LUISS

IRIS - Luiss Guido Carli Research Institutional Repository | Catalogo istituzionale della ricerca Luiss Guido Carli

Existence and Uniqueness of Semiparametric Projections

Pre-print version of the following publication: | Versione pre-print della seguente pubblicazione:

Original Citation/Citazione:
I., Komunjer; Ragusa, Giuseppe. (2009). Existence and Uniqueness of Semiparametric Projections. no. 0wg3j51c. University of California. http://escholarship.org/uc/item/0wg3j51c

Availability/Disponibilità:
This version is available at: 11385/7518 since: - Questa versione è disponibile alla pagina: 11385/7518 dal:

Publisher/Casa editrice:
University of California

Published version/Pubblicato:

License/Licenza:
DRM (Digital rights management) non definiti

Availability/Termini d'uso:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license. For all terms of use and more information see the publisher's website. | I termini e le condizioni relativi al riutilizzo della presente versione della pubblicazione sono disciplinati dalla politica editoriale. Le opere messe a disposizione con licenze Creative Commons possono essere utilizzate conformemente ai termini e alle condizioni previste da tali licenze. Per l'insieme delle condizioni di utilizzo e per ulteriori informazioni si rinvia al sito web dell'editore.

This item was downloaded from IRIS Luiss (https://iris.luiss.it/). When citing, please refer to the published version. | Questo documento è stato scaricato da IRIS Luiss (https://iris.luiss.it/). Per la citazione, fare riferimento alla versione pubblicata sul sito dell'editore.

(Article begins on next page | Il contributo inizia nella pagina successiva)

11 October 2026

Department of Economics, UCSD
UC San Diego

Title:
Existence and Uniqueness of Semiparametric Projections

Author:
Komunjer, Ivana, University of California - San Diego
Ragusa, Giuseppe, University of California, Irvine

Publication Date:
02-01-2009

Publication Info:
Department of Economics, UCSD, UC San Diego

Permalink:
http://escholarship.org/uc/item/0wg3j51c

Keywords:
semiparametric methods, least favorable submodels

Abstract:
In this paper we propose primitive conditions under which a projec- tion of a conditional density onto a set de#ned by conditional moment restric- tions exists and is unique. Moreover, we provide an analytic expression of the obtained projection. Our #rst result is to show the existence when the moment function is bounded. The result is as we would expect from the analogous results obtained in the unconditional case. Our second result relaxes the boundedness assumption and replaces it with a correct speci#cation condition. Showing that the correct speci#cation of the moment function is su#cient for the projection to exist is entirely new and not yet seen in the literature. #

eScholarship
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

Existence and Uniqueness of Semiparametric Projections

Ivana Komunjer and Giuseppe Ragusa*

University of California, San Diego and University of California, Irvine

Abstract: In this paper we propose primitive conditions under which a projection of a conditional density onto a set defined by conditional moment restrictions exists and is unique. Moreover, we provide an analytic expression of the obtained projection. Our first result is to show the existence when the moment function is bounded. The result is as we would expect from the analogous results obtained in the unconditional case. Our second result relaxes the boundedness assumption and replaces it with a correct specification condition. Showing that the correct specification of the moment function is sufficient for the projection to exist is entirely new and not yet seen in the literature.

*Affiliations and Contact Information. Komunjer (*Corresponding Author*): University of California, San Diego; 9500 Gilman Drive MC 0508; La Jolla, CA 92093-0508; E-mail: komunjer@ucsd.edu. Ragusa: University of California, Irvine; 3151 Social Science Plaza; Irvine, CA 92697-5100; E-mail: gragusa@uci.edu.

1. Introduction

Consider the problem of inferring a function p from a prior guess q, both elements of a space $\mathcal{P}$, when the only available information is that p belongs to some subset $\mathcal{Q}$ of $\mathcal{P}$. This problem is central in applications in statistics, probability theory, information theory, machine learning, physical chemistry, and other scientific fields. A familiar example is when p and q are probability distributions in $\mathcal{P}$, while $\mathcal{Q}$ is some known convex subset of that space. A general approach to the inference problem for p is to search for an element p^* in $\mathcal{Q}$ which minimizes a distance to q. When unique, the solution p^* is called the projection of q onto the set $\mathcal{Q}$. Of course, the form of p^* depends on the choice of the distance. By far the most employed is the Kullback-Leibler divergence (Kullback and Leibler, 1951) also referred to as: I-divergence, Kullback-Leibler distance, cross entropy, relative entropy, or information discrimination, depending on the field.

This paper is concerned with the problems of existence and characterization of p^* when $\mathcal{P}$ is the space of all conditional probability distributions, and the subset of interest $\mathcal{Q}$ is implicitly defined by a set of conditional moment restrictions. The applications of projections onto sets defined by moment restrictions are pervasive in many scientific fields. In statistics and econometrics they include: semiparametric efficient estimation (Tripathi and Kitamura, 2003; Kitamura et al., 2004), analysis of misspecified models (Sawa, 1978; White, 1982, 1994; Vuong, 1989; Chor-Yiu and White, 1996; Otsu et al., 2008), asset pricing estimation (Kitamura and Stutzer, 2002), optimal testing (Kitamura, 2001), methods of Bayesian prior determinations (Bernardo, 1979, 2005), as well as Bayesian inference in semiparametric models (Zellner, 1996, 2002, 2003; Zellner and Tobias, 2001; Kim, 2002). An extensive review of applications in other fields is given in Buck and Macaulay (1991).

Projecting a conditional distribution q onto a set $\mathcal{Q}$ involves a constrained optimization problem with infinite-dimensional variables. Hence, proving that a projection exists requires a demonstration that the optimization problem has a well defined solution. The literature offers several results dealing with the unconditional case, that is, the case in which $\mathcal{P}$ and $\mathcal{Q}$ are collections of unconditional probability distributions. A classical reference for the Kullback-Leibler distance is Csiszár (1975). For general distances indexed by convex functions see Liese (1975), Borwein and Lewis (1993), Csiszár (1995), and citations therein.

Establishing general existence results for projections is a non-trivial exercise. Borwein and Lewis (1991) exhibit simple examples of optimization problems in which the optimal value p^* is not attained. Showing that an optimal feasible p^* exists can be very difficult to justify depending on the probability space $\mathcal{P}$ and the distance employed. Proposed demonstrations entail many mathematical subtleties that are often overlooked in applications. A mathematical note by Borwein and Limber (1996) highlights often encountered errors.

Broadly speaking, known existence results require that set $\mathcal{Q}$ be *closed*. If $\mathcal{Q}$ is a set of probability distributions that satisfy some moment conditions, then the closedness of $\mathcal{Q}$ is typically obtained by assuming that the moment functions are bounded. The boundedness of moment functions is in turn obtained by assuming that the random variables under consideration have compact support (see, e.g., Borwein and Lewis, 1993).[1] In the case where the moment restrictions that define $\mathcal{Q}$ are unconditional, Csiszár (1995) gives a proof of existence that requires all the exponential moments of the underlying random variables to exist and be finite.

[1]Recently, in the context of Generalized Empirical Likelihood estimation, Otsu et al. (2008) use boundedness conditions to ensure existence (see their Corollary 3.3).

The main contribution of this paper is twofold. First, we extend known existence results to setups in which the projection set $\mathcal{Q}$ is defined by *conditional* moment restrictions. Second, we weaken the boundedness requirements on the moment functions defining the projection set $\mathcal{Q}$. Such requirements are typically employed by the literature dealing with the unconditional case.

Our proof of existence exploits special features of the projection problem usually encountered in a semiparametric setting in which the moment functions are parameterized by a finite dimensional parameter θ. Here, the projection set $\mathcal{Q}(\theta)$ is a collection of conditional probability distributions that satisfy the moment restriction when the parameter is set to θ. When there exists a value θ_0 of θ such that the true conditional distribution belongs to $\mathcal{Q}(\theta_0)$, the moment condition is correctly specified. Our main result is to show that under correct specification—and additional relatively mild assumptions—there exists a convex subset of Θ containing θ_0 such that for every θ in this subset the projection is guaranteed to exist.

We next discuss the form of the projection under the same set of assumptions. It is worth pointing out that while we assume the existence of a θ_0 that satisfies the conditional moment restriction, we do not assume that this θ_0 is unique. In other words, our existence result holds for conditional moment models whether or not they are identified, provided they remain correctly specified.

The conditional distribution projections that we characterize have some useful statistical properties. For instance, projections are a constructive way of obtaining the least favorable parametric submodels introduced by Stein (1956). In the context of efficient estimation, Komunjer and Vuong (2009) show that the least favorable distributions naturally lead to the semiparametric efficiency bounds based on the conditional moment restrictions. Another interesting feature of semiparametric projections is ob-

tained under the Kullback-Leibler distance. Then, the projection problem corresponds to a population counterpart of the smoothed empirical likelihood (EL) estimator for semiparametric models defined by conditional moment restrictions (Tripathi and Kitamura, 2003; Kitamura et al., 2004).

The paper is organized as follows. In Section 2 we present our setup and recall some well known concepts of convex analysis. The same section defines the projections on spaces of conditional probability densities. In Section 3, we focus on projections sets defined by conditional moment restrictions. The same section contains our main result which shows that the projection exists, and derives its analytic form. All our proofs are relegated to an Appendix.

