(\pi)$, $\langle f, g \rangle$ denotes the $l \times l-$matrix $\int f\overline{g}'$. By extension, for a vector g of elements of $L^2(\pi)$, $\|g\|^2$ denotes the matrix

$$\|g\|^2 = \int g(t)\, \overline{g(t)}' \pi(t)\, dt. \tag{2.3}$$

2.2. Examples

In this subsection, we give three motivating examples.

Example 1: Finite mixture of distributions

Finite mixture models are commonly used to model data from a population composed of a finite number of homogeneous subpopulations. Examples of applications are the estimation of a cost function in the presence of multiple technologies of production (Beard, Caudill, and Gropper, 1991) and the detection of sex bias in health outcomes in Bangladesh (Morduch and Stern, 1997). Ignoring heterogeneity may lead to seriously misleading results.

Consider the case where a population is supposed to be formed of two homogeneous subpopulations. Let λ be the unknown proportion of individuals of type 1. Individuals of type j have a density $f(x, \theta_j)$, $j = 1, 2$. The econometrician does not observe the type of the individuals, so that the likelihood for one observation is

$$\theta f(x; \theta_1) + (1 - \lambda) f(x; \theta_2).$$

Such models can be estimated using the EM algorithm or the method of moments, see Heckman, Robb, and Walker (1990), among others. The likelihood of a finite mixture of normal distributions with different variances being unbounded, maximum likelihood may break down. An alternative method is to use either the moment generating function (Quandt and Ramsey, 1978, Schmidt, 1982) or the characteristic function (Tran, 1998) which is equal to

$$\lambda \psi_{\theta_1}(t) + (1 - \lambda) \psi_{\theta_2}(t)$$

where $\psi_{\theta_j}(t) = \int e^{itx} f(x, \theta_j) dx$ with $j = 1, 2$.

Example 2: Convolution of distributions

Assume one observes i.i.d realizations of durations T and exogenous variables Z. Let T_0 be a latent duration and ε an unmeasured person-specific heterogeneity such that for an individual i, one observes

$$T_i = \exp\left(\beta' Z_i + \varepsilon_i\right) T_{0i} \tag{2.4}$$

where ε_i and T_{0i} are assumed to be independent. Lancaster's book (1990) gives many examples of (2.4). T_i may be, for instance, the unemployment spell. Taking the logarithm, we have the regression

$$X_i \equiv \ln T_i = \beta' Z_i + \ln T_{0i} + \varepsilon_i \equiv \beta' Z_i + \eta_i + \varepsilon_i \tag{2.5}$$

Models of this type have been more often specified in terms of hazard than in terms of regression. While (2.5) gives rise to a convolution problem, specification in terms of hazard gives rise to a mixture problem. Estimation by maximum likelihood where the mixing distribution is integrated out can be performed using the EM algorithm or by the Newton-Raphson algorithm. A convolution of the type (2.5) also appears in the random-effect models where η_i is the random, individual effect and ε_i is an error component (Hsiao, 1986). In most cases, the likelihood will have an intractable form whereas the characteristic function of $U = \eta + \varepsilon$ is easily obtained from

$$\psi_U = \psi_\eta \times \psi_\varepsilon$$

where ψ_η and ψ_ε are the characteristic functions of η and ε respectively. Our results will allow for the presence of covariates in the model.

Example 3. The stable distribution

The stable distribution is frequently applied to model financial data. Various examples can be found in the survey by McCulloch (1996). Except for a few cases, the likelihood of the stable distribution does not have a simple expression, however its characteristic function is well known. Many authors have used the characteristic function to estimate the parameters of the stable distribution, see Feuerverger and McDunnough (1981b) and the references in McCulloch, but none of them have efficiently used a continuum of moments.

3. Brief review of GMM when a continuum of moments is available

In this section, we summarize the results of CaFl and discuss their application to the estimation using the c.f.. Let $H = L^2(\pi)$ be the Hilbert space of reference. Let B be a bounded linear operator defined on H, or a subspace of H (such that $h \in H$), and B_n a sequence of random bounded linear operators converging to B. Let

$$h_n(t;\theta) = \frac{1}{n}\sum_{j=1}^{n} h(t, X_j;\theta).$$

The GMM estimator is such that

$$\hat{\theta}_n = \arg\min_{\theta} \|B_n h_n(.;\theta)\|.$$

Under a set of conditions listed in CaFl, this estimator is consistent and asymptotically normal. In the class of all weighting operators B, one yields an estimator with minimal variance. This optimal B is shown to be equal to $K^{-1/2}$ where K is the covariance operator associated with $h(t, X; \theta)$. The operator K is defined as

$$\begin{aligned} K &: f \in H \to g \in H \\ f(t) &\to g(s) = \int k(s,t) f(t) \pi(t) dt \end{aligned}$$

where

$$k(s,t) = E^{\theta_0}\left[h(s, X; \theta_0) \overline{h(t, X; \theta_0)}\right].$$

As $k(s,t) = \overline{k(t,s)}$, K is self-adjoint, which means that it satisfies $\langle Kf, g\rangle = \langle f, Kg\rangle$. Under some assumptions, K has a countable infinity of (positive) eigenvalues decreasing to zero; as a result, its inverse is not bounded. $K^{-1/2}g$ does not exist on the whole space H but on a subset of it, which corresponds to the so-called reproducing kernel Hilbert space (RKHS) associated with K denoted $\mathcal{H}(K)$. We use the notation

$$\left\|K^{-1/2}g\right\|^2 = \|g\|_K^2$$

where $\|.\|_K$ denotes the norm in $\mathcal{H}(K)$. Because the inverse of K is not bounded, we use a penalization term α_n to guarantee the existence of the inverse. The estimation of K and the choice of α_n will be discussed in Section 6. Let $K_n^{\alpha_n}$ denote the consistent estimator of K described in Section 6. The optimal GMM estimator of θ is obtained by

$$\hat{\theta}_n = \arg\min_{\theta} \left\|\left(K_n^{\alpha_n}\right)^{-1/2} h_n(.;\theta)\right\|. \quad (3.1)$$

Under the assumptions of Theorem 8 in CaFl, we have the following results

$$\hat{\theta}_n \to \theta_0 \text{ in probability,}$$

as n and $n\alpha_n^{3/2}$ go to infinity and α_n goes to zero and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \left(\left\|E^{\theta_0}(\nabla_\theta h)\right\|_K^2\right)^{-1}\right) \quad (3.2)$$

as n and $n\alpha_n^3$ go to infinity and α_n goes to zero. Remark that $\left\|E^{\theta_0}(\nabla_\theta h)\right\|_K^2$ denotes a matrix using the convention defined in (2.3).

The results of CaFl were proven in the case where π is the pdf of a Uniform distribution on $[0,T]$ while here π is a pdf on $\mathbf{R}^p$. Moreover in CaFl, H is a space of real-valued functions $L^2[0,T]$ while here $H = L^2(\pi)$ is a space of complex-valued functions. However, all the results on operator theory and reproducing kernel Hilbert space remain valid (see a treatment in a very general setting by Saitoh, 1997). Hence the results of CaFl can be transposed to our present setting by replacing $L^2[0,T]$ by $L^2(\pi)$ and adjusting the operations in these spaces wherever they appear. All the needed assumptions will be carefully checked.

4. Estimation using the characteristic function

Our decision to use the c.f. is motivated by efficiency considerations. Below, we first discuss which moment conditions permit to reach efficiency. Next, we study the properties of the GMM estimator.

4.1. Asymptotic efficiency

In this subsection, we assume that the result (4.3) holds. We discuss the choice of moment conditions h for which the asymptotic variance of $\hat{\theta}_n$ coincides with the Cramer Rao efficiency bound. Denote $L^2(X;\theta_0)$ the space $\left\{G(X)\,|E^{\theta_0}G(x)^2<\infty\right\}$. The covariance kernel k takes the form

$$k(s,t)=E^{\theta_0}\left[h(s,X;\theta_0)\overline{h(t,X;\theta_0)}\right] \tag{4.1}$$

where $\{h(t,.;\theta_0),t\in\mathbf{R}^p\}$ is a family of functions $L^2(X;\theta)$ that satisfy $E^{\theta_0}h(t,X;\theta_0)=0$.

For any $p\times 1$ vector g of functions in $L^2(\pi)$, we denote $C(g)$ the set of functions so that

$$C(g)=\left\{G\ p-\text{vectors of r.v. in } L^2(X)\ \text{ s.t. } \left\{\begin{array}{c} E^{\theta_0}G=0, \\ g(t)=E^{\theta_0}\left[G(X)\overline{h(t,X;\theta_0)}\right]\quad \pi-a.s. \end{array}\right\}\right\} \tag{4.2}$$

We want to characterize the norm $\|g\|_K^2$. Parzen (1970, page 25) gives a simple formula to calculate the norm of a function g in the RKHS associated with a covariance kernel k on $\mathbf{R}^2$. See Saitoh (1997) for the treatment of complex-valued functions and a very general space of reference. A proof of the following result can be found in Carrasco, Chernov, Florens, and Ghysels (2001) under very general conditions.

Proposition 4.1 (Parzen). *For any function g in $\mathcal{H}(K)$, the following relation holds*

$$\|g\|_K^2=\min_{G\in C(g)}\left[E^{\theta_0}(GG')\right].$$

Proposition 4.1 will be useful in the following to establish the asymptotic efficiency of our GMM estimator $\hat{\theta}_n$. Its asymptotic variance is given by $\left(\left\|E^{\theta_0}(\nabla_\theta h)\right\|_K^2\right)^{-1}$ the inverse of the norm of $g=E^{\theta_0}(\nabla_\theta h)$. Let S be the (not necessarily closed) space spanned by $\{h(t,X,\theta_0);t\in\mathbf{R}^p\}$. Let $\bar{S}$ be the closure of S. It consists of all finite combinations of the form $\sum_{l=1}^m\omega_l h(t_l,X;\theta_0)$ and limits in norm of these linear combinations. Assume that π satisfies Assumption 2.

Proposition 4.2. *The GMM estimator based on the moment conditions* $h(t, X; \theta)$ *reaches asymptotically the Cramer Rao efficiency bound*

(i) if and only if $\partial \ln f(X; \theta)/\partial\theta$ *belongs to* $\bar{S}$.

(ii) if $\{h(t, X; \theta_0)\}$ *is complete that is*

$$\begin{cases} E^{\theta_0}[G(X)] = 0 \\ E^{\theta_0}\left[G(X)\overline{h(t, X; \theta_0)}\right] = 0, \quad \pi - a.s \end{cases} \Rightarrow G = 0.$$

Note that if $\{h(t, X; \theta_0)\}$ is complete then the solution to Equation (4.2) is unique. Hence, the result of (ii) follows from the fact that the score always belongs to $C\left(E^{\theta_0}(\nabla_\theta h)\right)$. Remark that the condition satisfied by the elements of $C\left(E^{\theta_0}(\nabla_\theta h)\right)$ is basically the same condition as in Lemma 4.2 of Hansen (1985), see also Equation 5.8 in Newey and McFadden (1994). Hansen uses this condition to calculate the greatest lower bound for the asymptotic covariance matrix of instrumental variable estimators. His focus is on the optimal choice of instruments and not on efficiency. Proposition 4.2 gives conditions for the asymptotic efficiency of the estimator $\hat{\theta}_n$. What are the appropriate choices of h? Assume that $h(t, x; \theta)$ can be written as $h(t, x; \theta) = w(t'x) - E^\theta[w(t'X)]$. Then the completeness condition is equivalent to

$$\begin{aligned} E^{\theta_0}[G(X)w(t'X)] &= 0 \text{ for all } t \text{ in } \mathbf{R}^p \text{ except on a set of measure } 0 \\ &\Rightarrow G = 0. \end{aligned}$$

The choice of function $\{w(t'X)\}$ is closely related to the choice of a test function necessary to construct conditional moment specification tests (see Bierens, 1990 and Stinchcombe and White, 1998). According to Stinchcombe and White (Theorem 2.3), candidates for w are any analytic functions that are nonpolynomial, this includes $w(t'x) = \exp(it'x)$ (the characteristic function) and $w(t'x) = \exp(t'x)$ (the moment generating function) but also the logistic cumulative distribution function.

Note that it is not enough to have $E^{\theta_0}[G(X)\exp(it'X)] = 0$ for all t in an open subset of $\mathbf{R}^p$ to have $G = 0$. Indeed two characteristic functions may coincide on an open interval (even including zero) but may also correspond to two different distributions. Examples are provided by Lukacs (1970, page 85). Therefore, in general, we can not expect to reach efficiency if we restrict our attention to the c.f. on a subset of $\mathbf{R}^p$. On the other hand, there is a large class of c.f. that are uniquely defined by their values on an arbitrary interval I of $\mathbf{R}$ (assuming $p = 1$). This is the class of analytic c.f.[1] (see Remark 4.1.2, page 65 in Rossberg, Jesiak, and Siegel, 1985). As the continuation of these c.f. to the line is unique in the set of c.f, it is not necessary to use the full real line $\mathbf{R}$. Indeed, efficiency will be reached as long as π is positive on some arbitrary interval I of $\mathbf{R}$ (see Proposition 3.4.5. in Rossberg

[1] A c.f. ψ is said to be a a alytic c.f. if there exists a fu ctio $A(z)$ of the complex variable z which is a alytic i the circle $|z| < \rho$, A is co ti uous $[-\rho, \rho] \cap \mathbf{R}$ a d $\psi(t) = A(t)$ for $t \in \mathbf{R}$.

et al.). Examples of distributions with analytic c.f. include the Normal, Gamma, and Poisson distributions, while the c.f. of the Cauchy distribution is not analytic[2].

