Identification and information in models parametrized by a Banach space

Telmo J. Pérez-Izquierdo*

November 12, 2022

Abstract

This paper generalizes Escanciano (2021)'s identification result and Van Der Vaart (1991)'s information result to models parametrized by a Banach space. It provides a general framework to study identification and its regularity (regular if the Fisher information for the parameter is positive, irregular if it is zero) for a wider range of problems. In particular, this paper argues that non-parametric estimation of (a) the mean of a known transformation and (b) the density at a point naturally give raise to a Banach tangent space. Using the extended framework, the paper shows that Fisher Information for estimation of the mean of a transformation is zero if the transformation is not square-integrable. Information for estimation of the density at a point is zero if one considers an absolutely continuous (Lebesgue) density.

Keywords: Semiparametric identification; Regular identification; Fisher Information; Differentiability.

JEL classification: C13; C14; C18.

1 Introduction

It is well known that, even if a model is not identified, certain aspects of it may still be recovered from the data. Escanciano (2021) coins this as semiparametric identification. Moreover, Escanciano (2021) studies semiparametric identification for models generated by a (possibly infinite dimensional) parameter living in a Hilbert space. For such models, Van Der Vaart (1991) establishes a condition for positive Fisher information. These results provide a systematic approach to study identification and its regularity (regular if information is positive, irregular if it is zero) in models parametrized by a Hilbert space.

This framework, however, cannot be applied to certain problems. Consider the following two problems: (a) for a known transformation g, estimate

*Universidad Carlos III de Madrid. Email address: teperezi@eco.ucm3.es.

$E_0[g(X)]$; and (b) estimate the density of the random element X at a fixed point. A non-parametric treatment of these two problems usually refers to considering that the tangent set (that is, the space of all deviations from the "true" distribution of X, denoted p_0) spans any square-integrable function. In short, that the tangent space is $L_2(p_0)$, a Hilbert space.

To apply their results, both Escanciano (2021) and Van Der Vaart (1991) require the object of interest ($E[g(X)]$ and $p(x)$ in the above examples) to vary continuously with the deviations in the tangent space. However, when the tangent space is $L_2(p_0)$, (a) $E[g(X)]$ is a continuous functional if and only if $g \in L_2(p_0)$ (finite variance); and (b) $p(x)$ is not continuous. Therefore, the present framework needs to be extended to accommodate the above problems.

This paper generelizes Escanciano (2021)'s identification result and Van Der Vaart (1991)'s information result to models parametrized by a Banach space (Theorems 2.1 and 3.1, respectively). This extension allows to cover the two examples mentioned above. Indeed, this work argues that both problems are naturally parametrized by a Banach space, rather than by the Hilbert space of square-integrable functions.

I moreover show that the linear model studied here fits Van Der Vaart (1991)'s framework. This means that regular identification (identification and positive information) is equivalent to a parameter being identified and differentiable. Thus, following Theorem 2.1 in Van Der Vaart (1991), one has that regular identification is a necessary condition for the existence of a regular estimator (see Chamberlain, 1986, for an ealy reference in the issue).

This paper introduces a natural way to parametrize a probability model when the goal is to estimate $E_0[g(X)]$. The natural model contains every probability distribution for which $E[g(X)]$ is finite. The tangent space for this model is a Banach space. Using the extended framework, I show that $E_0[g(X)]$ is irregularly identified if $g \in L_q(p_0)$, $1 < q < 2$ (this result formalizes the discussion in Section 2.1 in Khan and Tamer, 2010).

I also argue that, when it comes to estimate the density of a random element X, a Banach space is the natural place to look at. In this case, letting μ denote the dominating measure, the model contains every density (w.r.t. μ) whose tails tend to zero. This paper provides the first proof, up to my knowledge, of the following fact: the Fisher information of $p_0(x)$ is zero if and only if $\mu(\{x\}) = 0$. Thus, estimation an absolutely continuous density at any point is an irregular identification problem.

This paper is organized as follows. Section 2 presents the identification result. The information result is derived in Section 3. Section 4 shows that the present setup fits the hypothesis in Van Der Vaart (1991). The following sections apply the results to the two problems descrived above: Section 5 treats estimation of the mean of a known transformation and Section 6 studies estimation of a density. The appendices collect several mathematical results that are used throughout the paper. No claim of originality is made for the results in the appendices.

2 Identification

Consider a probability model $\mathcal{P} \equiv \{P_\lambda : \lambda \in \Lambda\}$, where $\Lambda \subseteq V$, being V a normed vector space. The true distribution of the data X is $P_0 \in \mathcal{P}$. We assume that there is one parameter $\lambda_0 \in \Lambda$ that generated the data, i.e., $P_0 = P_{\lambda_0}$ (well-specified model). The goal is to identify and estimate the parameter $\beta_0 \equiv \phi(\lambda_0)$ for a functional $\phi : \Lambda \to \mathbb{R}$.

Definition 1. The parameter β_0 is **identified** if $P_\lambda = P_{\lambda_0} \Rightarrow \phi(\lambda) = \beta_0$ for every $\lambda \in \Lambda$.

Define the set of deviations from λ_0 as $B_0 \equiv \{b \in V : \lambda_0 + b \in \Lambda\}$. From now on, assume that the model $\mathcal{P}$ is dominated by a σ-finite measure μ, so that for each $\lambda \in \Lambda$, there is a density p_λ w.r.t. μ. The Score Operator $S : B_0 \to L_2(p_0)$ is defined as

$$Sb \equiv \frac{p_{\lambda_0+b} - p_{\lambda_0}}{p_{\lambda_0}}. \tag{2.1}$$

Let's restrict to the case where S is a linear map. In that case, it can be uniquely extended to $\langle B_0 \rangle$, the linear span of B_0. Moreover, we assume continuity.

Assumption 2.1. The Score Operator S is linear and continuous.

We do also restrict our attention to affine functionals ϕ. In that case, the "derivative" functional $\dot{\phi} : B_0 \to \mathbb{R}$, defined by $\dot{\phi}(b) \equiv \phi(\lambda_0 + b) - \phi(\lambda_0)$, is linear. Let's assume it is also continuous:

Assumption 2.2. $\dot{\phi}$ is linear and continuous.

Before giving the theorem, we need to introduce some notation and concepts from Banach space theory. The range of an operator between two vector spaces, $T : V \to U$, is the set $\mathcal{R}(T) \equiv \{T(v) : v \in V\} \subseteq U$. The kernel of the operator is the set $\mathcal{N}(T) \equiv \{v \in V : Tv = 0\}$. The closure $\overline{A}$ of a set A in a normed vector space is the smallest closed set, according to the norm topology, that contains A. The (normed) dual of a normed space V, denoted V^*, is the set of all continuous linear functionals $T : V \to \mathbb{R}$. For a linear and continuous operator between normed spaces $T : V \to U$, its adjoint $T^* : U^* \to V^*$ is defined by $\langle v, T^* u^* \rangle = \langle Tv, u^* \rangle$ for every $v \in V$, $u^* \in U^*$.[1]

[1] The above notation is usual in Banach space theory. For a linear operator between vector spaces $T : V \to U$, $\langle v, T \rangle \equiv Tv$. This notation avoids the confusing $T^* u^* v = u^* T v$, where the operations in the LHS are performed in a different order than in the RHS.

For a set $A \subseteq V$, its orthogonal or annihilator $A^\perp$ is a subset of the dual space V^*, defined by $A^\perp \equiv \{v^* \in V^* : \langle a, v^* \rangle = 0, \forall a \in A\}$. The orthogonal set of a subset of the dual space, $B \subseteq V^*$, is the set of vectors of V satisfying ${}^\perp B \equiv \{v \in V : \langle v, b^* \rangle = 0, \forall b^* \in B\}$.

The canonical embedding map from V to V^{**}, its second dual, denoted $Q : V \to V^{**}$, is defined by $\langle v^*, Qv \rangle = \langle v, v^* \rangle$ for every $v \in V$, $v^* \in V^*$. The weak* topology on V^* is the smallest topology that makes all the functions in $\mathcal{R}(Q)$ continuous. For a set $B \subseteq V^*$, its weak*-closure $\overline{B}^{w*}$ is the smallest closed set, according to the weak* topology, that contains B.

We also discuss here Assumption 1(iii) in Escanciano (2021). Proposition 4.1 in Escanciano (2021) shows that identification is equivalent to $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$. Sufficiency of $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$ can be proved under Assumptions 2.1 and 2.2. For necessity, Escanciano (2021) requires that for every $b \in \mathcal{N}(S)$, there exists a $\tau \neq 0$ such that $\lambda_0 + \tau b \in \Lambda$. We replace this assumption by the slightly stronger:

Assumption 2.3. The point $0 \in B_0 \subseteq \langle B_0 \rangle$ is an interior point of B_0 in the relative topology that $\langle B_0 \rangle$ inherits from the norm topology of V.

Note that the relative topology that $\langle B_0 \rangle$ inherits from the norm topology of V is the same as the topology that the restriction of the norm of V to elements in $\langle B_0 \rangle$ induces (see Lemma B.6). What Assumption 2.3 claims is that λ_0 is an interior point of Λ, when Λ is understood as a subset of the affine variety $\lambda_0 + \langle B_0 \rangle$. This is milder that requiring that λ_0 is in the interior of Λ relative to the norm topology of V.

Assumption 2.3 is satisfied if B_0 is a linear subspace of V, since $B_0 = \langle B_0 \rangle$ is then open for every topology of $\langle B_0 \rangle$ (note that, in this case, the interior of B_0 with respect to the norm topology of V is empty). Moreover, if Λ is already a linear subspace, then $B_0 = \Lambda$ and the assumption is satisfied (see Lemma B.8). The above assumption implies Assumption 1(iii) in Escanciano (2021), as we now show. We want to highlight here that Assumption 2.3 plays a more important role in Section 4, where it is necessary to guarantee that $\sqrt{p_\lambda}$ is well-defined in a neighborhood of λ_0.

Lemma 2.1. *Assumption 2.3 implies that for every $b \in \mathcal{N}(S)$, there exists a $\tau \neq 0$ such that $\lambda_0 + \tau b \in \Lambda$. The reverse implication, however, does not hold.*

Proof. We have that $0 \in B_0^\circ$, where the interior is with respect to the norm topology in $\langle B_0 \rangle$. An open neighborhood of 0 is absorbing (Megginson, 1998, Th. 2.2.9(f)). Therefore, for $b \in \mathcal{N}(S) \subseteq \langle B_0 \rangle$, there exists a $\tau_b > 0$ such that $tb \in B_0^\circ$ for $|t| \leq \tau_b$. This means that $\tau_b b \in B_0^\circ \subseteq B_0$, and thus $\lambda_0 + \tau_b b \in \Lambda$.

We provide a counterexample for the reverse implication. Let μ be the counting measure for $\mathbb{N}$ and p_0 be degenerated at 1, i.e., $p_0(1) = 1$ and

$p_0(m) = 0$ if $m \neq 1$. This means that $L_2(p_0) = \mathbb{R}$ (there is an isometric isomorphism between $L_2(p_0)$ and the one-dimensional Euclidean space). Let $V = \mathbb{R}^2$, $\Lambda = \{(x, y) \in \mathbb{R}^2 \colon 0 \leq x \leq 1, 0 \leq y \leq 1\}$, and $\lambda_0 = (0, 0)$. That is, $B_0 = \Lambda$. Consider $S(x, y) = y - x$, so that $\mathcal{N}(S) = \{(x, y) \in \mathbb{R}^2 \colon x = y\}$.

Let $(x, x) \in \mathcal{N}(S)$. If $x = 0$, $(0, 0) \in \Lambda$, so we may take any τ. If $x \neq 0$, let $\tau = x^{-1}$, so that $\tau(x, x) = (1, 1) \in \Lambda$. This shows that the conclusion of the lemma is met. However, Assumption 2.3 is not met in this case. Since $(0, 1), (1, 0) \in B_0$, we have that $\langle B_0 \rangle = \mathbb{R}^2$. We therefore work with the Euclidean topology in $\mathbb{R}^2$. For this topology, $(0, 0) \notin B_0^\circ$, as any ball centered around $(0, 0)$ contains points with strictly negative coordinates.

$\square$

Assumption 2.2 says that $\dot{\phi} \in \langle B \rangle^*$. That is, the functional can be understood as an element of the dual. Identification of β_0 depends on the position of the functional on this dual:

Theorem 2.1. *Consider Assumptions 2.1, 2.2, and 2.3. Then, β_0 is identified if and only if $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$.*

Proof. The proof closely follows Escanciano (2021). Lemma 2.1 guarantees that we can apply Proposition 4.1 in that paper, which easily extends to our setup. The result reads: Identification $\iff \mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$.

The first step is to proof that $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi}) \iff \mathcal{N}(\dot{\phi})^\perp \subseteq \mathcal{N}(S)^\perp$. For a normed space V, consider the sets $A, B \subseteq V$ and $A^*, B^* \subseteq V^*$. It is a basic fact that $A \subseteq B \Rightarrow B^\perp \subseteq A^\perp$ and $A^* \subseteq B^* \Rightarrow {}^\perp B^* \subseteq {}^\perp A^*$. Therefore,

$$\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi}) \Rightarrow \mathcal{N}(\dot{\phi})^\perp \subseteq \mathcal{N}(S)^\perp, \text{ and}$$
$$\mathcal{N}(\dot{\phi})^\perp \subseteq \mathcal{N}(S)^\perp \Rightarrow {}^\perp(\mathcal{N}(S)^\perp) \subseteq {}^\perp(\mathcal{N}(\dot{\phi})^\perp).$$

Since both $\mathcal{N}(S)$ and $\mathcal{N}(\dot{\phi})$ are closed subspaces of $\langle B_0 \rangle$, Proposition 1.10.15 in Megginson (1998) gives $\mathcal{N}(S) = {}^\perp(\mathcal{N}(S)^\perp)$ and $\mathcal{N}(\dot{\phi}) = {}^\perp(\mathcal{N}(\dot{\phi})^\perp)$. This concludes the step.