2. Setup

2.1. Preliminaries

Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose that $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Further, let $(\mathbb{E}, \mathcal{E})$ be a measurable space in which $\mathbb{E}$ is a complete separable metric space and $\mathcal{E}$ is the σ-algebra of Borel sets. Then, given an $\mathcal{F}$-measurable random element $X : \Omega \to \mathbb{E}$ we shall be interested in the regular conditional measure of X given $\mathcal{G}$, which we denote μ. That μ is a regular conditional measure means that $\mu : \Omega \times \mathcal{E} \to \mathbb{R}_+$ satisfies: (i) for each $B \in \mathcal{E}$, $\omega \mapsto \mu(\omega, B)$ is a version of $P(X(\omega) \in B|\mathcal{G})$, and (ii) for a.e. ω, $B \mapsto \mu(\omega, B)$ is a probability measure on $(\mathbb{E}, \mathcal{E})$. In particular, such measure exists for the spaces $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ $(n \in \mathbb{N})$ and $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ (see, e.g., Corollary on p.230 in Shiryaev, 1996). For simplicity, we shall focus on the case $\mathbb{E} = \mathbb{R}^n$.

We further assume that for a.e. ω, $\mu(\omega, \cdot)$ is absolutely continuous (with respect to Lebesgue measure). So by Radon-Nikodym theorem there exists $f : \Omega \times \mathbb{R}^n \to \mathbb{R}_+$

such that for a.e. ω we have:

$$\mu(\omega, B) = \int_B f(\omega, x)dx \tag{1}$$

i.e. f is a regular conditional density of X given $\mathcal{G}$. In what follows, we shall assume:

Assumption A1. For a.e. $\omega \in \Omega$, the function $x \mapsto f(\omega, x)$ is continuous on $\mathbb{R}^n$.

Under Assumption A1, the conditional density f is a Carathéodory function, and thus has the virtue of being jointly measurable, i.e. $(\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}))$-measurable (see, e.g., Lemma 4.51 in Aliprantis and Border, 2007). In particular, this implies by Tonelli's Theorem (see, e.g., Theorem 11.28 in Aliprantis and Border, 2007) that f is jointly integrable with respect to the product measure $P \times \lambda$ (λ being the Lebesgue measure on $\mathbb{R}^n$) and that:

$$\int f d(P \times \lambda) = \int_{\mathbb{R}^n} \int_\Omega f(\omega, x)dP(\omega)dx = \int_\Omega \int_{\mathbb{R}^n} f(\omega, x)dxdP(\omega) = 1 \tag{2}$$

where the last equality follows from (1).

Now, let $L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$ be the space of (equivalence classes of) functions $g : \Omega \times \mathbb{R}^n \to \mathbb{R}$ that are $(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}))$-measurable and $P \times \lambda$-integrable, i.e. $\int |g| d(P \times \lambda)$ exists and is finite. We say that two elements g_1 and g_2 of $L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$ belong to the same equivalence class—property which we denote $g_1 = g_2$ a.s.—if for a.e. ω we have $g_1(\omega, x) = g_2(\omega, x)$ for every $x \in \mathbb{R}^n$. For any $g \in L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$, the L_1-norm of g is defined by:

$$\|g\|_1 \equiv \int |g| d(P \times \lambda) = \int_{\mathbb{R}^n} \int_\Omega |g(\omega, x)| dP(\omega)dx$$

The $L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$ space equipped with the L_1-norm $\|\cdot\|_1$ is a Banach space, and the set of functions $h \in L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$ that are $(\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}))$-measurable forms a closed subspace of $L_1(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n))$ that we denote $L_1(\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^n))$. When the

conditioning is done with respect to a sub-σ-field generated by a subvector of X, then the above L_1-norm induces the metric of " integrated L_1-distance" used in Tang and Ghosal (2007). In particular, we shall be interested in those elements of $L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$ that are nonnegative valued, so we let $\mathcal{P} \equiv \{g \in L_1(\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^n)) : g(\Omega \times \mathbb{R}^n) \subseteq \mathbb{R}_+\}$. It follows from the property in Equation (2) that the conditional density f belongs to $\mathcal{P}$.

2.2. Distances, divergences and projections on $\mathcal{P}$

A distance $\mathcal{D}$ on the set $\mathcal{P}$ is any nonnegative valued function defined on $\mathcal{P} \times \mathcal{P}$ such that $\mathcal{D}(g_1, g_2) = 0$ if and only if $g_1 = g_2$ with probability one. In this paper, we further restrict the class of distances $\mathcal{D}$ and focus on the ϕ-divergences $\mathcal{D}_\phi$. The class of ϕ-divergences among probability distributions was first introduced by Ali and Silvey (1966) and Csiszár (1967); we now recall their definition.

Let $\mathcal{K}$ denote the class of all functions $\phi : [0, +\infty] \to [0, +\infty]$ with the following properties:

Assumption A2. (i) $\phi \in \mathcal{C}^4((0, +\infty))$; (ii) ϕ is strictly convex on $(0, +\infty)$; (iii) $\phi(1) = \phi'(1) = 0$, $\phi''(1) = 1$; (iv) $\lim_{u\to+\infty} \phi'(u) = +\infty$; (v) $\lim_{u\to 0} \phi'(u) = -\infty$.

In order to guarantee that ϕ is continuous on $[0, +\infty]$ we let $\phi(0) = \lim_{u\to 0} \phi(u)$ and $\phi(+\infty) = \lim_{u\to+\infty} \phi(u)$. Further, to deal with zero and infinity, we adopt the understanding that $\phi'(0) = \lim_{u\to 0} \phi'(u)$, $0 \cdot \phi\left(\frac{0}{0}\right) = 0$, and $0 \cdot \phi\left(\frac{v}{0}\right) = v \lim_{u\to+\infty} \phi'(u) = +\infty$ when $v > 0$.

Given a function $\phi \in \mathcal{K}$, a ϕ-divergence between g_1 and g_2 in $\mathcal{P}$, denoted $\mathcal{D}_\phi(g_1, g_2)$,

is then formally defined as:

$$\begin{aligned}\mathcal{D}_\phi(g_1, g_2) &\equiv \int g_2 \phi\left(\frac{g_1}{g_2}\right) d(P \times \lambda) \\ &= \int_{\mathbb{R}^n} \int_\Omega g_2(\omega, x) \phi\left(\frac{g_1(\omega, x)}{g_2(\omega, x)}\right) dP(\omega) dx \end{aligned} \tag{3}$$

Notice that the ϕ-divergence between g_1 and g_2 can also be expressed in terms of the corresponding conditional measures, $\nu_1(\omega, B) = \int_B g_1(\omega, x)dx$ and $\nu_2(\omega, B) = \int_B g_2(\omega, x)dx$, by defining $\mathcal{D}_\phi(\nu_1, \nu_2) \equiv \int \phi(d\nu_1/d\nu_2)d\nu_2$. This formulation is used by Kitamura and Stutzer (1997) and Kitamura (2001), for example. When one considers only measures $\nu_1 \ll \nu_2$, the two definitions are equivalent and $\mathcal{D}_\phi(\nu_1, \nu_2) = \mathcal{D}_\phi(g_1, g_2)$.

The class of ϕ-divergences $\mathcal{D}_\phi$ generally includes many distances used in statistics, such as the Kullback-Leibler distance (I-divergence) obtained when $\phi(u) = u\ln(u) - u + 1$ (see, e.g., Kullback and Khairat, 1966; Csiszár, 1975), and the χ^2 distance for which $\phi(u) = (u-1)^2$. In the econometric literature, an application of the Kullback-Leibler distance can be found in Kitamura and Stutzer (1997)'s Exponential Tilting estimator. Note that the requirement A2(iv) rules out the reverse I-divergence, $\phi(u) = -\ln u + u - 1$, and the Hellinger distance, $\phi(u) = (\sqrt{u} - 1)^2$, since for both cases $\lim_{u\to+\infty} \phi'(u) < +\infty$. The remaining assumptions A2(i)-(iii) are fairly standard. When combined with the continuity and convexity properties of ϕ, Assumptions A2(iv,v) guarantee that the map ϕ' is onto $[-\infty, +\infty]$. This property of ϕ' shall be particulary important when calculating the Legendre transform (or convex conjugate) of ϕ.

Before proceeding, we recall some useful concepts from convex analysis; for a detailed discussion, see, e.g., Rockafellar (1970) and Hiriart-Urruty and Lemarechal (1993). The convex conjugate (or Legendre-Fenchel transform) of ϕ is a real mapping $\phi^* : \mathbb{R} \to [-\infty, +\infty]$ which to every $v \in \mathbb{R}$ associates $\phi^*(v) \equiv \sup_{x\in\mathbb{R}}[vx -$

$\phi(x)]$. Under Assumption A2, ϕ is differentiable on $\mathbb{R}$ and its derivative ϕ' such that $\phi'([0,+\infty]) = [-\infty,+\infty]$, so the Legendre conjugate of ϕ equals:

$$\phi^*(\upsilon) = \upsilon(\phi')^{-1}(\upsilon) - \phi\left((\phi')^{-1}(\upsilon)\right).$$

The following lemma establishes several useful properties of ϕ^*.