4.2. Asymptotic properties of the GMM estimator

In this subsection, we focus on moment conditions based on the e.c.f.:

$$\begin{aligned} h_n(t;\theta) &= \frac{1}{n}\sum_{j=1}^{n}\left(e^{it'X_j}-\psi_\theta(t)\right) \\ &= \psi_n(t)-\psi_\theta(t). \end{aligned} \tag{4.3}$$

We will establish (in the Appendix) that $\sqrt{n}h_n(t;\theta_0)$ converges in $L^2(\pi)$ to a Gaussian process with mean zero and covariance

$$\begin{aligned} E^{\theta_0}\left[h_n(s;\theta_0)\overline{h_n(t;\theta_0)}\right] &= E^{\theta_0}\left[\left(e^{isX}-\psi_{\theta_0}(s)\right)\left(e^{-itX}-\overline{\psi_{\theta_0}(t)}\right)\right] \\ &= \psi_{\theta_0}(s-t)-\psi_{\theta_0}(s)\psi_{\theta_0}(-t). \end{aligned}$$

Hence the kernel of the covariance operator is given by

$$k(s,t)=\psi_{\theta_0}(s-t)-\psi_{\theta_0}(s)\psi_{\theta_0}(-t).$$

In Appendix A, we provide a set of primitive conditions under which the asymptotic normality of our GMM estimator can be established. Assumptions 1 and 2 define the problem. Assumption 3 is an identification assumption. Assumption 4 (i) to (iii) are standard. Our task will consist of showing that the assumptions of Theorem 8 in CaFl are either satisfied under our set of assumptions or not needed.

Proposition 4.3. *Let $\hat{\theta}_n$ be the estimator defined by (3.1) and (4.3). Under Assumptions 1 to 4, we have*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(0,I_{\theta_0}^{-1}\right) \tag{4.4}$$

where

$$I_{\theta_0}=E^{\theta}\left[\left(\frac{\partial\ln f_\theta}{\partial\theta}\right)\left(\frac{\partial\ln f_\theta}{\partial\theta}\right)'\right]\Bigg|_{\theta=\theta_0}$$

as n and $n\alpha_n^2$ go to infinity and α_n goes to zero.

Note that α_n is allowed to converge at a faster rate than that stated in Theorem 8 of CaFl. It is important to remark that the estimator will be asymptotically efficient for any choice of $\pi>0$ on $\mathbf{R}^p$ (see the discussion of Equation (4.11) below). However the choice of π might affect the small sample performance of $\hat{\theta}_n$.

To give an intuition of the proof of Proposition 4.3, we discuss some of the assumptions of CaFl. The functions Eh and $\nabla_\theta\psi$ are assumed to belong to $\mathcal{H}(K)$.

[2]Other examples a d sufficie t co ditio s for the a alycity of a c.f. ca be fou d i Rossberg, Jesia , a d Siegel (Chapter 4, 1985) a d Lu acs (Chapter 7, 1970).

Using Proposition 4.1, $\mathcal{H}(K)$ is easy to characterize. $\mathcal{H}(K)$ is the set of functions g such that $E^{\theta}G^2 < \infty$, where $G(X)$ has mean zero and satisfies

$$g(t) = \int G(x)\left(e^{-it'x} - \psi_\theta(-t)\right) f_\theta(x)\, dx. \tag{4.5}$$

Assuming that $f_\theta(x) > 0$ everywhere, the unique solution to Equation (4.5) is

$$G(x) = \frac{1}{2\pi} \frac{\int e^{it'x} g(t)\, dt}{f_\theta(x)}.$$

Hence $\mathcal{H}(K)$ consists of functions g in $L^2(\pi)$ which Fourier transforms $\mathcal{F}(g) \equiv \frac{1}{2\pi} \int e^{it'x} g(t)\, dt$ satisfy

$$\|g\|_K^2 = EG^2 = \int \frac{|\mathcal{F}(g)|^2}{f_\theta(x)} dx < \infty. \tag{4.6}$$

If f_θ vanishes, a similar conclusion holds. Let $D = \{x : f_\theta(x) > 0\}$. Then $\mathcal{H}(K)$ consists of functions of $L^2(\pi)$ whose Fourier transforms vanish on D^c and such that $\int_D \frac{|\mathcal{F}(g)|^2}{f_\theta(x)} dx < \infty$.

In the Appendix, we relax Assumption 8' of CaFl that was not satisfied, namely that $\frac{\partial \psi_\theta(t)}{\partial \theta_j}$ belongs to the domain of K^{-1} for $\theta = \theta_0$. This assumption would require that there exists a function g in $L^2(\pi)$ such that

$$\begin{aligned} \frac{\partial \psi_{\theta_0}(t)}{\partial \theta_j} &= (Kg)(t) \\ &= \int \left(\psi_{\theta_0}(s-t) - \psi_{\theta_0}(s)\psi_{\theta_0}(-t)\right) g(t)\, \pi(t)\, dt. \end{aligned}$$

Using the Fourier inversion formula several times (as in Feuerverger and McDunnough, 1981a), we obtain

$$g(t) = K^{-1}\left(\frac{\partial \psi_{\theta_0}}{\partial \theta_j}\right)(t) = \frac{1}{\pi(t)} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left. \frac{\partial \ln f(x;\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} e^{itx} dx. \tag{4.7}$$

As the score is not necessarily integrable, the integral does not always exist, and even if it exists, g will not belong to $L^2(\pi)$ in general. Note that this term is the optimal instrument of Feuerverger and McDunnough (1981a) and of Singleton (2001) (to be discussed later).

4.3. Comparison with Feuerverger and McDunnough

In Feuerverger and McDunnough (1981a) (to be referred to as FM), the space of reference is $L^2(\lambda)$, the Hilbert space of real-valued functions that are square integrable with respect to Lebesgue measure, λ. They treat the real part and imaginary part of the c.f. as two different moments, so that the moment function h is two-dimensional but real. They show that a way to reach the efficiency bound is to estimate θ (assumed for simplicity to be scalar) by solving

$$\int_{-\infty}^{+\infty} w(t)\left(\psi_n(t) - \psi_\theta(t)\right) dt = 0 \tag{4.8}$$

where $w(t) = g(t)\pi(t)$ and g is defined in (4.7). Below, we show that the expression (4.8) is equivalent (a) to the first order condition of our GMM objective function and (b) to the first order condition of the MLE. This shows, without using Parzen's results, that the moment conditions based on the c.f. deliver an efficient estimator. Our objective function is given by

$$Q_n = \left\langle K^{-1/2} h_n(.;\theta), K^{-1/2} h_n(.;\theta)\right\rangle,$$

where K is used instead of its estimator for simplicity. The first order condition is

$$\left\langle K^{-1/2}\frac{\partial}{\partial\theta} h_n(.;\theta), K^{-1/2} h_n(.;\theta)\right\rangle = 0. \tag{4.9}$$

Proceeding as if $\frac{\partial\psi_\theta}{\partial\theta}\left(= -\frac{\partial}{\partial\theta}h_n(.;\theta)\right)$ were in the range of K, (4.9) can be rewritten as

$$\left\langle K^{-1}\frac{\partial\psi_\theta}{\partial\theta}, h_n(.;\theta)\right\rangle = 0, \tag{4.10}$$

which coincides with (4.8). Replacing $w(t)$ by $g(t)\pi(t)$, we see from (4.7) that $\pi(t)$ cancels out. Hence (4.10) is equivalent to

$$\begin{aligned} \int \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left.\frac{\partial \ln f(y;\theta)}{\partial\theta}\right|_{\theta=\theta_0} e^{ity} dy \frac{1}{n}\sum_{j=1}^{n}\left(e^{-it'x_j} - \psi_\theta(-t)\right) dt &= 0 \Leftrightarrow \\ \sum_{j=1}^{n} \left.\frac{\partial \ln f(x_j;\theta)}{\partial\theta}\right|_{\theta=\theta_0} &= 0 \end{aligned} \tag{4.11}$$

by a property of Fourier Transform (Theorem 4.11.12 of Debnath and Mikunsinsky). Therefore, as long as $\pi > 0$, π will not affect the asymptotic efficiency. However, there is flaw in this argument as $\frac{\partial\psi_\theta(t)}{\partial\theta_j}$ does not belong to the domain of K^{-1}. FM are well aware that the score is not integrable with respect to Lebesgue measure and they propose to replace w by a truncated integral

$$w_m(t) = \frac{1}{2\pi}\int_{-m}^{m} \left.\frac{\partial \ln f(x;\theta)}{\partial\theta}\right|_{\theta=\theta_0} e^{-itx} dx.$$

While $w(t)$ alone is not well defined, the term in the left hand side of (4.8) is bounded for all θ (see proof of Assumption 9' in Appendix B). A major problem is that w depends on the likelihood function which is, of course, unknown. FM suggests to discretize an interval of $\mathbf{R}$ and to apply the usual GMM on the resulting set of moment conditions. However, discretization implies a loss of efficiency. They argue that letting the intervals between observations, Δt , go to zero, the GMM estimator will reach the efficiency bound. From CaFl, it is clear that the passage at the limit requires a lot of care and that, when Δt goes to zero, the dimension of the covariance matrix increases, and its inverse is not bounded. It is therefore necessary to stabilize the problem by including a penalization term.

5. Conditional characteristic function

In practice, models frequently include explanatory variables so that estimation has to rely on the conditional characteristic function (c.c.f.). In this section, we explain how to construct moment conditions without loss of efficiency.

Assume that an i.i.d. sample $X_i = (Y_i, Z_i)$ is available. Denote the characteristic function of Y conditional on Z by

$$\psi_\theta\left(t|Z\right) \equiv E^\theta\left(e^{itY}|Z\right).$$

In the sequel, θ is assumed to be identifiable from the conditional distribution of Y given Z or equivalently from the conditional c.f.. Therefore, we have the following identification condition

$$E^\theta\left[e^{itY} - \psi_\theta\left(t|Z\right)|Z\right] = 0 \text{ for all } t$$

implies $\theta = \theta_0$. But Z may or may not be weak exogenous[3] for θ. Denote $\psi_X\left(t, s\right) = E\left(e^{itY+isZ}\right)$ the joint characteristic function of (Y, Z), K^Z the conditional covariance operator with kernel

$$\begin{aligned} k^Z\left(t,s\right) &= E^\theta\left[\left(e^{itY} - \psi_\theta\left(t|Z\right)\right)\left(e^{-isY} - \psi_\theta\left(-s|Z\right)\right)|Z\right] \\ &= \psi_\theta\left(t-s|Z\right) - \psi_\theta\left(t|Z\right)\psi_\theta\left(-s|Z\right), \end{aligned}$$

and K the unconditional covariance operator with kernel

$$\begin{aligned} k\left(t,s\right) &= E^\theta\left[\left(e^{itY} - \psi_\theta\left(t|Z\right)\right)\left(e^{-isY} - \psi_\theta\left(-s|Z\right)\right)\right] \\ &= E^\theta\left[\psi_\theta\left(t-s|Z\right) - \psi_\theta\left(t|Z\right)\psi_\theta\left(-s|Z\right)\right]. \end{aligned}$$

[3] Z is said to be wea exoge ous for θ if its margi al distributio does ot depe d o θ. So that the joi t distributio of (Y, Z) ca be decomposed as

$$f_{Y,Z}\left(y,z;\theta,\lambda\right) = f_{Y|Z}\left(y|z;\theta\right) f_Z\left(z;\lambda\right).$$

Proposition 5.1. *(i) The GMM estimator $\hat{\theta}_n$ based on the continuum of moment*

$$h(t,s,X;\theta) = e^{itY+isZ} - \psi_X(t,s) \tag{5.1}$$

is asymptotically efficient.

(ii) Assume that Z is weak exogenous, we have

$$\left\langle \left(K^Z\right)^{-1} \frac{\partial \psi_\theta(t|Z)}{\partial \theta}, e^{ity} - \psi_\theta(t|Z) \right\rangle_{K^Z} = \frac{\partial \ln f(y|Z)}{\partial \theta}.$$

Hence solving

$$\frac{1}{n}\sum_{j=1}^{n} \left\langle \frac{\partial \psi_\theta(t|z_j)}{\partial \theta}, e^{ity_j} - \psi_\theta(t|z_j) \right\rangle_{K^Z} = 0$$

would deliver an asymptotically efficient estimator.

The first set of moment conditions is based on the joint characteristic function of $X = (Y, Z)$. Estimation can be implemented using results presented in CaFl and discussed in the next section. (ii) gives a method to obtain an efficient estimator by inverting the conditional operator K^Z. It might be tedious in practice because the inversion has to be done for each value of Z. One way would be to use a discretization of $\mathbf{R}$ and to inverse the covariance matrix instead of the covariance operator, as suggested by Singleton (2001, page 129). The inversion of the covariance matrix could be done using a regularization to avoid singularity when the grid is too fine.

We now discuss a special case. Assume that

$$Y = \beta' Z + U \tag{5.2}$$

where U is independent of Z. The c.f. of U denoted ψ_u is known and may depend on some parameter λ. Its p.d.f is denoted f_u. Denote $\theta = (\beta', \lambda')'$. Z is assumed to be weak exogenous for the parameter of interest θ.

Corollary 5.2. *Consider Model (5.2). The GMM estimator based on*

$$h(t,Y,Z) = e^{it(Y-\beta' Z)} - \psi_u(t) \tag{5.3}$$

is asymptotically efficient.