That $\mathcal{N}(S)^\perp = \overline{\mathcal{R}(S^*)}^{w*}$ follows from Lemma 3.1.16 and Proposition 2.6.6 in Megginson (1998). The lemma claims that $\mathcal{N}(S) = {}^\perp\mathcal{R}(S^*)$, so $\mathcal{N}(S)^\perp = ({}^\perp\mathcal{R}(S^*))^\perp$. Since $\mathcal{R}(\mathcal{S}^*)$ is a subspace, the proposition gives $({}^\perp\mathcal{R}(S^*))^\perp = \overline{\mathcal{R}(S^*)}^{w*}$.

We conclude by showing that $\mathcal{N}(\dot{\phi})^\perp \subseteq \overline{\mathcal{R}(S^*)}^{w*} \iff \dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$. First, note that $\mathcal{N}(\dot{\phi}) = {}^\perp\langle\{\dot{\phi}\}\rangle$, where $\langle\{\dot{\phi}\}\rangle \equiv \{\alpha\dot{\phi} : \alpha \in \mathbb{R}\}$ is the linear span of $\dot{\phi}$. Indeed,

$$b \in \mathcal{N}(\dot{\phi}) \Leftrightarrow \langle b, \dot{\phi} \rangle = 0 \Leftrightarrow \langle b, \alpha\dot{\phi} \rangle = 0, \forall \alpha \in \mathbb{R} \Leftrightarrow b \in {}^\perp\langle\{\dot{\phi}\}\rangle.$$

Thus, $\mathcal{N}(\dot{\phi})^\perp = ({}^\perp\langle\{\dot{\phi}\}\rangle)^\perp = \overline{\langle\{\dot{\phi}\}\rangle}^{w*}$ by Proposition 2.2.6 in Megginson (1998). This shows that $\mathcal{N}(\dot{\phi})^\perp$ is the smallest weak*-closed subspace containing $\dot{\phi}$. So, if $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$, then it must be that $\mathcal{N}(\dot{\phi})^\perp \subseteq \overline{\mathcal{R}(S^*)}^{w*}$, since

$\overline{\mathcal{R}(S^*)}^{w*}$ is a weak*-closed subspace containing $\dot{\phi}$. The other implication is trivial. $\square$

In general, we have that $\overline{\mathcal{R}(S^*)} \subseteq \overline{\mathcal{R}(S^*)}^{w*}$, since the weak* topology is a subtopology of the norm topology (Megginson, 1998, Th. 2.6.2). That is, $\overline{\mathcal{R}(S^*)}^{w*}$ is a closed set containing $\mathcal{R}(S^*)$. Both sets are equal for reflexive spaces:

Corollary 2.1. *Under the assumptions of Thereom 2.1, if $\overline{\langle B_0 \rangle}$ is reflexive, then β_0 is identified if and only if $\dot{\phi} \in \overline{\mathcal{R}(S^*)}$.*

Proof. If $\overline{\langle B_0 \rangle}$ is a reflexive space, the weak and weak* topologies in $\overline{\langle B_0 \rangle}^*$ are equal (Megginson, 1998, Th. 2.6.2). Moreover, since $\mathcal{R}(S^*)$ is a subspace, its weak closure equals its norm closure (Megginson, 1998, Cor. 2.5.17). That is, $\overline{\mathcal{R}(S^*)}^{w*} = \overline{\mathcal{R}(S^*)}$. $\square$

The above result generalizes the identification part in Theorem 4.1 from Escanciano (2021). Indeed, since S and $\dot{\phi}$ continous, they can be uniquely extended to $\overline{\langle B_0 \rangle}$ (Megginson, 1998, Th. 1.9.1). The result in Escanciano (2021) is stated for S and $\dot{\phi}$ defined in $\overline{\langle B_0 \rangle}$. So $\overline{\langle B_0 \rangle}$ is a Hilbert space, which are reflexive spaces.

3 Fisher Information

We now introduce the semiparametric Fisher information for $\beta_0 \equiv \phi(\lambda_0)$:

$$I_\phi \equiv \inf_{b \in \langle B_0 \rangle} \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}(b)|^2}.$$

We note that the extensions of S and $\dot{\phi}$ to $\overline{\langle B_0 \rangle}$ are continuous, hence so is $I_\phi(b) \equiv \|Sb\|^2_{L_2(p_0)} / |\dot{\phi}(b)|^2$ (except at the points where $\dot{\phi}(b) = 0$, which are not relevant for the infimum). Lemma B.1 then gives

$$I_\phi = \inf_{b \in \overline{\langle B_0 \rangle}} \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}(b)|^2}.$$

Therefore, for the results in this section, both S and $\dot{\phi}$ are considered functionals on $\overline{\langle B_0 \rangle}$. We introduce first the following definitions:

Definition 2. The parameter β_0 is **regularly identified** if it is identified and $I_\phi > 0$. It is **irregularly identified** if it is identified and $I_\phi = 0$.

Theorem 4.1 in Van Der Vaart (1991) shows that, for models parametrized by a Hilbert space, positive information is equivalent to $\dot{\phi} \in \mathcal{R}(S^*)$. This section generalizes the result for arbitrary normed spaces. We do this in several steps. We note that our proof greatly differs from that of Van Der Vaart (1991), which heavily relies on Hilbert space technology.

That $\dot{\phi}$ in the range of the adjoint Score operator implies positive information is immediate:

Proposition 3.1. *Under Assumptions 2.1 and 2.2,* $\dot{\phi} \in \mathcal{R}(S^*) \Rightarrow I_\phi > 0$.

Proof. If $\dot{\phi} \in \mathcal{R}(S^*)$, then there exists an $r^* \in L_2(p_0)^*$ such that $\dot{\phi} = S^* r^*$. Hence, $\dot{\phi}(b) = \langle b, \dot{\phi} \rangle = \langle b, S^* r^* \rangle = \langle Sb, r^* \rangle$. Thus, since r^* is bounded,

$$|\dot{\phi}(b)| = |r^*(Sb)| \leq \|r^*\|_{L_2(p_0)^*} \|Sb\|_{L_2(p_0)} .$$

That is, $I_\phi^{-1} \leq \|r^*\|^2_{L_2(p_0)^*} < \infty$. □

Moreover, if β_0 is not identified, there must be a problematic direction $b \in \overline{\langle B_0 \rangle}$. Through this problematic direction, there is no information for β_0. This means that information is zero for non-identified parameters:

Proposition 3.2. *Under Assumptions 2.1 and 2.2,* $\dot{\phi} \notin \overline{\mathcal{R}(S^*)}^{w*} \Rightarrow I_\phi = 0$.

Proof. From the proof of Theorem 2.1, $\mathcal{N}(S)^\perp = \overline{\mathcal{R}(S^*)}^{w*}$, which means that $\dot{\phi} \notin \mathcal{N}(S)^\perp$. By definition of the orthogonal set, this says that there exists at least one $b \in \mathcal{N}(S)$ such that $\langle b, \dot{\phi} \rangle \neq 0$. For this direction b,

$$I_\phi(b) \equiv \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}(b)|^2} = 0.$$

□

Proposition 3.2 and Theorem 2.1 imply that if information is positive, then the parameter must be identified. The above result, however, is not sufficient to claim that positive information is equivalent to $\dot{\phi} \in \mathcal{R}(S^*)$. The remainder of the section is devoted to show this implication (that is, the converse of Proposition 3.1). Before proving the result, we first discuss the two key ideas that lead to it.

Suppose first that S is one-to-one, i.e., that $\mathcal{N}(S) = \emptyset$. This means that every linear and continuous functional is identified (Prop. 4.1 in Escanciano, 2021, claims that identification is equivalent to $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$). By Theorem 2.1, we know that $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$. We do not know, however, whether it is regularly identified, $I_\phi > 0$, or irregularly identified, $I_\phi = 0$ and thus $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$ (see Proposition 3.1). It turns out that, in this case, we can characterize both scenarios.

Since S is one-to-one and onto $\mathcal{R}(S) \subseteq L_2(p_0)$, the inverse $S^{-1} : \mathcal{R}(S) \to \overline{\langle B_0 \rangle}$ is guaranteed to exists. Note that this inverse in linear (Lemma B.5). It turns out that positive information is intimatelly related to continuity of $\dot{\phi} \circ S^{-1} \colon \mathcal{R}(S) \to \mathbb{R}$:

Proposition 3.3. *Let Assumptions 2.1 and 2.2 hold. Assume further that S is one-to-one. Then, $I_\phi > 0$ if and only if $\dot{\phi} \circ S^{-1}$ is continuous.*

Proof. We first note that, since $\dot{\phi}$ is linear, so is $\dot{\phi} \circ S^{-1}$. Moreover, for every $b \in \overline{\langle B_0 \rangle}$, there exists and $r \in \mathcal{R}(S)$ such that $b = S^{-1}r$. Hence, information in direction b is given by

$$I_\phi(b) = \frac{\|r\|^2_{L_2(p_0)}}{|\dot{\phi} \circ S^{-1} r|^2}.$$

Consider that $I_\phi > 0$. Then, for every $r \in \mathcal{R}(S)$, taking $b = S^{-1}r$

$$I_\phi(b) \geq I_\phi \Rightarrow \frac{\|r\|^2_{L_2(p_0)}}{|\dot{\phi} \circ S^{-1} r|^2} \geq I_\phi \Rightarrow |\dot{\phi} \circ S^{-1} r| \leq \frac{1}{\sqrt{I_\phi}} \|r\|_{L_2(p_0)} .$$

Therefore, $\dot{\phi} \circ S^{-1}$ is continuous.

Equivalently, if $\dot{\phi} \circ S^{-1}$ is continuous, there exists a $C < \infty$ such that $|\dot{\phi} \circ S^{-1} r| \leq C \|r\|_{L_2(p_0)}$. Then, for every $b \in \overline{\langle B_0 \rangle}$, taking $r = Sb$,

$$|\dot{\phi}(b)| \leq C \|Sb\|_{L_2(p_0)} \Rightarrow I_\phi(b) = \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}(b)|^2} \geq \frac{1}{C^2},$$

and thus $I_\phi \geq 1/C^2 > 0$. □

This insight is key for several results to follow. It is what guarantees that, if $\mathcal{R}(S)$ is closed, identification is equivalent to positive information (see Section 3.1). Moreover, we will see that continuity of $\dot{\phi} \circ S^{-1}$ implies $\dot{\phi} \in \mathcal{R}(S^*)$. Then, that positive information leads to $\dot{\phi} \in \mathcal{R}(S^*)$ will follow.

What remains is to deal with the possibility of S not being one-to-one. However, if β_0 is identified, we can work with a suitable notion of "inverse" of S. Identification of β_0 means that S is "sufficiently invertible", in the sense that we can assign a unique value of the parameter $\dot{\phi}$ for each $r \in \mathcal{R}(S)$, even if S is not one-to-one. The following construction formalizes this fact.

Define the quotient space $Q \equiv \overline{\langle B_0 \rangle}/\mathcal{N}(S)$ containing all equivalence classes $[b] = b + \mathcal{N}(S) = \{b + n \colon n \in \mathcal{N}(S)\}$. Elements in Q are sets containing indistinguishable directions: if $b' \in [b]$, then $b' = b + n$, with $n \in \mathcal{N}(S)$. That is, $S(b' - b) = Sn = 0$, so $Sb = Sb'$. This set can be turned into a normed space by defining:

- $[b] + [b'] \equiv (b + b') + \mathcal{N}(S) = [b + b']$.
- $\alpha[b] \equiv (\alpha b) + \mathcal{N}(S) = [\alpha b]$.
- $\|[b]\|_Q \equiv \inf_{n \in \mathcal{N}(S)} \|b + n\|_{\overline{\langle B_0 \rangle}}$, which is a norm since $\mathcal{N}(S)$ is closed (Megginson, 1998, Th. 1.7.4).

We define the following operators in the quotien space: $\tilde{S}\colon Q \to L_2(p_0)$, given by $\tilde{S}[b] = Sb$; and $\tilde{\dot{\phi}}\colon Q \to \mathbb{R}$, given by $\tilde{\dot{\phi}}[b] = \dot{\phi}(b)$. The next lemma ensures that they are well defined and satisfy the expected properties.

Lemma 3.1. *Let the assumptions of Theorem 2.1 hold. Assume further that β_0 is identified. Then,*

1. *$\tilde{S}$ is well-defined, linear, continuous, and one-to-one. Moreover, $\tilde{S}$ is onto $\mathcal{R}(S) \subseteq L_2(p_0)$.*

2. *$\tilde{\dot{\phi}}$ is well-defined, linear, and continuous.*

Proof. For $\tilde{S}$:

- Well-defined: For $b' \in [b]$, $b' \neq b$, we have that $b' - b \in \mathcal{N}(S)$ and thus $Sb = Sb'$.

- Linear: $\tilde{S}(\alpha[b] + \beta[b']) = \tilde{S}[\alpha b + \beta b'] = S(\alpha b + \beta b') = \alpha Sb + \beta Sb' = \alpha S[b] + \beta S[b']$.

- Continuous: Let $\pi\colon \overline{\langle B_0 \rangle} \to Q$ be the quotient map $\pi(b) = [b]$. Consider an open set $U \subseteq L_2(p_0)$. We have that $\pi(S^{-1}(U)) \subseteq \tilde{S}^{-1}(U)$. Indeed, if $[b] \in \pi(S^{-1}(U))$, then $b \in S^{-1}(U)$ and thus $Sb \in U$. This implies that $\tilde{S}[b] \in U$ so $[b] \in \tilde{S}^{-1}(U)$. Moreover, by continuity of S, $S^{-1}(U)$ is open. This, on top of Lemma 1.7.11 in Megginson (1998), means that, $\forall [b] \in \tilde{S}^{-1}(U)$, we can find an open ball centered at $[b]$ and completely contained in $\tilde{S}^{-1}(U)$.

