



Article Identification in Parametric Models: The Minimum Hellinger Distance Criterion

David Pacini

School of Economics, University of Bristol, Bristol BS8 1TU, UK; david.pacini@bristol.ac.uk

Abstract: This note studies the criterion for identifiability in parametric models based on the minimization of the Hellinger distance and exhibits its relationship to the identifiability criterion based on the Fisher matrix. It shows that the Hellinger distance criterion serves to establish identifiability of parameters of interest, or lack of it, in situations where the criterion based on the Fisher matrix does not apply, like in models where the support of the observed variables depends on the parameter of interest or in models with irregular points of the Fisher matrix. Several examples illustrating this result are provided.

Keywords: identification; Hellinger distance; Fisher matrix

1. Introduction

There are values of unknown parameters of interest in data analysis that cannot be determined even in the most favorable situation where the maximum amount of data is available, i.e., when the distribution of the population is known. This difficulty has been tackled by either introducing criteria securing that the parameter of interest is (local) identifiable or by delineating the set of observationally equivalent values of the parameter of interest; for a review of these approaches, see, e.g., (Paulino and Pereira 1998) or (Lewbel 2019). This note contributes to these efforts by studying the criterion for identifiability based on the minimization of the Hellinger distance, which was introduced by Beran (1977), and exhibiting its relationship to the criterion for local identifiability based on the non-singularity the Fisher matrix, which was introduced by Rothenberg (1971). The similarities and differences between these two criteria for identifiability have so far not been studied.

The main result in this note is to show that the Hellinger distance criterion can be used to verify the (local) identifiability of a parameter of interest, or lack of it, either in models or points in the parameter space where the Fisher matrix criterion does not apply. This note illustrates this result with several examples, including a parametric procurement auction model, the uniform, normal squared, and Laplace location models. These models are either *irregular* because the support of the observed variables depends on the parameter of interest or the parameter space has *irregular points* of the Fisher matrix. Additional examples of *irregular models* and models with *irregular points* of the Fisher matrix are referenced below after defining the concepts of a *regular point* of the Fisher matrix and a *regular model* according to conventional usage, see, e.g., Rothenberg (1971).

Let *Y* be a vector-valued random variable in $\mathcal{Y} \subseteq \mathbb{R}^L$ with probability function P_o . Let the available data be a sample $\{Y_i\}_{i=1}^N$ of independent and identically distributed replications of *Y*. Consider a family \mathcal{F} of probability density functions $f : \mathcal{Y} \to [0, \infty)$ defined with respect to a common dominating measure μ , which will allow us to dispense with the need to distinguish between continuous and discrete random variables.¹ Let \mathcal{F}_{Θ} denote a subset of densities in \mathcal{F} indexed by $\theta \in \Theta$, where the parameter space Θ is a subset of \mathbb{R}^K , with *K* a positive integer. Let f_{θ} denote an element of \mathcal{F}_{Θ} .

Definition 1 (Identifiability). A parameter point θ_o in Θ is said to be identifiable if there is no other θ in Θ such that $f_{\theta}(y) = f_{\theta_o}(y)$ for μ -a.s y.



Citation: Pacini, David. 2022. Identification in Parametric Models: The Minimum Hellinger Distance Criterion. *Econometrics* 10: 10. https://doi.org/10.3390/ econometrics10010010

Academic Editor: Martin Wagner

Received: 5 October 2021 Accepted: 15 February 2022 Published: 21 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). **Definition 2** (Local Identifiability). A parameter point θ_o in Θ is said to be locally identifiable if there exists an open neighborhood $\Theta_{\star} \subseteq \Theta$ of θ_o containing no other θ such that Yes, it has been $f_{\theta}(y) = f_{\theta_o}(y)$ for μ -a.s y.

Definition 3 (Regular Points). *The Fisher matrix* $\mathcal{I}(\theta)$ *is the variance-covariance of the score* $\nabla \ell(\theta) := \nabla \ln f_{\theta}$,

 $\mathcal{I}(\theta) := \mathbb{E}[\nabla \ell(\theta) \nabla \ell(\theta)^{\top}] - \mathbb{E}[\nabla \ell(\theta)] \mathbb{E}[\nabla \ell(\theta)^{\top}].$

The point $\theta_o \in \Theta$ is said to be a regular point of the Fisher matrix if there exists an open neighborhood of θ_o in which $\mathcal{I}(\theta)$ has constant rank.

The (local) identifiability of regular points of the Fisher matrix in parametric models has been extensively studied, see, e.g., Rothenberg (1971). In contrast, the identifiability of irregular points has been less studied and the literature is rather unclear about what may happen about (local) identifiability of irregular points of the Fisher matrix. The study of irregular points is worthy of consideration because, first, there are several models of interest with this type of point in the parameter space (see the list below), and second, because irregular points may either correspond to:

- points in the parameter space that are not locally identifiable and for which a consistent estimator cannot not exist, e.g., the measurement error model studied by Reiersol (1950), or a consistent estimator can only exist after a normalization; or
- points in the parameter space that are locally identifiable and for which a \sqrt{N} consistent estimator cannot exist (and some algorithms, e.g., Newton–Raphson method
 based on the Fisher matrix, will face difficulties in converging) or a \sqrt{N} -consistent
 estimator can only exist after a reparametrization of the model, see, e.g., the bivariate
 probit model in Han and McCloskey (2019).

Hinkley (1973) noted that an irregular point of the Fisher matrix arises in the normal unsigned location model when the location parameter is zero. Sargan (1983) constructed simultaneous equation models with irregular points of the Fisher matrix. Lee and Chesher (1986) showed that the normal regression model with non-ignorable non-response has irregular points of the Fisher matrix in the vicinity of ignorable non-response. Li et al. (2009) noted that finite-mixture density models have irregular points of the Fisher matrix in the vicinity of homogeneity. Hallin and Ley (2012) showed that skew-symmetric density models have irregular points of the Fisher matrix in the vicinity of symmetry. We use below the normal squared location model (see Example 3) to illustrate in a transparent way the notion of an irregular point of the Fisher matrix.

The next Section shows that the criterion for local identifiability based on minimizing the Hellinger distance, unlike the criterion based on the non-singularity of the Fisher matrix, does apply to both regular and irregular points of the Fisher matrix and to *regular* and *irregular models*, to be defined below in Section 3. Section 3 shows that, for regular points of the Fisher matrix in the class of *regular models* studied by Rothenberg (1971), the criterion based on the Fisher matrix is a particular case of the criterion based on minimizing the Hellinger distance (but not for *irregular models* or irregular points of the Fisher matrix). Section 4 relates the minimum Hellinger distance criterion with the criterion based on the reversed Kullback–Liebler criterion, introduced by Bowden (1973), by showing that both are particular cases of the criterion for identifiability based on the minimization of a φ -divergence.