6. Implementation

6.1. Estimation of the covariance operator

Let f be some element of $L^2(\pi)$. It is natural to estimate K by K_n the integral operator

$$(K_n f)(t) = \int k_n(t,s) f(s) \pi(s) ds$$

with kernel

$$k_n(t,s) = \frac{1}{n}\sum_{i=1}^{n} k\left(x^i, t, s\right) = \frac{1}{n}\sum_{i=1}^{n} h_t\left(x^i, \theta^0\right)\overline{h_s\left(x^i, \theta^0\right)}$$

Let $\hat{\theta}_n^1$ be a $n^{1/2}-$consistent first step estimate of θ_0. This first step estimator can be obtained by minimizing $\left\|\overline{h}(\theta)\right\|$. If h_t^i denotes $h_t(x^i, \hat{\theta}_n^1)$, our estimate satisfies :

$$(K_n f)(t) = \frac{1}{n}\sum_{i=1}^{n} h_t^i \int \overline{h_s^i} f(s)\,\pi(s)\,ds.$$

When s is a vector, the integral is actually multiple and π may be selected as the product of univariate p.d.f. The operator K_n is degenerate and has at most n eigenvalues and eigenfunctions. Let $\hat{\phi}_l$ and $\hat{\mu}_l$, $l = 1, ..., n$ denote the eigenfunctions and eigenvalues of K_n. As explained in CaFl (Section 3), $\hat{\phi}_l$ and $\hat{\mu}_l$ are easy to calculate. Let M be the $n \times n$ matrix with principal elements $\frac{1}{n} m_{ij}$ where

$$m_{ij} = \int \overline{h_s^i} h_s^j \,\pi(s)\,ds.$$

Denote $\underline{\beta}^l = [\beta_1^l, ..., \beta_n^l]'$ and $\hat{\mu}_l$, $l = 1, ..., n$ the eigenvectors and eigenvalues of M respectively. Note that M is self-adjoint, semi-positive definite, hence its eigenvalues are real positive. The eigenfunctions of K_n are $\hat{\phi}_l(t) = \underline{h_t \beta}^l$ where $\underline{h_t} = [h_t^1, h_t^2, ..., h_t^n]$ and its eigenvalues are $\hat{\mu}_l$. Note that eigenfunctions of a self-adjoint operator associated with different eigenvalues are necessarily orthogonal, therefore $\hat{\phi}_l$ need not be orthogonalized, only normed. From now on, $\hat{\phi}_l$ will denote the orthonormalized eigenfunctions associated with the eigenvalues, $\hat{\mu}_l$, ranked in decreasing order.

Researchers, who have previously used characteristic functions to do inference, worked with two sets of moments: those corresponding to the real part of the c.f. and those corresponding to the imaginary part, see for instance Feuerverger and McDunnough (1981a, b), Koutrouvelis and Kellermeier (1981), and also the recent papers by Singleton (2001), Chacko and Viceira (2001), Jiang and Knight (1999). Nowadays, most software packages (GAUSS among others) allow for complex numbers so that moment conditions (2.1) can be handled directly.

6.2. Expression of the objective function

The calculation of the objective function involves the computation of $K^{-1/2}$ where $K^{-1/2}$ can be seen as $(K^{-1})^{1/2}$. We first study the properties of K^{-1}. $K^{-1}f$ is solution of

$$Kg = f \tag{6.1}$$

This problem is typically ill-posed. In particular, the solution, g, of (6.1) is unstable to small changes in f. To stabilize the solution, we introduce a regularization parameter α_n that goes to zero when n goes to infinity. This method is called Tikhonov approximation and is explained in Groetsch (1993, page 84) and CaFl. Equation (6.1) is approximated by

$$(K^2 + \alpha_n I) g_{\alpha_n} = Kf.$$

The square root of the generalized inverse of K is

$$\left(\hat{K}_n^{\alpha_n}\right)^{-1/2} f = \sum_{j=1}^{n} \frac{\sqrt{\hat{\mu}_j}}{\sqrt{\hat{\mu}_j^2 + \alpha_n}} \left\langle f, \phi_j \right\rangle \phi_j$$

for f in $\mathcal{H}(K)$. Clearly, the solution g_{α_n} should converge to $K^{-1}f$ when α_n goes to zero, but for α_n too close to zero, the solution becomes unstable. There is a trade-off between the accuracy of the solution and its stability. Therefore, the right choice of α_n is crucial. The optimal GMM estimator is given by

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{j=1}^{n} \frac{\hat{\mu}_j}{\hat{\mu}_j^2 + \alpha_n} \left| \left\langle h_n(\theta), \hat{\phi}_j \right\rangle \right|^2.$$

Note that the regularization is not the only way to circumvent the instability problem. Another solution consists in truncating the sum in the objective function. Because the eigenvalues, $\hat{\mu}_j$, converge to zero, the sum $\sum_{j=1}^{L} \frac{1}{\hat{\mu}_j} \left| \left\langle h_n(\theta), \hat{\phi}_j \right\rangle \right|^2$ might diverge if the truncation parameter, L, is too large, hence the resulting estimators may not be consistent. Therefore the choice of the optimal truncation depends on the decay rate of the eigenvalues, which is difficult to assess, see Cardot, Ferraty, and Sarda (1999).

6.3. Choice of the regularization parameter α_n

From Proposition 4.3, α_n should converge to 0 such that $n\alpha_n^2 \to \infty$ as n goes to infinity. This is however not informative on how to choose α_n in practice. In this section, we propose a data-driven method for choosing α_n. Let denote $\hat{\theta}^{\alpha}$ the solution of

$$\arg\min_{\theta} \sum_{j=1}^{n} \frac{\hat{\mu}_j}{\hat{\mu}_j^2 + \alpha} \left| \left\langle h_n(\theta), \hat{\phi}_j \right\rangle \right|^2$$

for any $\alpha > 0$. This corresponds to minimizing $\langle A_n h_n, h_n \rangle$ where A_n is a bounded weighting operator

$$A_n = (K_n^2 + \alpha I)^{-1} K_n.$$

Define

$$A = (K^2 + \alpha I)^{-1} K.$$

We plan to select α_n that minimizes the Mean Square Error (MSE) of $\hat{\theta}^{\alpha}$. Because the estimator $\hat{\theta}^{\alpha}$ is unbiased for any α (constant and positive), we need to compute the higher order asymptotic bias of $\hat{\theta}^{\alpha}$. In that respect, we follow a recent paper by Newey and Smith (2001).

Proposition 6.1. *Consider moment conditions of the type (2.1). The MSE of* $\hat{\theta}^{\alpha}$ *satisfies*

$$\begin{aligned} MSE\left(\hat{\theta}^{\alpha}\right) &= E^{\theta_0}\left[\left(\hat{\theta}^{\alpha}-\theta_0\right)'\left(\hat{\theta}^{\alpha}-\theta_0\right)\right] \\ &= Bias^2 + Var. \end{aligned}$$

where $Bias^2 = \frac{1}{n^2}E^{\theta_0}(Q_1)' E^{\theta_0}(Q_1)$ *and* $Var = \frac{1}{n}trace(V)$ *with*

$$V = M^{-1}\langle A\bigtriangledown_\theta \psi_\theta, KA\bigtriangledown_\theta \psi_\theta\rangle M^{-1}$$

and

$$\begin{aligned} Q_1 &= M^{-1}\left\langle \bigtriangledown_\theta\psi_\theta, (K^\alpha)^{-2} h_i(.)\right\rangle \int |h_i(t)|^2 \pi(t)\, dt && (6.2)\\ &+ \langle \bigtriangledown_{\theta\theta}\psi_\theta, Ah_i\rangle M^{-1}\langle \bigtriangledown_\theta\psi_\theta, Ah_i\rangle && (6.3)\\ &- M^{-1}\left\langle \bigtriangledown_\theta\psi_\theta, (K^\alpha)^{-2} h_i(.)\right\rangle \overline{\langle \bigtriangledown_\theta\psi_\theta, h_i\rangle}' M^{-1}\langle \bigtriangledown_\theta\psi_\theta, Ah_i\rangle && (6.4)\\ &- \frac{M^{-1}}{2}\left\langle \bigtriangledown_\theta\psi_\theta, \langle \bigtriangledown_\theta\psi_\theta, Ah_i\rangle' M^{-1} A\bigtriangledown_{\theta\theta}\psi_\theta M^{-1}\langle \bigtriangledown_\theta\psi_\theta, Ah_i\rangle\right\rangle. && (6.5) \end{aligned}$$

and $M = \langle \bigtriangledown_\theta\psi_\theta, A\bigtriangledown_\theta\psi_\theta\rangle$ *and* $h_i = h(., X_i; \theta_0)$.

Note that

$$\begin{aligned} (K^\alpha)^{-2} f &= \sum \frac{\mu_j^2}{\left(\mu_j^2+\alpha\right)^2}\left\langle f, \phi_j\right\rangle \phi_j, \\ Af &= \sum \frac{\mu_j}{\mu_j^2+\alpha}\left\langle f, \phi_j\right\rangle \phi_j. \end{aligned}$$

Both $(K^\alpha)^{-2}$ and A can be estimated by replacing μ_j and ϕ_j by their estimators. To estimate the bias, the expectation is replaced by a sample mean. As usual, there is a trade-off between the bias (decreasing in α) and the variance (increasing in α). The term V is bounded for any α while the term $E^{\theta_0}(Q_1)' E^{\theta_0}(Q_1)$ will diverge when α goes to zero. More precisely, the terms (6.2) to (6.4) are $O_p(1/\alpha)$ and the term (6.5) is $O_p(1)$ and therefore can be neglected in the calculation of α. Hence $E^{\theta_0}(Q_1)' E^{\theta_0}(Q_1) = O_p(1/\alpha^2)$. Matching the terms $Bias^2$ which is $O_p(1/(n^2\alpha^2))$ and Var which is $O_p(1/n)$, we obtain $n\alpha^2 = O_p(1)$ which implies $n\alpha^{3/2} \to \infty$ (as needed for the consistency of the estimator) but not $n\alpha^2 \to \infty$ as required in Proposition 4.3. It means that using the MSE as criterion will deliver an α_n that converges to zero at a slightly faster rate than desired. Remark that if the moment conditions include exogenous variables as in (5.3), $E^\theta \bigtriangledown_\theta h_n = E^\theta \bigtriangledown_\theta h_i \neq \bigtriangledown_\theta\psi_\theta$, the bias includes an extra term corresponding to (B.20) in the appendix. In Q_1, one needs to add:

$$-\frac{M^{-1}}{n}\left\langle \bigtriangledown_\theta h_i - E^\theta \bigtriangledown_\theta h_i, Ah_i\right\rangle.$$

7. Goodness of fit test for convolution of distributions

Consider a model

$$X_i = \beta' Z_i + \eta_i + \varepsilon_i$$

where $\{X_i, Z_i\}$ are observable. Z_i is independent of η_i and ε_i. η_i is an unobserved random effect and ε_i is an unobserved random variable. η_i and ε_i are mutually independent, i.i.d. and have zero mean. In the following, we will assume that the distribution of ε_i is known and that of η_i is unknown. Nonparametric estimation of the density of η is possible using the deconvolution kernel, see Carroll and Hall (1988), Stefanski and Carroll (1990), Li and Vuong (1998) among others, and also Horowitz and Markatou (1996) and Li, Perrigne, and Vuong (2000) for applications. However, the resulting estimator of the density has a very slow speed of convergence especially if ε is normally distributed. It might be more meaningful to select a parametric form for the characteristic function of η, denoted $\psi_\eta(\alpha, t)$ and to perform a goodness-of-fit test.

The set of parameters $\theta = (\alpha', \beta')'$ is jointly estimated from the moment conditions given by

$$E^\theta h(\theta, t) \equiv E^\theta \left[e^{it(X - \beta' Z)} - \psi_\eta(\alpha, t)\, \psi_\varepsilon(t) \right] = 0.$$

Since the model is clearly overidentified, a simple specification test is to check whether the overidentifying restrictions are close to zero. This test, similar to Hansen (1982)'s J test, is based on the GMM objective function where the parameters have been replaced by their estimators. The only specificity here is that the number of moment restrictions being infinite, Hansen's test needs to be rescaled. This test appears in CaFl in the context of unconditional moment conditions.

Define

$$p_n = \sum_{j=1}^{n} \frac{\mu_j^2}{\mu_j^2 + \alpha_n}, \quad q_n = 2 \sum_{j=1}^{n} \frac{\mu_j^4}{\left(\mu_j^2 + \alpha_n\right)^2}. \tag{7.1}$$

Proposition 7.1. *Assume $q_n \sqrt{\alpha_n} \to \infty$ as n goes to infinity. Under Assumptions 1 to 4, we have*

$$\tau_n = \frac{\| \sqrt{n} h_n\left(.; \hat{\theta}\right) \|_{K_n^{\alpha_n}}^2 - \hat{p}_n}{\sqrt{\hat{q}_n}} \overset{n\to\infty}{\to} \mathcal{N}(0,1)$$

as α_n goes to zero and $n\alpha_n^3$ goes to infinity. $\hat{p}_n$ and $\hat{q}_n$ are the counterparts of p_n and q_n where the μ_j have been replaced by their estimators $\hat{\mu}_j$.