Lemma 1. *Under Assumption A2, we have: (i)* $\phi^* \in \mathcal{C}^2(\mathbb{R})$*, (ii)* ϕ^* *is strictly convex on* $\mathbb{R}$*, (iii)* $\phi^* > 0$ *on* $(0,+\infty)$*, (iv)* $\phi^{*'} > 0$ *on* $\mathbb{R}$*, (v)* $\phi^{*'}(\upsilon) = (\phi')^{-1}(\upsilon)$ *for any* $\upsilon \in \mathbb{R}$*, (vi)* $\phi^{*''}(\upsilon) = [\phi''((\phi')^{-1}(\upsilon))]^{-1}$ *for any* $\upsilon \in \mathbb{R}$*.*

We are now ready to introduce the concept of projection of a conditional density. With the ϕ-divergence given in Equation (3), the $\mathcal{D}_\phi$-projection of f onto a subset $\mathcal{Q}$ of $\mathcal{P}$ is defined as follows:

Definition 1. The $\mathcal{D}_\phi$-projection of f onto a set $\mathcal{Q} \subseteq \mathcal{P}$ is (when it exists) a $g^* \in \mathcal{Q}$ satisfying: $\mathcal{D}_\phi(g^*, f) = \mathcal{D}_\phi(\mathcal{Q}, f)$, where $\mathcal{D}_\phi(\mathcal{Q}, f) \equiv \inf_{g\in\mathcal{Q}} \mathcal{D}_\phi(g, f)$.

The next section discusses conditions under which the $\mathcal{D}_\phi$-projection of f is guaranteed to exist.

3. Projection Existence and Characterization

3.1. Projection Set

In most statistical and econometric applications, the projection set $\mathcal{Q}$ is defined by a set of either unconditional or conditional moment restrictions. While the unconditional problem has been extensively studied in the literature, little is known about the conditional one. Here we focus on sets $\mathcal{Q}$ defined by conditional moment restrictions.

Let $\Theta \subseteq \mathbb{R}^k$ ($k \in \mathbb{N}$, $k < \infty$) and consider some known moment function $a : \Omega \times \mathbb{R}^n \times \Theta \to \mathbb{R}^m$ ($m \in \mathbb{N}$, $m < \infty$) parameterized by $\theta \in \Theta$. We further assume

that for every $\theta \in \Theta$, $a(\cdot, \cdot, \theta)$ is $(\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^m))$-measurable and such that for a.e. ω the conditional expectation

$$E[a(X, \theta)|\mathcal{G}] \equiv \int_{\mathbb{R}^n} a(\omega, x, \theta) f(\omega, x) dx \tag{4}$$

exists and is finite. Note that the number of restrictions m can be greater than one.

We now focus on $\mathcal{D}_\phi$-projecting f onto a set of conditional densities that satisfy with probability one the conditional moment restrictions $E[a(X, \theta)|\mathcal{G}] = 0$ for a given value of θ. The projection set $\mathcal{Q}$ is then parameterized by θ and we denote it $\mathcal{Q}(\theta)$. The set $\mathcal{Q}(\theta)$ can be characterized as follows:

$$\mathcal{Q}(\theta) \equiv \left\{ g \in \mathcal{P} : \int_{\mathbb{R}^n} a(\omega, x, \theta) g(\omega, x) dx = 0 \text{ and } \int_{\mathbb{R}^n} g(\omega, x) dx = 1, \text{ for a.e. } \omega \right\}.$$

From a statistical point of view, the set $\mathcal{Q}(\theta)$ is a component of a semiparametric model $\mathcal{Q}_\Theta$ defined as a collection of all densities in $\mathcal{Q}(\theta)$ obtained by letting θ vary in Θ. More formally, we have $\mathcal{Q}_\Theta \equiv \bigcup_{\theta \in \Theta} \mathcal{Q}(\theta)$.

In what follows we establish the existence of the $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta)$ under alternative assumptions on the moment function a.

3.2. Bounded Case

When the projection set $\mathcal{Q}(\theta)$ is non-empty, one way to establish the existence of the $\mathcal{D}_\phi$-projection is to verify that $\mathcal{Q}(\theta)$ is compact in the topology induced by $\| \cdot \|_1$. For this, we first establish the lower semi-continuity of the distance $\mathcal{D}_\phi(\cdot, f)$.

Lemma 2. *Under Assumption A2, $\mathcal{D}_\phi(\cdot, f)$ is lower semi-continuous on $\mathcal{P}$.*

The lower semi-continuity of $\mathcal{D}_\phi(\cdot, f)$ on $\mathcal{P}$ allows us to establish the existence of the $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta) \subseteq \mathcal{P}$ when $\mathcal{Q}(\theta)$ satisfies a simple topological

condition—that of compactness. Indeed, when $\mathcal{Q}(\theta)$ is compact, we can apply a well-known result that a real-valued lower semi-continuous function on as compact space attains a minimum value (see, e.g., Theorem 2.40 in Aliprantis and Border, 2007).

However, establishing the compactness of $\mathcal{Q}(\theta)$ is generally non trivial. Often $\mathcal{Q}(\theta)$ only satisfies a weaker topological condition—that of being closed under $\|\cdot\|_1$. A sufficient condition for $\mathcal{Q}(\theta)$ to be closed is given in the following:

Assumption A3. For every $\theta \in \Theta$, there exists a positive constant $M(\theta)$ such that for every $x \in \mathbb{R}^n$ and a.e. ω, $|a(\omega, x, \theta)| \leqslant M(\theta) < \infty$.

Under our assumptions on ϕ (in particular, under the assumption $\lim_{u\to+\infty} \phi'(u) = +\infty$) the closedness of $\mathcal{Q}(\theta)$ is sufficient to guarantee that a $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta)$ exists. The proof of the following theorem adapts the arguments used by Liese (1975) to models with conditional moment restrictions.

Theorem 1. *Let Assumptions A1-A3 hold. Then, for every $\theta \in \Theta$, a $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta)$ exists.*

In most statistical and econometric applications, Assumption A3 is too strong and it is often ruled out by the nature of the model itself. For instance, simple models with conditional mean restrictions on random variables that have full support lead to unbounded moment functions. Of course, depending on the particular application, it may possible to replace Assumption A3 with an alternative sufficient condition for $\mathcal{Q}(\theta)$ to be closed.

3.3. Unbounded Case

When the moment function a in (4) is not necessarily bounded, it is not a trivial exercise to establish that a projection exists and to then characterize it. Known results

dealing with the projection in the unconditional settings (Teboulle and Vajda, 1993; Csiszár, 1995) cannot be extended to the conditional setting considered here.

Our approach to establishing the existence of the projection is based on the following intuitive argument. If f satisfies the conditional moment restriction for some $\theta_0 \in \Theta$, i.e. if $\int_{\mathbb{R}^n} a(\omega, x, \theta_0) f(\omega, x) dx = 0$ a.s., then, clearly, $f \in \mathcal{Q}(\theta_0)$. In addition, with probability one we have $\mathcal{D}_\phi(f, f) = 0$. Hence, when $\theta = \theta_0$ the $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta)$ exists and is unique: it is f itself. Provided we can invoke the Implicit Function Theorem, it should then hold that for small deviations of θ around θ_0 the projection of f onto $\mathcal{Q}(\theta)$ continues to exist. We now provide a more formal treatment of this argument.

We start by assuming that the moment function a and the conditional density f have the following property:

Assumption A4. There exists $\theta_0 \in \Theta$ such that for a.e. ω, we have $E[a(X, \theta_0)|\mathcal{G}] = 0$.

Assumption A4 states that the statistical model $\mathcal{Q}_\Theta = \bigcup_{\theta \in \Theta} \mathcal{Q}(\theta)$ defined by the conditional moment restrictions based on the moment function a is correctly specified, i.e. $f \in \mathcal{Q}_\Theta$. Note that A4 does not impose the value θ_0 to be unique. In other words, the statistical model $\mathcal{Q}_\Theta$ need not be identified, and we can have $\theta_1 \in \Theta \backslash \{\theta_0\}$ such that both $f \in \mathcal{Q}(\theta_0)$ and $f \in \mathcal{Q}(\theta_1)$ hold. It is worth pointing out that we do not even require θ_0 to be locally identified, i.e. the moment function a is allowed to be such that $E[a(X, \theta)|\mathcal{G}]$ remains zero on connected open subsets of Θ.

In what follows, we restrict our attention to cases in which a is continuously differentiable with respect to θ.

Assumption A5. For every $x \in \mathbb{R}^n$ and a.e. ω, the mapping $\theta \mapsto a(\omega, x, \theta)$ is in $\mathcal{C}^1(\Theta, \mathbb{R}^m)$.

For mappings that satisfy Assumption A5 we let $D_\theta a \in L(\mathbb{R}^k, \mathbb{R}^m)$ denote the partial derivative of the moment function a with respect to θ.

We first restrict the behavior of the Legendre conjugate ϕ^* and its derivative $\phi^{*'}$ by imposing several local integrability conditions. In what follows, $\mathcal{U}(\theta_0, \varepsilon) \equiv \mathcal{B}((\theta_0', 0, 0')', \varepsilon)$ is an open ball in $\mathbb{R}^{k+m+1}$, centered at $(\theta_0', 0, 0')' \in \Theta \times \mathbb{R}^{m+1}$ and with radius $\varepsilon > 0$.

