Journal of Statistical and Econometric Methods, Vol.11, No.4, 2022, 1-15
ISSN: 2241-0384 (print), 2241-0376 (online)
https://doi.org/10.47260/jsem/1141
Scientific Press International Limited

A Goodness-of-Identifiability Criterion for Parametric Statistical Models

David Pacini[1]

Abstract

This note sets out a goodness-of-identifiability criterion. This criterion quantifies the identifying strength of a parametric statistical model. Unlike the qualitative criterion for identifiability based only on the Fisher matrix, it applies to both regular and irregular points of the Fisher matrix. Unlike the qualitative criterion based only on the Hellinger distance, it quantifies set-identification.

JEL classification numbers: C10, C50.
Keywords: Parametric statistical model, Identifiability, Fisher matrix, Hellinger distance.

---

[1]School of Economics, University of Bristol.

Article Info: *Received:* July 22, 2022. *Revised:* October 19, 2022.
Published online: October 24, 2022.

1. Introduction

The identifying strength of a statistical model is the ability of the model to distinguish different points in the parameter space from hypothetical knowledge of the population. It is a key input in the process of design, evaluating and choosing estimators for the parameters in the model. There is no quantitative criterion in the literature for measuring identifying strength. This gap often leaves fundamental questions -such as how the identifying strength quantitatively varies across the parameter space in a model- answered.

This note proposes a goodness-of-identifiability criterion for quantifying the identifying strength of parametric statistical models. The criterion measures the difference between the log of (one plus) the minimum eigenvalue of the Fisher matrix and the log of (one plus) the diameter of the set of observational equivalent parameter points. We characterize the minimum eigenvalue of the Fisher matrix and the diameter of the set of observational equivalent parameter points in terms of optimization problems only involving transformations of the Hellinger pseudo-metric function. This characterization synthetizes existing qualitative criteria for identifiability and also uncovers a novel method for formalizing the notion of identifying strength using techniques from convex analysis.

The goodness-of-identifiability criterion seems the first measure available for quantifying the identifying strength of a parametric statistical model. The other two existing criteria for identifiability in parametric models, namely the Fisher matrix and Hellinger distance criteria, fall short as quantitative measures of identifying strength. The Fisher matrix criterion cannot quantify differences between point-from set-identifiability when the parameter is an irregular point of the Fisher matrix. Models with irregular points in the parameter space include the normal instrumental variable model (Hausman, 1974), the normal sample selection model (Lee and Chesher, 1986), and the skew-normal location- scale model (Hallin and Ley, 2012). Unlike the Fisher matrix criterion, the goodness-of-identifiability criterion quantifies identifying power at regular and irregular points of the Fisher matrix. The Hellinger distance criterion, in turn, cannot quantify differences between different degrees of set identifiability. Models with different degrees of set-identifiability include the normal switching regression model (Vijverberg, 1993) and parametric finite mixture models (Tamer, Chen, and Ponomareva, 2014). Unlike the Hellinger distance criterion, the goodness-of-identifiability criterion can quantify different levels of set-identifying strength.

We now place the goodness-of-identifiability criterion in the context of the existing literature. The Fisher matrix criterion for identifiability was introduced by Rothenberg (1971). It was related to the Kullback-Liebler divergence by Bowden (1973). The inability of the Fisher matrix criterion for distinguishing identifiability from lack of it when the parameter is an irregular point was noticed by, e.g., Stoica and Soderstrom (1982) and Sargan (1983). The Hellinger distance criterion for identifiability was introduced by Beran (1977). It was related to the Kullback-Liebler divergence by Pacini (2022). Unlike the Fisher matrix criterion, the

Hellinger distance criterion distinguishes identifiability from lack of it when the parameter is an irregular point, see e.g., Pacini (2022). The Hellinger distance criterion, however, cannot distinguish different degrees of set identifiability. The goodness-of-identifiability criterion introduced in this note fills this gap and relates the Fisher matrix and Hellinger distance criteria.

2. Definitions and Methods

2.1 Parametric Statistical Models

Let $\boldsymbol{Y_i}$ denote a random vector taking values on a sample space $\boldsymbol{\mathcal{Y}}$. The available data $\{\boldsymbol{Y_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$ are $\boldsymbol{N}$ independent and identically distributed replications of $\boldsymbol{Y_i}$. Let $\boldsymbol{P_\theta}$ be a probability function defined on the measurable space $(\boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{A}})$ and parametrized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The set $\boldsymbol{\mathcal{A}}$ is the $\boldsymbol{\sigma}$-field of Borel subsets $\boldsymbol{A} \in \boldsymbol{\mathcal{Y}}$. The parameter space $\boldsymbol{\Theta}$ is a subset of the Euclidean space $\mathbb{R}^{\boldsymbol{K}}$ for a positive integer $\boldsymbol{K}$.

We assume that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{P_\theta}$ is absolutely continuous with respect to a $\boldsymbol{\sigma}$-finite measure $\boldsymbol{\mu}$. Let $\boldsymbol{f_\theta} = \mathrm{d}\boldsymbol{P_\theta}/\mathrm{d}\boldsymbol{\mu}$ denote the density of $\boldsymbol{P_\theta}$ with respect to $\boldsymbol{\mu}$. The parametric statistical model is $\boldsymbol{\mathcal{F}_\Theta} = \{\boldsymbol{f_\theta}\}_{\boldsymbol{\theta \in \Theta}}$. We now impose the regularity conditions considered by Rothenberg (1971). We maintain them through the rest of this note.