2. The Minimum Hellinger Distance Criterion

Identifying θ_o is the problem of distinguishing f_{θ_o} from the other members of \mathcal{F}_{Θ} . It is then convenient to begin by introducing a notion of how densities differ from each other. The squared Hellinger distance for the pair of densities f_{θ} , f_{θ_0} in \mathcal{F}_{Θ} is the square of the $L_2(\mu)$ -norm of the difference between the squared-root of the densities:

$$\rho(\theta) := \frac{1}{2} \left\| f_{\theta}^{1/2} - f_{\theta_0}^{1/2} \right\|_{L_2(\mu)}^2 = \frac{1}{2} \int \left(f_{\theta}^{1/2} - f_{\theta_0}^{1/2} \right)^2 d\mu = 1 - \int f_{\theta}^{1/2} f_{\theta_0}^{1/2} d\mu.$$

The squared Hellinger distance has the following well-known properties (see, e.g., Pardo 2005, p. 51), which are going to be used later.

Lemma 1. ρ can take values from 0 to 1, which are independent of the choice of the dominating measure μ , and $\rho(\theta) = 0$ if and only if $f_{\theta}(y) = f_{\theta_0}(y)$ for μ -a.s y.

(All the proofs are in Appendix A) Alternative notions of divergence between densities, other than the squared Hellinger distance, are studied in the Section 4. Since $\rho(\theta)$ is equal to zero if and only if f_{θ} and f_{θ_0} are equal, one has the following characterization of identifiability.

Lemma 2. The parameter $\theta_o \in \Theta$ is identifiable in the model \mathcal{F}_{Θ} if and only if, for all $\theta \in \Theta$ such that $\rho(\theta) = 0, \theta = \theta_o$.

Moreover, since $\theta \mapsto \rho(\theta)$ is non-negative and reaches a minimum at $\theta = \theta_0$, one obtains the following criterion for identifiability based on minimizing the squared Hellinger distance.

Proposition 1. The parameter $\theta_o \in \Theta$ is identifiable in the model \mathcal{F}_{Θ} if and only if

$$\theta_o = \arg\min_{\theta\in\Theta} \rho(\theta)$$

This criterion applies to models where:

- the support of Y depends on the parameter of interest (see Examples 1 and 2 below);
- θ_o is not a regular point of the Fisher matrix (see Example 3 below);
- some elements of the Fisher matrix $\mathcal{I}(\theta_o)$ are not defined (see Example 5 below);
- $\theta \mapsto \mathcal{I}(\theta)$ is not continuous (see Example 6 below);
- Θ is infinite-dimensional, as in semiparametric models (which are out of the scope of this note).²

The following examples illustrate the use of Proposition 1 and the definitions introduced so far. They are also going to illustrate, in the next section, the regularity conditions employed by Rothenberg (1971) to obtain a criterion for local identifiability based on the Fisher matrix. In these examples, μ denotes the Lebesgue measure. The Supplementary Materials presents step-by-step calculations of the squared Hellinger distance in Examples 1–5.

Example 1 (Uniform Location Model). *Set* $\mathcal{Y} = (0, \infty)$ *and* $\Theta = (0, \infty)$ *. Consider the uniform location model*

$$f_{\theta}(y) = \theta^{-1} 1(0 \le y \le \theta).$$

The Hellinger distance is

$$ho(heta) = 1 - rac{ig(heta+ heta_o - | heta- heta_o|ig)}{2\sqrt{ heta heta_o}}.$$

Since the unique solution to $\frac{1}{2\sqrt{\theta\theta_o}}(\theta + \theta_o - |\theta - \theta_o|) = 1$ is $\theta = \theta_o$, see Figure 1*a*, one has arg $\min_{\theta \in \Theta} \rho(\theta) = \theta_o$.

The Fisher matrix is $\mathcal{I}(\theta_o) = 0$ *, which is a singular matrix.*

Example 2 (First-Price Auction Model). Consider the first-price procurement auction model with m bidders introduced in (Paarsch 1992, Section 4.2.2). For bidders with independent private

valuations, following an exponential distribution with parameter θ (*Paarsch* 1992, *Display* 4.18) shows that the density of the wining bid Y_i in the *i*-th auction is

$$f_{\theta}(y) = \frac{m}{\theta} \exp\left(\frac{m}{m-1} - \frac{m}{\theta}y\right) \mathbb{1}\left(y \ge \frac{\theta}{m-1}\right)$$

Set $\mathcal{Y} = \mathbb{R}_+$ and $\Theta = (0, \infty)$. The Hellinger distance in this case is

$$\rho(\theta) = 1 - \frac{2\sqrt{\theta\theta_o}}{(\theta + \theta_o)} \exp\left(1 - \frac{(\theta + \theta_o)}{2\theta\theta_o} \max(\theta, \theta_o)\right)^{\frac{m}{m-1}}.$$

For $\theta < \theta_0$ (resp. $\theta > \theta_0$), one has $\nabla \rho(\theta) < 0$ ($\nabla \rho(\theta) > 0$), see Figure 1b. Hence, by continuity of $\theta \mapsto \rho(\theta)$, $\arg \min_{\theta \in \Theta} \rho(\theta) = \theta_0$. The Fisher matrix is $\mathcal{I}(\theta_0) = 0$, which is a singular matrix.



Figure 1. The Hellinger distance in Examples 1 and 2. (a) Example 1 ($\theta_0 = 4$); (b) Example 2 ($\theta_0 = 4$, m = 5).

Example 3 (Normal Squared Location Model). *Set* $\mathcal{Y} = \mathbb{R}$ *and* $\Theta = \mathbb{R}$ *. Consider the normal squared location model*

$$f_{\theta}(y) = (\sqrt{2\pi})^{-1} \exp\left[-(y-\theta^2)^2/2\right].$$

This model would arise, for example, if Y is the difference between a matched pair of random variables whose control and treatment labels are not observed. The Hellinger distance is

$$\rho(\theta) = 1 - \exp(-(\theta^2 - \theta_o^2)^2/8).$$

The parameter point $\theta_o = 0$ is identifiable because $\theta \mapsto 1 - \exp(-\theta^4/8)$ is a strictly convex function, see Figure 2a. The parameter points $\theta_o \neq 0$ are not identifiable because $\rho(\theta_o) = \rho(-\theta_o) = 0$, see Figure 2b. The Fisher matrix is $\mathcal{I}(\theta_o) = 4\theta_o^2$, which implies that $\mathcal{I}(0) = 0$ is a singular matrix and $\theta_o = 0$ is an irregular point of the Fisher matrix.