This test has the advantage of being an omnibus test, which asymptotic distribution is free of nuisance parameters. Moreover, it has power against a wide range of alternatives. Since it is not designed for a specific alternative, it is likely to have a low power in small samples. Various goodness-of-fit tests based on the e.c.f. have been previously proposed, although none of the following references deal with the problem of convolution. The closest test to ours (Koutrouvelis and Kellermeier, 1981) is based on the moments

$$\psi_n(t) - \psi_0(t;\theta)$$

for values $t = t_1, t_2, ..., t_m$ "suitably chosen around 0". The statistic is given by the GMM quadratic form where θ has been replaced by its estimator. This is the usual J-test and its asymptotic distribution is a chi-square with $2m - k$ degrees of freedom. Epps and Pulley (1983) use as test statistic

$$T = \int \left|\psi_n(t) - \psi_0\left(t;\hat{\theta}\right)\right|^2 d\omega(t)$$

where $\hat{\theta}$ is the MLE estimator of θ and $d\omega$ is an appropriate weight. The distribution of this statistic is not standard and actually depends on nuisance parameters. Epps, Singleton, and Pulley (1982) propose specification tests based on the empirical moment generating function (e.g.f.). It uses the difference between the e.g.f. and the true one evaluated at a single value of t, which is chosen in order to maximize the power against a specific alternative. This statistic is asymptotically normal.

8. Simulations

8.1. The method

To assess the small sample performance of our estimation procedure, we simulate samples of size $n = 100$. The number of repetitions is $rep = 2000$ for Table 1 and $rep = 1000$ for the subsequent tables. For each estimator, we report the mean, the standard deviation, and the root mean square error (RMSE). The true values of the parameters are written below the title.

The first step estimator is given by

$$\hat{\theta}_n^1 = \arg\min_\theta \|h_n\|^2 = \arg\min_\theta \int h_n(t;\theta)\,\overline{(h_n(t;\theta))}\pi(t)\,dt.$$

where $h_n(t;\theta) = \frac{1}{n}\sum_{j=1}^n h(t, X_j;\theta)$. The objective function is minimized using the procedure *optmum* of GAUSS. Then, we compute all the eigenvalues and eigenvectors of the matrix M defined in Section 6. However since the eigenvalues decrease very quickly to 0, we calculate only $\sqrt{n}$ eigenfunctions $\hat{\phi}_j$ that correspond to the largest eigenvalues. Because numerical integrations are time consuming, we replace the inner products by sums on an equally-spaced grid with intervals 1/25. Finally we use again *optmum* to compute $\hat{\theta} = \arg\min_\theta \|h_n\|_K^2$. The starting values of this algorithm are set equal to the first step estimators.

The columns called "First step" and "Continuum GMM" correspond to a choice of π as the p.d.f. of the standard normal distribution, $\pi(t) = e^{-t^2/2}/\sqrt{2\pi} \equiv \phi(t)$.

To avoid integrating over $\mathbf{R}$, we use a tan transformation so that all integrations are on $[-\pi/2, \pi/2]$, for instance

$$\|h\|^2 = \int_{-\infty}^{+\infty} h(t)\,\overline{h(t)}\phi(t)\,dt = \int_{-\pi/2}^{\pi/2} h(\tan(u))\,\overline{h(\tan(u))}\phi(\tan(u))\,\frac{1}{\cos(u)^2}du.$$

The estimator in the first column of "Second Step" uses the penalization α_n that minimizes the MSE, using as starting values 10^{-6} for Tables 1 and 2 and 10^{-4} for Tables 4a to c. Although the bias is continuously decreasing and the variance at first increases and then flattens after a certain point, the sum of the two is not convex and it is important to choose a starting value that is sufficiently small. α_n is obtained using the Steepest Descent algorithm of the *optmum* procedure of GAUSS. The median, mean, and standard deviation of α_n are reported in the bottom of the table. The next three columns correspond to various values of α_n chosen fixed a priori. The columns called Unif correspond to a choice of π as the p.d.f. of the Uniform distribution on $[-1, 1]$ or $[-2, 2]$ associated with the optimal α_n for Tables 1 and 2 and with the first step estimator for Tables 4a to c. Discrete GMM is the usual efficient GMM estimator based on 5 (complex) moment conditions corresponding to $t = 1, 3, 5, 7, 9, 11$. We choose a coarse grid because a finer grid would yield a singular covariance matrix for some of the simulations. A similar grid ($t = 1, 2, 3, 4, 5$) was chosen by Chacko and Viceira (2001) in their simulations. The starting values for the discrete GMM are set equal to the first step continuum GMM estimator. The starting values for MLE are the same as those used for the first step GMM estimator.

8.2. Estimation of a normal distribution

First we estimate a normal distribution $\mathcal{N}(\nu, \sigma)$ with $\nu = 1$ and $\sigma = 0.5$. This exercise is intended to show how well the method works in a simple case. The starting values for the optimization of the first step estimator are drawn in a Uniform distribution around the true values: U[-1,3] for ν and U[.1,.9] for σ.

From Table 1, we see that:

(i) GMM is dominated by MLE.

(ii) GMM is not very sensitive to the choice of the penalization term α_n.

(iii) Using the Uniform on $[-1, 1]$ gives an estimator that seems to be more efficient than the GMM based on a normal π. This conclusion is reversed if one uses a Uniform on $[-2, 2]$.

(iv) Discrete GMM performs very poorly relatively to the continuum GMM.

8.3. Estimation of a stable distribution

We focus on a stable distribution where the location parameter δ is known and equal to 0 and we estimate the scale c, the index ν, and the skewness parameter β. The characteristic function of a stable distribution can be parametrized as (DuMouchel, 1975)

$$\begin{aligned}\psi_\theta(t) &= \exp\left\{-c|t|^\nu\left(1 - i\beta sign(t)\tan\left(\frac{\pi\nu}{2}\right)\right)\right\},\ \nu \neq 1,\\ &= \exp\left\{-c|t|^\nu\left(1 + i\beta\frac{2}{\pi}sign(t)\ln(|t|)\right)\right\},\ \nu = 1.\end{aligned}$$

where $0 < \nu \leq 2$, $-1 \leq \beta \leq 1$, $0 < c < \infty$. We consider a special case of a stable distribution that is easy to simulate (Devroye, 1986, page 456)[4] and corresponds to $\nu = 1/4, \beta = 0$ and $c = 1$. In the estimation, we impose all the restrictions on the parameters. The starting values for the first step GMM estimator are set equal to the true values.

From Table 2, we draw two main conclusions:

(i) The second step estimator using the optimal α_n is clearly better than the first step and is at least as good as the other continuum GMM using fixed α_n.

(ii) The continuum GMM estimators using a uniform distribution on [-1,1] or [-2,2] for π are outperformed by that using the normal distribution on $\mathbf{R}$ for π. This result goes in the opposite direction from what we found in Table 1. One likely explanation is that the c.f. of the normal is analytic while the c.f. of a stable distribution with exponent $\nu < 2$ is not analytic (Lukacs, 1970, page 191).

8.4. Estimation of a convolution

In this Monte Carlo experiment, we estimate a convolution of the type (2.5) given in Example 2. In the random-effect literature, it is common to assume that η_i and ε_i are both normally distributed and the identification of the respective variances is guaranteed by the observation of panel data. However, using panel data on annual earnings, Horowitz and Markatou (1996) show that, although η_i is likely to be normal, the distribution of ε_i exhibits thicker tails than the normal distribution. Here we consider the case where η_i is normal and ε_i is Laplace. As β is easy to estimate by ordinary least-squares, we do not estimate β (β is set equal to 0) and we focus our attention on the estimation of the other parameters (except for Table 3 where $\beta = 1$).

Assume that $\varepsilon_i \sim iid\ Laplace(\lambda)$ with density

$$f(x) = \frac{1}{2\lambda}\exp\left\{-\frac{|x|}{\lambda}\right\},\ \lambda > 0,\ x \in R,$$

$E(\varepsilon_i) = 0$ and $V(\varepsilon_i) = 2\lambda^2$ and $\eta_i \sim iid\mathcal{N}(\nu,\sigma)$. Denote $\theta = (\lambda,\nu,\sigma^2)'$. Let $u = \varepsilon + \eta$. Since ε and η are independent, we have $\psi_u = \psi_\varepsilon \times \psi_\eta$ with

$$\psi_\varepsilon(t) = \frac{1}{\lambda^2t^2 + 1} \text{ and } \psi_\eta(t) = e^{i\nu t}e^{-\sigma^2t^2/2}.$$

[4]It correspo ds to $N_3/(2N_1N_2^3)$ where N_1, N_2, a d N_3 are i depe de t $\mathcal{N}(0,1)$.

The likelihood of u has a closed-form expression. Here we give it in terms of ν, λ, and $\delta = \sigma/\lambda$:

$$\frac{1}{2\lambda}\exp\left(\frac{\delta^2}{2}\right)\left\{\exp\left(-\frac{(u-\nu)}{\lambda}\right)\Phi\left(\frac{(u-\nu)}{\lambda\delta}-\delta\right)+\exp\left(\frac{u-\nu}{\lambda}\right)\Phi\left(\frac{(u-\nu)}{\lambda\delta}+\delta\right)\right\}.$$

We maximize the likelihood of $X_j - \hat{\nu}$ using the procedure *maxlik* in GAUSS. The starting values for MLE and first step GMM are chosen so that ν is equal to the OLS estimator, $\sigma = S/\sqrt{2}$ and $\lambda = S/2$ where S is the standard deviation of the residual of the regression of X on 1. This corresponds to the case where the contributions of ε and η to the total variance are the same.

In Table 3, we compare the performance of GMM estimators based on a finite grid for various intervals and discretizations Δt. The exogenous i.i.d. variable Z takes only two values 1 and 2, with probabilities 1/2 and 1/2. Note that l values of t give $2l$ moments because each $h(t)$ is composed of a real and imaginary part. With $l = 2$, the model is exactly identified. In this case, we report the first step estimator because, for numerical reasons, the second step estimator is generally not as good as the first one. The best performance is obtained for a discretization with width 1/2 of the interval $[0, b]$ for which case the model is exactly identified. This result is worth commenting on. If the grid is fine, the moments are strongly correlated and the GMM estimator has a large variance. Eventually, when Δt goes to zero, the covariance matrix becomes singular and the GMM procedure breaks down. Actually, even for a fixed Δt, too many moments relative to the sample size will render the covariance matrix singular. Adding more moments does not necessarily decrease the standard error of the estimators, contrary to what is often believed (Schmidt, 1982). A theory on the optimal number of moments as a function of the sample size and the properties of the weighting matrix could be developed but is beyond the scope of this paper.

The columns Unif of Tables 4a to c report the first step estimator because there is no gain in using the second step estimator. For Table 4b, the MLE had some difficulty in converging, therefore we report the statistics on 1000 successful estimations.

From Tables 4a to c, we can draw the following conclusions:

(i) The second step estimator turns out to be worse than the first step. This might be due to the difficulty in estimating α_n in small samples.

(ii) In general, MLE is slightly better than GMM.

(iii) GMM based on a continuum of moment conditions performs generally better than GMM based on a discrete grid.

(iv) Although the c.f. of u is analytic[5], there is no clear advantage of taking a small interval over integrating on $\mathbf{R}$.

[5] The c.f. of a ormal is a alytic with strip of regularity $\mathbf{R}$ while the cf of the Laplace is a alytic with strip of regularity $[-1/\lambda, 1/\lambda]$. By Theorem 5 of Wu sch (1994), the product of these two c.f. is a alytic with strip of regularity $[-1/\lambda, 1/\lambda]$.

9. Conclusion

This paper uses the empirical characteristic function for estimation. The proposed GMM method takes advantage of the continuum of moment conditions provided by the c.f. so that our estimator is asymptotically efficient. We showed that the choice of the norm (π) does not affect the efficiency of the estimator as long as the interval of integration is $\mathbf{R}^p$. In the usual GMM, an important open question is the choice of the number of the fixed grid points. Our method circumvents this problem but instead we then need to choose a penalization term α_n. We propose a data-driven method for selecting α_n.

From our limited Monte Carlo experiment, it seems that:

(i) Our GMM procedure produces estimates with good finite sample properties.

(ii) The estimation is not very sensitive to the penalization term α_n.

(iii) The GMM estimator based on an equispaced grid is less precise than that based on a continuum of moment conditions.

(iv) When the c.f. is analytic (ex: Normal), using a continuum of moment conditions on a small interval gives better results than using the full real line, whereas, when the c.f. is not analytic (ex: Stable), using $\mathbf{R}$ is better.

We illustrated our method on a random-effect model. This type of model is frequently encountered in microeconometrics. However, the use of the characteristic function is not limited to a cross-sectional setting and has recently received a surge of interest in finance. While the likelihood of an asset pricing model is not easily tractable, its c.c.f. has often a closed-form expression, and offers a way to estimate the parameters (Chacko and Viceira, 1999, Jiang and Knight, 1999, Singleton, 2001). We will consider the estimation of diffusions using GMM in another paper (Carrasco, Chernov, Florens, and Ghysels 2000).