Assumption A6. There exists $\mathcal{U}(\theta_0, \varepsilon_1) \subset \Theta \times \mathbb{R}^{m+1}$ such that for every $(\theta', \eta, \lambda')' \in \mathcal{U}(\theta_0, \varepsilon_1)$ and a.e. ω we have:

(i) $\int_{\mathbb{R}^n} \phi^*(\eta + \lambda' a(\omega, x, \theta)) f(\omega, x) dx < \infty$

(ii) $\int_{\mathbb{R}^n} \phi^{*'}(\eta + \lambda' a(\omega, x, \theta)) f(\omega, x) dx < \infty$

(iii) $\int_{\mathbb{R}^n} |\phi^{*'}(\eta + \lambda' a(\omega, x, \theta))| f(\omega, x) dx < \infty$

Assumption A6 effectively imposes restrictions on the true conditional density f. We now give an interpretation of A6(i,ii) in the case of the Kullback-Leibler distance (I-divergence) obtained when $\phi(u) = u \ln u - u + 1$. The Legendre conjugate of ϕ then equals $\phi^*(v) = \exp v - 1$, so the properties in A6(i,ii) hold under a conditional version of a *weak Cramér condition*: for every θ in a neighborhood of θ_0 and every λ close to $0 \in \mathbb{R}^m$, we have $\int_{\mathbb{R}^m} \exp\big(\lambda' a(\omega, x, \theta)\big) f(\omega, x) dx < \infty$ with probability one. The Cramér condition restricts the generating function for the conditional moments of f—when θ is close to θ_0—to be finite on a neighborhood of zero, at which the restriction is obviously satisfied.

The following conditions ensure that one can differentiate under the integral sign:

Assumption A7. There exists $\mathcal{U}(\theta_0, \varepsilon_2) \subset \Theta \times \mathbb{R}^{m+1}$ such that for a.e. ω we have:

(i) $\int_{\mathbb{R}^n} \sup \phi^{*''}\big(\eta + \lambda' a(\omega, x, \theta)\big)\big(1 + |a(\omega, x, \theta)|^2\big) f(\omega, x) dx < \infty$

(ii) $\int_{\mathbb{R}^n} \sup \phi^{*''}\big(\eta + \lambda' a(\omega, x, \theta)\big) \| D_\theta a(\omega, x, \theta)' \lambda a(\omega, x, \theta)' \| f(\omega, x) dx < \infty$

(iii) $\int_{\mathbb{R}^n} \sup \phi^{*''}\big(\eta + \lambda' a(\omega, x, \theta)\big)|D_\theta a(\omega, x, \theta)'\lambda| f(\omega, x)dx < \infty$

(iv) $\int_{\mathbb{R}^n} \sup \phi^{*'}\big(\eta + \lambda' a(\omega, x, \theta)\big)\|D_\theta a(\omega, x, \theta)\| f(\omega, x)dx < \infty$

where sup stands for $\sup_{(\theta', \eta, \lambda')' \in \mathcal{U}(\theta_0, \varepsilon_2)}$.

Assumption A7 is used to ensure that Lebesgue Dominated Convergence Theorem applies, i.e. that we can interchange the order of integration and differentiation in the first order conditions that characterize the projection g^* in Definition 1. In order to apply the Implicit Function Theorem to those conditions obtained when $\theta = \theta_0$, we require the following assumption:

Assumption A8. For a.e. ω, $\int_{\mathbb{R}^n} a(\omega, x, \theta_0)a(\omega, x, \theta_0)' f(\omega, x)dx$ is invertible.

Note that Assumption A8 does not say anything about the properties of $D_\theta a(x, \theta_0)$ which are important in establishing that θ_0 is locally identified. The main reason why only the invertibility of $E[a(X, \theta_0)a(X, \theta_0)'|\mathcal{G}]$ is needed is because our proof establishes local existence of a mapping $\theta \mapsto (\eta(\theta), \lambda(\theta))$ around the point θ_0 at which $(\eta(\theta_0), \lambda(\theta_0)')' = 0 \in \mathbb{R}^{m+1}$, and where $\eta \in \mathbb{R}$ and $\lambda \in \mathbb{R}^m$ are the Lagrange multipliers defined in Theorem 2 below. As a such, we only need the matrix of derivatives with respect to η and λ to be invertible (see Step 3 in the proof of Theorem 2).

We are now ready to state the main result of this section. Similar to previously, $\mathcal{B}(\theta_0, \varepsilon)$ is an open ball in $\mathbb{R}^k$, centered at $\theta_0 \in \Theta$ and with radius $\varepsilon > 0$.

Theorem 2. *Let Assumptions A1-A8 hold. Then there exists $\mathcal{B}(\theta_0, \varepsilon) \subset \Theta$ such that for every $\theta \in \mathcal{B}(\theta_0, \varepsilon)$, a $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}_\theta$ exists. Moreover, the projection denoted g^* is P a.s. unique and given by:*

$$g^*(\omega, x, \theta) \equiv \phi^{*'}\big(\eta(\omega, \theta) + \lambda(\omega, \theta)' a(\omega, x, \theta)\big) f(\omega, x) \tag{5}$$

for every $x \in \mathbb{R}^n$ and a.e. ω, with $(\eta(\omega, \theta), \lambda(\omega, \theta)) \equiv \arg\inf_{(\eta, \lambda')' \in \mathbb{R}^{m+1}} \int \phi^*\big(\eta + \lambda' a(\omega, x, \theta)\big) f(\omega, x)dx - \eta$.

We first comment on the strength of the assumptions used in Theorem 2. Similar to us, Csiszár (1995) gives a proof of the existence of the $\mathcal{D}_\phi$-projection that does not make the boundedness assumptions on the moment function a. In particular, Corollary to Theorem 3 in Csiszár (1995) is based on a moment condition on the convex conjugate ϕ^* of ϕ. Under the Kullback-Leibler distance (I-divergence), this condition translates into a *strong Cramér condition*, whereby "strong" we mean that the finiteness of the generating function for the conditional moments of $a(X, \theta)$ (when θ is close to θ_0) needs to hold for all $\lambda \in \mathbb{R}^m$. This condition is obviously stronger than our "weak" version imposed in Assumption A6, which only needs to hold for λ in some neighborhood of $0 \in \mathbb{R}^m$.

Theorem 2 establishes two important results. First, it shows that the $\mathcal{D}_\phi$-projection of f onto $\mathcal{Q}(\theta)$ exists and is a.s. P unique. As pointed out previously, this result exploits the existence of the $\mathcal{D}_\phi$-projection when $\theta = \theta_0$ and extends it by means of the Implicit Function Theorem. It is worth noting that the proof of Theorem 2 establishes in a direct way that there exists g^* in $\mathcal{Q}(\theta)$ with density given in Equation (5). An early suggestion of such direct approach can be found in Csiszár (1975) (see a discussion on p.156 in Csiszár, 1975, for the unconditional case).

The second key result of Theorem 2 is to derive the analytic expression of g^*. The density of the $\mathcal{D}_\phi$-projection obtained in Equation (5) reveals an interesting property: it is parameterized by two random *finite dimensional* Lagrange multipliers η and λ, both of which are $\mathcal{G}$-measurable and depend on θ. In other words, projecting onto the semiparametric set $\mathcal{Q}(\theta)$ reduces the problem to the one in which the density g^* can be written as a product of two terms: a first one $\phi^{*'}(\eta(\omega, \theta) + \lambda(\omega, \theta)' a(\omega, x, \theta))$ that is finitely parameterized by θ, and a second one that is the true density $f(\omega, x)$ which does not depend on θ.

A number of interesting properties can be derived from the expression of the $\mathcal{D}_\phi$-projected density g^* obtained in Theorem 2. First, we can note that for a.e. ω, we have: $\eta(\omega, \theta_0) = 0 \in \mathbb{R}$ and $\lambda(\omega, \theta_0) = 0 \in \mathbb{R}^m$. We are now interested in the values that successive derivatives of g^* with respect to the parameter θ (when they exist) take at a true value θ_0. Under the same set of conditions as in Theorem 2, we have the following result:

Corollary 3. *Assume the conditions of Theorem 2 hold. Then, for a.e.* ω, $\eta(\omega, \cdot)$ *and* $\lambda(\omega, \cdot)$ *are continuously differentiable on* $\mathcal{B}(\theta_0, \varepsilon)$, *and we have:* $D_\theta \eta(\omega, \theta_0) = 0$, *and* $D_\theta \lambda(\omega, \theta_0) = E\big[D_\theta a(X, \theta_0)\big|\mathcal{G}\big]\big\{E\big[a(X, \theta_0)a(X, \theta_0)'|\mathcal{G}\big]\big\}^{-1}$.

In particular, Corollary 3 implies that the projected densities g^* in Theorem 2 are continuously differentiable with respect to θ.