Example 4 (Demand-and-Supply Model). Let Y = (P, Q) denote the observed price and quantity of a good transacted in a market at a given period of time. Linear approximations to the demand and supply functions are

$$D = \alpha + \beta P + V$$
 (Demand)

$$S = \gamma + \delta P + U$$
 (Supply)

$$Q = D = S,$$
 (Equilibrium)

where α , β , γ , δ are unknown parameters and (U, V) is an unobserved random vector. Assume that U and V are independent and jointly normal distributed with mean zero and unknown variance σ_{11} and σ_{22} , respectively. Set $\theta = (\alpha, \beta, \gamma, \delta, \sigma_{11}, \sigma_{22})$. The density of the observed variables is then

$$f_{\theta}(y) = \frac{\exp\left(-\frac{1}{2}(y-\mu)'\Omega^{-1}(y-\mu)\right)}{\sqrt{(2\pi)^{2}\det(\Omega)}},$$

where $det(\cdot)$ is the determinant of the matrix in the parenthesis and

$$\mu = \begin{pmatrix} \frac{\alpha - \gamma}{\delta - \beta} \\ \frac{\delta \alpha - \beta \gamma}{\delta - \beta} \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \frac{\sigma_{11} + \sigma_{22}}{(\alpha - \gamma)^2} & \frac{\delta \sigma_{11} + \beta \sigma_{22}}{(\alpha - \gamma)^2} \\ \frac{\delta \sigma_{11} + \beta \sigma_{22}}{(\alpha - \gamma)^2} & \frac{\delta^2 \sigma_{11} + \beta^2 \sigma_{22}}{(\alpha - \gamma)^2} \end{pmatrix}.$$

The squared Hellinger distance is

$$\rho(\theta) = 1 - \frac{\det(\Omega)^{1/4}\det(\Omega_o)^{1/4}}{\det\left(\frac{\Omega+\Omega_o}{2}\right)^{1/2}}\exp\bigg(-\frac{1}{8}(\mu-\mu_o)'\bigg(\frac{\Omega+\Omega_o}{2}\bigg)^{-1}(\mu-\mu_o)\bigg).$$

To show that θ_0 is not identifiable, by Proposition 1, it suffices to verify that $\arg \min_{\theta} \rho(\theta)$ is not a singleton. We elaborate on this point in the Supplementary Material.



Figure 2. The Hellinger distance in Example 3. (a) Example 3 ($\theta_o = 0$); (b) Example 3 ($\theta_o = 1$). **Example 5** (Laplace Location Model). Set $\mathcal{Y} = \mathbb{R}$, $\Theta = \mathbb{R}$. Consider the Laplace location model

$$f_{\theta}(y) = \frac{1}{2} \exp(-|y - \theta|).$$

The squared Hellinger distance is

$$ho(heta) = 1 - \exp\left(-rac{| heta - heta_o|}{2}
ight) - rac{\delta}{2} \left[1(heta - heta_o > 0)\exp\left(-rac{ heta - heta_o}{2}
ight) - 1(heta - heta_o < 0)\exp\left(rac{ heta - heta_o}{2}
ight)
ight].$$

For any $\theta - \theta_o < 0$ ($\theta - \theta_o < 0$), one has $\nabla \rho(\theta) > 0$ ($\nabla \rho(\theta) > 0$). By continuity, $\theta \mapsto \rho(\theta)$ has then a unique minimizer at $\theta = \theta_o$, and, by Proposition 1, θ_o is identifiable. The Fisher matrix is $\mathcal{I}(\theta) = 1$, which is a non-singular matrix.

Example 6 (Exponential Mixture). Set $\mathcal{Y} = [0, \infty]$, $\theta = (\theta_1, \theta_2, \theta_3)'$ and $\Theta = [0, 1] \times [0, \infty) \times [0, \infty)$. Consider the finite mixture of exponential model

$$f_{\theta}(y) = (1 - \theta_1) \exp(\ln \theta_2 - \theta_2 y) + \theta_1 \exp(\ln \theta_3 - \theta_3 y).$$

Consider $\theta_o = (1/2, 1, 2)$ and $\theta_{\star} = (1/2, 2, 1)$. Since $f_{\theta_o}^{1/2} = f_{\theta_{\star}}^{1/2}$, one has $\rho(\theta_{\star}) = 0$ and $\theta_{\star} \in \arg \min_{\theta \in \Theta} \rho(\theta)$. Since $\theta_o \neq \theta_{\star}$, it follows from Proposition 1 that θ_o is not identifiable.

The previous examples also illustrate the difference between identifiable and local identifiable points in the parameter space.

Example 7 (Normal Squared Location Model, Continued). In this example, any $\theta_o \in \Theta$ is locally identifiable—even the irregular point $\theta_o = 0$ to the Fisher matrix—and only $\theta_o = 0$ is identifiable, see Figure 2.

We also have the following criterion for local identifiability based on minimizing the squared Hellinger distance.

Proposition 2. The parameter $\theta_o \in \Theta$ is locally identifiable in the model \mathcal{F}_{Θ} if and only if there exists an open set $\Theta_* \ni \theta_o$ such that

$$\theta_o = \arg\min_{\theta\in \Theta_\star} \rho(\theta)$$

This criterion, unlike the criterion based on the Fisher matrix by Rothenberg (1971) *and re-stated below as Lemma 3 for the sake of completeness, applies to the case when:*

- *the support of Y depends on the parameter of interest;*
- θ_o is not a regular point of the Fisher matrix;
- some elements of the Fisher matrix $\mathcal{I}(\theta_o)$ are not defined;
- $\theta \mapsto \mathcal{I}(\theta)$ is not continuous;
- Θ is infinite-dimensional.

Proposition 2 reduces local identifiability to a unique solution of a well-defined minimization problem. One general criterion, and, as argued, e.g., (Rockafellar and Wets 1998), virtually the only available one, to check in advance for the uniqueness of a minimizer of an optimization problem is the strict convexity of the objective function. The application of this general criterion to the characterization of local identifiability in Proposition 2 yields the following result:

Proposition 3. If $\theta \mapsto \rho(\theta)$ is a locally strictly convex function around θ_0 (i.e., if there is an open convex set $\Theta_* \ni \theta_0$ such that $\rho : \Theta_* \mapsto [0,1]$ is a strictly convex function), then θ_0 is locally identifiable.