- One-to-one: If $\tilde{S}[b] = \tilde{S}[b']$, then $Sb = Sb'$, so $b - b' \in \mathcal{N}(S)$. Therefore, $[b] = [b']$.

- Onto $\mathcal{R}(S) \subseteq L_2(p_0)$: For $r \in \mathcal{R}(S)$, there exists a $b \in \overline{\langle B_0 \rangle}$ such that $Sb = r$. Then, $\tilde{S}[b] = r$.

For $\tilde{\dot{\phi}}$:

- Well-defined: For $b' \in [b]$, $b' \neq b$, we had that $b' - b \in \mathcal{N}(S)$. Since β_0 is identified, Proposition 4.1 in Escanciano (2021) gives that $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$ and thus $b' - b \in \mathcal{N}(\dot{\phi})$. That is, $\dot{\phi}(b) = \dot{\phi}(b')$.

- Linearity and continuity follow as above.

$\square$

The preceding discussion shows that, when β_0 is identified, if there are two indistingushable directions $b, b' \in \overline{\langle B_0 \rangle}$, then $\dot{\phi} b = \dot{\phi} b'$. Not surprisingly, information in both directions is also the same. We can thus define

$$\tilde{I}_\phi([b]) \equiv \frac{\left\| \tilde{S}[b] \right\|^2_{L_2(p_0)}}{\left| \tilde{\dot{\phi}}[b] \right|^2},$$

and claim that $I_\phi(b) = \tilde{I}_\phi([b])$ for every $b \in \overline{\langle B_0 \rangle}$. Moreover, we have the following lemma:

Lemma 3.2. *Let the assumptions of Theorem 2.1 hold. Assume further that β_0 is identified. Then,*

$$I_\phi = \inf_{[b] \in Q} \tilde{I}_\phi([b]) \equiv \tilde{I}_\phi.$$

Proof. For every $[b] \in Q$, $\tilde{I}_\phi([b]) = I_\phi(b) \geq I_\phi$, since I_ϕ is the infimum. Therefore, I_ϕ is a lower bound for $\tilde{I}_\phi([b])$ and thus $\tilde{I}_\phi \leq I_\phi$. Likewise, for every $b \in \overline{\langle B_0 \rangle}$, $I_\phi(b) = \tilde{I}_\phi([b]) \geq \tilde{I}_\phi$. So $\tilde{I}_\phi \geq I_\phi$. □

With these results, we are ready to state the main theorem of the section. The proof strategy is to apply Proposition 3.3 to $\tilde{S}$ and $\tilde{\dot{\phi}}$. Then, we show that $\tilde{\dot{\phi}} \circ \tilde{S}^{-1}$ being continuous implies $\dot{\phi} \in \mathcal{R}(S^*)$.

Theorem 3.1. *Under the assumptions of Theorem 2.1,*

$$I_\phi > 0 \Rightarrow \dot{\phi} \in \mathcal{R}(S^*).$$

Proof. Combining Proposition 3.2 and Theorem 2.1 gives that β_0 is identified. Thus, Lemma 3.1 guarantees that we can apply Proposition 3.3 to $\tilde{I}_\phi$ and $\tilde{\dot{\phi}} \circ \tilde{S}^{-1} : \mathcal{R}(S) \to \mathbb{R}$. Therefore, since $\tilde{I}_\phi = I_\phi > 0$ (Lemma 3.2), we have that $\tilde{\dot{\phi}} \circ \tilde{S}^{-1}$ is continuous.

By the Hahn-Banach Extension Theorem (Megginson, 1998, Th. 1.9.6), there exists a linear and continuous functional $r^* : L_2(p_0) \to \mathbb{R}$ such that $\langle r, r^* \rangle = \langle r, \tilde{\dot{\phi}} \circ \tilde{S}^{-1} \rangle$ for every $r \in \mathcal{R}(S)$. That is, $r^* \in \underline{L_2(p_0)}^*$. We conclude by showing that $S^* r^* = \dot{\phi}$, so $\dot{\phi} \in \mathcal{R}(S^*)$. For every $b \in \overline{\langle B_0 \rangle}$,

$$\langle b, S^* r^* \rangle = \langle Sb, r^* \rangle = \langle \tilde{S}[b], r^* \rangle = \langle \tilde{S}[b], \tilde{\dot{\phi}} \circ \tilde{S}^{-1} \rangle = \langle [b], \tilde{\dot{\phi}} \rangle = \langle b, \dot{\phi} \rangle.$$

□

We can put together the results in Theorem 2.1, Theorem 3.1 and Proposition 3.1 to conclude that:

Corollary 3.1. *Under the assumptions in Theorem 2.1,*

1. *β_0 is regularly identified iff $\dot{\phi} \in \mathcal{R}(S^*)$,*
2. *β_0 is irregularly identified iff $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$, and*
3. *β_0 is not identified iff $\dot{\phi} \notin \overline{\mathcal{R}(S^*)}^{w*}$.*

Figure 1 gives a graphic depiction of Corollary 3.1:

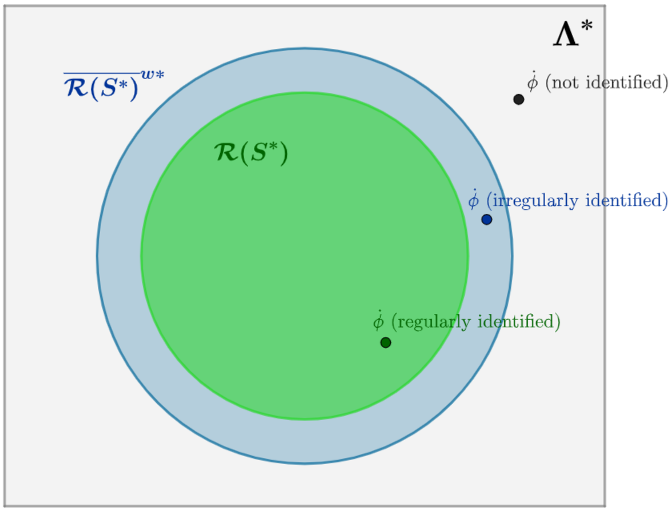

Figure 1: Identification and information for $\beta_0 \equiv \phi(\lambda_0)$ depending on the position of $\dot{\phi}$ in the dual space of $\overline{\langle B_0 \rangle}$.

3.1 Closed range

For models parametrized by a Hilbert space it is well established that, if $\mathcal{R}(S)$ is closed, regular identification is equivalent to identification (see Van Der Vaart, 1991, Coro. 3.2). This result easily extends to models parametrized by a Banach space.

Corollary 3.2. *Under the assumptions of Corollary 3.1, if $\mathcal{R}(S)$ is closed, then β_0 is identified if and only if $I_\phi > 0$.*

Proof. $\mathcal{R}(S)$ is closed if and only if $\mathcal{R}(S^*)$ is weakly* closed (Megginson, 1998, Th. 3.1.21). Therefore, $\overline{\mathcal{R}(S^*)}^{w*} \setminus \mathcal{R}(S^*) = \emptyset$ and only two cases are possible in Corollary 3.1. Case 1, where β_0 is regularly identified; or Case 3, where β_0 is not identified. □

When $\mathcal{R}(S)$ is closed, we can also proof the assertion "identification implies positive information" relying in Proposition 3.3. This enriches the discussion, as it discloses the important implication of having a closed range: that S^{-1} is continuous (assuming S is one-to-one).

Consider a direction $b \in \overline{\langle B_0 \rangle}$ with small numerator in $I_\phi(b)$. That is, $\|Sb\|^2_{L_2(p_0)}$ is small. Then, since S^{-1} is continuous, it must be that $\|b\|_{\overline{\langle B_0 \rangle}}$ is also small. Finally, continuity of $\dot{\phi}$ implies that the denominator of $I_\phi(b)$ is also small. This is enough to keep information for β_0 bounded away from zero. As we did for Theorem 3.1, the fact that β_0 is identified allows us to work with a suitable notion of "inverse" of S, even if the map is not one-to-one.

The following alternative proof fills the gaps:

Alternative proof. We prove that, for V a Banach space, identification implies positive information when $\mathcal{R}(S)$ is closed. When β_0 is identified, Lemma 3.1 guarantees that we can apply Proposition 3.3 to $\tilde{I}_\phi$ and $\tilde{\dot{\phi}} \circ \tilde{S}^{-1}$.

Now, since $\overline{\langle B_0 \rangle}$ is a Banach space, Q is also a Banach space (Megginson, 1998, Th. 1.7.7). Moreover, since $\mathcal{R}(S)$ is closed, it is Banach. Banach's Inverse Mapping Theorem ensures that $\tilde{S}^{-1}$ is continuous (Luenberger, 1997, Th. 1 in Sec. 6.4). Therefore, $\tilde{\dot{\phi}} \circ \tilde{S}^{-1}$ is continuous and, by Proposition 3.3, $\tilde{I}_\phi > 0$. By Lemma 3.2, $I_\phi = \tilde{I}_\phi > 0$.

□

3.2 Estimation of irregularly identified parameters

Suppose that β_0 is irregularly identified. That is, $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$. To estimate $\beta_0 = \phi(\lambda_0)$, it is standard to construct a sequence of functionals $\phi_h : \Lambda \to \mathbb{R}$ such that

1. ϕ_n is regularly identified. That is, $\dot{\phi}_n(b) \equiv \phi_n(\lambda_0 + b) - \phi_n(\lambda_0)$ satisfies $\dot{\phi}_n \in \mathcal{R}(S^*)$.

2. For every $\lambda \in \Lambda$, $\phi_n(\lambda) \to \phi(\lambda)$ as $n \to \infty$.

A first example is estimation of a density at a point. Here Λ is a set of densities w.r.t. a σ-finite measure μ and $\phi(\lambda) \equiv \lambda(x_0)$, for a fixed x_0. Here ϕ is replaced by $\phi_n(\lambda) \equiv n \, \mathrm{E}_\lambda[K(n(x_0 - X))]$. Equivalently, setting $h = 1/n$, one estimates $\phi_h(\lambda) \equiv h^{-1} \, \mathrm{E}_\lambda[K((x_0 - X)/h)]$. This sequence converges to $\phi(\lambda)$ as $h \to 0$ if the kernel K satisfies certain properties.

Khan and Tamer (2010) study estimation of $\phi(\lambda_0) \equiv \mathrm{E}_0[g(X)]$ for $g \notin L_2(p_0)$ (and thus irregularly identified, see Section 5 bellow). They propose to introduce a trimming factor $\tau_n(x)$, with $\tau_n \to 1$ a.s. p_0, so that $\mathrm{E}_0[\tau_n(X)^2 g(X)^2] < \infty$. Here, the functional is replace by the sequence $\phi_n(\lambda) = \mathrm{E}_\lambda[\tau_n(X) g(X)]$.

In general, $\phi_n(\lambda) \to \phi(\lambda)$ for every $\lambda \in \Lambda$ implies $\dot{\phi}_n(b) \to \dot{\phi}(b)$ for every $b \in B_0$. By linearity and continuity of $\dot{\phi}_n$ (i.e., by $\dot{\phi}_n \in \mathcal{R}(S^*)$), convergence holds for every $b \in \overline{\langle B_0 \rangle}$.[2] Thus, this construction leads to $\dot{\phi}_n \xrightarrow{w*} \dot{\phi}$ in $\overline{\langle B_0 \rangle}^*$. It is interesting to realize that, under some circumstances, one can always construct a sequence $\dot{\phi}_n$ that converges weakly* to $\dot{\phi}$.

Since $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$, we can always find a net $\dot{\phi}_\alpha \in \mathcal{R}(S^*)$ such that $\dot{\phi}_\alpha \xrightarrow{w*} \dot{\phi}$ (Megginson, 1998, Prop. 2.1.18). However, the existence of a

[2] For $b \in \overline{\langle B_0 \rangle} \backslash \langle B_0 \rangle$, take $b_m \to b$ as $m \to \infty$. Thus, $\lim_n \dot{\phi}_n(b) \equiv \lim_n \lim_m \dot{\phi}_n(b_m) = \lim_m \lim_n \dot{\phi}_n(b_m) = \lim_m \dot{\phi}(b_m) \equiv \dot{\phi}(b)$, as long as the limits can be interchanged.

sequence converging weakly* to $\dot{\phi}$ is not guaranteed. The sequential closure of a set A is defined as the collection of all limit points of sequences in A. It is known that there are Hausdorff topologies for which a set A can be sequentially closed but not closed (see Exercise 2.3 in Megginson, 1998). For such a set A, its closure contains points that are not the limit of a sequence in A.

It turns out that if V is a separable Banach space, then we are save. In this case, $\overline{\langle B_0 \rangle}$ is also a separable Banach space, so $\overline{\mathcal{R}(S^*)}^{w*}$ is also weakly* sequentially closed (Megginson, 1998, Cor. 2.7.13). That is, there is a sequence $\dot{\phi}_n$ in $\mathcal{R}(S^*)$ that converges weakly* to $\dot{\phi}$. For non-separable spaces, one can require that $\dot{\phi} \in \overline{\mathcal{R}(S^*)} \subseteq \overline{\mathcal{R}(S^*)}^{w*}$. This leads to a stronger result: there exists a sequence $\dot{\phi}_n$ in $\mathcal{R}(S^*)$ such that $\dot{\phi}_n \to \dot{\phi}$ in norm.

We now assume that there exists a sequence $\dot{\phi}_n \in \mathcal{R}(S^*)$ such that $\dot{\phi}_n$ converges weakly* (i.e., pointwise) to $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$. Define

$$I_{\phi_n} \equiv \inf_{b \in \overline{\langle B_0 \rangle}} \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}_n(b)|^2}.$$

Proposition 3.1 states that $I_{\phi_n} > 0$ for every n. However, the fact that β_0 is irregularly identified ($I_\phi = 0$) implies that the information for ϕ_n must tend to zero. Thus, the Cramér-Rao Lower Bound gets larger as n grows.