Appendix A: Proofs

Proof of Lemma 1. First, note that from the expression of the Legendre conjugate, ϕ^* is continuous and differentiable on $\mathbb{R}$. In addition, the derivative of ϕ^* is given by:

$$\phi^{*'}(\upsilon) = (\phi')^{-1}(\upsilon), \text{ for any } \upsilon \in \mathbb{R}.$$

Given the strict convexity of ϕ in Assumption A2(ii), ϕ' is continuous and strictly increasing on $(0, +\infty)$ with $\phi'(0) = -\infty$ from A2(v), and $\phi'(+\infty) = +\infty$ from A2(iv); so its inverse $\phi^{*'}$ is continuous and strictly increasing on $\mathbb{R}$. Hence, ϕ^* is strictly convex. Since $\lim_{\upsilon \to -\infty} \phi^{*'}(\upsilon) = 0$, we have $\phi^{*'} > 0$ in $\mathbb{R}$. Moreover, from A2(iii) $\phi^*(0) = 0$ which combined with the previous property gives $\phi^* > 0$ on $(0, +\infty)$. Finally, A2(ii) implies $\phi'' > 0$ on $(0, +\infty)$ so $\phi^{*'}$ is continuously differentiable on $\mathbb{R}$ with derivative:

$$\phi^{*''}(\upsilon) = \frac{1}{\phi''((\phi')^{-1}(\upsilon))}.$$

This completes the proof of Lemma 1. □

Proof of Lemma 2. The lower semi-continuity can be formulated as a property of a sequence $\{\mathcal{D}_\phi(g_i, f)\}$ when $\{g_i\}$ is a sequence in $\mathcal{P}$. It suffices to prove that $\liminf_i \mathcal{D}_\phi(g_i, f) \geqslant \mathcal{D}_\phi(g, f)$ whenever $\lim_{i \to \infty} \|g_i - g\|_1 = 0$; then by Lemma 2.41 in Aliprantis and Border (2007), $\mathcal{D}_\phi(\cdot, f)$ is lower semicontinuous. So let $g_i \to g$ in L_1-norm. The function ϕ is continuous on $[0, +\infty]$ hence it is lower semicontinuous. Moreover, it is bounded below by 0, so Theorem 3.13 in Aliprantis and Border (2007) applies and there exists a sequence of Lipschitz continuous functions $\{\phi_k\}$ such that as $k \to \infty$, $\phi_k(u) \uparrow \phi(u)$ for all $u \in [0, +\infty]$. That ϕ_k are Lipschitz continuous means that there exists some real number c such that for every $(u, v) \in [0, +\infty]^2$, we have $|\phi_k(u) - \phi_k(v)| \leqslant c|u - v|$. Now, for any $(\omega, x) \in \Omega \times \mathbb{R}^n$ let

$$u \equiv \frac{g_i(\omega, x)}{f(\omega, x)} \text{ and } v \equiv \frac{g(\omega, x)}{f(\omega, x)}$$

Then we have

$$\left|\phi_k\left(\frac{g_i(\omega,x)}{f(\omega,x)}\right)f(\omega,x)-\phi_k\left(\frac{g(\omega,x)}{f(\omega,x)}\right)f(\omega,x)\right|\leqslant c|g_i(\omega,x)-g(\omega,x)|$$

so by using the triangle inequality

$$\begin{aligned}&|\mathcal{D}_{\phi_k}(g_i,f)-\mathcal{D}_{\phi_k}(g,f)|\\ &\quad\leqslant\int_\Omega\int_{\mathbb{R}^n}\left|\phi_i\left(\frac{g_i(\omega,x)}{f(\omega,x)}\right)f(\omega,x)-\phi_i\left(\frac{g(\omega,x)}{f(\omega,x)}\right)f(\omega,x)\right|dP(\omega)dx\\ &\quad\leqslant c\|g_i-g\|_1\end{aligned}$$

Hence for every $k\in\mathbb{N}$,

$$\mathcal{D}_{\phi_k}(g_i,f)\to\mathcal{D}_{\phi_k}(g,f)\text{ as }i\to\infty \tag{6}$$

Using a reasoning similar to that above shows that for every $i\in\mathbb{N}$,

$$\mathcal{D}_{\phi_k}(g_i,f)\leqslant\mathcal{D}_{\phi}(g_i,f) \tag{7}$$

The remainder of the proof is similar to that of Theorem 15.5 in Aliprantis and Border (2007). From Equations (6) and (7) we see that $\mathcal{D}_{\phi_k}(g,f)\leqslant\liminf_i\mathcal{D}_\phi(g_i,f)$ for every k. Hence,

$$\mathcal{D}_\phi(g,f)=\lim_{k\to\infty}\mathcal{D}_{\phi_k}(g,f)\leqslant\liminf_i\mathcal{D}_\phi(g_i,f)$$

which establishes the lower semicontinuity of $\mathcal{D}_\phi(\cdot,f)$. □

Proof of Theorem 1. We proceed in two steps.

STEP 1: We first show that under Assumption A3, the projection set $\mathcal{Q}(\theta)$ is closed under $\|\cdot\|_1$. For this, fix $\theta\in\Theta$, let $\{q_i\}$ be any convergent sequence in $\mathcal{Q}(\theta)$, and denote by $\bar{q}$ its limit, $\lim_{i\to\infty}\|q_i-\bar{q}\|_1=0$. We now show that then $\bar{q}\in\mathcal{Q}(\theta)$, i.e. the

set $\mathcal{Q}(\theta)$ is closed. We have:

$$\begin{aligned}\int_\Omega \left| \int_{\mathbb{R}^n} a(\omega,x,\theta)\bar{q}(\omega,x)dx \right| dP(\omega) &\leqslant \int_\Omega \left| \int_{\mathbb{R}^n} a(\omega,x,\theta)[\bar{q}(\omega,x) - q_i(\omega,x)]dx \right| dP(\omega) \\ &\quad + \int_\Omega \left| \int_{\mathbb{R}^n} a(\omega,x,\theta) q_i(\omega,x)dx \right| dP(\omega) \\ &= \int_\Omega \left| \int_{\mathbb{R}^n} a(\omega,x,\theta)[\bar{q}(\omega,x) - q_i(\omega,x)]dx \right| dP(\omega) \\ &\leqslant \int_\Omega \int_{\mathbb{R}^n} |a(\omega,x,\theta)| \cdot |\bar{q}(\omega,x) - q_i(\omega,x)| dx dP(\omega) \\ &\leqslant M(\theta)\|q_i - \bar{q}\|_1\end{aligned}$$

where the first equality uses $q_i \in \mathcal{Q}(\theta)$, and the last inequality follows by Assumption A3. Taking the limit of the above as $i \to \infty$ it then follows that

$$\int_\Omega \left| \int_{\mathbb{R}^n} a(\omega,x,\theta)\bar{q}(\omega,x)dx \right| dP(\omega) = 0$$

and since the quantity inside the first integral is everywhere non-negative, the above implies that for a.e. ω,

$$\int_{\mathbb{R}^n} a(\omega,x,\theta)\bar{q}(\omega,x)dx = 0$$

Hence, $\bar{q} \in \mathcal{Q}(\theta)$.

STEP 2: With θ fixed as in Step 1, assume that $\inf_{q\in\mathcal{Q}(\theta)} \mathcal{D}_\phi(q,f) = d < +\infty$, for if $\inf_{q\in\mathcal{Q}(\theta)} \mathcal{D}_\phi(q,f) = +\infty$ there is nothing to prove as any $q \in \mathcal{Q}(\theta)$ is a $\mathcal{D}_\phi$-projection. It suffices to show that there exists $q^* \in \mathcal{Q}(\theta)$ such that $\mathcal{D}_\phi(q^*,f) = d$. For this, let

$$\mathcal{Q}_d(\theta) = \big\{q \in \mathcal{Q}(\theta) : \mathcal{D}_\phi(q,f) \leqslant 2d\big\}.$$

The set $\mathcal{Q}_d(\theta)$ is a convex and non-empty subset of the Banach space $L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$. We start by showing that $\mathcal{Q}_d(\theta)$ is weakly sequentially compact in $L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$. For this, note that every $q \in \mathcal{Q}_d(\theta)$ has the same support as f, since otherwise we

would have $\mathcal{D}_\phi(q,f) = +\infty$. Then,

$$\lim_{b\to\infty}\int_{\{q>b\}} qd(P\times\lambda) \leqslant \lim_{c\to\infty}\int_{\{q/f>c\}} qd(P\times\lambda)$$

where $c \equiv b/\operatorname{ess\,sup} f$, and $\operatorname{ess\,sup} f \equiv \inf\{a\in\mathbb{R} : \int_{\{f(\omega,x)>a\}} d(P\times\lambda) = 0\}$ which is finite because $f \in L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$. Now, note that Assumption A2(ii) implies that whenever $c \geqslant 1$ and for any $x > c$ we have $(x-1)\phi'(x) > \phi(x)-\phi(1)$, so $\phi(x)-x\phi'(x) < -\phi'(x) \leqslant 0$ where the second inequality combines Assumptions A2(ii) and (iii). Hence, the mapping $x \mapsto x/\phi(x)$ is decreasing on $[c,+\infty]$ and $\lim_{c\to\infty} c/\phi(c) = 0$. This together with nonnegativity of ϕ implies that, uniformly for all $q \in \mathcal{Q}_d(\theta)$,

$$\begin{aligned}\lim_{c\to\infty}\int_{\{q/f>c\}} qd(P\times\lambda) &= \lim_{c\to\infty}\int_{\{q/f>c\}} \frac{q/f}{\phi(q/f)}\phi(q/f)fd(P\times\lambda)\\ &\leqslant \lim_{c\to\infty}\frac{c}{\phi(c)}\sup_{q\in\mathcal{Q}_d(\theta)}\int \phi(q/f)fd(P\times\lambda) = 0,\end{aligned}$$

that is, the family of densities in $\mathcal{Q}_d(\theta)$ is uniformly integrable. Thus, since a set is weakly compact if and only if its elements are uniformly integrable, the set $\mathcal{Q}_d(\theta)$ is weakly sequentially compact in $L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$ (see, e.g., Doob, 1984).