Proposition 3 leads to the observation that local identifiability can be seen to be related to the local convexity of the Hellinger distance. As with our earlier propositions, it holds when the support of Y depends on the parameter of interest, θ_0 is not a regular point of the Fisher matrix, some elements of the Fisher matrix $\mathcal{I}(\theta_0)$ are not defined or $\theta \mapsto \mathcal{I}(\theta)$ is not continuous.

3. The Fisher Matrix Criterion

Rothenberg (1971) gives a criterion for local identifiability in terms of the non-singularity of the Fisher matrix. Additional insight about the relevance—and limitations—of the Fisher matrix criterion for local identifiability may then be gained by relating it to the criterion based on minimizing the Hellinger distance. To study this relationship, we now focus on the *regular models* studied by Rothenberg (1971).

Assumption 1 (R (Regular Models)). \mathcal{F}_{Θ} is such that:

(A1) Θ is an open set in \mathbb{R}^{K} . (A2) $f_{\theta} \geq 0$ and $\int f_{\theta} d\mu = 1$ for all $\theta \in \Theta$. (A3) $supp(f_{\theta}) := \{y \in \mathcal{Y} : f_{\theta}(y) > 0\}$ is the same for all $\theta \in \Theta$. (A4) For all θ in a convex set containing Θ and for all $y \in supp(f_{\theta})$, the functions $\theta \mapsto f_{\theta}$ and $\theta \mapsto \ell(\theta) := \ln f_{\theta}$ are continuously differentiable. (A5) The elements of the matrix $\mathbb{E}[\nabla \ell(\theta) \nabla \ell(\theta)^{\top}]$ are finite and are continuous functions of θ everywhere in Θ .

We now replicate the characterization of local identifiability by Rothenberg (1971) Theorem 1 based on the non-singularity of the Fisher matrix.

Lemma 3. Let the regularity conditions in Assumption R hold. Let θ_o be a regular point of the Fisher matrix $\mathcal{I}(\theta_o)$. Then, θ_o is locally identifiable if and only if $\mathcal{I}(\theta_o)$ is non-singular.

This characterization of local identifiability only applies to the regular models defined by Assumption R and to the regular points of the Fisher matrix, which may be a subset of the parameter space (see Example 3). These conditions do not have themselves any direct statistical or economic interpretation: their role is just to permit a characterization of local identifiability.³ We have already referenced in the introduction a list of models with irregular points of the Fisher matrix, for which the characterization in Lemma 3 does not apply. We now use Examples 1–5 to illustrate the notions of regular and irregular models and their implications for the analysis of identifiability. The richness of the possibilities that follow is a recall of the care needed in using the Fisher matrix criterion for showing local identifiability (or lack of it). It also highlights the convenience of the identifiability criterion based on minimizing the Hellinger distance as a unifying approach to study the identifiability of regular or irregular points of the Fisher matrix in either regular or irregular models. Specifically:

- The uniform location model in Example 1 and the first-price auction model in Example 2 have, respectively, $supp(f_{\theta}) = [0, 1/\theta]$ and $supp(f_{\theta}) = [\theta/[m-1], \infty)$, which means that these models violate the regularity condition (*A*3). We have seen that θ_o is identifiable in Examples 1 and 2, which implies that (*A*3) is not necessary for identifiability. These models also have a singular Fisher matrix, which implies that, in irregular models violating (*A*3), the non-singularity of the Fisher matrix is not a necessary condition for (local) identifiability.
- One can verify that the normal squared location model in Example 3 and the normal supply-and-demand model in Example 4 both satisfy the regularity conditions in Assumption R. We have seen that in Example 3 the parameter of interest is locally identifiable while in Example 4 it is not, which means that the regularity conditions in Assumption R are not sufficient or necessary for (local) identifiability, they are just convenient. In Example 3, moreover, $\theta_0 = 0$ is not a regular point of the Fisher matrix and is locally identifiable, which implies that, for irregular points of the Fisher matrix, the non-singularity of the Fisher matrix is not a necessary condition for (local) identifiability.
- In Example 5, the function $\theta \mapsto \ln(1/2) |y \theta|$ is not differentiable when $y = \theta$, which means that the Laplace location model is an irregular model because it violates (*A*4).
- To illustrate a failure of (A1) and (A5), consider the finite mixture of exponential model in Example 6 with $\theta_1 = 0$, $\theta_2 = 1$ and $\theta_3 = 0.5$. In this case, $\mathbb{E}[\nabla \ell(0) \nabla \ell(0)^\top] = (1-2)^2/[2(2-2)]$, which is not finite.

We also have the following result linking the Hellinger distance to the Fisher matrix, which we are going to use to show that, in regular models with irregular points to the Fisher matrix, the non-singularity of the Fisher matrix is only a sufficient condition for local identifiability.

Lemma 4. Let the regularity conditions in Assumption R hold and assume that $\theta \mapsto \ell(\theta) := f_{\theta}^{1/2}$ is continuously differentiable μ -a.e. Then, the Hellinger distance and the Fisher matrix are related by

$$\nabla^2 \rho(\theta_o) = c \mathcal{I}(\theta_o)$$
, where $c = 1/4$.

Though this result is known, see, e.g., (Borovkov 1998), its implications for local identifiability have so far not been drawn.

Since the Fisher matrix is a variance-covariance matrix, one has that $\mathcal{I}(\theta)$ is, under (*A*5), a real symmetric semi-definite positive matrix for every $\theta \in \Theta$, and then the following result follows from Lemma 4 and the characterization of a convex function in terms of its Hessian, see, e.g., (Rockafellar and Wets 1998, Theorem 2.14).

Proposition 4. Let the regularity conditions in Assumption R hold and assume that $\theta \mapsto f_{\theta}^{1/2}$ is continuously differentiable μ -a.e. Then, $\theta \mapsto \rho(\theta)$ is a locally convex function around θ_0 . Furthermore, if $\mathcal{I}(\theta_0)$ is non-singular, then $\theta \mapsto \rho(\theta)$ is a locally strict convex function around θ_0 and θ_0 is locally identifiable.

Two remarks are in order. First, notice that, unlike Lemma 3, the result in Proposition 4 also applies when θ_o is not a regular point of the Fisher matrix and the non-singularity of the Fisher matrix becomes only sufficient for local identifiability. Second, if $\mathcal{I}(\theta_o)$ is singular, the function $\rho : \Theta_* \rightarrow [0, 1]$ is still locally convex (because $\mathcal{I}(\theta_o)$ is positive semi-definite) and arg $\min_{\theta \in \Theta_*} \rho(\theta)$ is a convex, but not necessarily bounded, set, which is a result that can be used to delineate the set of observational equivalent values of θ_o . This note does not pursue this interesting direction.