Proposition 3.4. *Let the assumptions of Theorem 2.1 hold. Consider a sequence $\dot{\phi}_n \in \mathcal{R}(S^*)$ converging weakly* to $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$. Then, $I_{\phi_n} \to 0$ as $n \to \infty$.*

Proof. For $b \in \overline{\langle B_0 \rangle}$, we have that $\dot{\phi}_n(b) \to \dot{\phi}(b)$. Thus,

$$I_{\phi_n}(b) \equiv \frac{\|Sb\|^2_{L_2(p_0)}}{|\dot{\phi}_n(b)|^2} \to I_\phi(b) \text{ as } n \to \infty,$$

as long as $\dot{\phi}(b) \neq 0$. By Theorem 3.1, $\inf_{b \in \overline{\langle B_0 \rangle}} I_\phi(b) = I_\phi = 0$. The conclusion follows from Lemma C.2:

$$\lim_{n \to \infty} \inf_{b \in \overline{\langle B_0 \rangle}} I_{\phi_n}(b) \leq \inf_{b \in \overline{\langle B_0 \rangle}} \lim_{n \to \infty} I_{\phi_n}(b) = 0.$$

□

4 Differentiability

Let $\beta_0 = \phi(\lambda_0)$ be an identified parameter. That is, $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*}$. Identification of λ_0 would mean that the map $\lambda \mapsto p_\lambda$ is invertible. Semiparametric identification of β_0 means that the map is "sufficiently invertible", so that we can assign a unique parameter to each density (see the discussion in Section 3).

Consider a measure $P \in \mathcal{P}$ with density p. We will define a map $\psi : \mathcal{P} \to \mathbb{R}$ by the following procedure. Since $P \in \mathcal{P}$, there exist a $\lambda \in \Lambda$ such that $p = p_\lambda$. Take $\psi(P) = \phi(\lambda)$.

Lemma 4.1. *Under Assumptions 2.1 and 2.2, ψ is a well-defined map if β_0 is identified.*

Proof. Identification implies $\mathcal{N}(S) \subseteq \mathcal{N}(\dot{\phi})$ (Escanciano, 2021, Prop. 4.1, note that Assumption 2.3 is needed only for necessity). We need to show that if, for $P \in \mathcal{P}$, there are two parameters $\lambda, \lambda' \in \Lambda$ such that $p_\lambda = p_{\lambda'} = p$, then it must be that $\phi(\lambda) = \phi(\lambda')$.

Define $b \equiv \lambda - \lambda_0 \in B_0$ and $b' \equiv \lambda' - \lambda_0 \in B_0$. We have that $Sb = Sb'$, so, by linearity, $b - b' \in \mathcal{N}(S)$. Since the parameter is identified, this means that $b - b' \in \mathcal{N}(\dot{\phi})$. Then, by linearity,

$$\dot{\phi}(b) = \dot{\phi}(b') \Rightarrow \phi(\lambda_0 + b) = \phi(\lambda_0 + b') \Rightarrow \phi(\lambda) = \phi(\lambda').$$

$\square$

Lemma 4.1 shows that the parameter of interest is given by a functional $\phi(\lambda) = \psi(P_\lambda)$. It is interesting to analyze whether ψ is a differentiable functional, as this is a necessary condition for root-n estimation of β_0 (Van Der Vaart, 1991, Th. 2.1). Theorem 3.1 in Van Der Vaart (1991), which also applies to Banach spaces, gives the answer. We show that Van Der Vaart (1991)'s result applies to our setup. First, we introduce the relevant definitions:

Definition 3.

- Let U be an arbitrary set. A **path in U** is a map $t \mapsto u_t$ from $(0, \tau) \subset \mathbb{R}$ to U.
- A path in $\mathcal{P}$ is **regular** if there exists an $r \in L_2(p_0)$ such
$$\int \left[\frac{1}{t}(\sqrt{p_t} - \sqrt{p_0}) - \frac{1}{2} r \sqrt{p_0}\right]^2 d\mu \to 0 \text{ as } t \to 0.$$
- Let $\mathcal{C}$ be a collection of regular paths in $\mathcal{P}$. A parameter $\psi : \mathcal{P} \to \mathbb{R}$ is **differentiable at P_0 relative to $\mathcal{C}$** if there exists a continuous and linear map $\dot{\psi} : L_2(p_0) \to \mathbb{R}$ such that
$$\frac{\psi(P_t) - \psi(P_0)}{t} \to \dot{\psi}(r) \text{ as } t \to 0,$$
for every regular path P_t in $\mathcal{C}$.

To study differentiability of the map $\psi(P_\lambda)$, Van Der Vaart (1991) considers a collection of paths in V such that

$$\frac{\lambda_t - \lambda_0}{t} \to v \text{ as } t \to 0, \tag{4.1}$$

where convergence is in the norm of V. The tangent set $\mathcal{T}$ at λ_0 is the set of all v thus obtained. To fit our setup in Van Der Vaart (1991)'s framework, we must define a suitable collection of paths.

Under the linear and continuous structure of S, it is natural to work with its extension $S\colon \overline{\langle B_0 \rangle} \to L_2(p_0)$.[3] However, Λ may not have an affine structure (i.e., B_0 may not be a linear subspace), so paths in Λ may not be able to span any $b \in \overline{\langle B_0 \rangle}$. Nevertheless, the linear and continuous structure of S and $\dot{\phi}$ leads to an extension of p_λ and ϕ to[4]

$$\tilde{\Lambda} = \{\lambda \in V\colon \lambda - \lambda_0 \in \overline{\langle B_0 \rangle}\} = \lambda_0 + \overline{\langle B_0 \rangle}.$$

For $\lambda \in \tilde{\Lambda}$, we have that $\lambda - \lambda_0 \in \overline{\langle B_0 \rangle}$, so both $\dot{\phi}(\lambda - \lambda_0)$ and $S(\lambda - \lambda_0)$ are well defined. We thus define $\phi(\lambda) \equiv \phi(\lambda_0) + \dot{\phi}(\lambda - \lambda_0)$ and $p_\lambda \equiv p_0(1 + S(\lambda - \lambda_0))$ for every $\lambda \in \tilde{\Lambda}$. The next result shows that paths in $\tilde{\Lambda}$ span $\overline{\langle B_0 \rangle}$.

Lemma 4.2. *Let $\mathcal{T}$ be the set of all $v \in V$ generated by paths $\lambda_t \in \tilde{\Lambda}$ satisfying* (4.1). *Then $\mathcal{T} = \overline{\langle B_0 \rangle}$.*

Proof. If $v \in \mathcal{T}$, there is a path λ_t in $\tilde{\Lambda}$ such that $\xi_t \equiv t^{-1}(\lambda_t - \lambda_0) \to v$ as $t \to 0$. Since $\lambda_t \in \tilde{\Lambda}$ for any t, $\lambda_t - \lambda_0 \in \overline{\langle B_0 \rangle}$, hence $\xi_t \in \overline{\langle B_0 \rangle}$ for any t. As $\overline{\langle B_0 \rangle}$ is closed, this means $v \in \overline{\langle B_0 \rangle}$.

Now take $b \in \overline{\langle B_0 \rangle}$. Then, $tb \in \overline{\langle B_0 \rangle}$. Therefore, $\lambda_t \equiv \lambda_0 + tb$ is a path in $\tilde{\Lambda}$ that satisfies $t^{-1}(\lambda_t - \lambda_0) \to b$ as $t \to 0$. □

The key condition in Van Der Vaart (1991) is the existence of a continuous and linear functional $A : \mathcal{T} \to L_2(p_0)$ such that

$$\int \left[\frac{1}{t}(\sqrt{p_{\lambda_t}} - \sqrt{p_0}) - \frac{1}{2} A v \sqrt{p_0} \right]^2 d\mu \to 0 \text{ as } t \to 0, \tag{4.2}$$

for every path λ_t in the collection that generated $\mathcal{T}$ (i.e., satisfying (4.1)). We show that in the affine case, where $p_\lambda = p_0(1 + S(\lambda - \lambda_0))$ for $\lambda \in \tilde{\Lambda}$, the Score operator defined in equation (2.1) satisfies the above condition. Since we consider all paths $\lambda_t \in \tilde{\Lambda}$, condition (4.2) is equivalent to Hadamard differentiability of $\lambda \mapsto \sqrt{p_\lambda}$ (see p. 182 in Van Der Vaart, 1991; and also result 1.2.7 in Yamamuro, 1974).

[3]It some abuse of notation, we denote the Score operator defined in B_0 and its extension to $\overline{\langle B_0 \rangle}$ by S. This extension exists since $L_2(p_0)$ is complete (Megginson, 1998, Th. 1.9.1).

[4]Given a point $v \in V$ and a set $A \subseteq V$ in a vector space, $v + A \equiv \{v + a\colon a \in A\}$.

Before jumping to the result, we must introduce a warning. For $\lambda \in \Lambda$, one has that $P_\lambda \in \mathcal{P}$, a probability model. Thus, $p_\lambda \geq 0$ a.s. μ, and $\sqrt{p_\lambda}$ is well-defined. This may not hold for $\lambda \in \tilde{\Lambda}$. Then, to discuss differentiability of $G(\lambda) \equiv \sqrt{p_\lambda}$, we must assume that the function is well defined for λ sufficiently close to λ_0. To do so, we must slightly strengthen Assumption 2.3:[5]

Assumption 4.1. The point $0 \in B_0 \subseteq \overline{\langle B_0 \rangle}$ is an interior point of B_0 in the relative topology that $\overline{\langle B_0 \rangle}$ inherits from the norm topology of V.

This assumption guarantees that λ_0 is an interior point of Λ (in the relative topology of $\tilde{\Lambda}$). G is then well-defined in an open neighborhood around λ_0. In the following proposition, Λ° denotes the interior of Λ with respect to the relative topology of $\tilde{\Lambda}$ (note that this may be different from the interior of Λ is the norm topology of V).

Proposition 4.1. *Let $G : \Lambda^\circ \to L_2(\mu)$ be given by*

$$G(\lambda) \equiv \sqrt{p_\lambda} = \sqrt{p_0(1 + S(\lambda - \lambda_0))}.$$

Under Asumptions 2.1 and 4.1, G is Hadamard differentiable at λ_0 with derivative $1/2S(\lambda - \lambda_0)\sqrt{p_0}$.

Proof. Under Assumption 4.1 $0 \in B_0^\circ \subseteq \overline{\langle B_0 \rangle}$, where the interior is in the relative topology of $\overline{\langle B_0 \rangle}$. Hence, we can equivalently work with $F : B_0^\circ \to L_2(\mu)$ given by $F(b) \equiv \sqrt{p_0(1 + Sb)}$, and show that it is Hadamard differentiable at 0 with derivative $1/2Sb\sqrt{p_0}$. To do so, we first show that F is Gateaux differentiable at 0 and then proceed to show that F is locally Lipschitz at 0.

For $b \in \overline{\langle B_0 \rangle}$, define the remainder

$$\mathrm{Re}(tb) \equiv F(tb) - F(0) - \frac{t}{2} Sb\sqrt{p_0} = \sqrt{p_0(1 + tSb)} - \sqrt{p_0} - \frac{t}{2} Sb\sqrt{p_0}.$$

We must show that $\mathrm{Re}(tb)/t \xrightarrow{L_2(\mu)} 0$ as $t \to 0$ (see Definition 3.1.1 in Fernholz, 1983).

For a fixed $b \in \overline{\langle B_0 \rangle}$, the Mean Value Theorem applied to $\sqrt{x}$ gives that, for all x,

$$\frac{\sqrt{p_0(x)(1 + tSb(x))} - \sqrt{p_0(x)}}{t} = \frac{1}{2\sqrt{\eta(x)}} Sb(x)p_0(x), \tag{4.3}$$

for $\min\{p_0(x), p_0(x)(1 + tSb(x))\} \leq \eta(x) \leq \max\{p_0(x), p_0(x)(1 + tSb(x))\}$.

Now, since B_0° is an open neighborhood of 0, it is absorbing (Megginson, 1998, Th. 2.2.9(f)). Thus, there exists a τ_b such that $tb \in U$ for $|t| < \tau_b$.

[5] Assumption 4.1 implies Assumption 2.3. However, the converse is not true. The interior of B_0 in the relative topology of $\langle B_0 \rangle$ may be a proper subset of the interior of B_0 in the relative topology of $\overline{\langle B_0 \rangle}$ (see Lemma B.7).

Therefore, $p_0(1+tSb) \geq 0$ a.s. μ; and we have that $tSb \geq -1$ a.s. μ, for $|t| < \tau_b$. Fix $c \in (0,1)$ and let $|t| < c\tau_b$. Then, $t/cSb \geq -1$ a.s. μ and thus $tSb \geq -c$ a.s. μ. This leads to $p_0(1+tSb) \geq p_0(1-c)$ a.s. μ. Hence, for $|t| < c\tau_b$, we have that $\eta(x) \geq p_0(x)(1-c)$ for μ-almost all x. This, on top equation (4.3), means that, for $|t| < c\tau_b$,

$$\frac{\sqrt{p_0(1+tSb)} - \sqrt{p_0}}{t} \leq \frac{1}{2\sqrt{1-c}} Sb\sqrt{p_0} \text{ a.s. } \mu.$$

Therefore, by the triangle inequality, for $|t| < c\tau_b$,

$$\left|\frac{\mathrm{Re}(tb)}{t}\right| \leq \left(\frac{1}{2\sqrt{1-c}} + \frac{1}{2}\right) |Sb|\sqrt{p_0} \text{ a.s. } \mu,$$

with $|Sb|\sqrt{p_0} \in L_2(\mu)$ since $Sb \in L_2(p_0)$. Since $p_0(1+tSb) \to p_0$ pointwise as $t \to 0$, equation (4.3) implies pointwise convergence of $\mathrm{Re}(tb)/t$ to 0 as $t \to 0$. Then, by the Dominated Convergence Theorem, for every $b \in \overline{\langle B_0 \rangle}$,

$$\left\|\frac{\mathrm{Re}(tb)}{t}\right\|_{L_2(\mu)} \to 0 \text{ as } t \to 0,$$

and thus F is Gateaux differentiable at 0 with derivative $1/2Sb\sqrt{p_0}$.