Let $\{q_i\}$ be a sequence in $\mathcal{Q}_d(\theta)$ for which

$$\lim_{i\to\infty}\int \phi(q_i/f)fd(P\times\lambda) = \inf_{q\in\mathcal{Q}_d(\theta)}\mathcal{D}_\phi(q,f) = d.$$

Weak sequential compactness of $\mathcal{Q}_d(\theta)$ implies that there exists a subsequence q_{i_k} tending weakly to some $q^* \in L_1(\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^n))$. Then, lower semicontinuity of $\mathcal{D}_\phi(\cdot,f)$ established in Lemma 2 leads to

$$\int \phi(q^*/f)fd(P\times\lambda) \leqslant \lim_{k\to\infty}\int \phi(q_{i_k}/f)fd(P\times\lambda) = d,$$

i.e. $\mathcal{D}_\phi(q^*,f) = d \leqslant 2d$ and so $q^* \in \mathcal{Q}_d(\theta)$. Since $\mathcal{Q}_d(\theta) \subseteq \mathcal{Q}(\theta)$ and from Step 1 $\mathcal{Q}(\theta)$ is closed, the limit q^* of the subsequence must be in $\mathcal{Q}(\theta)$. □

Proof of Theorem 2 and Corollary 3. The proof is done in five steps.

STEP 1: For every $(\theta', \eta, \lambda')' \in \Theta \times \mathbb{R}^{1+m}$ let:

$$I(\omega, \theta, \eta, \lambda) \equiv \int_{\mathbb{R}^n} \phi^*\big(\eta + \lambda' a(\omega, x, \theta)\big) f(\omega, x) dx - \eta$$

Fix $\theta_0 \in \Theta$. We start by showing that that for a.e. ω we have: $\inf_{(\eta,\lambda')' \in \mathbb{R}^{m+1}} I(\omega, \theta_0, \eta, \lambda)$ is attained, $(\eta(\omega, \theta_0), \lambda(\omega, \theta_0)')' = 0 \in \mathbb{R}^{m+1}$ is optimal, and $I(\omega, \theta_0, \eta(\omega, \theta_0), \lambda(\omega, \theta_0)) < \infty$. For this, we use the strict convexity of ϕ^* (from Lemma 1(ii)) which implies that for any $\upsilon \in \mathbb{R}$, $\phi^*(\upsilon) - \phi^*(0) \geqslant \upsilon \phi^{*'}(0)$. From Lemma 1(v) and Assumption A2(iii) we know that $\phi^{*'}(0) = 1$ and $\phi^*(0) = 0$, so for any $(\eta, \lambda')' \in \mathbb{R}^{m+1}$ and a.e. ω we have: $I(\omega, \theta_0, \eta, \lambda) \geqslant \lambda' E\big[a(X, \theta_0)\big|\mathcal{G}\big] = 0$. So for any $(\eta, \lambda')' \in \mathbb{R}^{m+1}$ and a.e. ω it holds that $I(\omega, \theta_0, \eta, \lambda) \geqslant I(\omega, \theta_0, \eta(\omega, \theta_0), \lambda(\omega, \theta_0)) = 0$, which shows that $(\eta(\omega, \theta_0), \lambda(\omega, \theta_0)')' = 0 \in \mathbb{R}^{m+1}$ is optimal and that $\inf_{(\eta,\lambda')' \in \mathbb{R}^{m+1}} I(\omega, \theta_0, \eta, \lambda)$ is attained. Moreover, under Assumption A6(i) we have that for a.e. ω, $I(\omega, \theta_0, \eta, \lambda) < \infty$ for any $(\eta, \lambda')' \in \mathbb{R}^{m+1} \cap \mathcal{U}_{\theta_0, \varepsilon_1}$ which is open. (Recall that $\mathcal{U}(\theta_0, \varepsilon_1)$ is an open ball in $\Theta \times \mathbb{R}^{m+1}$ with radius $\varepsilon_1 > 0$ and centered at $(\theta_0', 0, 0')' \in \Theta \times \mathbb{R}^{m+1}$.) Hence, $(\eta(\omega, \theta_0), \lambda(\omega, \theta_0)')'$ is an interior optimum, and we have for a.e. ω, $D_\eta I(\omega, \theta_0, \eta(\omega, \theta_0), \lambda(\omega, \theta_0)) = 0$ and $D_\lambda I(\omega, \theta_0, \eta(\omega, \theta_0), \lambda(\omega, \theta_0)) = 0$, where $D_\eta I(\omega, \theta, \eta, \lambda)' \in L(\mathbb{R}, \mathbb{R})$ and $D_\lambda I(\omega, \theta, \eta, \lambda)' \in L(\mathbb{R}^m, \mathbb{R})$ denote the partial derivatives of I with respect to η and λ, respectively.

STEP 2: We now use the results of Step 1 to derive the set of first order conditions satisfied by $(\eta(\omega, \theta_0), \lambda(\omega, \theta_0)')'$. For this, we use Lebesgue Dominated Convergence Theorem to be able to take the limit into the expectation in:

$$D_\eta I(\omega, \bar{\theta}, \eta(\omega, \theta_0), \lambda(\omega, \theta_0)) = \lim_{h \to 0} \int_{\mathbb{R}^n} \frac{\phi^*\big(\eta(\omega, \theta_0) + h + \lambda(\omega, \theta_0)' a(\omega, x, \theta_0)\big) - \phi^*\big(\eta(\omega, \theta_0) + \lambda(\omega, \theta_0)' a(\omega, x, \theta_0)\big)}{h} f(\omega, x) dx - 1$$

Under Assumption A2, Lemma 1 applies and ϕ^* is in $\mathcal{C}^2(\mathbb{R}, \mathbb{R})$ so by mean value

theorem and for a.e. ω:

$$\frac{\phi^*\big(\eta(\omega,\theta_0)+h+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)-\phi^*\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)}{h}=$$
$$\phi^{*\prime}\big(\eta(\omega,\theta_0)+\dot{h}+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)$$

with $\dot{h}\in(\min\{0,h\},\max\{0,h\})$. Given that $\phi^{*\prime}$ is positive and strictly increasing on $\mathbb{R}$ (see Lemma 1(iv)), we have for a.e. ω:

$$0<\phi^{*\prime}\big(\eta(\omega,\theta_0)+\dot{h}+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)\leqslant\phi^{*\prime}\big(\eta(\omega,\theta_0)+\max\{0,h\}+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big).$$

Now, for $h\in\mathbb{R}$ such that $(\theta_0',\eta(\omega,\theta_0)+h,\lambda(\omega,\theta_0)')'=(\theta_0',h,0')'\in\mathcal{U}(\theta_0,\varepsilon_1)$, the upper bound above is integrable with respect to $f(\omega,x)$; we can therefore exchange limit and expectation to get that for a.e. ω:

$$D_\eta I(\omega,\theta_0,\eta(\omega,\theta_0),\lambda(\omega,\theta_0))=\int_{\mathbb{R}^n}\phi^{*\prime}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)f(\omega,x)dx-1$$

The same reasoning shows that for any $(\theta',\eta,h,\lambda')'\in\Theta\times\mathbb{R}^{m+2}$ such that $(\theta',\eta,\lambda')'\in\mathcal{U}(\theta_0,\varepsilon_1)$ and $(\theta',\eta+h,\lambda')'\in\mathcal{U}(\theta_0,\varepsilon_1)$, we have for a.e. ω:

$$\lim_{h\to0}\int_{\mathbb{R}^n}\phi^{*\prime}\big(\eta+h+\lambda'a(\omega,x,\theta)\big)f(\omega,x)dx=\int_{\mathbb{R}^n}\phi^{*\prime}\big(\eta+\lambda'a(\omega,x,\theta)\big)f(\omega,x)dx$$

so that $\eta\mapsto\int_{\mathbb{R}^n}\phi^{*\prime}\big(\eta+\lambda'a(\omega,x,\theta)\big)f(\omega,x)dx$ is continuous on $\mathbb{R}\cap\mathcal{U}(\theta_0,\varepsilon_1)$.

Similarly, fix any $1\leqslant j\leqslant m$ and consider the partial derivative of $I(\omega,\theta_0,\eta,\lambda)$ with respect to λ_j, when evaluated at $(\eta(\omega,\theta_0),\lambda(\omega,\theta_0))$. We have for a.e. ω:

$$\frac{\phi^*\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)+ha^j(\omega,x,\theta_0)\big)-\phi^*\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big)}{h}$$
$$=\phi^{*\prime}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)+\dot{h}a^j(\omega,x,\theta_0)\big)a^j(\omega,x,\theta_0),$$

where a^j denotes the jth component of a, and $\dot{h}\in(\min\{0,h\},\max\{0,h\})$. Now, using again the convexity of ϕ^* we have for a.e. ω:

$$\Big|\phi^{*\prime}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)+\dot{h}a^j(\omega,x,\theta_0)\big)a^j(\omega,x,\theta_0)\Big|\leqslant\big|a^j(\omega,x,\theta_0)\big|\times$$
$$\max\Big\{\phi^{*\prime}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)\big),\phi^{*\prime}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)'a(\omega,x,\theta_0)+ha^j(\omega,x,\theta_0)\big)\Big\}$$

Both terms of the right hand side of the above inequality are integrable with respect to f, so using again Lebesgue's Dominated Convergence theorem, we get for a.e. ω:

$$D_{\lambda_j} I(\omega, \theta_0, \eta(\omega,\theta_0), \lambda(\omega,\theta_0)) = \int_{\mathbb{R}^n} \phi^{*'}\big(\eta(\omega,\theta_0)+\lambda(\omega,\theta_0)' a(\omega,x,\theta_0)\big) a^j(\omega,x,\theta_0) f(\omega,x) dx$$

Same reasoning as previously shows that, moreover, for any $(\theta', \eta, \lambda')' \in \mathcal{U}(\theta_0, \varepsilon_1)$ and a.e. ω we have $\lambda_j \mapsto \int_{\mathbb{R}^n} \phi^{*'}\big(\eta + \lambda' a(\omega,x,\theta_0)\big) a^j(\omega,x,\theta_0) f(\omega,x) dx$ continuous on $\mathbb{R} \cap \mathcal{U}(\theta_0, \varepsilon_1)$.