Results of the Monte Carlo experiments

Table 1. Normal Distribution

$(\nu = 1.0, \sigma = 0.5)$

		First step	Second step continuum on $\mathbf{R}$				MLE	GMM discrete	Unif. $[-1,1]$	Unif. $[-2,2]$
mean	ν	1.0020	1.0022	1.0020	1.0020	1.0021	1.0020	0.9794	1.0020	1.0021
	σ	0.4962	0.4938	0.4974	0.4972	0.4942	0.4957	0.4926	0.4963	0.4957
std	ν	0.0509	0.0511	0.0510	0.0509	0.0508	0.0498	0.8049	0.0505	0.0547
	σ	0.0359	0.0346	0.0358	0.0357	0.0352	0.0343	0.0592	0.0348	0.0361
RMSE	ν	0.0509	0.0511	0.0510	0.0510	0.0508	0.0499	0.8049	0.0505	0.0547
	σ	0.0361	0.0351	0.0359	0.0358	0.0356	0.0346	0.0597	0.0350	0.0363
α_n	med		1.6e-6	0.1	0.001	1e-5			1.8e-6	0.0006
	mean		0.0015	0.1	0.001	1e-5			0.0201	0.0794
	std		0.0297	0	0	0			0.1136	0.2346

Table 2. Stable Distribution

$(c = 1, \nu = 0.25, \beta = 0)$

		First step	Second step continuum on $\mathbf{R}$				GMM discrete	Unif. $[-1,1]$	Unif. $[-2,2]$
mean	c	1.1156	1.1121	1.1138	1.1105	1.1195	0.5506	1.1052	1.1125
	ν	0.3060	0.2768	0.2744	0.2763	0.2808	0.3454	0.2568	0.2580
	β	0.0138	-0.0040	-0.0074	-0.0042	0.0025	-0.0051	-0.0039	0.0183
std	c	0.1739	0.1528	0.1559	0.1521	0.1558	5.8205	0.1684	0.1666
	ν	0.1057	0.0744	0.0799	0.0755	0.0779	0.1643	0.0930	0.0865
	β	0.2626	0.2050	0.2300	0.2085	0.2263	0.2779	0.2830	0.2469
RMSE	c	0.2087	0.1894	0.1930	0.1879	0.1963	5.8349	0.1985	0.2010
	ν	0.1195	0.0790	0.0835	0.0799	0.0838	0.1899	0.0932	0.0868
	β	0.2628	0.2049	0.2300	0.2085	0.2262	0.2778	0.2829	0.2475
α_n	med		0.0022	0.1	0.001	1e-5		0.0115	0.0121
	mean		0.0569	0.1	0.001	1e-5		0.0221	0.0156
	std		0.2058	0	0	0		0.0918	0.0450

Table 3. Sensitivity of GMM to discretization

$(\nu = 1.0, \sigma = 0.5, \lambda = 0.5, \beta = 1.0, b = \max\{x\}/2\pi \simeq 0.9)$

		GMM	discrete GMM					
interval		[0,b]	[0,b]	[0,b]	[0,b]	[0,2b]	[0,2b]	[0,3b]
Δt		$\alpha_n = 0.1$	1/2	1/4	1/5	1	1/2	1
t		continu.	.5,1	.25,.5,.75,1	0.2,...,0.8	1,2	.5,1,1.5,2	1,2,3
RMSE	ν	0.279	0.278	0.337	0.420	0.282	0.304	0.333
	σ	0.154	0.211	0.226	0.234	0.199	0.221	0.226
	λ	0.116	0.153	0.255	0.315	0.165	0.247	0.237
	β	0.176	0.177	0.215	0.277	0.180	0.198	0.218

Table 4a. Convolution

($\nu = 1.0, \sigma = 0.5, \lambda = 0.5$)

		First step	Second step continuum on **R**				MLE	GMM discrete	Unif. $[-1,1]$	Unif. $[-2,2]$
mean	ν	1.0063	1.0047	1.0051	1.0054	1.0043	1.0064	1.0012	1.0075	1.0057
	σ	0.5037	0.4751	0.5015	0.4910	0.4780	0.4748	0.4876	0.5217	0.4888
	λ	0.4286	0.4466	0.4109	0.4282	0.4453	0.4821	0.4373	0.4209	0.4453
std	ν	0.0891	0.0903	0.0903	0.0900	0.0904	0.0889	0.0958	0.0888	0.0912
	σ	0.2102	0.2353	0.2414	0.2371	0.2291	0.1801	0.2501	0.2313	0.2149
	λ	0.1983	0.2072	0.2253	0.2121	0.2099	0.1021	0.1804	0.1703	0.1776
RMSE	ν	0.0893	0.0904	0.0904	0.0901	0.0905	0.0891	0.0958	0.0890	0.0914
	σ	0.2101	0.2365	0.2413	0.2371	0.2300	0.1818	0.2503	0.2322	0.2151
	λ	0.2107	0.2139	0.2422	0.2238	0.2168	0.1036	0.1909	0.1877	0.1857
α_n	med		0.0001	0.1	0.001	1e-5				
	mean		0.0003	0.1	0.001	1e-5				
	std		0.0035	0	0	0				

Table 4b. Convolution

($\nu = 1.0, \sigma = 0.5, \lambda = 0.25$)

		First step	Second step continuum on **R**				MLE	GMM discrete	Unif. $[-1,1]$	Unif. $[-2,2]$
mean	ν	1.0053	1.0046	1.0056	1.0051	1.0046	0.9984	0.9955	1.0054	1.0053
	σ	0.4966	0.4884	0.5122	0.5115	0.4851	0.3201	0.4649	0.5025	0.5020
	λ	0.1863	0.1946	0.1571	0.1557	0.1973	0.3657	0.2216	0.2020	0.1792
std	ν	0.0661	0.0676	0.0678	0.0676	0.0676	0.0634	0.0772	0.0650	0.0669
	σ	0.1249	0.1376	0.1268	0.1294	0.1414	0.1197	0.1875	0.1128	0.1228
	λ	0.1479	0.1586	0.1507	0.1561	0.1612	0.0558	0.1480	0.1081	0.1460
RMSE	ν	0.0663	0.0677	0.0680	0.0677	0.0678	0.0634	0.0773	0.0651	0.0671
	σ	0.1249	0.1380	0.1273	0.1298	0.1421	0.2161	0.1906	0.1128	0.1227
	λ	0.1609	0.1679	0.1769	0.1823	0.1695	0.1285	0.1507	0.1182	0.1622
α	med		1.1e-5	0.1	0.001	1e-5				
	mean		0.0003	0.1	0.001	1e-5				
	std		0.0031	0	0	0				

Table 4c. Convolution
$(\nu = 1.0, \sigma = 0.25, \lambda = 0.5)$

		First step	Second step continuum on $\mathbf{R}$				MLE	GMM discrete	Unif. $[-1,1]$	Unif. $[-2,2]$
mean	ν	1.0046	1.0056	1.0019	1.0039	1.0065	1.0038	1.0011	1.0061	1.0040
	σ	0.2399	0.2229	0.2494	0.2458	0.2151	0.2538	0.2400	0.3085	0.2447
	λ	0.4677	0.4686	0.4477	0.4519	0.4704	0.4685	0.4506	0.4376	0.4611
std	ν	0.0705	0.0699	0.0704	0.0701	0.0699	0.0700	0.0752	0.0743	0.0709
	σ	0.1848	0.2013	0.2200	0.2127	0.2017	0.1720	0.1970	0.2129	0.1972
	λ	0.1033	0.1166	0.1434	0.1342	0.1247	0.0936	0.1222	0.1123	0.1051
RMSE	ν	0.0706	0.0701	0.0704	0.0701	0.0702	0.0700	0.0752	0.0746	0.0710
	σ	0.1849	0.2030	0.2199	0.2126	0.2046	0.1720	0.1972	0.2207	0.1971
	λ	0.1082	0.1207	0.1526	0.1425	0.1282	0.0987	0.1318	0.1284	0.1120
α	med		8.8e-5	0.1	0.01	1e-5				
	mean		8.5e-5	0.1	0.001	1e-5				
	std		4.7e-5	0	0	0				

A. Appendix: Assumptions

Assumption 1: The observed data $\{x^1, ..., x^n\}$ are independent and identically distributed realizations of X. The r.v. X takes its values in $\mathbf{R}^p$ and has a p.d.f. $f(x;\theta)$ with $\theta \in \Theta \subset \mathbf{R}^q$ and Θ compact. The characteristic function of X is denoted $\psi_\theta(t)$.

Assumption 2: π is the p.d.f. of a distribution that is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^p$. $\pi(x) > 0$ for all $x \in \mathbf{R}^p$. $L^2(\pi)$ is the Hilbert space of complex-valued functions that are square integrable with respect to π. Denote $\langle .,. \rangle$ and $\| \, . \, \|$ the inner product and the norm defined on $L^2(\pi)$.

Assumption 3: The equation

$$E^{\theta_0}(e^{it'X}) - \psi_\theta(t) = 0 \ \pi - a.s.$$

has a unique solution θ_0 which is an interior point of Θ.

Assumption 4: (i) $f(x;\theta)$ is continuously differentiable with respect to $\theta = (\theta_1, ..., \theta_q)$ on Θ,

(ii) $\int \sup_{\theta\in\Theta} \left\| \frac{\partial f(x;\theta)}{\partial\theta} \right\| dx < \infty$,

(iii) $I_{\theta_0} = E^\theta \left[\left(\frac{\partial \ln f_\theta}{\partial\theta} \right) \left(\frac{\partial \ln f_\theta}{\partial\theta} \right)' \right] \Big|_{\theta=\theta_0}$ is positive definite.

B. Appendix: Proofs of propositions

Proof of Proposition 4.2. In the following, we denote $C\left(\overline{E^{\theta_0}\left(\nabla_\theta h\right)}\right)$ by C. The proof has three steps.

1. First we show that $G_0 = \frac{\partial \ln f(X;\theta_0)}{\partial \theta}$ belongs to C.

2. Second we show that G_0 is the element of C with minimal norm if and only if $\partial \ln f(X;\theta_0)/\partial\theta \in \bar{S}$.

3. Finally, we show that $\{h(t,X;\theta_0)\}$ complete is a sufficient condition.

1. Note that
$$\int h(t,x;\theta) f(x;\theta)\, dx = 0.$$
Differentiating with respect to θ, we obtain
$$\int \frac{\partial h(t,x;\theta)}{\partial\theta} f(x;\theta)\, dx + \int \frac{\partial f(x;\theta)}{\partial\theta} h(t,x;\theta)\, dx = 0$$
by Assumption 4 (ii). Hence
$$\begin{aligned} E\left[G_0 \overline{h(t,X;\theta_0)}\right] &= \int \frac{\partial f(x;\theta_0)}{\partial\theta} \overline{h(t,x;\theta_0)} dx && \text{(B.1)}\\ &= -\int \frac{\partial \overline{h(t,x;\theta_0)}}{\partial\theta} f(x;\theta_0)\, dx && \text{(B.2)}\\ &= -\overline{E \bigtriangledown_\theta h}. && \text{(B.3)}\end{aligned}$$

2. Let $G = G_0 + G_1$. $G \in C \Rightarrow E\left[G_1(X)\overline{h(t,X;\theta_0)}\right] = 0$ for all $t \in \mathbf{R}^p$. Here and in the sequel, we omit "$\pi - a.s.$". Note also that
$$E[GG'] = E[G_0G_0'] + E[G_1G_0'] + E[G_0G_1'] + E[G_1G_1'].$$

G_0 has minimal norm
$\Longleftrightarrow (E[G_1(X)h(t,X;\theta_0)] = 0$ for all $t \in \mathbf{R}^p \Rightarrow E[G_1G_0'] = 0)$
$\Longleftrightarrow (E[G_1(X)h(t,X;\theta_0)] = 0$ for all $t \in \mathbf{R}^p \Rightarrow E[G_1\partial \ln f(X;\theta_0)/\partial\theta] = 0)$
$\Longleftrightarrow (\forall G_1 \perp S \Rightarrow G_1 \perp \partial \ln f(X;\theta_0)/\partial\theta)$
$\Longleftrightarrow (\partial \ln f(X;\theta_0)/\partial\theta \perp S^\perp)$
$\Longleftrightarrow (\partial \ln f(X;\theta_0)/\partial\theta \in \left(S^\perp\right)^\perp = \overline{S})$.

3. If $\{h(t,X;\theta_0)\}$ is complete, the solution to Equation (4.2) is unique. Hence, the result of (ii) follows from the fact that the score is solution by point 1. ■

Before proving Proposition 4.3, we need to establish some preliminary results. In the following, Assumptions CF1, CF2, etc refer to the assumptions of CaFl, while Assumptions 1, 2 etc refer to our assumptions. Assumptions CF8' and CF9 are replaced by the following.

Assumption 8": (a) h_n is continuously differentiable with respect to θ.
(b) There exists a function ψ such that $\bigtriangledown_\theta h_n = \bigtriangledown_\theta \psi = E \bigtriangledown_\theta \psi$.
(c) $\bigtriangledown_\theta \psi \in \mathcal{H}(K)$ for $\theta = \theta_0$.

Assumption 9': For all integer n, the following relationship holds
$$\frac{\partial}{\partial\theta'} \|h_n(\theta)\|^2_{K_n^\alpha} = 2\left\langle \frac{\partial h_n(\theta)}{\partial\theta'}, h_n(\theta)\right\rangle_{K_n^\alpha}.$$

Lemma B.1. *Under Assumptions CF1, CF2, CF3, CF4', CF5, CF6, 8", 9', CF11', CF12, CF13, CF14, we have*

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) \xrightarrow{L} \mathcal{N}\left(0,\langle\nabla_\theta\psi,\nabla_\theta\psi\rangle_K\right)$$

as n and $n\alpha_n^2$ go to infinity and α_n goes to zero.