We now show that F is locally Lipschitz around 0. Let $b, b' \in B_0^\circ$. Then, $Sb, Sb' \geq -1$ a.s. μ. Fix $c \in (0,1)$. For an x such that $Sb(x) \geq Sb'(x)$, Lemma C.1 gives

$$\begin{aligned} F(b)(x) - F(b')(x) &\leq \frac{1}{c}\big(F(cb)(x) - F(cb')(x)\big) = \\ &= \frac{\sqrt{p_0(x)(1+cSb(x))} - \sqrt{p_0(x)(1+cSb'(x))}}{c}. \end{aligned}$$

We can now apply the Mean Value Theorem again to get

$$F(b)(x) - F(b')(x) \leq \frac{1}{2\sqrt{\eta(x)}} S(b-b')(x)p_0(x),$$

for $\eta(x) \geq p_0(x)(1+cSb'(x))$. Then, since $Sb' \geq -1$ a.s. μ, $\eta(x) \geq p_0(x)(1-c)$ for μ-almost all x. Hence, if $Sb(x) \geq Sb'(x)$,

$$F(b) - F(b') \leq \frac{1}{2\sqrt{1-c}} S(b-b')\sqrt{p_0} \leq \frac{1}{2\sqrt{1-c}} |S(b-b')|\sqrt{p_0} \text{ a.s. } \mu.$$

For an x such that $Sb'(x) \geq Sb(x)$, an analogous procedure leads to

$$F(b') - F(b) \leq \frac{1}{2\sqrt{1-c}} |S(b-b')|\sqrt{p_0} \text{ a.s. } \mu.$$

Therefore,

$$|F(b) - F(b')| \leq \frac{1}{2\sqrt{1-c}} |S(b-b')|\sqrt{p_0} \text{ a.s. } \mu.$$

Then, squaring and integrating with respect to μ gives us

$$\left\|F(b) - F(b')\right\|_{L_2(\mu)} \leq \frac{1}{4(1-c)} \left\|S(b-b')\right\|_{L_2(p_0)}.$$

To conclude, recall that S is linear and continuous, so

$$\left\|F(b) - F(b')\right\|_{L_2(\mu)} \leq \frac{\|S\|}{4(1-c)} \left\|b-b'\right\|_{\overline{\langle B_0\rangle}}.$$

That is, F is locally Lipschitz at 0. Since it is also Gateaux differentiable at 0, we conclude that it is Hadamard differentiable at 0 (see Prop. 3.5 in Shapiro, 1990, and references therein). □

Comment: If $M : L_1(\mu) \to L_2(\mu)$ given by $M(p) = \sqrt{p}$ was Hadamard differentiable, differentiability of G in the affine setup will readily follow from a Chain Rule (Yamamuro, 1974, Result 1.2.9). However, M is not Hadamard differentiable at an arbitrary point (see Lemma B.3).

To conclude, we note that the fact that $\dot{\phi}$ is continuous implies that ϕ is differentiable in the sense that $t^{-1}(\phi(\lambda_t) - \phi(\lambda_0)) \to \dot{\phi}(b)$ as $t \to 0$, for any path in $\tilde{\Lambda}$ such that $t^{-1}(\lambda_t - \lambda_0) \to b$ as $t \to 0$.[6] Indeed, by linearity and continuity of $\dot{\phi}$,

$$\frac{\phi(\lambda_t) - \phi(\lambda_0)}{t} = \frac{1}{t}\dot{\phi}(\lambda_t - \lambda_0) = \dot{\phi}\left(\frac{\lambda_t - \lambda_0}{t}\right) \to \dot{\phi}(b).$$

We are then ready to apply Theorem 3.1 in Van Der Vaart (1991). Let $\mathcal{C}$ be the collection of paths $t \mapsto p_{\lambda_t}$ such that λ_t is a path in $\tilde{\Lambda}$ satisfying (4.1). Then, differentiability of ψ relative to $\mathcal{C}$ depends on the relative position of $\dot{\phi}$ and $\mathcal{R}(S^*)$:

Proposition 4.2. *Suppose V is Banach space. Consider that β_0 is identified and let Assumptions 2.1, 2.2, and 4.1 hold. Then, $\psi(P)$ is differentiable at P_0 relative to $\mathcal{C}$ if and only if $\dot{\phi} \in \mathcal{R}(S^*)$.*

Proof. Lemma 4.1 guarantees that $\psi(P_\lambda) = \phi(\lambda)$ is well defined. Moreover, by Lemma 4.2, $\mathcal{T} = \overline{\langle B_0\rangle}$ is a closed linear subspace. We can also verify that if $t \mapsto \lambda_t$ is a path in $\tilde{\Lambda}$ such that $t^{-1}(\lambda_t - \lambda_0) \to b \in \overline{\langle B_0\rangle}$, then $t \mapsto \lambda_{ht}$ is also a valid path when $h > 0$. For $\varepsilon > 0$, let δ be such that $s < \delta$ implies $\left\|s^{-1}(\lambda_s - \lambda_0) - b\right\|_{\overline{\langle B_0\rangle}} < \varepsilon/h$. Then, for $\delta' \equiv \delta/h$, $t < \delta'$ implies $ht < \delta$ and thus

$$\left\|\frac{\lambda_{ht} - \lambda_0}{t} - hb\right\|_{\overline{\langle B_0\rangle}} = h\left\|\frac{\lambda_{ht} - \lambda_0}{ht} - b\right\|_{\overline{\langle B_0\rangle}} < \varepsilon.$$

[6] We follow here Van Der Vaart (1991)'s wording, even if this notion of differentiability is slightly different to the one of a differentiable parameter.

Proposition 4.1 shows that S satisfies (4.2) and that ϕ is differentiable follows from the discussion preceding this result. We can then apply Theorem 3.1 in Van Der Vaart (1991). In its Banach space version, it claims that ψ is differentiable if and only if $\mathcal{R}(\dot{\phi}^*) \subseteq \mathcal{R}(S^*)$, for $\dot{\phi}^* : \mathbb{R}^* \to \overline{\langle B_0 \rangle}^*$ given by $\langle b, \dot{\phi}^* \alpha^* \rangle = \langle \dot{\phi} b, \alpha^* \rangle$, for $b \in \overline{\langle B_0 \rangle}$ and $\alpha^* \in \mathbb{R}^*$ (Van Der Vaart, 1991, works with functionals that take values in arbitrary normed spaces, thus the formulation in terms of the adjoint of $\dot{\phi}$).

We conclude the proof by clarifying that, for real valued functionals, $\mathcal{R}(\dot{\phi}^*) = \langle \dot{\phi} \rangle$, the linear span of $\dot{\phi} \in \overline{\langle B_0 \rangle}^*$. Note that linear and continuous functionals $\alpha^* : \mathbb{R} \to \mathbb{R}$ are given by $\langle x, \alpha^* \rangle = \alpha^*(x) = \alpha x$, for $\alpha \in \mathbb{R}$ (here we identify the linear functional with the slope, as $\mathbb{R}^* = \mathbb{R}$). Then, following the definition of the adjoint, $\langle b, \dot{\phi}^* \alpha^* \rangle = \langle \dot{\phi} b, \alpha^* \rangle = \langle b, \alpha \dot{\phi} \rangle$ for $\alpha \in \mathbb{R}$ such that $\alpha^*(x) = \alpha x$. We then have that ψ is differentiable if and only if $\langle \dot{\phi} \rangle \subseteq \mathcal{R}(S^*)$, which is equivalent to $\dot{\phi} \in \mathcal{R}(S^*)$. $\square$

The following corollary puts together the results from Theorem 2.1 and Proposition 4.2. It is worth noting that both S and $\dot{\phi}$ are understood as functionals on $\overline{\langle B_0 \rangle}$.

Corollary 4.1. *Suppose V is Banach space. Let Assumptions 2.1, 2.2, and 4.1 hold. Then,*

1. *β_0 is identified and differentiable iff $\dot{\phi} \in \mathcal{R}(S^*)$,*
2. *β_0 is identified but not differentiable iff $\dot{\phi} \in \overline{\mathcal{R}(S^*)}^{w*} \backslash \mathcal{R}(S^*)$, and*
3. *β_0 is not identified iff $\dot{\phi} \notin \overline{\mathcal{R}(S^*)}^{w*}$.*

We note here that a combination of Theorems 3.1 and 4.1 in Van Der Vaart (1991) shows that, in models parametrized by a Hilbert space, regular identification is equivalent to differentiability of an identified parameter.[7] By comparing the above corollary with Corollary 3.1, we can see that this can be extended to Banach spaces:

Corollary 4.2. *Under the assumptions of Corollary 4.1, β_0 is regularly identified if and only if it is identified and differentiable.*

Hence, in light of Van Der Vaart (1991, Th. 2.1), regular identification is necessary for root-n estimation of the parameter in models parametrized by a Banach space.

[7] One cannot talk about differentiability of non-identified parameters, as the definition of differentiability assumes that the parameter is given by a map $\psi : \mathcal{P} \to \mathbb{R}$. Therefore, one can consider that if a parameter is not identified, then it is not differentiable.

5 Estimation of the average of a known transformation

Let g be a real valued measurable function. The goal is to estimate $\beta_0 \equiv \mathrm{E}_0[g(X)]$. Consider that the probability distrubutions in our model are dominated by a σ-finite measure μ. The natural space to look at is $\mathcal{P}_g \equiv \{p \in L_1(\mu) : gp \in L_1(\mu)\}$, so that the parameter is well defined.

Fix a true distribution $p_0 \in \mathcal{P}_g$ (which satisfies $p_0 \geq 0$ and $\|p_0\|_1 = 1$, although not all functions in $\mathcal{P}_g$ are required to). In what follows, E_0 means expectation with respect to p_0. We can reparametrize the problem. Define

$$\Lambda \equiv \{\lambda \in L_1(\mu) : p_0(1+\lambda) \in \mathcal{P}_g\},$$

so that now the true parameter is $\lambda_0 = 0$. One has that, indeed,

$$\Lambda = \{\lambda \in L_1(p_0) : \mathrm{E}_0[|\lambda g|] < \infty\}.$$

This is true since, being $p_0 \in L_1(\mu)$ and $gp_0 \in L_1(\mu)$, the condition $p_0(1+\lambda) \in \mathcal{P}_g$ is equivalent to

$$\int p_0|1+\lambda||g| < \infty \iff \int |\lambda g| p_0 < \infty, \text{ and}$$
$$\int p_0|1+\lambda| < \infty \iff \int |\lambda| p_0 < \infty.$$

Note now that, since $\lambda_0 = 0$, $B_0 \equiv \{b \in L_1(p_0) : \lambda_0 + b \in \Lambda\} = \Lambda$. Moreover, B_0 is linear, so $\langle B_0 \rangle = B_0$. That is, $B_0 = \langle B_0 \rangle = \Lambda = \tilde{\Lambda}$. The map from the parameter $\lambda \in \Lambda$ to the data $\mathcal{P}_g$ is $\lambda \mapsto p_\lambda \equiv p_0(1+\lambda)$. Hence, the Score Operator $S : \langle B_0 \rangle \to L_2(p_0)$ is

$$Sb \equiv \frac{p_{\lambda_0+b} - p_{\lambda_0}}{p_{\lambda_0}} = \frac{p_b - p_0}{p_0} = b.$$

Suppose that $\langle B_0 \rangle = B_0 \subseteq L_2(p_0)$ (we will later see that this is the case). Then S is the identity between B_0 and $L_2(p_0)$. S is then continuous when B_0 is considered with the topology of the restriction of L_2-norm. Moreover, S is linear.

The parameter of interest is $\beta_0 = \int gp_0$. The parameter is given by the functional $\phi : \Lambda \to \mathbb{R}$, being $\phi(\lambda) \equiv \int g(1+\lambda)p_0$, for $\lambda_0 = 0$. This functional is affine, so the functional $\dot{\phi} : B_0 \to \mathbb{R}$, given by

$$\dot{\phi}(b) \equiv \phi(\lambda_0 + b) - \phi(\lambda_0) = \phi(b) - \phi(0) = \int gbp_0 = \mathrm{E}_0[gb],$$

is linear.

There are two well established facts in the literature about identification and estimation of β_0:

1. β_0 is identified if $g \in L_1(p_0)$. That is, β_0 is identified as long as it is well-defined.

2. Root-n estimation of β_0 requires that $g \in L_2(p_0)$ (to apply the usual Central Limit Theorem). If $g \notin L_2(p_0)$, then root-n estimation of β_0 is not possible.

Our results allow to reach these conclusions from a general framework. Consider that $g \in L_q(p_0)$ for $1 < q \leq 2$. By Lemma A.2 bellow, notice that $B_0 = \Lambda = L_{q'}(p_0)$, with $1/q + q/q' = 1$, so $2 \leq q' < \infty$. Hence, Lemma A.1 gives that $B_0 \subseteq L_2(p_0)$ and that $Sb = b$, the identity, is continuous when B_0 is equipped with the $L_{q'}$-norm topology. Moreover, $\dot{\phi}$ is also continuous with respect to the the $L_{q'}$-norm topology according to Lemma A.2. In conclusion, when we consider $V = \Lambda = L_{q'}(p_0)$, we have that S and $\dot{\phi}$ are both linear and continuous. It is worth highlighting that for $q \neq 2$, V is not a Hilbert space. However, for any $1 < q \leq 2$, V is a Banach space.

We now study the adjoint operator $S^* : (L_2(p_0))^* \to L_{q'}(p_0)^*$. Note that $(L_2(p_0))^* = L_2(p_0)$ and $L_{q'}(p_0)^* = L_q(p_0)$, $1 < q \leq 2$. It is easy to see that S^* is the identity between $L_2(p_0)$ and $L_q(p_0)$, i.e., for every $u \in L_2(p_0)$, $S^*u = u$. The range of the adjoint is thus $\mathcal{R}(S^*) = L_2(p_0) \subseteq L_q(p_0)$.