Table 1 summarizes the information in this note about the necessity and sufficiency of the non-singularity of the Fisher matrix for local identifiability.

Tabl	e 1.	For	local	ide	ntifia	bility	, the	e non-singul	larity	of the	e Fisher	Matrix is	• •
------	------	-----	-------	-----	--------	--------	-------	--------------	--------	--------	----------	-----------	-----

	Regular Points	Irregular Points		
Regular Models	necessary and sufficient (Lemma 3)	only sufficient (Proposition 4 and Example 3)		
Irregular Models	not necessary (Examples 1, 2, and 5)			

We conclude this section by mentioning that, in response to the misbehavior of the Fisher matrix when informing about the difficulty to estimate parameters of interest in parametric models, alternative notions of information, other than the Fisher matrix, have been proposed in the literature (see, e.g., Donoho and Liu 1987). Without further elaboration, these alternative notions of information are not directly applicable to construct new criteria of identifiability. In particular, the geometric information based on the modulus of continuity of $\theta_0 \mapsto \arg \min_{\theta} \rho(\theta)$ with respect to the Hellinger distance, introduced by Donoho and Liu (1987) to geometrize convergence rates, cannot be used to construct a criterion for local identifiability because this modulus of continuity, in its current format, is not defined for parameters that are not locally identifiable.⁴

4. The Kullback–Liebler Divergence and Other Divergences

Some of the examples where we have had success in using the Hellinger distance to analyze identifiability share the same structure: the Hellinger distance is a locally convex function, see Figure 2, and so the results from convex optimization become available. If the Hellinger distance proves to be difficult to analyze, one can set out a criterion for identifiability based on another divergence function, such as the reversed Kullback–Liebler divergence (see, e.g., Bowden 1973)

$$\kappa(heta) = -H(heta)$$
, where $H(heta) := \int \ln\left(rac{f_ heta}{f_{ heta_o}}
ight) f_{ heta_o} d\mu$.

One can unify the identification criteria based on the Hellinger distance and the reversed Kullback–Liebler divergence by using the family of φ -divergences defined as

where f_{θ}/f_{θ_0} is the likelihood ratio and $\varphi : \mathbb{R} \to [0, +\infty]$ is a proper closed convex function with $\varphi(1) = 0$ and such that $x \mapsto \varphi(x)$ is strictly convex on a neighborhood of x = 1. The squared Hellinger distance corresponds to the member of this family with $\varphi(x) = \frac{1}{2}(1 - \sqrt{x})^2$, whereas the reversed Kullback–Liebler divergence corresponds to $\varphi(x) = -\ln x + x - 1$. The following result is an immediate consequence of the property that δ_{φ} is non-negative and it is equal to zero if and only if $f_{\theta} = f_{\theta_0}$ (see, e.g., Pardo 2005, Proposition 1.1)).

Proposition 5. The parameter $\theta_o \in \Theta$ is locally identifiable in the model \mathcal{F}_{Θ} if and only if there exists an open set $\Theta_* \ni \theta_o$ such that

$$\theta_o = \arg\min_{\theta\in\Theta_\star} \delta_{\varphi}(\theta).$$

This result, which is a generalization of Proposition 2, shows that the choice of a φ -divergence for analyzing the identifiability of a parameter of interest only hinges on the difficulty to characterize the set $\arg \min_{\theta} \delta_{\varphi}(\theta)$ for a given φ -divergence. The choice of the Hellinger distance over the reversed Kullback–Liebler divergence is, however, not inconsequential when choosing φ -divergence to construct an estimator for the parameter of interest. The use of the Hellinger distance may lead to an estimator that is more robust than the maximum likelihood estimator and equally efficient, see, e.g., Beran (1977) and Jimenez and Shao (2002).⁵

We conclude this section with the following result showing that, for the regular models analyzed by Rothenberg (1971), the Hellinger distance and the reversed Kullback–Liebler divergence are both locally convex around a minimizer.

Lemma 5. Let the regularity conditions in Assumption R hold and assume that $\theta \mapsto f_{\theta}^{1/2}$ is continuously differentiable μ -a.e. Let us assume, furthermore, that, in a neighborhood of θ_o , f_{θ} and $\ln f_{\theta}$ are twice differentiable in θ , with derivatives continuous in $y \in supp(f_{\theta})$. Then, the Hellinger distance and the Kullback–Liebler divergence are related by

$$\nabla^2 \rho(\theta_o) = c \nabla^2 \kappa(\theta_o) forc = 1/4.$$

Supplementary Materials: The following are available at https://www.mdpi.com/article/10.3390/econometrics10010010/s1, Auxiliary calculations in Examples 1–5.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: I would like to thank Sami Stouli and Vincent Han for offering constructive suggestions on previous versions of this paper. All errors are mine.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof

Proof of Lemma 1. Write

$$\rho(\theta) = \frac{1}{2} \|f_{\theta}^{1/2} - f_{\theta_o}^{1/2}\|_{L^2(\mu)} = \frac{1}{2} \int (f_{\theta}^{1/2} - f_{\theta_o}^{1/2})^2 d = 1 - \int f_{\theta}^{1/2} f_{\theta_o}^{1/2} d\mu.$$

Hence, $\rho(\theta) = 0$ if and only if $f_{\theta} = f_{\theta_0}$ and $\rho(\theta) = 1$ if and only if $f_{\theta}f_{\theta_0} = 0$. To show that $\rho(\theta)$ does not depend on the choice of the dominating measure μ , let g_{θ} and g_{θ_0} denote

the densities of P_{θ} and P_{θ_0} relative to another dominating measure ν . Let h and k denote the densities of μ,ν relative to $\mu + \nu$. The density of P_{θ} relative to $\mu + \nu$ is $f_{\theta}h$ and also $g_{\theta}k$. Thus, $f_{\theta}h = g_{\theta}k$ and also $f_{\theta_0}h = g_{\theta_0}k$. Hence, $(f_{\theta}f_{\theta_0})^{1/2}h = (g_{\theta}g_{\theta_0})^{1/2}k$ and

$$\int (g_{\theta}g_{\theta_{o}})^{1/2}d\nu = \int (g_{\theta}g_{\theta_{o}})^{1/2}kd(\nu+\mu) = \int (g_{\theta}g_{\theta_{o}})^{1/2}hd(\nu+\mu) = \int (f_{\theta}f_{\theta_{o}})^{1/2}d\mu$$

which completes the proof. \Box

Proof of Lemma 2. In the text. \Box

Proof of Lemma 3. We replicate the proof by Rothenberg (1971) Theorem 1. By the mean value theorem, there is θ_{\star} between θ and θ_{o} such that

$$\ell_{ heta} - \ell_{ heta_o} =
abla \ell_{ heta} (heta_\star)^{ op} (heta - heta_o).$$