Proof of Lemma B.1. The consistency follows directly from Theorem 8 of CaFl. We need to prove the asymptotic normality under the new assumption 8" which replaces CF8'. Using a Taylor expansion around θ_0 of the first order condition of the GMM minimization problem and by Assumption 9', we obtain

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) &= -\left\langle (K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right),(K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\overline{\theta}_n\right)\right\rangle^{-1} && \text{(B.4)}\\ &\left\langle (K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right),(K_n^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle && \text{(B.5)}\end{aligned}$$

where $\overline{\theta}_n$ is a mean value. Assuming that $\hat{\theta}_n$ converges at the speed $\sqrt{n}$ (which will be confirmed later), the continuity of $\nabla_\theta\psi$ and Theorem 7 of CaFl yield

$$\left\langle (K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right),(K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\overline{\theta}_n\right)\right\rangle \xrightarrow{P} \left\langle\nabla_\theta\psi\left(\theta_0\right),\nabla_\theta\psi\left(\theta_0\right)\right\rangle_K$$

as n and $n\alpha_n^{3/2}$ go to infinity. Next, we consider (B.5):

$$\begin{aligned}&\left\langle (K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right),(K_n^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle\\ =\;&\left\langle (K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right)-K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right),(K_n^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle && \text{(B.6)}\\ &+\left\langle K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right),(K_n^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle && \text{(B.7)}\end{aligned}$$

$$(B.6)\leq\left\|(K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right)-K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right)\right\|\left\|(K_n^{\alpha_n})^{-1/2}\right\|\left\|\sqrt{n}h_n\left(\theta_0\right)\right\|.$$

By the proof of Theorem 7 in CaFl, we have

$$\begin{aligned}\left\|(K_n^{\alpha_n})^{-1/2}\nabla_\theta\,\psi\left(\hat{\theta}_n\right)-K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right)\right\| &= O_p\left(\frac{1}{\sqrt{n\alpha_n^{3/2}}}\right),\\ \left\|(K_n^{\alpha_n})^{-1/2}\right\| &= O_p\left(\frac{1}{\alpha_n^{1/4}}\right).\end{aligned}$$

As moreover $\left\|\sqrt{n}h_n\left(\theta_0\right)\right\|=O_p\left(1\right)$, the term $(B.6)$ is $O_p\left(1/\sqrt{n\alpha_n^2}\right)=o_p\left(1\right)$ as $n\alpha_n^2$ goes to infinity by assumption.

The term $(B.7)$ can be decomposed as

$$\begin{aligned}&\left\langle K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right),(K_n^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle\\ =\;&\left\langle K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right),\left((K_n^{\alpha_n})^{-1/2}-(K^{\alpha})^{-1/2}\right)\sqrt{n}h_n\left(\theta_0\right)\right\rangle && \text{(B.8)}\\ &+\left\langle K^{-1/2}\nabla_\theta\,\psi\left(\theta_0\right),(K^{\alpha_n})^{-1/2}\sqrt{n}h_n\left(\theta_0\right)\right\rangle. && \text{(B.9)}\end{aligned}$$

$$
\begin{aligned}
(B.8) &\leq \left\| K^{-1/2} \nabla_\theta \psi (\theta_0) \right\| \left\| (K_n^{\alpha_n})^{-1/2} - (K^{\alpha_n})^{-1/2} \right\| \left\| \sqrt{n} h_n (\theta_0) \right\| \\
&= O_p \left(\frac{1}{\sqrt{n \alpha_n^{3/2}}} \right)
\end{aligned}
$$

by the proof of Theorem 7 of CaFl, therefore the term (B.8) goes to zero. It remains to show that (B.9) is asymptotically normal.

$$
\begin{aligned}
(B.9) &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\mu_j^2 + \alpha_n}} \left\langle \nabla_\theta \psi (\theta_0), \phi_j \right\rangle \overline{\left\langle \sqrt{n} h_n (\theta_0), \phi_j \right\rangle} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{\infty} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\mu_j^2 + \alpha_n}} \left\langle \nabla_\theta \psi (\theta_0), \phi_j \right\rangle \overline{\left\langle h_i (\theta_0), \phi_j \right\rangle} \\
&\equiv \sum_{i=1}^{n} Y_{in}.
\end{aligned}
$$

In the following, θ_0 is dropped to simplify the notation. Note that Y_{in} is i.i.d., $E(Y_{in}) = 0$ and

$$
\begin{aligned}
V(Y_{in}) &= \frac{1}{n} \sum_j \frac{1}{\mu_j^2 + \alpha_n} \left| \left\langle \nabla_\theta \psi, \phi_j \right\rangle \right|^2 var \left[\left\langle h_i, \phi_j \right\rangle \right] \\
&\quad + \frac{2}{n} \sum_{j<k} \frac{1}{\sqrt{\mu_j^2 + \alpha_n}} \frac{1}{\sqrt{\mu_k^2 + \alpha_n}} \left\langle \nabla_\theta \psi, \phi_j \right\rangle \overline{\left\langle \nabla_\theta \psi, \phi_k \right\rangle} cov \left(\left\langle h_i, \phi_j \right\rangle, \left\langle h_i, \phi_k \right\rangle \right).
\end{aligned}
$$

We have

$$
\begin{aligned}
cov \left(\left\langle h_i, \phi_j \right\rangle, \left\langle h_i, \phi_k \right\rangle \right) &= E \left[\int h_i(t) \overline{\phi_j(t)} \pi(t) dt \overline{\int h_i(s) \overline{\phi_k(s)} \pi(s) ds} \right] \\
&= \int \int E \left[h_i(t) \overline{h_i(s)} \right] \overline{\phi_j(t)} \pi(t) dt \phi_k(s) \pi(s) ds \\
&= \left\langle K \phi_k, \phi_j \right\rangle \\
&= \begin{cases} 0 \text{ if } k \neq j, \\ \mu_j \text{ if } k = j. \end{cases}
\end{aligned}
$$

Hence

$$
V(Y_{in}) = \frac{1}{n} \sum_j \frac{\mu_j}{\mu_j^2 + \alpha_n} \left| \left\langle \nabla_\theta \psi, \phi_j \right\rangle \right|^2
$$

and

$$
s_n^2 \equiv \sum_i V(Y_{in}) = \sum_j \frac{\mu_j}{\mu_j^2 + \alpha_n} \left| \left\langle \nabla_\theta \psi, \phi_j \right\rangle \right|^2 \underset{n \to \infty}{\to} \left\langle \nabla_\theta \psi, \nabla_\theta \psi \right\rangle_K
$$

by Step (i)1 of the proof of Theorem 7 in CaFl. To establish the asymptotic normality, we need to check the Liapunov condition:

$$
\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} E \left[|Y_{in}|^{2+\delta} \right] \to 0.
$$

Let $\delta = 2$,

$$E\left[|Y_{in}|^4\right] \leq \frac{1}{n^2} E\left\{\left(\sum_j \frac{1}{\sqrt{\mu_j^2 + \alpha_n}} \left|\left\langle \nabla_\theta \psi, \phi_j \right\rangle\right| \left|\left\langle h_i, \phi_j \right\rangle\right|\right)^4\right\}$$

By Cauchy-Schwartz, we have

$$\begin{aligned} \sum_j \left|\left\langle \nabla_\theta \psi, \phi_j \right\rangle\right| \left|\left\langle h_i, \phi_j \right\rangle\right| &\leq \left(\sum_j \left|\left\langle \nabla_\theta \psi, \phi_j \right\rangle\right|^2\right)^{1/2} \left(\sum_j \left|\left\langle h_i, \phi_j \right\rangle\right|^2\right)^{1/2} \\ &= \|\nabla_\theta \psi\| \, \|h_i\| \,. \end{aligned}$$

Using $\sqrt{\mu_j^2 + \alpha_n} \geq \sqrt{\alpha_n}$, we have

$$\sum_{i=1}^n E\left[|Y_{in}|^4\right] \leq \frac{1}{n\alpha_n^2} E\left[\|h_i\|^4\right] \|\nabla_\theta \psi\|^4$$

$\|\nabla_\theta \psi\| < \infty$ because $\|\nabla_\theta \psi\|_K < \infty$ and $E\left[\|h_i\|^4\right] < \infty$ by Assumption CF13. Hence, the upper bound converges to 0 as $n\alpha_n^2$ goes to infinity. By Lindeberg Theorem (Davidson, 1994), we have

$$\frac{\sum_i Y_{in}}{s_n^2} \xrightarrow{L} \mathcal{N}(0,1)$$

or equivalently

$$\left\langle K^{-1/2} \, \nabla_\theta \, \psi(\theta_0), \left(K^{\alpha_n}\right)^{-1/2} \sqrt{n} h_n(\theta_0) \right\rangle \xrightarrow{L} \mathcal{N}\left(0, \langle \nabla_\theta \psi, \nabla_\theta \psi \rangle_K\right).$$

Proof of Proposition 4.3. The asymptotic normality follows directly from Lemma B.1. The efficiency follows from Proposition 4.2. We just need to verify the assumptions of Lemma B.1.

Assumptions CF1 and 1 are the same.

Assumption CF2: $h(t, X; \theta_0) = e^{it'X} - \psi_{\theta_0}(t)$ belongs to $L^2(\pi)$ under Assumption 2 and is continuous in θ by Assumption 4(i).

Assumptions CF2 and 3 are the same.

Assumptions CF4' requires that $\psi_{\theta_0}(t) - \psi_\theta(t) \in \mathcal{H}(K) + \mathcal{H}(K)^\perp$ for any $\theta \in \Theta$. It is equivalent to check that $\overline{\psi_{\theta_0}(t) - \psi_\theta(t)} \in \mathcal{H}(K) + \mathcal{H}(K)^\perp$. Note that the Fourier transform of $\psi_{\theta_0}(-t) - \psi_\theta(-t)$ is

$$\frac{1}{2\pi} \int e^{-it'x} \left(\psi_{\theta_0}(t) - \psi_\theta(t)\right) dt = \left[f_{\theta_0}(x) - f_\theta(x)\right].$$

Denote

$$\tilde{G} = \frac{f_{\theta_0}(x) - f_\theta(x)}{f_{\theta_0}(x)}. \tag{B.10}$$

Using the characterization of $\mathcal{H}(K)$ given in (4.6), $\psi_{\theta_0}(-t) - \psi_\theta(-t) \in \mathcal{H}(K)$ iff

$$\left\|\psi_{\theta_0}(t) - \psi_\theta(t)\right\|_K^2 = E\tilde{G}^2 = \int \frac{\left(f_{\theta_0}(x) - f_\theta(x)\right)^2}{f_{\theta_0}(x)} dx < \infty$$

which is satisfied as $f_\theta(x)/f_{\theta_0}(x) \leq 1$. We have proven that $E^{\theta_0}(h(t, X; \theta))$ belongs to the domain of $K^{-1/2}$ also denoted $\mathcal{H}(K)$. Assumptions CF4' is satisfied.

Assumption CF5 stipulates that if $Eh \in \mathcal{N}\left(K^{-1/2}\right)$ then $Eh = 0$. This is an identification condition. A stronger requirement would be $\|g\|_K = 0 \Rightarrow g = 0$. We know that there is an element of $C(g)$ such that $\|g\|_K^2 = \overline{EG^2}$. Hence $\|g\|_K^2 = 0 \Rightarrow G = 0$ with probability 1. Moreover we have $EG(x)\overline{h(t,x)} = 0$ for all t hence $g(t) = 0$ for all t. We have shown that $\mathcal{N}\left(K^{-1/2}\right) = \{0\}$. Assumption CF5 is therefore satisfied.

Assumption CF6: (a) $h_n(\theta) = \frac{1}{n}\sum_{j=1}^n \left(e^{it'X_j} - \psi_\theta(t)\right) \in \mathcal{D}\left(\left(K_n^\alpha\right)^{-1/2}\right)$ is satisfied because, for small n, $\left(K_n^\alpha\right)^{-1/2}$ is bounded so its domain is $L^2(\pi)$ and for n large, it is satisfied because $E^{\theta_0}(h(t, X; \theta))$ belongs to the domain of $K^{-1/2}$ as shown in the proof of Assumption CF4'.

(b) Next we need to check that $Q_n(\theta) = \|h_n(\theta)\|_{K_n^\alpha}^2$ is continuous in θ. For n small, the term

$$\|h_n(\theta)\|_{K_n^\alpha}^2 = \sum_{j=1}^n \frac{\left\langle h_n(\theta), \hat{\phi}_j\right\rangle^2}{\hat{\mu}_j} \tag{B.11}$$

is continuous because $h_n(\theta)$ is continuous (as $\psi_\theta(t)$ is continuous by Assumption 4(i)). For n large, we have shown in the proof of Assumptions CF4' that

$$\left\|E^{\theta_0}(h(t, X; \theta))\right\|_K^2 = E^{\theta_0}\left[\left(1 - \frac{f(x;\theta)}{f(x;\theta_0)}\right)^2\right] \tag{B.12}$$

which is also continuous in θ.

Assumption 8': (a) h differentiable is satisfied under Assumption 4(i).

(b) Note that $E^{\theta_0}\left[\frac{\partial h(t,X;\theta)}{\partial \theta_j}\right] = \frac{\partial \psi_\theta(t)}{\partial \theta_j}$ and $\frac{\partial E^{\theta_0}[h(t,X;\theta)]}{\partial \theta_j} = \frac{\partial \psi_\theta(t)}{\partial \theta_j}$, hence $E^{\theta_0}\left[\frac{\partial h(t,X;\theta)}{\partial \theta_j}\right] = \frac{\partial E^{\theta_0}[h(t,X;\theta)]}{\partial \theta_j}$.