The weak* closure of the range of the adjoint, $\overline{\mathcal{R}(S^*)}^{w*}$, is easy to compute in this case. Note that $L_q(p_0)$ is reflexive if $1 < q \leq 2$. Thus, $\overline{L_2(p_0)}^{w*} = \overline{L_2(p_0)}$ as subset of $L_q(p_0)$, $1 < q \leq 2$. It is well known that continuous functions with compact support are dense in $L_q(p_0)$, $1 \leq q < \infty$. Hence, $\overline{L_2(p_0)}^{w*} = \overline{L_2(p_0)} = L_q(p_0)$ as a subset of $L_q(p_0)$, $1 < q \leq 2$.

What is the discussion of the preceding discussion? Note that the representer of $\dot{\phi}$ as an element of $L_{q'}(p_0)^*$ is clearly g.[8] Since $g \in L_q(p_0)$ and $\overline{\mathcal{R}(\mathcal{S}^*)}^{w*} = L_q(p_0)$, Theorem 2.1 gives that β_0 is identified for $1 < q \leq 2$, which is the expected conclusion (the $q = 1$ must be studied separately, as the space is not reflexive).

For $q = 2$, $\overline{\mathcal{R}(S^*)}^{w*} = L_2(p_0) = \mathcal{R}(S^*)$. Therefore, $g \in \mathcal{R}(S^*)$ and Proposition 3.1 states that β_0 is regularly identified (an thus differentiable). Root-n estimation of β_0 may be possible (indeed, it is possible by the Central Limit Theorem).

For $1 < q < 2$, $\mathcal{R}(S^*) = L_2(p_0)$ is a proper subset of $\overline{\mathcal{R}(S^*)}^{w*} = L_q(p_0)$. In this case, $g \notin \mathcal{R}(S^*)$ and therefore, by Theorem 3.1, β_0 is irregularly identified. This means that root-n estimation of β_0 is ruled out.

6 Estimation of the density at a point

Consider a locally compact Hausdorff space X and let $C_0(X)$ denote the set of continuous functions $f : X \to \mathbb{R}$ that vanish at infinity. A function

[8] By this we mean that $Tg = \dot{\phi}$, being T the well known isometric isomorphism $T : L_q(p_0) \to L_{q'}(p_0)^*$ (see, for instance, Ex. 1.10.2 in Megginson, 1998).

vanishes at infinity if for every $\varepsilon > 0$ there exists a compact set $K \subseteq X$ such that $|f(x)| < \varepsilon$ if $x \notin K$ (see Def. 3.16 in Rudin, 1987).

According to Theorem 3.17 in Rudin (1987), $C_0(X)$ is a Banach space when endowed with the norm

$$\|f\|_\infty \equiv \sup_{x \in X} |f(x)|.$$

It is worth noting that elements in $C_0(X)$ are functions, not equivalence classes of functions (Rudin, 1987, Remark 3.15).

The probability model $\mathcal{P}$ is parametrized by $\Lambda = V = C_0(X)$, so that the map $\Lambda \to \mathcal{P}$, is the identity. That is, assuming that there is a locally finite and σ-finite measure μ that dominates $P_\lambda \in \mathcal{P}$, then the density of P_λ w.r.t. μ is λ. Call $\lambda_0 = p_0$ to the true density. For a fixed $x \in X$, the parameter of interest is $\beta_0 \equiv p_0(x) = \lambda_0(x)$, which is given by the functional $\phi_x : \Lambda \to \mathbb{R}$, $\phi_x(\lambda) \equiv \lambda(x)$.

Note that Λ is a linear space and thus $B_0 = \Lambda$. Moreover, ϕ_x is already linear so

$$\dot{\phi}_x(b) \equiv \phi_x(\lambda_0 + b) - \phi_x(\lambda_0) = \phi_x(b),$$

for every $b \in B_0$. Moreover, $\dot{\phi}$ is continuous as

$$|\dot{\phi}_x b| = |b(x)| \leq \sup_{x \in X} |b(x)| = \|b\|_\infty .$$

The Score operator is given by $Sb = b/p_0$. Which conditions ensure that $Sb \in L_2(p_0)$ and that S is continuous? A first approach would be to consider that the true density is bounded away from zero in its support K: $p_0(x) \geq C > 0$ if $x \in K$. It is easy to see that $K \equiv \overline{\{x \in X \colon p_0(x) > 0\}} = \{x \in X : p_0(x) \geq C\}$, since the later set is closed (inverse image of a closed set by a continuous function). Moreover, since $p_0 = \lambda_0 \in C_0(X)$, K is compact. To see this, recall that for $C > 0$ there exists a compact set K' such that $x \notin K' \Rightarrow p_0(x) < C$. Therefore, $p_0(x) \geq C \Rightarrow x \in K'$, so $K \subset K'$. Since K is a closed subset of a compact set, then K is compact. Moreover, K being compact and μ locally finite implies that $\mu(K) < \infty$.

We can then show that

$$\int_X (Sb)^2 p_0 d\mu = \int_K \frac{b^2}{p_0} d\mu \leq \frac{\mu(K)}{C} \|b\|_\infty^2 ,$$

so $Sb \in L_2(p_0)$ and $\|Sb\|_{L_2(p_0)} \leq \sqrt{\mu(K)/C}\, \|b\|_\infty$. That is, the Score operator is continuous.

It is well known that the dual space of $B_0 = C_0(X)$ is the set regular Borel measures on X (this is the Riesz-Markov-Kakutani Representation Theorem). That is, for every linear and continuous functional $F : C_0(X) \to \mathbb{R}$, there exists a regular Borel measure ν such that

$$Fb = \int b d\nu, \quad \forall b \in C_0(X).$$

Denote this set by $\text{rca}(X) \equiv C_0(X)^*$. Recall that the adjoint $S^* : L_2(p_0) \to \text{rca}(X)$ is characterized by

$$\langle b, S^* r\rangle = \langle Sb, r\rangle = \int Sbrp_0 d\mu = \int brd\mu.$$

Now, let ν^* be the regular Borel measure representing S^*r. The above equation implies that $d\nu^* = rd\mu$, i.e., ν^* has a density with respect to μ. We can then conclude that $\nu \in \mathcal{R}(S^*)$ if and only ν is absolutely continuous w.r.t. μ.

Notice that the regular measure representing $\dot{\phi}_x \in C_0(X)^*$ is δ_x, the Dirac measure on $x \in X$:

$$\delta_x(A) \equiv \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

Indeed, $\dot{\phi}_x(b) = b(x) = \int bd\delta_x$ for every $b \in B_0 = C_0(X)$.

Then, in light of Corollary 3.1, we can conclude the following. $\beta_0 \equiv p_0(x)$ is regularly identified if and only if $\delta_x \in \mathcal{R}(S^*)$. That is, δ_x must be absolutely continuous w.r.t. μ. Since $\delta_x(\{x\}) = 1 > 0$, regular identification of β_0 is equivalent to $\mu(\{x\}) > 0$.[9]

The preceding discussion shows that regular identification of β_0 is equivalent to μ having a positive mass on $x \in X$. Thus, if μ is the Lebesgue measure, estimation of a density at a point is always irregular. On the other hand, when μ has a positive mass at point x, $p_0(x)$ is regularly identified (e.g., a discrete model or if there is bunching at the point). In that case, with an iid sample $(X_1, ..., X_n)$, one can then estimate $p_0(x)$ at root-n rate by:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i = x) \xrightarrow{P} P_0(X_i = x) = \int_{\{x\}} p_0 d\mu = p_0(x)\mu(\{x\}).$$

Note that we can set w.l.o.g. $\mu(\{x\}) = 1$, since $\mu(\{x\}) > 0$.

We now show that, for any $x \in X$, $\dot{\phi}_x \in \overline{R(S^*)}^{w*}$. Therefore, Theorem 2.1 allows us to conclude that $\beta_0 \equiv p_0(x)$ is identified. It suffices to construct a net in $\mathcal{R}(S^*)$ that converges to $\dot{\phi}_x$ (Megginson, 1998, Prop. 2.1.18).

Let $\mathcal{O}_x \equiv \{U \subseteq X : x \in U, U \text{ open}\}$ be the set of all open neighborhoods of x. We turn $\mathcal{O}_x$ into a directed set by defining $V \preceq U$ if and only if $U \subseteq V$. Note that $\preceq$ is a preorder: it is reflexive since $U \subseteq U$ and transitive since $U \subseteq V$ and $V \subseteq W$ implies $U \subseteq W$. Moreover, it is upward directed as for two sets $U, V \in \mathcal{O}_x$, we have that $U \cap V \in \mathcal{O}_x$ satisfies $U \preceq U \cap V$ and $V \preceq U \cap V$.

[9] If $\mu(\{x\}) > 0$, then δ_x is absolutely continuos w.r.t. μ. The density is $r(s) = 1$ if $s = x$ and 0 otherwise. This density satisfies $r \in L_2(p_0)$ since $\int r^2 p_0 d\mu = p_0(x)\mu(\{x\}) < \infty$. This is the other requirement to be in the range of the adjoint.

We need to assume that for every open set $U \subseteq X$, $\mu(U) > 0$ (as is the case, for instance, with the Lebesgue measure). Then, we define the following net in $\mathcal{R}(S^*)$. For any index $U \in \mathcal{O}_x$, define

$$r_U(x) \equiv \begin{cases} \mu(U)^{-1} & x \in U \\ 0 & x \notin U \end{cases}.$$

Note that $r_U \in L_2(p_0)$ for every $U \in \mathcal{O}_x$, as

$$\int r_U^2 p_0 d\mu = \int_U p_0 d\mu = \int_{K \cap U} p_0 d\mu \leq \mu(K) \left\| p_0 \right\|_\infty .$$

Therefore, the measure $d\nu_U = r_U d\mu$ (i.e., the measure with density r_U w.r.t. μ) is in $\mathcal{R}(S^*)$. This measure is simply the uniform distribution in U.

We conclude by showing that $\nu_U \xrightarrow{w*} \dot{\phi}_x$ in $\text{rca}(X) = C_0(X)^*$. This is equivalent to (Megginson, 1998, p. 224)

$$\int b d\nu_U \to \int b d\dot{\phi}_x = b(x), \quad \forall b \in C_0(X). \tag{6.1}$$

We have that

$$\left| \int b(s) d\nu_U(s) - b(x) \right| \leq \frac{1}{\mu(U)} \int_U |b(s) - b(x)| d\mu(s) \leq \sup_{s \in U} |b(s) - b(x)|.$$

Since $B(b(x), \varepsilon/2)$, the open ball centered at $b(x)$ with radius $\varepsilon/2$, is an open neighborhood of $b(x)$, by continuity of b we have that there exists an open neighborhood of x, $\bar{U} \in \mathcal{O}_x$, such that $s \in \bar{U} \Rightarrow b(s) \in B(b(x), \varepsilon/2)$. Then, for every $U \in \mathcal{O}_x$ satisfying $\bar{U} \preceq U$, we have the following. If $s \in U \subseteq \bar{U}$, then $|b(s) - b(x)| < \varepsilon/2$. Hence,

$$\left| \int b(s) d\nu_U(s) - b(x) \right| \leq \varepsilon/2 < \varepsilon.$$

That is, for all $U \in \mathcal{O}_x$, $\bar{U} \preceq U$, we have that $\int b d\nu_U \in B(b(x), \varepsilon)$. This shows (6.1).

References

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *journal of Econometrics*, 32(2):189–218.

Escanciano, J. C. (2021). Semiparametric identification and fisher information. *Econometric Theory*, pages 1–38.

Fernholz, L. T. (1983). *Von Mises calculus for statistical functionals*, volume 19. Springer Science & Business Media.

Khan, S. and Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, 78(6):2021–2042.

Luenberger, D. G. (1997). *Optimization by vector space methods.* John Wiley & Sons.

Megginson, R. E. (1998). *An Introduction to Banach Space Theory*, volume 183. Springer Science & Business Media.

Rudin, W. (1987). *Real and Complex Analysis.* McGraw-Hill, Third edition.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of optimization theory and applications*, 66(3):477–487.

Stein, E. M. and Shakarchi, R. (2011). Functional analysis: introduction to further topics in analysis. *Princeton Lectures in Analysis*, 4.

Van Der Vaart, A. (1991). On differentiable functionals. *The Annals of Statistics*, pages 178–204.

Yamamuro, S. (1974). *Differential calculus in topological linear spaces*, volume 374. Springer.

Appendix A Measure theoretical results

Lemma A.1. *Let (X, m) be a finite measure space. Then, for $p', p \in [1, \infty]$ with $p' \geq p$, $L_{p'}(m) \subseteq L_p(m)$ and there exists a K such that $\|f\|_p \leq K \|f\|_{p'}$ for every $f \in L_{p'}(m)$.*

Proof. Let $\mathbf{1}$ be the real valued function $\mathbf{1}(x) = 1, \forall x \in X$. Since $m(X) < \infty$, $\mathbf{1} \in L_{q'}(m)$ for every $q' \in [1, \infty]$. Consider $f \in L_{p'}(m)$. Then, $f^p \in L_q(m)$ for $q \equiv p'/p \geq 1$. By Hölder's inequality, for $q' = 1 - 1/q$,

$$\|f^p \mathbf{1}\|_1 \leq \|f^p\|_q \|\mathbf{1}\|_{q'} \Rightarrow \|f^p\|_1^{1/p} \leq \|f^p\|_q^{1/p} \|\mathbf{1}\|_{q'}^{1/p}.$$

Define $K \equiv \|\mathbf{1}\|_{q'}^{1/p}$. The fact that

$$\|f^p \mathbf{1}\|_1^{1/p} = \left(\int f^p dm\right)^{\frac{1}{p}} = \|f\|_p \text{ and}$$
$$\|f^p\|_q^{1/p} = \left(\int (f^p)^{\frac{p'}{p}} dm\right)^{\frac{p}{p'}\frac{1}{p}} = \|f\|_{p'}$$

concludes the proof. $\square$

Lemma A.2. *Let (X, μ) be a finite measure space (will also work with σ-finite) and define $Tf = \int gf$ for $f \in L_p$,*[10] *$1 \leq p \leq \infty$. The following statements are equivalent:*

1. *$fg \in L_1, \forall f \in L_p$,*
2. *T is L_p-continuous, and*
3. *$g \in L_q$, with $1/p + 1/q = 1$.*

Proof.