Assume that θ_o is not locally identifiable. Then, there is a sequence $\{\theta_j\}_j$ converging to θ_o such that $\ell_{\theta_j} = \ell_{\theta_o}$. This implies $\nabla \ell_{\theta}(\theta_{\star})^{\top} q_j = 0$, where $q_j = (\theta_j - \theta_o) / \|\theta_j - \theta_o\|$. The sequence $\{q_j\}_j$ belongs to the unit sphere and therefore is convergent to a limit q_o . As θ_j approaches θ_o , q_j approaches q_o and in the limit $q_o^{\top} \nabla \ell_{\theta}(\theta_o)$. However, this implies that

$$q^{\top} \mathcal{I}_{\theta_o} q = q^{\top} \mathbb{E}[\nabla \ell_{\theta}(\theta_o) \nabla \ell_{\theta}(\theta_o)^{\top}] q = 0,$$

and, hence, \mathcal{I}_{θ_0} must be singular.

To show the converse, suppose that \mathcal{I}_{θ} has constant rank r < K in a neighborhood of θ_o . Consider then the eigenvector v_{θ} associated to one of the zero eigenvalues of \mathcal{I}_{θ} . Since $0 = v_{\theta}^{\top} \mathcal{I}_{\theta} v_{\theta}$, we have for all θ near θ_o

$$v_{ heta}^{ op}
abla \ell_{ heta} = 0$$

Since \mathcal{I}_{θ} is continuous and has constant rank, the function $\theta \mapsto v_{\theta}$ is continuous in a neighborhood of θ_0 . Consider now the curve $\gamma : [0, t_{\star}] \to \mathbb{R}^K$ defined by the function $\theta(t)$, which solves the differential equation $\frac{\partial \theta(t)}{\partial t} = v_{\theta}$ with $\theta(0) = \theta_0$ for $0 \le t \le t^*$. The log density function is differentiable in t with

$$rac{\partial \ell_{ heta(t)}}{\partial t} = v_{ heta(t)}^{ op}
abla \ell_{ heta}(heta(t))$$

However, by the preceding display this is zero for all $0 \le t \le t^*$. Thus $\theta \mapsto \ell_{\theta}$ is constant on the curve γ and θ_o is not locally identifiable. \Box

Proof of Lemma 4. Assume first that θ is a scalar, i.e., K = 1. Re-write

$$\rho(\theta) := \frac{1}{2} \left\| f_{\theta}^{1/2} - f_{\theta_o}^{1/2} \right\|_{L_2(\mu)}^2 = \frac{1}{2} \left[\int \left(f_{\theta}^{1/2} - f_{\theta_o}^{1/2} \right)^2 d\mu \right] = 1 - \int f_{\theta}^{1/2} f_{\theta_o}^{1/2} d\mu.$$

Differentiating $\theta \mapsto \rho(\theta)$, one has that

$$\nabla \rho(\theta) = -\frac{1}{2} \int f_{\theta_o}^{1/2} f_{\theta}^{-1/2} \nabla f_{\theta}(\theta) d\mu = \frac{1}{2} \int \frac{(f_{\theta}^{1/2} - f_{\theta_o}^{1/2}) \nabla f_{\theta}(\theta)}{f_{\theta}^{1/2}} d\mu,$$

where Assumptions (A3) and (A4) allow us to pass the derivative under the integral sign. Since $\theta \mapsto \rho(\theta)$ reaches a minimum at θ_o , one has $\nabla \rho(\theta_o) = 0$ and so

$$\frac{\nabla \rho(\theta) - \nabla \rho(\theta_o)}{(\theta - \theta_o)} = \frac{1}{2} \int \frac{(f_{\theta}^{1/2} - f_{\theta_o}^{1/2}) \nabla f_{\theta}(\theta)}{(\theta - \theta_o) f_{\theta}^{1/2}} d\mu,$$

which, by the Lebesgue dominated convergence theorem, satisfies

$$abla^2
ho(heta_o):=\lim_{ heta
ightarrow heta_o}rac{
abla
ho(heta)-
abla
ho(heta_o)}{ heta- heta_o}=rac{1}{4}\mathcal{I}(heta_o),$$

because the integrand converges point-wise

$$\frac{(f_{\theta}^{1/2} - f_{\theta_o}^{1/2}) \nabla f_{\theta}(\theta)}{(\theta - \theta_o) f_{\theta}^{1/2}} \rightarrow \frac{\nabla f_{\theta}(\theta_o) \nabla f_{\theta}(\theta_o)^{\top}}{2f_{\theta_o}} = \frac{f_{\theta_o} \nabla \ln f_{\theta}(\theta_o) \nabla \ln f_{\theta}(\theta_o)^{\top} f_{\theta_o}}{2f_{\theta_o}} \\ = \frac{1}{2} \nabla \ln f_{\theta}(\theta_o) \nabla \ln f_{\theta}(\theta_o)^{\top} f_{\theta_o},$$

and it is dominated by a sum of, under (A5), integrable functions

$$\left|\frac{(f_{\theta}^{1/2} - f_{\theta_o}^{1/2})\nabla f_{\theta}(\theta)}{(\theta - \theta_o)f_{\theta}^{1/2}}\right| \leq \frac{(f_{\theta}^{1/2} - f_{\theta_o}^{1/2})^2}{(\theta - \theta_o)^2} + \frac{\nabla f_{\theta}(\theta)\nabla f_{\theta}(\theta)^{\top}}{f_{\theta}}.$$

To extend this proof to the case when θ is a vector, one applies the argument above element-wise to the components of $\nabla^2 \rho(\theta_o)$. \Box

Proof of Lemma 5. If $\nabla^2 H(\theta) = -\mathcal{I}(\theta_0)$, the claim then follows from Proposition 1 after replacing $c\mathcal{I}(\theta_0) = \nabla^2 \rho(\theta_0)$. To verify that $\nabla^2 H(\theta_0) = -\mathcal{I}(\theta_0)$, we follow (Bowden 1973, Section 2). Recall that $\nabla \ell_{\theta} = \nabla \ln f_{\theta} = \frac{\nabla f_{\theta}}{f_{\theta}}$ and, since we have assumed that $\int f_{\theta} d\mu = 1$ for any $\theta \in \Theta$, one has that

$$\int \nabla f_{\theta}(\theta) d\mu = \underset{K \times K}{0} \text{ and } \int \nabla^2 f_{\theta}(\theta) d\mu = \underset{K \times K}{0} \text{ for any } \theta \in \Theta.$$