(c) We show that $\overline{\nabla_\theta \psi} \in \mathcal{H}(K)$ for $\theta = \theta_0$. We apply Proposition 4.1. We need to find G_2 the centered solution of

$$\begin{aligned}
\nabla_\theta \psi_\theta(-t) &= E^{\theta_0}\left[G_2(X) e^{-it'X}\right] \\
&= \int e^{-it'x} G_2(x) f(x;\theta_0)\, dx.
\end{aligned}$$

$$\begin{aligned}
G_2(x) f(x;\theta_0) &= \frac{1}{2\pi}\int e^{it'x} \nabla_\theta \psi_\theta(-t)\, dt \\
&= \frac{1}{2\pi}\int e^{-it'x} \nabla_\theta \psi_\theta(t)\, dt \\
&= \frac{\partial f(x;\theta)}{\partial \theta'}.
\end{aligned}$$

So that

$$G_2(x) = \frac{\partial f(x;\theta)}{\partial \theta'} \frac{1}{f(x;\theta_0)} \tag{B.13}$$

and

$$EG_2G_2' = E^{\theta_0}\left[\frac{\partial f(x;\theta)}{\partial \theta'}\left(\frac{\partial f(x;\theta)}{\partial \theta'}\right)'\left(\frac{f(x;\theta)}{f(x;\theta_0)}\right)^2\right]$$

is bounded by I_{θ_0} for $\theta \neq \theta_0$ and is equal to I_{θ_0} for $\theta = \theta_0$ (this way we have proven the asymptotic efficiency again).

Assumption 9': We need to show that

$$\frac{\partial}{\partial \theta'}\|h_n(\theta)\|^2_{K_n^\alpha} = 2\left\langle \frac{\partial h_n(\theta)}{\partial \theta'}, h_n(\theta)\right\rangle_{K_n^\alpha} \tag{B.14}$$

For n small, this follows from the expression of (B.11). We now turn our attention to the limit. Note that by Equation (B.12), we have

$$\begin{aligned}\frac{\partial}{\partial \theta'}\left\|E^{\theta_0}(h(t,X;\theta))\right\|^2_K &= \frac{\partial}{\partial \theta'}E^{\theta_0}\left[\left(1-\frac{f(x;\theta)}{f(x;\theta_0)}\right)^2\right] \\ &= -2\int \frac{\partial f(x;\theta)}{\partial \theta'}\left(1-\frac{f(x;\theta)}{f(x;\theta_0)}\right)dx.\end{aligned}$$

Now we want to calculate $\left\langle \frac{\partial E^{\theta_0}(h(t,X;\theta))}{\partial \theta'}, E^{\theta_0}(h(t,X;\theta))\right\rangle_K$ and show that both sides of Equation (B.14) are the same. Using the expression of G_2 given in (B.13) and $\tilde{G}$ in (B.10), Parzen (1970, page 6) implies:

$$\begin{aligned}\left\langle \frac{\partial E^{\theta_0}(h(t,X;\theta))}{\partial \theta'}, E^{\theta_0}(h(t,X;\theta))\right\rangle_K &= EG_2\tilde{G} \\ &= -\int \frac{\partial f(x;\theta)}{\partial \theta'}\left(1-\frac{f(x;\theta)}{f(x;\theta_0)}\right)dx.\end{aligned}$$

This proves (B.14) at the limit.

Assumption CF11': We need to prove that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n\left(e^{it'X_j}-\psi_\theta(t)\right)$$

converges to a Gaussian process $\mathcal{N}(0,K)$ in $L^2(\pi)$. This result follows from Theorem 1.8.4 of Van der Vaart and Wellner (1996, page 50) under the condition

$$E^\theta\left[\|h(.,X;\theta)\|^2\right] < \infty$$

which is satisfied because $\|h(.,X;\theta)\|$ is bounded. Indeed

$$\begin{aligned}
\|h(.,X;\theta)\|^2 &= \int \left|e^{it'X} - \psi_\theta(t)\right|^2 \pi(t)\,dt \\
&= \int \left(1 - 2\operatorname{Re}\left(e^{it'X}\psi_\theta(-t)\right) + |\psi_\theta(t)|^2\right)\pi(t)\,dt \\
&= 1 + \int |\psi_\theta(t)|^2 \pi(t)\,dt \\
&\leq 2 \qquad \text{(B.15)}
\end{aligned}$$

Here we see the importance that π be a p.d.f.

Assumption CF12: We need to check that

$$\int\int |k(s,t)|^2 \pi(s)\pi(t)\,dsdt < \infty.$$

Replacing k by its expression, we obtain

$$\begin{aligned}
&\int\int \left|\psi_{\theta_0}(s-t) - \psi_{\theta_0}(s)\psi_{\theta_0}(-t)\right|^2 \pi(s)\pi(t)\,dsdt \\
\leq &\int\int \left(\left|\psi_{\theta_0}(s-t)\right| + \left|\psi_{\theta_0}(s)\right|\left|\psi_{\theta_0}(-t)\right|\right)^2 \pi(s)\pi(t)\,dsdt \\
\leq &\ 2
\end{aligned}$$

because c.f. are necessarily bounded by 1 in absolute value.

Assumption CF13: We need to check that $E\|h\|^4 < \infty$ which holds because of (B.15).

Assumption CF14: (a) First check that

$$\left\|h_n - E^{\theta_0}(h(\theta))\right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in θ on Θ. Note that

$$h_n - E^{\theta_0}(h(\theta)) = \frac{1}{n}\sum_{j=1}^n \left(e^{it'X_j} - \psi_{\theta_0}(t)\right)$$

does not depend on θ. We have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{j=1}^n \left(e^{it'X_j} - \psi_{\theta_0}(t)\right)\right\|^2 \\
= &\int \left|\frac{1}{n}\sum_{j=1}^n \left(e^{it'X_j} - \psi_{\theta_0}(t)\right)\right|^2 \pi(t)\,dt \\
= &\frac{1}{n^2}\sum_{j=1}^n \int \left|e^{it'X_j} - \psi_{\theta_0}(t)\right|^2 \pi(t)\,dt \\
&+\frac{2}{n^2}\sum_{j<k}\int \operatorname{Re}\left(e^{it'X_j} - \psi_{\theta_0}(t)\right)\left(e^{-it'X_k} - \psi_{\theta_0}(-t)\right)\pi(t)\,dt
\end{aligned}$$

which is $O_p\left(\frac{1}{n}\right)$ because $\{X_j\}$ are iid.

(b) Second check that

$$\left\|\bigtriangledown_\theta h_n - E^{\theta_0}\left(\bigtriangledown_\theta h\left(\theta\right)\right)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in θ on Θ. Note that

$$\bigtriangledown_\theta h_n - E^{\theta_0}\left(\bigtriangledown_\theta h\left(\theta\right)\right) = -\bigtriangledown_\theta \psi + E^{\theta_0}\left[\bigtriangledown_\theta \psi\right] = 0.$$

So this condition is trivially satisfied. ■

Proof of Proposition 5.1. (i) The set of moment conditions (5.1) is complete hence the efficiency follows from Theorem 4.2.

(ii) We compute the inverse of the operator K^Z for each Z at the function $\partial\psi_\theta\left(t|Z\right)/\partial\theta$. Denote $w\left(s,Z\right)$ this inverse, it is such that

$$\frac{\partial\psi_\theta\left(t|Z\right)}{\partial\theta} = \int\left[\psi_\theta\left(t-s|Z\right)-\psi_\theta\left(t|Z\right)\psi_\theta\left(-s|Z\right)\right]w\left(s,Z\right)\pi\left(s\right)ds.$$

Assume that $\int\psi_\theta\left(-s|Z\right)w\left(s,Z\right)\pi\left(s\right)ds = 0$. Replacing $\psi_\theta\left(t-s|Z\right)$ by its expression and interchanging the order of integration yields

$$\begin{aligned}\frac{\partial\psi_\theta\left(t|Z\right)}{\partial\theta} &= \int e^{ity}\int e^{-isy}w\left(s,Z\right)\pi\left(s\right)ds f\left(y|Z\right)dy\\ &= \int e^{ity}\psi_{w\pi}\left(y,Z\right)f\left(y|Z\right)dy\end{aligned}$$

where $\psi_{w\pi}\left(y,Z\right) = \int e^{-isy}w\left(s,Z\right)\pi\left(s\right)ds$. Using the Fourier inversion formula yields

$$\begin{aligned}\psi_{w\pi}\left(y,Z\right)f\left(y|Z\right) &= \frac{1}{2\pi}\int e^{-ity}\frac{\partial\psi_\theta\left(t|Z\right)}{\partial\theta}dt\\ &= \frac{\partial f\left(y|Z\right)}{\partial\theta}.\end{aligned}$$

Using again a Fourier inversion, we obtain

$$w\left(t,Z\right) = \frac{1}{\pi\left(t\right)}\int e^{ity}\frac{\partial\ln f\left(y|Z\right)}{\partial\theta}dy.$$

Note that the condition $\int\psi_\theta\left(-s|Z\right)w\left(s,Z\right)\pi\left(s\right)ds = 0$ is indeed satisfied. The result follows. ■

Proof of Corollary 5.2. The score function $\partial\ln f\left(y|z\right)/\partial\theta = \partial\ln f_u\left(y-\beta' z\right)/\partial\theta$ is spanned by the set of moment conditions S. ■

Proof of Proposition 6.1.

Denote

$$\nabla_{\theta\theta} h_n = \frac{\partial^2 h_n}{\partial\theta\partial\theta'}$$

Derivation of the asymptotic bias of $\hat{\theta}^{\alpha}$

We follow the same steps as in Newey and Smith (2001). The GMM objective function is given by

$$\langle h_n(\theta), A_n h_n(\theta) \rangle .$$

$\hat{\theta}^{\alpha}$ satisfies the first order condition is

$$\left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n h_n\left(\hat{\theta}^{\alpha}\right) \right\rangle = 0.$$

A Taylor expansion gives

$$\begin{aligned}
0 &= \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n h_n(\theta_0) \right\rangle \\
&+ \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n \nabla_\theta\, h_n(\theta_0) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle \\
&+ \frac{1}{2} \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n \left(\hat{\theta}^{\alpha} - \theta_0\right)' \nabla_{\theta\theta}\, h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle
\end{aligned}$$

where $\overline{\theta}$ is a mean value. Denote

$$\begin{aligned}
M_n^* &= \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n \nabla_\theta\, h_n(\theta_0) \right\rangle, \\
M_n &= \langle \nabla_\theta h_n(\theta_0), A_n \nabla_\theta\, h_n(\theta_0) \rangle, \\
M &= \left\langle E^{\theta_0} \nabla_\theta\, h(\theta_0), A E^{\theta_0} \nabla_\theta\, h(\theta_0) \right\rangle.
\end{aligned}$$

$$\begin{aligned}
0 &= \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n h_n(\theta_0) \right\rangle \\
&+ (M_n^* - M + M) \left(\hat{\theta}^{\alpha} - \theta_0\right) \\
&+ \frac{1}{2} \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n \left(\hat{\theta}^{\alpha} - \theta_0\right)' \nabla_{\theta\theta}\, h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle.
\end{aligned}$$

$$\begin{aligned}
\left(\hat{\theta}^{\alpha} - \theta_0\right) &= -M^{-1} \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n h_n(\theta_0) \right\rangle \\
&\quad - M^{-1} (M_n^* - M) \left(\hat{\theta}^{\alpha} - \theta_0\right) \\
&\quad - \frac{1}{2} M^{-1} \left\langle \nabla_\theta h_n\left(\hat{\theta}^{\alpha}\right), A_n \left(\hat{\theta}^{\alpha} - \theta_0\right)' \nabla_{\theta\theta}\, h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle \\
&= -M^{-1} \langle \nabla_\theta h_n(\theta_0), A_n h_n(\theta_0) \rangle \\
&\quad - M^{-1} \left\langle \nabla_{\theta\theta} h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right), A_n h_n(\theta_0) \right\rangle \\
&\quad - M^{-1} (M_n^* - M) \left(\hat{\theta}^{\alpha} - \theta_0\right) \\
&\quad - \frac{M^{-1}}{2} \left\langle \nabla_\theta h_n(\theta_0), A_n \left(\hat{\theta}^{\alpha} - \theta_0\right)' \nabla_{\theta\theta}\, h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle \\
&\quad - \frac{M^{-1}}{2} \left\langle \nabla_{\theta\theta} h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right), A_n \left(\hat{\theta}^{\alpha} - \theta_0\right)' \nabla_{\theta\theta}\, h_n\left(\overline{\theta}\right) \left(\hat{\theta}^{\alpha} - \theta_0\right) \right\rangle.
\end{aligned}$$