(1) $\Rightarrow$ (2): Define

$$g_n(x) \equiv \begin{cases} n & \text{if } |g(x)| \geq n \\ g(x) & \text{if } |g(x)| < n \end{cases} \tag{A.1}$$

and $T_n f \equiv \int g_n f$. Let q such that $1/p + 1/q = 1$. Then, since $|g_n(x)| \leq n$,

$$\int |g_n|^q \leq \int n^q = \mu(X) n^q < \infty.$$

That is, $g_n \in L_q$ and by Holder's inequality $|T_n f| \leq ||g_n||_q ||f||_p$, so T_n is continuous in L_p.

[10] All integrals and L_p spaces are with respect to μ.

Since $|g_n(x)| \leq |g(x)|$, for each $f \in L_p$, $g_n f$ is dominated by gf. By (1), $gf \in L_1$, so we can apply the Dominated Convergence Theorem to obtain that $T_n f \to Tf$ for each $f \in L_p$. Moreover,

$$|T_n f| \leq \int |f g_n| \leq \int |fg| < \infty.$$

One can then apply the Uniform Boundedness Principle. The principle states that for a collection $\mathcal{T}$ of linear bounded functionals on a Banach space,

$$\sup_{T \in \mathcal{T}} |Tf| < \infty, \forall f \Rightarrow \sup_{T \in \mathcal{T}} ||T|| < \infty,$$

being the last norm the one of the functional in the dual space. In this particular case, $\mathcal{T} \equiv \{T_n : n \in \mathbb{N}\}$, so we get that there exists a $C < \infty$ such that $||T_n|| \leq C$ for all $n \in \mathbb{N}$.

We can then conclude that T is bounded (continuous) since

$$|Tf| = \lim_{n\to\infty} |T_n f| \leq \lim_{n\to\infty} ||T_n|| ||f||_p \leq \lim_{n\to\infty} C||f||_p = C||f||_p.$$

(2) $\Rightarrow$ (3): For $1/p + 1/q = 1$, consider $g_n \in L_q$ as defined in (A.1). We prove first the $p > 1$ (so $q < \infty$) case.

Define $f_n \equiv |g_n|^{q-1} \operatorname{sign}(g)/||g_n||_q^{q-1}$. Then,

$$\int |f_n|^p = \frac{\int |g_n|^{p(q-1)}}{||g_n||_q^{p(q-1)}} = \frac{||g_n||_q^q}{||g_n||_q^q} = 1,$$

since $p(q-1) = q$. That is, $f_n \in L_p$. Now, $|g_n(x)| \leq |g(x)|$ and by (2), $\exists C < \infty$ such that $|Tf| \leq C||f||_p$. Then, for every $n \in \mathbb{N}$,

$$\begin{aligned} ||g_n||_q &= \frac{||g_n||_q^q}{||g_n||_q^{q-1}} = \int \frac{|g_n|^q}{||g_n||_q^{q-1}} = \int \frac{|g_n|^{q-1}}{||g_n||_q^{q-1}} |g_n| \leq \int \frac{|g_n|^{q-1}}{||g_n||_q^{q-1}} |g| = \\ &= \int \frac{|g_n|^{q-1}}{||g_n||_q^{q-1}} \operatorname{sign}(g) g = \int f_n g \leq C||f_n||_p = C. \end{aligned}$$

To conclude, as $|g_n|^q \to |g|^q$ pointwise, we can apply Fatou's lemma:

$$\int |g|^q \leq \lim_{n\to\infty} \int |g_n|^q \leq \lim_{n\to\infty} C^q \leq C^q.$$

So $g \in L_q$.

If $p = 1$, note that $g_n \in L_\infty$ as they are bounded. For $\varepsilon > 0$, consider E_n a set where $|g_n(x)| \geq ||g_n||_\infty - \varepsilon$ for $x \in E_n$. Define

$f_n = \chi_{E_n} \operatorname{sign}(g)/\mu(E_n)$, where χ_{E_n} is the indicator function of set E_n. Then, since $|g_n(x)| \le |g(x)|$,

$$\int f_n g = \frac{1}{\mu(E_n)} \int_{E_n} \operatorname{sign}(g) g = \frac{1}{\mu(E_n)} \int_{E_n} |g| \ge \frac{1}{\mu(E_n)} \int_{E_n} |g_n| \ge ||g_n||_\infty - \varepsilon.$$

Now, $||f_n||_1 = 1$ and by (2), $|\int fg| \le C||f||_1$ for every $f \in L_1$. This leads to $||g_n||_\infty \le C + \varepsilon$, or, since ε is arbitrary, $||g_n|| \le C$ for every $n \in \mathbb{N}$. Since $g_n \to g$, this implies that $||g||_\infty \le C$, i.e., $g \in L_\infty$.

(3) $\Rightarrow$ (1): Take $f \in L_p$ and, by (3), $g \in L_q$, with $1/p + 1/q = 1$. Holder's inequality gives the result.

$\square$

Appendix B Functional analysis results

B.1 Extending functionals

Lemma B.1. *Let $B \subseteq V$ be a subset of a normed space. Then, for a continuous function $f : \overline{B} \to \mathbb{R}$,*

$$\inf_{b \in B} f(b) = \inf_{b \in \overline{B}} f(b).$$

Proof. Call $\overline{I} \equiv \inf_{b \in \overline{B}} f(b)$. We have that $\overline{I} \leq f(b)$ for every $b \in \overline{B}$. This means that $\overline{I}$ is a lower bound of $\{f(b) : b \in B\}$.

To show that it is the greatest lower bound, it suffices find a $b \in B$, for each $\varepsilon > 0$, satisfying $f(b) < \overline{I} + \varepsilon$. Since $\overline{I}$ is the infimum of $\{f(b) : b \in \overline{B}\}$, take $\tilde{b} \in \overline{B}$ such that

$$f(\tilde{b}) < \overline{I} + \frac{\varepsilon}{2}.$$

As $\tilde{b} \in \overline{B}$, there is a a sequence $b_n \in B$ converging to $\tilde{b}$. By continuity of f, the sequence $f(b_n)$ converges to $f(\tilde{b})$. Thus, there is an n_0 such that

$$|f(b_n) - f(\tilde{b})| < \frac{\varepsilon}{2} \text{ for } n \geq n_0.$$

Thus,

$$f(b_{n_0}) < f(\tilde{b}) + \frac{\varepsilon}{2} < \overline{I} + \varepsilon.$$

□

Let V be a normed space and W a Banach space. Consider a subspace $B \subseteq V$. Assume that $T \colon B \to W$ is linear and continuous. Then, we know that there is a unique linear and continuous extension $\tilde{T} \colon \overline{B} \to W$ such that $\tilde{T}b = Tb$ if $b \in B$ (Megginson, 1998, Th. 1.9.1). This extension is defined by

$$\tilde{T}b \equiv \lim_{n \to \infty} Tb_n,$$

for any sequence $b_n \to b \in V$, $b_n \in B$. We have the following result:

Lemma B.2. *Let $T \colon B \subseteq V \to W$, where B is a subspace of V and W is Banach. Then,*

$$\mathcal{R}(T) \subseteq \mathcal{R}(\tilde{T}) \subseteq \overline{\mathcal{R}(T)}.$$

Proof. For the first inclusion, let $w \in \mathcal{R}(T)$. Then, there exists a $b \in B$ such that $Tb = w$. This means that $\tilde{T}b = Tb = w$, so $w \in \mathcal{R}(\tilde{T})$.

For the second inclusion, let $w \in \mathcal{R}(\tilde{T})$. Then, there is a sequence $b_n \in B$, $b_n \to b$, such that $Tb_n \to w$. As $b_n \in B$ implies that $Tb_n \in \mathcal{R}(T)$, we have that $w \in \overline{\mathcal{R}(T)}$. □

An interesting corollary to the above lemma is that $\overline{\mathcal{R}(\tilde{T})} = \overline{\mathcal{R}(T)}$. Indeed, since $\overline{\mathcal{R}(\tilde{T})}$ is the smallest closed set containing $\mathcal{R}(\tilde{T})$, the fact that $\overline{\mathcal{R}(T)}$ is closed and contains $\mathcal{R}(\tilde{T})$ implies $\overline{\mathcal{R}(\tilde{T})} \subseteq \overline{\mathcal{R}(T)}$. The other inclusion in immediate from $\mathcal{R}(T) \subseteq \mathcal{R}(\tilde{T})$.

B.2 Differentiability

We translate here the definition of $\mathcal{S}$-differentiability from Fernholz (1983). Let V and W be topological vector spaces and consider $A \subseteq V$ be open.

Definition 4. Let $\mathcal{S}$ be a collection of sets in V. $F : A \to W$ is **$\mathcal{S}$-differentiable at** $f \in A$ if there exist a linear and continuous functional $\dot{F} : V \to W$ such that, for every $H \in \mathcal{S}$,

$$\frac{F(f+th) - F(f) - \dot{F}(th)}{t} \to 0 \text{ as } t \to 0$$

uniformly in $h \in H$.

Comment: The fact that F is defined in an open neighborhood A of f ensures that eventually $f + th \in A$. Indeed, $-f + A \equiv \{g - f : g \in A\}$ is an open neighborhood of 0 (Megginson, 1998, Th. 2.2.9(c)), and thus absorbing (Megginson, 1998, Th. 2.2.9(f)). This means that for every $h \in V$, there exists a τ_h such that $th \in -f + A$ for $t < \tau_h$. Then, for $t < \tau_h$, $f + th \in A$ since $f + th - f = th \in -f + A$.

When $\mathcal{S}$ is the set of all singletons (or finite sets), then we say that F is **Gateaux differentiable**. Note that in this case $\mathcal{S}$-differentiability reduces to pointwise differentiability. When $\mathcal{S}$ is the set of all compact sets of V, then the above definition is equivalent to Hadamard differentiability, and is therefore referred to as it (see Def. 2.2 and Prop. 3.3 in Shapiro, 1990; result 1.2.7 in Yamamuro, 1974):

Definition 5. $F : A \to W$ is **Hadamard differentiable at** $f \in A$ if there exist a linear and continuous functional $\dot{F} : V \to W$ such that, for every continuous path $\varphi : [0, 1] \to A$ with $\varphi(0) = f$ and $t^{-1}(\varphi(t) - \varphi(0)) \to v \in V$ as $t \to 0$, one has that

$$\frac{F(\varphi(t)) - F(f)}{t} \to \dot{F}v \text{ as } t \to 0.$$

From Definition 4, one can easily see that Hadamard differentiability implies Gateaux differentiability. Hence, the following lemma implies that $f \mapsto \sqrt{f}$ is not Hadamard differentiable.

Lemma B.3. *$F : L_1(\mu) \to L_2(\mu)$ given by $F(f) = \sqrt{f}$ is not Gateaux differentiable at an arbitrary $f \in \{f \in L_1(\mu) : f > 0$ a.s. $\mu\}$.*

Proof. Note that $F(f) = \varphi \circ f$ for $\varphi(x) \equiv \sqrt{x}$. Assume that F is Gateaux differentiable at an arbitrary $f \in \{f \in L_1(\mu) : f > 0 \text{ a.s. } \mu\}$. We will arrive to a contradiction.

By Lemma B.4, we have that for every $h \in L_1(\mu)$, $h/\sqrt{f} = 2\dot{F}h$ a.s. μ, so $h/\sqrt{f} \in L_2(\mu)$. However, when μ is the Lebesgue measure on $[1,\infty)$, and for $f(x) = x^{-4}$ and $h(x) = x^{-2}$, one has that

$$\int_1^\infty \frac{h(x)^2}{f(x)} d\mu(x) = \int_1^\infty d\mu(x) = \mu([1,\infty)) = \infty.$$

That is, $h/\sqrt{f} \notin L_2(\mu)$, a contradiction. □

Lemma B.4. *Let $p, p' \in [1,\infty)$. Suppose that $F : L_p(\mu) \to L_{p'}(\mu)$ is given by $F(f) = \varphi \circ f$, for a continuously differentiable $\varphi : U \to \mathbb{R}$, $U \subseteq \mathbb{R}$ open. Then, if F is Gateaux differentiable at f, its derivative is $\dot{F}h = (\dot{\varphi} \circ f)h$ a.s. μ, where $\dot{\varphi}$ is the derivative of φ.*

Proof. Fix $h \in L_p(\mu)$. Since F is Gateaux differentiable at f,

$$\frac{F(f+th) - F(f)}{t} \xrightarrow{L_{p'}(\mu)} \dot{F}h, \text{ as } t \to 0.$$

This means that $\dot{F}_n h \equiv n(F(f + h/n) - F(f)) \to \dot{F}h$ in $L_{p'}(\mu)$ as $n \to \infty$. Therefore, there exist a subsequence $\dot{F}_{n_k}h$ converging to $\dot{F}h$ a.s. μ as $k \to \infty$ (this is a by-product of the proof of completeness of $L_p(\mu)$, see Th. 1.3 in Stein and Shakarchi, 2011, and references therein).

By the Mean Value Theorem for φ,

$$\dot{F}_n h(x) = n\left(\varphi\left(f(x) + \frac{1}{n}h(x)\right) - \varphi(f(x))\right) = \dot{\varphi}(g_n(x))h(x),$$

for $\min\{f(x), f(x) + h(x)/n\} \leq g_n(x) \leq \max\{f(x), f(x) + h(x)/n\}$. As $f + h/n \to f$ pointwise as $n \to \infty$, $g_n \to f$ pointwise as $n \to \infty$. Thus, continuity of $\dot{\varphi}$ implies that $\dot{F}_n h \to (\dot{\varphi} \circ f)h$ pointwise.