In particular, the first order conditions satisfied by $(\eta(\omega,\theta_0), \lambda(\omega,\theta_0))$ can then be written for a.e. ω as:

$$\begin{aligned} 0 &= \int_{\mathbb{R}^m} \phi^{*'}\big(\eta(\omega,\theta_0) + \lambda(\omega,\theta_0)' a(\omega,x,\theta_0)\big) f(\omega,x) dx - 1 \\ 0 &= \int_{\mathbb{R}^m} \phi^{*'}\big(\eta(\omega,\theta_0) + \lambda(\omega,\theta_0)' a(\omega,x,\theta_0)\big) a(\omega,x,\theta_0) f(\omega,x) dx \end{aligned} \tag{8}$$

Step 3: We now invoke the Implicit Function Theorem around the first order condition satisfied by the Lagrange multipliers $(\eta(\omega,\theta_0), \lambda(\omega,\theta_0))$ to extend the results to a neighborhood of θ_0. For this, let $\tau \equiv (\eta, \lambda')' \in \mathbb{R}^{m+1}$ and $\tau_0 \equiv 0 \in \mathbb{R}^{m+1}$. For any $(\theta, \tau) \in \Theta \times \mathbb{R}^{m+1}$ and a.e. ω consider then:

$$\tilde{F}(\omega, \theta, \tau) \equiv \int_{\mathbb{R}^n} F(\omega, x, \theta, \tau) f(\omega, x) dx,$$

where for any $x \in \mathbb{R}^n$ we define:

$$F(\omega, x, \theta, \tau) \equiv \begin{pmatrix} \phi^{*'}\big(\eta + \lambda' a(\omega,x,\theta)\big) - 1 \\ \phi^{*'}\big(\eta + \lambda' a(\omega,x,\theta)\big) a(\omega,x,\theta) \end{pmatrix}.$$

Note that under A6, Step 2 shows that for a.e. ω the mapping $\tau \mapsto \tilde{F}(\omega,\theta,\tau)$ is continuous on $\mathbb{R}^{m+1} \cap \mathcal{U}(\theta_0, \varepsilon_1)$. Continuity of $\theta \mapsto \tilde{F}(\omega,\theta,\tau)$ on $\Theta \cap \mathcal{U}(\theta_0, \varepsilon_1)$ for a.e. ω follows from continuity of ϕ^* (Lemma 1(i)) and $a(\omega, x, \cdot)$ (Assumption A5), and from Assumption A6(ii,iii) by using the same reasoning as in Step 2.

We now establish that for a.e. ω, the mapping $(\theta,\tau) \mapsto \tilde{F}(\omega,\theta,\tau)$ is also continuously differentiable in a neighborhood of (τ_0,θ_0). Under Assumptions A2 and A5, the mapping $(\theta,\tau) \mapsto F(\omega,x,\theta,\tau)$ is continuously differentiable on $\Theta \times \mathbb{R}^{m+1}$. Let then $D_\tau F(\omega,x,\theta,\tau)' \in L(\mathbb{R}^{m+1},\mathbb{R}^{m+1})$ and $D_\theta F(\omega,x,\theta,\tau)' \in L(\mathbb{R}^k,\mathbb{R}^{m+1})$ denote the derivatives of F with respect to τ and θ, respectively. Writing a for $a(\omega,x,\theta)$ we have for every $x \in \mathbb{R}^n$ and a.e. ω:

$$\begin{aligned} D_\tau F(\omega,x,\theta,\tau) &= \phi^{*''}(\eta+\lambda' a)\begin{pmatrix} 1 & a' \\ a & aa' \end{pmatrix} \\ D_\theta F(\omega,x,\theta,\tau) &= \Big(\phi^{*''}(\eta+\lambda' a)D_\theta a\lambda \quad \phi^{*''}(\eta+\lambda' a)D_\theta a\lambda a' + \phi^{*'}(\eta+\lambda' a)D_\theta a\Big) \end{aligned}$$

where $D_\theta a' \in L(\mathbb{R}^k,\mathbb{R}^m)$ denotes a partial derivative of $a(\omega,x,\theta)$ with respect to θ. Using the fact that ϕ^* is convex, we then have for every $x \in \mathbb{R}^n$ and a.e. ω:

$$\|D_\tau F(\omega,x,\theta,\tau)\| = \phi^{*''}(\eta+\lambda' a)\big(1+|a|^2\big),$$

and

$$\|D_\theta F(\omega,x,\theta,\tau)\| \leqslant \phi^{*''}(\eta+\lambda' a)|D_\theta a\lambda| + \phi^{*''}(\eta+\lambda' a)\|D_\theta a\lambda a'\| + \phi^{*'}(\eta+\lambda' a)\|D_\theta a\|.$$

Given the continuity of $a(\omega,x,\cdot)$, $\phi^{*\prime}$, and $\phi^{*\prime\prime}$, and the moment assumptions in A7, both $\|D_\tau F(\omega,x,\theta,\tau)\|$ and $\|D_\theta F(\omega,x,\theta,\tau)\|$ are bounded on $\mathcal{U}(\theta_0,\varepsilon_1) \cap \mathcal{U}(\theta_0,\varepsilon_2)$ by quantities that are integrable with respect to f. So by Lebesgue Dominated Convergence Theorem we can exchange limits and integration to get that for a.e. ω:

$$\begin{aligned} &D_\tau \tilde{F}(\omega,\theta,\tau) \\ &\quad = \begin{pmatrix} \int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a)f(\omega,x)dx & \int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a)af(\omega,x)dx' \\ \int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a)af(\omega,x)dx & \int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a)aa'f(\omega,x)dx \end{pmatrix} \end{aligned}$$

and

$$
\begin{aligned}
&D_\theta \tilde{F}(\omega,\theta,\tau) \\
&= \left(\int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a) D_\theta a \lambda f(\omega,x)dx \quad \int_{\mathbb{R}^n} \phi^{*''}(\eta+\lambda' a) D_\theta a \lambda a' + \phi^{*'}(\eta+\lambda' a) D_\theta a f(\omega,x)dx \right)
\end{aligned}
$$

for all $(\theta,\tau) \in \mathcal{U}(\theta_0,\varepsilon_1) \cap \mathcal{U}(\theta_0,\varepsilon_2)$. Same assumptions suffice to show that for a.e. ω the mapping $(\theta,\tau) \mapsto D_\tau \tilde{F}(\omega,\theta,\tau)$ and $(\theta,\tau) \mapsto D_\theta \tilde{F}(\omega,\theta,\tau)$ are continuous on $\mathcal{U}(\theta_0,\varepsilon_1) \cap \mathcal{U}(\theta_0,\varepsilon_2)$, following a reasoning similar to that in the Step 2. In particular, under Assumption A4 we have for a.e. ω:

$$
\begin{aligned}
D_\tau \tilde{F}(\omega,\theta_0,\tau_0) &= \begin{pmatrix} 1 & 0 \\ 0 & \int_{\mathbb{R}^n} a(\omega,x,\theta_0) a(\omega,x,\theta_0)' f(\omega,x)dx \end{pmatrix} \\
D_\theta \tilde{F}(\omega,\theta_0,\tau_0) &= \begin{pmatrix} 0 & \int_{\mathbb{R}^n} D_\theta a(\omega,x,\theta_0) f(\omega,x)dx \end{pmatrix}
\end{aligned}
$$

Finally, we invoke the Implicit Function Theorem for (θ,τ) in a neighborhood of (θ_0,τ_0), which by Equation (8) are known to solve $\tilde{F}(\omega,\theta_0,\tau_0) = 0$ for a.e. ω. Under Assumption A8, $D_\tau \tilde{F}(\omega,\theta_0,\tau_0)$ is invertible for a.e. ω. Then the Implicit Function Theorem (e.g. Theorem 9.28 in Rudin (1976)) applies and there exists $\mathcal{B}(\theta_0,\varepsilon)$ in which to any $\theta \in \mathcal{B}(\theta_0,\varepsilon) \subset \Theta$ there corresponds a unique $\tau = \tau(\omega,\theta)$ such that for a.e. ω:

$$
(\theta,\tau) \in \mathcal{U}(\theta_0,\varepsilon_1) \cap \mathcal{U}(\theta_0,\varepsilon_2) \text{ and } \tilde{F}(\omega,\theta,\tau(\omega,\theta)) = 0
$$

Note that since we are interested in solving for τ as a function of θ in $\tilde{F}(\omega,\theta,\tau) = 0$, we only need the partial derivative of $\tilde{F}$ with respect to τ to be invertible. No restrictions are placed on the partial derivatives of $\tilde{F}$ with respect to θ.