Differentiating $\theta \mapsto H(\theta)$, one obtains

$$abla H(heta) = \int rac{f_{ heta_o}}{f_ heta} rac{
abla f_{ heta(heta)}}{f_{ heta_o}} f_{ heta_o} d\mu,$$

and differentiating again

$$\nabla^2 H(\theta) = \int \frac{\partial}{\partial \theta^{\top}} \left[\frac{\nabla f_{\theta}(\theta)}{f_{\theta}} \right] f_{\theta_o} d\mu = \int \left[\frac{\nabla^2 f_{\theta}(\theta)}{f_{\theta}} - \nabla \ln f_{\theta}(\theta) \nabla \ln f_{\theta}(\theta)^{\top} \right] f_{\theta_o} d\mu.$$

Evaluating at $\theta = \theta_o$, and using $\int \nabla^2 f_{\theta}(\theta_o) d\mu = 0$, one obtains $\nabla^2 H(\theta_o) = -\mathcal{I}(\theta_o)$. \Box

Proof of Proposition 1. The sufficiency has already been established by Beran (1977), Theorem 1(iii) and it is an immediate consequence of the definition of identifiability. The necessity is in the text and it follows immediately from Lemmas 1 and 2. \Box

Proof of Proposition 2. It is immediate from Proposition 1 and the definition of local identifiability (Definition 2). \Box

Proof of Proposition 3. This result follows from the uniqueness of a solution for strictly convex problems (see, e.g., Rockafellar and Wets 1998, Theorem 2.6) after noticing, from Lemma 1, that $\theta \mapsto \rho(\theta)$ is bounded, and hence a proper function. \Box

Proof of Proposition 4. The proof for the claim that $\theta \mapsto \rho(\theta)$ is a locally convex function around θ_o is in the text. It only remains to show that, if the Fisher matrix is non-singular, then θ_o is locally identifiable. When the Fisher matrix is non-singular, by Lemma 4 and the characterization of convex functions in (Rockafellar and Wets 1998, Theorem 2.14), the Hellinger distance is a strictly locally convex function. The claim then follows Proposition 3.

Proof of Proposition 5. In the text. \Box

Appendix B. Variational Representation and Estimation

It is well-known, see, e.g., Beran (1977), that the estimator based on minimizing the Hellinger distance between the density postulated by the model for the observed variables and a kernel nonparametric estimator of the density of these variables can be more robust (to ρ -perturbations of density of the observed variables) than the maximum likelihood estimator and still asymptotically efficient in regular models. This minimum Hellinger distance-to-kernel estimator requires smoothing, which becomes an inconvenient requirement in models with observable variables with mixed support, such as the normal regression model with non-response in the dependent variable, or with support depending on the unknown parameter, such as the parametric auction model in Example 2, or in models with high-dimensional observable variables, due to the curse of dimensionality. This Appendix derives the variational representation of the Hellinger distance. This variational representation serves to construct the minimum dual Hellinger distance estimator, which unlike the minimum Hellinger distance-to-kernel estimator, does not require the use of a smooth estimator of the density of the observable variables.

Recall first that the squared Hellinger distance is

$$\rho(\theta) = \frac{1}{2} \int \left(f_{\theta}^{1/2} - f_{\theta_0}^{1/2} \right)^2 d\mu.$$
 (A1)

We are going to verify that

$$\rho(\theta) = \frac{1}{2} \sup_{\tilde{\theta} \in \Theta} \left\{ \int \left(f_{\theta} - f_{\theta}^{1/2} f_{\tilde{\theta}}^{1/2} \right) d\mu - \int \frac{f_{\theta}}{f_{\tilde{\theta}}} dP_{\theta_o} - 1 \right\}$$
(A2)

The expression in the last display, unlike (A1), admits, under a bracketing number condition on the family of likelihood ratios $\{y \to f_{\theta}(y)/f_{\bar{\theta}}(y), \theta, \tilde{\theta} \in \Theta\}$, a consistent sample analog estimator not depending on smoothing parameters. The minimum dual Hellinger distance estimator of θ_o is the set of minimizers of the sample analog of (A2):

$$\hat{\theta} = \underset{\substack{\theta \in \Theta \\ generator}}{\arg\min_{\theta \in \Theta}} \underset{\substack{\tilde{\theta} \in \Theta \\ discriminator}}{\sup} \left\{ \int \left(f_{\theta} - f_{\theta}^{1/2} f_{\tilde{\theta}}^{1/2} \right) d\mu - N^{-1} \sum_{i=1}^{N} \frac{f_{\theta}(Y_i)}{f_{\tilde{\theta}}(Y_i)} \right\}.$$

One could use a simulator to approximate f_{θ} or $f_{\tilde{\theta}}$ if these densities have an untractable form.⁶

To verify (A2), define the functions

$$\varphi(x) := \frac{1}{2}(1 - x^{1/2})^2 \text{ and } \bar{\varphi}(x) := \sup_{\bar{x} \in \mathbb{R}} \bar{x}x - \varphi(x)$$

and write the squared Hellinger distance in (A1) as

$$\rho(\theta) = \int \varphi\left(\frac{f_{\theta}}{f_{\theta_o}}\right) d\mu.$$
 (A3)

The function $\bar{\varphi}$ is the convex conjugate of $\varphi(x)$. We first show that

$$\rho(\theta) = \sup_{\tilde{\theta}\in\Theta} \left\{ \int \dot{\varphi}\left(\frac{f_{\theta}}{f_{\tilde{\theta}}}\right) d\mu - \int \bar{\varphi}\left(\dot{\varphi}\left(\frac{f_{\theta}}{f_{\tilde{\theta}}}\right)\right) f_{\theta_{o}} d\mu \right\}$$
(A4)

where $\mathcal{G}_{\theta} = \{y \mapsto \dot{\varphi}(f_{\theta}(y)/f_{\tilde{\theta}}(y)), \tilde{\theta} \in \Theta\}$ and $\dot{\varphi}(x) = \frac{1}{2}(1-x^{-1/2})$ is the derivative of $x \mapsto \varphi(x) = \frac{1}{2}(1-x^{1/2})^2$. For all $x \in (0, +\infty)$, one has that $\ddot{\varphi}(x) = \frac{1/2}{x}^{-3/2} > 0$ and

then $x \mapsto \varphi(x) = \frac{1}{2}(1 - x^{1/2})^2$ is strictly convex on $(0, +\infty)$. By strict convexity, for any $x, \tilde{x} \in (0, +\infty)$, it holds that

$$\varphi(x) \ge \varphi(x) + \dot{\varphi}(\tilde{x})(x - \tilde{x})$$

with equality if and only if $x = \tilde{x}$. Fix two values $\theta, \tilde{\theta}$ in the parameter space and set

$$x = rac{dP_{ heta}}{dP_{ heta_o}} ext{ and } ilde{x} = rac{dP_{ heta}}{dP_{ ilde{ heta}}}$$