To conclude, uniqueness of the limit implies that $\dot{F}h(x) = \dot{\varphi}(f(x))h(x)$ for those x such that $\dot{F}_{n_k}h(x) \to \dot{F}h(x)$. That is, the equality holds for μ-almost all x. □

B.3 Others

Lemma B.5. *Let $S\colon V \to W$ be a one-to-one and onto linear map between two normed spaces. Then $S^{-1}\colon W \to V$, such that $SS^{-1}w = w$ and $S^{-1}Sv = v$ for any $v \in V$ and $w \in W$, is linear.*

Proof. For $w, w' \in W$, there exists $v, v' \in V$ such that $Sv = w$ and $Sv' = w'$. Then, by linearity of S,

$$\begin{aligned} S^{-1}(\alpha w + \beta w') = S(\alpha Sv + \beta Sv') = S^{-1}S(\alpha v + \beta v') = \alpha v + \beta v' = \\ = \alpha S^{-1}w + \beta S^{-1}w'. \end{aligned}$$

□

Lemma B.6. *Let $(V, \|\cdot\|_V)$ be a normed vector space endowed with the topology induced by the norm. Let $W \subseteq V$ be a linear subspace. The following two topologies for W are identical:*

- *$\mathcal{T}_n$, the topology induced by the norm $\|\cdot\|_W$, the restriction of $\|\cdot\|_V$ to elements of W (i.e., $\|w\|_W = \|w\|_V$ for every $w \in W$); and*
- *$\mathcal{T}_r$, the relative topology of W as a subset of V. That is, $U \in \mathcal{T}_r$ iff there exist an open (with respect to the norm topology of V) set $O \subseteq V$ such that $U = O \cap W$.*

Proof. We introduce the following notation. For $v \in V$, $w \in W$ and $r > 0$, $B_V(v, r) \equiv \{x \in V : \|x - v\|_V < r\}$ and $B_W(w, r) \equiv \{x \in W : \|x - w\|_W < r\}$. The result follows from the fact that $\|w\|_W = \|w\|_V$ for $w \in W$. This implies that $B_W(w, r) = B_V(w, r) \cap W$ for $w \in W$.

To see this, note that if $x \in B_W(w, r)$, then $x \in W$ and $\|x - w\|_V = \|x - w\|_W < r$. Thus, $x \in B_V(w, r) \cap W$. Moreover, if $x \in B_V(w, r) \cap W$, then $x \in W$, which means that $\|x - w\|_W = \|x - w\|_V < r$. Note that we have used that W is a linear space, so that $x - w \in W$ if both points are in W.

Let $U \in \mathcal{T}_n$. Then, for every $u \in U$, there exists a $r_u > 0$ such that $B_W(u, r_u) \subseteq U$. This means that $U = \cup_{u \in U} B_W(u, r_u)$. Define the set $O = \bigcup_{u \in U} B_V(u, r_u)$, which is open in V as it is the union of open sets. We have that

$$O \cap W = \left(\bigcup_{u \in U} B_V(u, r_u) \right) \cap W = \bigcup_{u \in U} (B_V(u, r_u) \cap W) = \bigcup_{u \in U} B_W(u, r_u) = \\ = U,$$

so $U \in \mathcal{T}_r$.

Now, let $U \in \mathcal{T}_r$. That is, there exists an open set $O \subseteq V$, with respect to the norm topology in V, such that $U = O \cap W$. Since $u \in U$ implies that $u \in O$, we have that there exists a $r > 0$ such that $B_V(u, r) \subseteq O$. Thus, $B_V(u, r) \cap W \subseteq O \cap W = U$. Since $B_V(u, r) \cup W = B_W(u, r)$, this shows that $U \in \mathcal{T}_n$. □

Lemma B.7. *Let $W \subset V$ be a subspace of a normed space $(V, \|\cdot\|_V)$. Let $\overline{W}$ be the closure of W in the norm topology of V. Equip W and $\overline{W}$ with the topologies induced by the restriction of the norm of V ($\|w\|_W = \|w\|_V$ for $w \in W$, and $\|w\|_{\overline{W}} = \|w\|_V$ for $w \in \overline{W}$). Consider a subset $B \subseteq W \subseteq \overline{W}$. We have that the interior of B w.r.t. the topology of $\overline{W}$ is a subset of the interior of B w.r.t. the topology of W. Furthermore, the inclusion may be proper.*

Proof. Denote the interior of B w.r.t. the topology of W by $B^{\circ w}$ and the interior of B w.r.t. the topology of $\overline{W}$ by $B^{\circ \overline{w}}$. Let E stand for and arbitrary normed space. For $e \in E$ and $r > 0$, define the ball $B_E(e,r) \equiv \{e \in E \colon \|x - e\|_E < r\}$.

Let $x \in B^{\circ\overline{W}}$. There exists then a ball $B_{\overline{W}}(x,r) \subseteq B \subseteq W$. Since $B_W(w,r) = B_{\overline{W}}(w,r) \cap W$ for every $w \in W$ (see the proof of Lemma B.6), we have that $B_W(x,r) = B_{\overline{W}}(x,r)$ and thus $B_W(x,r) \subseteq B$. That is, $x \in B^{\circ W}$.

To see that the inclusion may be proper, let ℓ_∞ be the space of bounded sequences equipped with the supremum norm:

$$\|x\|_\infty \equiv \sup_{i \in \mathbb{N}} |x_i|.$$

Consider the subspace $c_f \subseteq \ell_\infty$ of finitely nonzero sequences. It is well known that its closure in the supremum norm is $\overline{c_f} = c_0$, the space of sequences that converge to 0.[11]

We have that 0 is a point in the interior of c_f w.r.t. the norm topology of c_f. Indeed, c_f is open in any topology of c_f, so every point is interior. However, 0 is not an interior point of c_f in the norm topology of c_0.

Consider that 0 was an interior point of c_f in the norm topology of c_0. That is, there exists an $r > 0$ such that $B_{c_0}(0, 2r) \subseteq c_f$. Note now that $x \equiv (x_i)_i \equiv r/i \in B_{c_0}(0, 2r)$, as $x \in c_0$ and $\|x - 0\|_\infty = r < 2r$. This means that $x \in c_f$, which is a contradiction. $\square$

Lemma B.8. *Let $\Lambda \subseteq V$ be subset of a vector space V. For $\lambda_0 \in \Lambda$, define $B_0 \equiv \{b \in V \colon \lambda_0 + b \in \Lambda\} = -\lambda_0 + \Lambda$. If Λ is a linear subspace, $B_0 = \Lambda$.*

Proof. For $b \in B_0$, we have that there exists a $\lambda \in \Lambda$ such that $b = \lambda - \lambda_0$. Thus, $b \in \Lambda$ as Λ is a linear subspace.

Let $\lambda \in \Lambda$. Since Λ is a linear subspace, $\lambda + \lambda_0 \in \Lambda$ and thus $\lambda = \lambda + \lambda_0 - \lambda_0 \in B_0$. $\square$

Definition 6. Let $C \subset V$ be a subset of a vector space.

- C is a **cone** if for all $\alpha > 0$, $x \in C \Rightarrow \alpha x \in C$.
- A cone C is a **convex** if $\alpha x + \beta y \in C$ for every $x, y \in C$ and $\alpha, \beta > 0$.

The linear span of convex cones has a particular characterization:

Lemma B.9. *Let $C \subseteq V$ be a convex cone in a vector space. Then,*

$$\langle C \rangle = \{x - y \colon x, y \in C\}.$$

[11] Let $x \in \overline{c_f}$. Then there exists a sequence $x_n \in c_f$ such that $x_n \to x$. For $\varepsilon > 0$, let n_0 be such that $\|x - x_n\|_\infty < \varepsilon$ for $n \geq n_0$. Since $x_{n_0} \in c_f$, let i_0 be such that $x_{in_0} = 0$ if $i \geq i_0$. Then $|x_i| < \varepsilon$ for all $i \geq i_0$, so $x \in c_0$.

Let $x \in c_0$. Define $x_{in} \equiv x_i \mathbb{1}(i \leq n)$. We have that $x_n \in c_f$ and $\|x_n - x\|_\infty = \sup_{i \geq n} |x_i|$, which tends to zero since $|x_i| \to 0$ as $i \to \infty$. Thus, $x_n \to x$ and $x \in \overline{c_f}$.

Proof. If $v = x - y$ with $x, y \in C$, then clearly $x \in \langle C \rangle$. Thus, one inclusion is trivial.

To see that $\langle C \rangle \subseteq \{x - y \colon x, y \in C\}$, let $v = \sum_{i=1}^{k} \alpha_i v_i$ with $v_i \in C$ and $\alpha_i \neq 0$. Let $I^+ \equiv \{i \colon \alpha_i > 0\}$ and $I^- \equiv \{i \colon \alpha_i < 0\}$. Then,

$$v = \underbrace{\sum_{i \in I^+} \alpha_i v_i}_{\in C} - \underbrace{\sum_{i \in I^-} |\alpha_i| v_i}_{\in C}.$$

□

We can use the above lemma to see that, for Riesz spaces, the linear span of any convex cone is the whole space.

Definition 7. A **Riesz space** is a vector space V equipped with a partial order $\leq$ (reflexive, anti-symmetric and transitive) such that:

1. For any $z \in V$, $x \leq y \Rightarrow x + z \leq y + z$.
2. For any $\alpha > 0$, $0 \leq x \Rightarrow 0 \leq \alpha x$.
3. Lattice-ordering: Any par of elements $x, y \in V$ have a supremum $\sup\{x, y\} \in V$ and an infumum $\inf\{x, y\} \in V$.

By Property 1 above, we have that

$$\sup\{z + x, z + y\} = z + \sup\{x, y\}.$$

To see this, call S_L and S_R to the suprema in the LHS and the RHS, respectivelly. By definition, $x + z \leq S_L$ and $z + y \leq S_L$. Thus, by Property 1, $S_L - z$ is an upper bound of $\{x, y\}$ and $S_R \leq S_L - z \Rightarrow z + S_R \leq S_L$ (by P1). For the other inequality, note that $x \leq S_R$ and $y \leq S_R$ implies that $z + S_R$ is an upper bound of $\{z + x, z + y\}$ (by P1). Thus, $S_L \leq z + S_R$.

Moreover, Property 3 allows us to define $x^+ \equiv \sup\{x, 0\}$ (positive part of x) and $x^- \equiv \sup\{-x, 0\}$ (negative part of x). With these definitions, we have that

$$x = x^+ - x^-.$$

To see this, note that $x^+ = \sup\{x, 0\} = x + \sup\{x - x, -x\} = x + \sup\{0, -x\} = x + x^-$. This results leads to the following lemma:

Lemma B.10. *Let V be a Riesz space and define $Q \equiv \{x \in V : 0 \leq x\}$. We have that $\langle Q \rangle = V$.*

Proof. Q is a cone by Property 2. Let $x, y \in Q$. Property 1 and $0 \leq y$ gives $x \leq x + y$. Since $0 \leq x$, transitivity gives $0 \leq x + y$, so $x + y \in Q$. This means that Q is a conex cone and $\langle Q \rangle = \{x - y \colon x, y \in Q\}$.

By the preceding discussion, any $x \in V$ can be expressed as $x = x^+ - x^-$. By definition of the supremum, $0 \leq x^+$ and $0 \leq x^-$. Therefore, $x^+ \in Q$ and $x^- \in Q$, so $\{x - y \colon x, y \in Q\} = V$. □

Appendix C Other results

Lemma C.1. *Let $t \in (0,1)$. For $-1 \leq y \leq x$,*

$$t\left(\sqrt{1+x} - \sqrt{1+y}\right) \leq \sqrt{1+tx} - \sqrt{1+ty}.$$

Proof. For $s \geq -1$, define $f(s) = \sqrt{1+ts} - t\sqrt{1+s}$. Then,

$$\frac{df}{ds}(s) = \frac{t}{2}\left(\frac{1}{\sqrt{1+ts}} - \frac{1}{\sqrt{1+s}}\right).$$

Then, since $t < 1$, $\sqrt{1+ts} < \sqrt{1+s}$ and $df/ds(s) > 0$. As f is increasing, for $x \geq y$:

$$\begin{aligned} f(x) \geq f(y) \iff \sqrt{1+tx} - t\sqrt{1+x} \geq \sqrt{1+ty} - t\sqrt{1+y} \iff \\ \sqrt{1+tx} - \sqrt{1+ty} \geq t\sqrt{1+x} - t\sqrt{1+y}. \end{aligned}$$

□

Lemma C.2. *Let $B \subseteq V$ be a subset of a normed space and consider a function $f\colon B \to \mathbb{R}$. Suppose that $f_n\colon B \to \mathbb{R}$ converges pointwise to f. Then, if the limit exists,*

$$\lim_{n\to\infty} \inf_{b\in B} f_n(b) \leq \inf_{b\in B} f(b).$$

Proof. Denote the infimums inside the limit in the LHS by I_n and the infimum in the RHS by I. By definition, for $\varepsilon/2 > 0$, we have that there exists a $\bar{b} \in B$ such that $f(\bar{b}) < I + \varepsilon/2$. Moreover, by pointwise convergence, there exists and $n_{\bar{b}}$ such that $|f_n(\bar{b}) - f(\bar{b})| < \varepsilon/2$ for $n \geq n_{\bar{b}}$. Thus, for $n \geq n_{\bar{b}}$, $f_n(\bar{b}) < I + \varepsilon$. This means that, for $n \geq n_{\bar{b}}$, $I_n \leq f_n(\bar{b}) < I + \varepsilon$. Taking limits in both sides yields

$$\lim_{n\to\infty} I_n \leq I + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this concludes the proof. □