Note that the last term is $O_p\left(n^{-3/2}\right)$ and $M_n^* = M_n + O_p\left(n^{-1/2}\right)$. We obtain

$$
\begin{aligned}
& \left(\hat{\theta}^{\alpha} - \theta_0\right) \\
= \; & -M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle && \text{(B.16)} \\
& -M^{-1}\left[\left\langle \bigtriangledown_\theta h_n\left(\theta_0\right), A_n h_n\left(\theta_0\right)\right\rangle - \left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle\right] \\
& -M^{-1}\left\langle \bigtriangledown_{\theta\theta} h_n\left(\theta_0\right)\left(\hat{\theta}^{\alpha} - \theta_0\right), A_n h_n\left(\theta_0\right)\right\rangle \\
& -M^{-1}\left\langle\left(\bigtriangledown_{\theta\theta} h_n\left(\overline{\theta}\right) - \bigtriangledown_{\theta\theta} h_n\left(\theta_0\right)\right)\left(\hat{\theta}^{\alpha} - \theta_0\right), A_n h_n\left(\theta_0\right)\right\rangle && \text{(B.17)} \\
& -M^{-1}\left(M_n - M\right)\left(\hat{\theta}^{\alpha} - \theta_0\right) \\
& -\frac{M^{-1}}{2}\left\langle \bigtriangledown_\theta h_n\left(\theta_0\right), A_n\left(\hat{\theta}^{\alpha} - \theta_0\right)' \bigtriangledown_{\theta\theta} h_n\left(\theta_0\right)\left(\hat{\theta}^{\alpha} - \theta_0\right)\right\rangle \\
& -\frac{M^{-1}}{2}\left\langle \bigtriangledown_\theta h_n\left(\theta_0\right), A_n\left(\hat{\theta}^{\alpha} - \theta_0\right)'\left(\bigtriangledown_{\theta\theta} h_n\left(\overline{\theta}\right) - \bigtriangledown_{\theta\theta} h_n\left(\theta_0\right)\right)\left(\hat{\theta}^{\alpha} - \theta_0\right)\right\rangle && \text{(B.18)} \\
& +O_p\left(n^{-3/2}\right)
\end{aligned}
$$

As (B.17) and (B.18) are $O_p\left(n^{-3/2}\right)$ they can be included in the rest. Replacing $\hat{\theta}^{\alpha} - \theta_0$ by $-M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle$, we obtain

$$
\begin{aligned}
& \left(\hat{\theta}^{\alpha} - \theta_0\right) \\
= \; & -M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle && \text{(B.19)} \\
& -M^{-1}\left\langle \bigtriangledown_\theta h_n\left(\theta_0\right) - E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), A_n h_n\left(\theta_0\right)\right\rangle && \text{(B.20)} \\
& -M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), \left(A_n - A\right) h_n\left(\theta_0\right)\right\rangle && \text{(B.21)} \\
& +M^{-1}\left\langle E^{\theta_0} \bigtriangledown_{\theta\theta} h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle && \text{(B.22)} \\
& +M^{-1}\left(M_n - M\right) M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle && \text{(B.23)} \\
& -\frac{M^{-1}}{2}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right),\right. && \text{(B.24)} \\
& \qquad \left.\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle' M^{-1} A E^{\theta_0} \bigtriangledown_{\theta\theta} h\left(\theta_0\right) M^{-1}\left\langle E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right), Ah_n\left(\theta_0\right)\right\rangle\right\rangle \\
& +O_p\left(n^{-3/2}\right)
\end{aligned}
$$

We want to give an expression of the bias $E^{\theta_0}\left(\hat{\theta}^{\alpha} - \theta_0\right)$ that is easy to estimate. Note that $\bigtriangledown_\theta h_n\left(\theta_0\right) = \bigtriangledown_\theta \psi_\theta|_{\theta=\theta_0} = E^{\theta_0} \bigtriangledown_\theta h\left(\theta_0\right)$ hence (B.20) is equal to zero. To simplify the notation, we omit θ_0 everywhere in the sequel. Denote $h\left(t, X_i; \theta_0\right)$ as h_i, $K_n^\alpha = A_n^{-1} = (K_n^2 + \alpha I)K_n^{-1}$, and $K^\alpha = A^{-1} = (K^2 + \alpha I)K^{-1}$. The expectation of the first term (B.19) is zero.

Term (B.21): Using a Taylor expansion we have

$$
A_n - A = \left(K_n^\alpha\right)^{-1} - \left(K_n^\alpha\right)^{-1} \simeq -\left(K^\alpha\right)^{-2}\left(K_n^\alpha - K^\alpha\right) \simeq -\left(K^\alpha\right)^{-2}\left(K_n - K\right).
$$

$$
\begin{aligned}
& E\left[\left(K_n - K\right) h_n\right] \\
= \; & E \int \left[\frac{1}{n}\sum_{j=1}^{n} h_j(s)\,\overline{h_j(t)} - k(s,t)\right] h_i(t)\,\pi(t)\,dt \\
= \; & \frac{1}{n}\int E\left[h_i(s)\,|h_i(t)|^2\right]\pi(t)\,dt.
\end{aligned}
$$

Hence the expectation of term (B.21) is given by

$$
\frac{1}{n}M^{-1}\left\langle \nabla_\theta \psi_\theta, (K^\alpha)^{-2}\left[\int E\left[h_i(.)\,|h_i(t)|^2\right]\pi(t)\,dt\right]\right\rangle.
$$

Term (B.22): The expectation of this term is simply

$$
\frac{1}{n}E\left[\langle \nabla_{\theta\theta}\psi_\theta, Ah_i\rangle M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right].
$$

Term (B.23): Note that

$$
M_n - M = \langle \nabla_\theta\psi_\theta, (A_n - A)\nabla_\theta\,\psi_\theta\rangle
$$

$$
\begin{aligned}
& E\left[(M_n - M)\,M^{-1}\left\langle E^{\theta_0}\nabla_\theta\, h(\theta_0), Ah_n(\theta_0)\right\rangle\right] \\
= \; & E\left[\langle \nabla_\theta\psi_\theta, (A_n - A)\nabla_\theta\,\psi_\theta\rangle M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right] \\
\simeq \; & -\frac{1}{n}E\left[\int \nabla_\theta\psi_\theta\,(K^\alpha)^{-2}\left\{\overline{\int [k_i(.,s) - k(.,s)]\,\nabla_\theta\,\psi_\theta(s)\,\pi(ds)}\right\}\pi(dt)\,M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right] \\
= \; & -\frac{1}{n}E\left[\left\langle \nabla_\theta\psi_\theta, (K^\alpha)^{-2}h_i(.)\right\rangle\left(\overline{\int \overline{h_i(s)}\,\nabla_\theta\,\psi_\theta(s)\,\pi(ds)}\right)' M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right].
\end{aligned}
$$

The expectation of (B.23) is given by

$$
-\frac{1}{n}M^{-1}E\left[\left\langle \nabla_\theta\psi_\theta, (K^\alpha)^{-2}h_i(.)\right\rangle\left(\int h_i(s)\,\nabla_\theta\,\psi_\theta(-s)\,\pi(ds)\right)' M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right].
$$

Term (B.24): The expectation of this term is simply

$$
-\frac{1}{n}\frac{M^{-1}}{2}\left\langle \nabla_\theta\psi_\theta, E\left[\langle \nabla_\theta\psi_\theta, Ah_i\rangle' M^{-1}A\,\nabla_{\theta\theta}\,\psi_\theta M^{-1}\langle \nabla_\theta\psi_\theta, Ah_i\rangle\right]\right\rangle. \qquad \text{(B.25)}
$$

Derivation of the asymptotic variance of $\hat{\theta}^\alpha$

When α is fixed, the operator A is bounded, hence we can use the expression of the variance given in Theorem 2 of CaFl where $B = A^{1/2}$. The asymptotic variance of $\sqrt{n}\left(\hat{\theta}^\alpha - \theta_0\right)$ is given by

$$
\begin{aligned}
& M^{-1}\left\langle A^{1/2}E^{\theta_0}\nabla_\theta\, h(\theta_0), \left(A^{1/2}KA^{1/2}\right)A^{1/2}E^{\theta_0}\nabla_\theta\, h(\theta_0)\right\rangle M^{-1} \\
= \; & M^{-1}\left\langle AE^{\theta_0}\nabla_\theta\, h(\theta_0), KAE^{\theta_0}\nabla_\theta\, h(\theta_0)\right\rangle M^{-1}.
\end{aligned}
$$

■

C. References

Beard, R., S. Caudill, and D. Gropper (1991) "Finite Mixture Estimation of Multiproduct Cost Functions", *Review of Economics and Statistics*, 654-664.

Bierens, H. (1990) "A consistent Conditional Moment Test of Functional Form", *Econometrica*, Vol. 58, 1443-1458.

Cardot, H., F. Ferraty, and P. Sarda (1999) "Functional Linear Model", *Statistics and Probability Letters*, Vol. 15, 11-22.

Carrasco, M., M. Chernov, J. P. Florens and E. Ghysels (2000) "Estimating Diffusions with a Continuum of Moment Conditions", mimeo.

Carrasco, M. and J. P. Florens (2000) "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, 16, 797-834.

Carroll, R. and P. Hall (1988) "Optimal Rates of Convergence for Deconvolving a Density", *Journal of American Statistical Association*, 83, No.404, 1184-1186.

Chacko, G. and L. Viceira (2001) "Spectral GMM Estimation of Continuous-Time Processes", mimeo, Harvard University.

Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

Debnath, L. and L. Mikunsinski (1999) *Introduction to Hilbert Spaces with Applications*, Academic Press, San Diego.

Devroye, L. (1986) *Non-Uniform Random Variate Generation*, Springer Verlag, NY.

DuMouchel, W. (1975) "Stable Distributions in Statistical Inference: 2. Information from Stably Distributed Samples", *Journal of the American Statistical Association*, 70, 386-393.

Epps, T. and L. Pulley (1983) "A test for normality based on the empirical characteristic function", *Biometrika*, 70, 3, 723-6.

Epps, T., K. Singleton and L. Pulley (1982) "A test of separate families of distributions based on the empirical moment generating function", *Biometrika*, 69, 2, 391-399.

Feuerverger, A. (1990) "An efficiency result for the empirical characteristic function in stationary time-series models", *The Canadian Journal of Statistics*, Vol. 18, N.2, 155-161.

Feuerverger, A. and P. McDunnough (1981a) "On the Efficiency of Empirical Characteristic Function Procedures", *J. R. Statist. Soc.* B, 43, 20-27.

Feuerverger, A. and P. McDunnough (1981b) "On Some Fourier Methods for Inference", *Journal of the American Statistical Association*, Vol. 78, 379-387.

Feuerverger, A. and R. Mureika (1977) "The Empirical Characteristic Function and its Applications", *The Annals of Statistics*, Vol. 5, 88-97.

Groetsch, C. (1993) *Inverse Problems in the Mathematical Sciences*, Vieweg, Wiesbaden.

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

Hansen, L. (1985) "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators", *Journal of Econometrics*, 30, 203-238.

HECKMAN, J., R. ROBB, AND J. WALKER (1990) "Testing the Mixture of Exponentials and Estimating the Mixing distribution by the Method of Moments", *Journal of the American Statistical Association*, Vol. 85, 582-589.

HOROWITZ, J. AND M. MARKATOU (1996) "Semiparametric Estimation of Regression Models for Panel Data", *Review of Economic Studies*, 63, 145-168.

HSIAO, C. (1986) *Analysis of Panel Data*, Econometric Society Monograph, Cambridge University Press, Cambridge, UK.

JIANG, G. AND J. KNIGHT (1999) "Efficient Estimation of the continuous Time Stochastic Volatility Model Via the Empirical Characteristic Function", forthcoming in *Journal of Business & Economic Statistics*.

KOUTROUVELIS, I. AND J. KELLERMEIER (1981) "A Goodness-of-fit Test based on the Empirical Characteristic Function when Parameters must be Estimated" *J.R. Statist. Soc.* B. **43**, 173-176.

LANCASTER, T. (1990) *The Econometric Analysis of Transition data*, Cambridge University Press, UK.

LI, T. AND Q. VUONG (1998) "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators", *Journal of Multivariate Analysis*, 65, 139-165.

LI, T., I. PERRIGNE, AND Q. VUONG (2000) "Conditionally independent private information in OCS wildcat auctions", *Journal of Econometric*s, 98, 129-161.

LUKACS, E. (1970) *Characteristic Functions*, Griffin's statistical monographs and courses. London.

MCCULLOCH, J. H. (1996) "Financial Applications of Stable Distributions", *Handbook of Statistics*, Vol. 14.

MORDUCH, J. AND H. STERN (1997) "Using mixture models to detect sex bias in health outcomes in Bangladesh", *Journal of Econometrics*, 77, 259-276.

NEWEY, W. AND D. MCFADDEN (1994) "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics*, Vol. IV, eds R.F. Engle and D. McFadden, Elsevier.

NEWEY, W. AND R. SMITH (2001) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", mimeo, MIT.

PARZEN, E. (1970) "Statistical Inference on time series by RKHS methods", *12th Biennial Seminar Canadian Mathematical Congress Proc.*, R. Pyke, ed., Canadian Mathematical Society, Montreal.

QUANDT, R. AND J. RAMSEY (1978) "Estimating Mixtures of Normal Distributions and Switching Regressions", *Journal of the American Statistical Association*, Vol. 73, 730-738.

ROSSBERG, H-J., B. JESIAK, AND G. SIEGEL (1985) *Analytic Methods of Probability Theory*, Akademie Verlag, Berlin.

SAITOH, S. (1997) *Integral transforms, reproducing kernels and their applications*, Longman.

Schmidt, P. (1982) "An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions", *Econometrica*, 50, 501-516

Singleton, K. (2001) "Estimation of Affine Pricing Models Using the Empirical Characteristic Function", *Journal of Econometrics*, 102, 111-141.

Stefanski, L. and R. Carroll (1990) "Deconvoluting Kernel Density Estimators", *Statistics*, 2, 169-184.

Stinchcombe, M. and H. White (1998) "Consistent specification testing with nuisance parameters present only under the alternative", *Econometric Theory*, 14, 295-325.

Tran, K. C. (1998) "Estimating mixtures of normal distributions via the empirical characteristic function". *Econometric Reviews*, 17, 167-183.

Wunsch, A. D. (1994) *Complex Variables with Applications*, Addison-Wesley, Massachusetts.