Inserting these values in the last inequality and integrating with respect to P_{θ_0} yields

$$ho(heta) \geq \int \left[\dot{\varphi}\left(rac{dP_{ heta}}{dP_{ ilde{ heta}}}
ight) dP_{ heta} - ilde{arphi}\left(rac{dP_{ heta}}{dP_{ ilde{ heta}}}
ight)
ight] dP_{ heta_o},$$

where $\tilde{\varphi}(x) := \dot{\varphi}(x)x - \phi(x)$, which in turn implies

$$ho(heta) \geq \int \dot{arphi} igg(rac{dP_ heta}{dP_{ heta}} igg) dP_ heta - \int ilde{arphi} igg(rac{dP_ heta}{dP_{ heta}} igg) dP_{ heta_o}.$$

When $\tilde{\theta} = \theta_o$, this inequality turns to equality, which yields (A4) after noticing that $\dot{\phi}(x)x - \phi(x) = \bar{\phi}(\dot{\phi}(x))$.

To conclude the verification, since $x \mapsto \varphi(x) = \frac{1}{2}(1 - x^{1/2})^2$ is differentiable for all $x \in (0, +\infty)$, one has

$$\begin{split} \bar{\varphi}(\dot{\varphi}(x)) &= \dot{\varphi}(x)x - \varphi(x) \\ &= \frac{1}{2}(1 - x^{-1/2})x - \frac{1}{2}(1 - x^{1/2})^2 = \frac{1}{2}[(x - x^{1/2}) - (1 - x^{1/2})^2] \\ &= \frac{1}{2}(x^{1/2} - 1) \end{split} \tag{A5}$$

By replacing (A5) back in (A4), one obtains (A2).

Notes

- ¹ We use \rightarrow in ' $f : \mathcal{Y} \rightarrow [0, \infty)$ ' to declare the domain (\mathcal{Y}) and codomain ($[0, \infty)$) of the function f and we use the arrow notation ' \mapsto ' to define the rule of a function inline. We use ':=' to indicate that an expression is 'defined to be equal to'. This notation is in line with conventional usage.
- ² See, e.g., Escanciano (2021) for a systematic approach to identification in semiparametric models.
- ³ As a referee has pointed out, necessary and sufficient conditions for (local) identification require different assumptions. Some of the conditions in R are not necessary if we only seek sufficient conditions: differentiability of the score function and non-singularity of the Fisher matrix would suffice.
- ⁴ A related modulus of continuity has been introduced by Escanciano (2021, Online Supplementary Materials, Lemma 1.3) to provide sufficient conditions for (local) identification in semiparametric models. The analysis of these models is out of the scope of this paper.
- ⁵ Appendix B elaborates more on this point by using the variational representation of the Hellinger distance to construct a minimum distance estimator which does not require a non-parametric estimator of the density of the data.
- ⁶ One could also replace the space Θ in the *discriminator* model by a family of compositional functions—as in neural network models—to gain, if needed, flexibility when fitting $f_{\tilde{\theta}}$ by introducing, again, tuning parameters.

References

Beran, R. 1977. Minimum Hellinger Distance Estimates for Parametric Models. *The Annals of Statistics* 5: 445–63. [CrossRef] Borovkov, A. 1998. *Mathematical Statistics*. Amsterdam: Gordon and Breach.

Bowden, R. 1973. The Theory of Parametric Identification. Econometrica 41: 1069. [CrossRef]

Donoho, D., and R. Liu. 1987. Geometrizing Rates of Convergence, I. Technical Report 137. Berkeley: University of California.

Escanciano, J. C. 2021. Semiparametric Identification and Fisher Information. Econometric Theory 1–38. [CrossRef]

Hallin, M., and C. Ley. 2012. Skew-Symmetric Distributions and Fisher Information—A Tale of Two Densities. *Bernoulli* 18: 747–63. [CrossRef]

- Han, S., and A. McCloskey. 2019. Estimation and Inference with a (Nearly) Singular Jacobian. *Quantitative Economics* 10: 1019–68. [CrossRef]
- Hinkley, D. 1973. Two-Sample Tests with Unordered Pairs. *Journal of the Royal Statistical Association (Series B: Methodological)* 36: 2466–80. [CrossRef]
- Jimenez, R., and Y. Shao. 2002. On Robustness and Efficiency of Minimum Divergence Estimators. Test 10: 241-48. [CrossRef]
- Lee, L., and A. Chesher. 1986. Specification Testing when Score Test Statistics are Identically Zero. *Journal of Econometrics* 31: 121–49. [CrossRef]
- Lewbel, A. 2019. The Identification Zoo: Meanings of Identification in Econometrics. *Journal of Economic Literature* 57: 835–903. [CrossRef]
- Li, P., J. Chen, and P. Marriot. 2009. Non-Finite Fisher Information and Homogeneity: And EM Approach. *Biometrika* 96: 411–26. [CrossRef]
- Paarsch, H. 1992. Deciding between the Common and Private Value Paradigms in Empirical Models of Auctions. *Journal of Econometrics* 51: 191–215. [CrossRef]
- Pardo, L. 2005. Statistical Inference Based on Divergence Measures. New York: Chapman & Hall/CRC Press.
- Paulino, C., and C. Pereira. 1994. On Identifiability of Parametric Statistical Models. *Journal Italian Statistical Society* 3: 125–51. [CrossRef]
- Reiersol, O. 1950. The Identifiability of a Linear Reation between Variables which are Subject to Error. *Econometrica* 18: 375–89. [CrossRef]
- Rockafellar, T., and R. Wets. 1998. Variational Analysis. Berlin: Springer.
- Rothenberg, R. 1971. Identification in Parametric Models. Econometrica 39: 577-91. [CrossRef]
- Sargan, J. 1983. Identification and Lack of Identification. Econometrica 51: 1605–33. [CrossRef]