STEP 4: For every $\theta \in \mathcal{B}(\theta_0,\varepsilon)$, every $x \in \mathbb{R}^n$ and a.e. ω let then:

$$
g^*(\omega,x,\theta) = \phi^{*'}\big(\eta(\omega,\theta) + \lambda(\omega,\theta)' a(\omega,x,\theta)\big) f(\omega,x) \tag{9}
$$

where $(\eta(\omega,\theta),\lambda(\omega,\theta)')' = \tau(\omega,\theta)$ and $\tau(\omega,\theta)$ was defined in Step 3. We now show that $g^*(\omega,x,\theta)$ defined in Equation (9) is in $\mathcal{Q}(\theta)$. Then, we show that it is optimal. We have for a.e. ω, $\tilde{F}(\omega,\theta,\tau(\omega,\theta)) = 0$ which using the definition of $\tilde{F}$ gives for a.e. ω:

$$\begin{aligned} 0 &= \int_{\mathbb{R}^n} g^*(\omega,x,\theta)dx - 1 \\ 0 &= \int_{\mathbb{R}^n} a(\omega,x,\theta)g^*(\omega,x,\theta)dx \end{aligned}$$

so $g \in \mathcal{Q}(\theta)$. We now show that g is indeed optimal. Let π_θ be any other probability density belonging to $\mathcal{Q}(\theta)$. As consequence of Assumption A2, we have that for all $(\upsilon,u) \in \mathbb{R}^2$ (see Hiriart-Urruty and Lemarechal (1993)):

$$\phi^*(\upsilon) = \upsilon(\phi')^{-1}(\upsilon) - \phi\left((\phi')^{-1}(\upsilon)\right) \geqslant \upsilon u - \phi(u).$$

When evaluated at $u \equiv \pi_\theta(\omega,x)/f(\omega,x)$ and $\upsilon \equiv \eta(\omega,\theta)+\lambda(\omega,\theta)'a(\omega,x,\theta)$, the above inequality becomes for a.e. ω:

$$\begin{aligned} \big(\eta(\omega,\theta) + \lambda(\omega,\theta)'a(\omega,x,\theta)\big)\phi^{*'}\big(\eta(\omega,\theta) + \lambda(\omega,\theta)'a(\omega,x,\theta)\big)f(\omega,x) & \\ - \phi\left(\phi^{*'}\big(\eta(\omega,\theta) + \lambda(\omega,\theta)'a(\omega,x,\theta)\big)\right) f(\omega,x) & \\ \geqslant \pi_\theta(\omega,x)\big(\eta(\omega,\theta) + \lambda(\omega,\theta)'a(\omega,x,\theta)\big) - \phi\Big(\frac{\pi_\theta(\omega,x)}{f(\omega,x)}\Big)f(\omega,x) & \end{aligned}$$

where we have used the fact that $(\phi')^{-1} = \phi^{*'}$ shown in Lemma 1(v). Integrating over $\Omega \times \mathbb{R}^n$, using Equation (8) and feasibility of the probability density π_θ then gives:

$$\begin{aligned} \mathcal{D}_\phi(g^*,f) &= \int_\Omega \int_{\mathbb{R}^m} \phi\left(\phi^{*'}\big(\eta(\omega,\theta) + \lambda(\omega,\theta)'a(\omega,x,\theta)\big)\right) f(\omega,x)dxdP(\omega) \\ &\leqslant \int_\Omega \int_{\mathbb{R}^n} \phi\Big(\frac{\pi_\theta(\omega,x)}{f(\omega,x)}\Big)f(\omega,x)dxdP(\omega) = \mathcal{D}_\phi(\pi_\theta,f), \end{aligned}$$

so g^* is optimal.

STEP 5: In addition, the mapping $\theta \mapsto \tau(\omega, \theta)$ is continuously differentiable on $\mathcal{B}(\theta_0, \varepsilon)$ and when $\theta = \theta_0$ we have:

$$D_\theta \tau(\omega, \theta_0) = D_\theta \tilde{F}(\omega, \theta_0, \tau(\omega, \theta_0)) \big[D_\tau \tilde{F}(\omega, \theta_0, \tau(\omega, \theta_0)) \big]^{-1}, \text{ for a.e. } \omega$$

In particular, for a.e. ω we have:

$$\begin{aligned} D_\tau \tilde{F}(\omega, \theta_0, \tau(\omega, \theta_0)) &= \begin{pmatrix} 1 & 0 \\ 0 & E\big[a(X, \theta_0) a(X, \theta_0)' | \mathcal{G}\big] \end{pmatrix} \\ D_\theta \tilde{F}(\omega, \theta_0, \tau(\omega, \theta_0)) &= \begin{pmatrix} 0 & E[D_\theta a(X, \theta_0) | \mathcal{G}] \end{pmatrix} \end{aligned}$$

which shows that

$$D_\theta \tau(\omega, \theta_0) = \begin{pmatrix} 0 \\ E[D_\theta a(X, \theta_0) | \mathcal{G}] \big\{ E\big[a(X, \theta_0) a(X, \theta_0)' | \mathcal{G}\big] \big\}^{-1} \end{pmatrix} \text{ with probability 1.}$$

This completes the proof of Theorem 2 and its Corollary 3. □

References

ALI, S. M. AND S. D. SILVEY (1966): "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society Ser. B*, 28, 131–142.

ALIPRANTIS, C. D. AND K. C. BORDER (2007): *Infinite Dimensional Analysis*, Berlin: Springer-Verlag, 3rd ed.

BERNARDO, J. (1979): "Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41, 113–147.

——— (2005): "Reference analysis," *Handbook of Statistics*, 25, 17–90.

Borwein, J. M. and A. S. Lewis (1991): "Duality Relationships for Entropy-Like Minimization Problems," *SIAM Journal on Control and Optimization*, 29, 325–338.

——— (1993): "Partially-finite Programming in L_1 and the Existence of Maximum Entropy Estimates," *SIAM Journal on Optimization*, 3, 248–267.

Borwein, J. M. and M. A. Limber (1996): "On Entropy Maximization via Convex Programming," Mimeo, Centre for Experimental and Construtive Mathematics, Department of Mathematics and Statistics, Simon Fraser University.

Buck, B. and V. Macaulay (1991): *Maximum entropy in action: a collection of expository essays*, Oxford University Press.

Chor-Yiu, S. and H. White (1996): "Information criteria for selecting possibly misspecified parametric models," *Journal of Econometrics*, 71, 207–225.

Csiszár, I. (1967): "Information-type measures of the difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, 299–318.

——— (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–58.

——— (1995): "Generalized projections for non-negative functions," *Acta Mathematica Hungarica*, 68, 161–186.

Doob, J. (1984): *Classical Potential Theory and Its Probabilistic Counterpart*, Springer.

HIRIART-URRUTY, J.-B. AND C. LEMARECHAL (1993): *Convex Analysis and Minimization Algorithms I: Fundamentals (Grundlehren Der Mathematischen Wissenschaften)*, Springer.

KIM, J. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107, 175–193.

KITAMURA, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, 1661–1672.

KITAMURA, Y. AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

——— (2002): "Connections between entropic and linear projections in asset pricing estimation," *Journal of Econometrics*, 107, 159–174.

KITAMURA, Y., G. TRIPATHI, AND H. AHN (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667–1714.

KOMUNJER, I. AND Q. VUONG (2009): "Semiparametric Efficiency Bound in Time-Series Models for Conditional Quantiles," *Econometric Theory*, forthcoming.

KULLBACK, S. AND M. A. KHAIRAT (1966): "A Note on Minimum Discrimination Information," *The Annals of Mathematical Statistics*, 37, 279–280.

KULLBACK, S. AND R. LEIBLER (1951): "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86.

LIESE, F. (1975): "On the existence of $f-$projections," *Colloq. Math. Soc. J. Bolyai*, 16, 431–446, budapest.

Otsu, T., M. H. Seo, and Y.-J. Whang (2008): "Testing for Non-Nested Conditional Moment Restrictions using Unconditional Empirical Likelihood," Cowles Foundation Discussion Paper No. 1660.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press.

Rudin, W. (1976): *Principles of Mathematical Analysis*, McGraw-Hill.

Sawa, T. (1978): "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46, 1273–1291.

Shiryaev, A. (1996): *Probability*, New York: Springer-Verlag, second ed.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, vol. 1, 187–196.

Tang, Y. and S. Ghosal (2007): "Posterior Consistency of Dirichlet Mixtures for Estimating a Transition Density," *Journal of Statistical Planning and Inference*, 137, 1711–1726.

Teboulle, M. and I. Vajda (1993): "Convergence of Best ϕ-Entropy Estimates," *IEEE Transactions on Information Theory*, 39, 297–301.

Tripathi, G. and Y. Kitamura (2003): "Testing conditional moment restrictions," *Annals of Statistics*, 31, 2059–2095.

Vuong, Q. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

——— (1994): *Estimation, Inference and Specification Analysis*, Cambridge University Press.

ZELLNER, A. (1996): "Models, prior information, and Bayesian analysis," *Journal of Econometrics*, 75, 51–68.

——— (2002): "Information processing and Bayesian analysis," *Journal of Econometrics*, 107, 41–50.

——— (2003): "Some Recent Developments in Econometric Inference," *Econometric Reviews*, 22, 203–215.

ZELLNER, A. AND J. TOBIAS (2001): "Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model," *International Economic Review*, 42, 121–139.