Assumption 2.1 *(Regularity Conditions).* $\boldsymbol{\mathcal{F}_\Theta}$ *is such that: (i)* $\boldsymbol{\Theta}$ *is an open subset in* $\mathbb{R}^{\boldsymbol{K}}$ *; (ii)* $\boldsymbol{f_\theta} \geq \mathbf{0}$ *and* $\int \boldsymbol{f_\theta}\, \boldsymbol{d\mu} = \mathbf{1}$ *for every* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *; (iii)* $\boldsymbol{supp}(\boldsymbol{f_\theta}) := \{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}} : \boldsymbol{f_\theta} > \mathbf{0}\}$ *is the same for every* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*; (iv) for all* $\boldsymbol{\theta}$ *in a convex set containing* $\boldsymbol{\Theta}$ *and for all* $\boldsymbol{y} \in \boldsymbol{supp}(\boldsymbol{f_\theta})$*, the functions* $\boldsymbol{\theta} \mapsto \boldsymbol{f_\theta}$ *and* $\boldsymbol{\theta} \mapsto \boldsymbol{\ell}(\boldsymbol{\theta}) := \boldsymbol{ln}\, \boldsymbol{f_\theta}$ *are continuously differentiable* $\boldsymbol{\mu}$*-a.e.; (v) the elements of the matrix* $\mathbb{E}[\boldsymbol{\nabla\ell}(\boldsymbol{\theta})\boldsymbol{\nabla\ell}(\boldsymbol{\theta})']$ *are finite and continuous functions of* $\boldsymbol{\theta}$ *everywhere in* $\boldsymbol{\Theta}$.

Pacini (2022) presents examples and counterexamples illustrating Assumption 2.1

2.2 Local Identifiability and Regular Points

The Fisher matrix is the variance-covariance matrix of the score:

$$\boldsymbol{\nabla\ell}(\boldsymbol{\theta}) := \boldsymbol{\nabla}\mathbf{ln}\, \boldsymbol{f_\theta},\ \ \boldsymbol{where}\ \boldsymbol{\mathcal{J}}(\boldsymbol{\theta}): \\ = \mathbb{E}[\boldsymbol{\nabla\ell}(\boldsymbol{\theta})\boldsymbol{\nabla\ell}(\boldsymbol{\theta})'] - \mathbb{E}[\boldsymbol{\nabla\ell}(\boldsymbol{\theta})]\mathbb{E}[\boldsymbol{\nabla\ell}(\boldsymbol{\theta})]'$$

The following definitions, of local identifiability and regular point to the Fisher matrix, are from Rothenberg (1971).

Definition 2.1 *(Local Identifiability). A parameter point* $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$ *is locally identifiable if there exists an open neighborhood of* $\boldsymbol{\theta_0}$ *containing no other* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *such that* $\boldsymbol{f_\theta} = \boldsymbol{f_{\theta_0}}$.

Definition 2.2 *(Regular Point). A parameter point* $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$ *is a regular point of the matrix* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta_0})$ *if there exists an open neighborhood of* $\boldsymbol{\theta_0}$ *in which* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})$ *has constant rank.*

Lewbel (2019) presents examples illustrating the concept of local identifiability. The next example illustrates the concept of regular point, which has been less illustrated in the literature.

Example 2.1 *(Normal Squared Location Model). Set* $\boldsymbol{\mathcal{Y}} = \mathbb{R}$ *and* $\boldsymbol{\Theta} = \mathbb{R}$. *Consider the normal squared location model*

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

This model would, for example, arise if $\boldsymbol{Y_i}$ *is the difference between a matched pair of random variables whose control and treatment labels are not observed. The Fisher matrix is* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) = \boldsymbol{4\theta^2}$, *see e.g., Pacini (2022). For* $\boldsymbol{\theta} \neq \boldsymbol{0}$, $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})$ *has rank one. For* $\boldsymbol{\theta} = \boldsymbol{0}$, $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})$ *has rank zero. We deduce that* $\boldsymbol{\theta} = \boldsymbol{0}$ *is an irregular point of the Fisher matrix.*

Other models with irregular points of the Fisher matrix include, as already mentioned, the normal instrumental variable model (Hausman, 1974), the normal sample selection model (Lee and Chesher, 1986), and the skew-normal location-scale model (Hallin and Ley, 2012).

2.3 Fisher Matrix Criterion

For regular points of the Fisher matrix, we have the following characterization of local identifiability, see e.g., Rothenberg (1971, Theorem 1).

Lemma 2.1 *(Rothenberg, 1971, Theorem 1) Let* $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$ *be a regular point of* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta_0})$. *The point* $\boldsymbol{\theta_0}$ *is locally identifiable if and only if* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta_0})$ *is non-singular.*

All the proofs are in the Appendix. Lemma 1 does not apply to irregular points in the parameter space.

We are going to investigate how to obtain a characterization of local identifiability applying to all the points of the parameter space. We find convenient to make use of the following three concepts in this investigation: *Hellinger distance*, the *diameter* of a set in $\mathbb{R}^{\boldsymbol{K}}$, and *equivalent class*. We next review these concepts for the sake of completeness.

2.4 Hellinger Distance

The Hellinger distance between the densities $\boldsymbol{f_\theta}$ and $\boldsymbol{f_{\theta_0}}$ is

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

Fix $\boldsymbol{\theta_0}$ in $\boldsymbol{\Theta}$. The Hellinger pseudo-metric function is $\boldsymbol{\theta} \mapsto \boldsymbol{\rho(\theta)} := \boldsymbol{\rho(\theta,\theta_0)}$. Since $\boldsymbol{\rho(\theta)}$ can be zero for some $\boldsymbol{\theta}$ different from $\boldsymbol{\theta_0}$ is not a metric function. We have the following result

Lemma 2.2 *(Hellinger Pseudo-metric Function). $\boldsymbol{\rho}$ can take values between 0 and 1, which are independent of the choice of the dominating measure $\boldsymbol{\mu}$. $\boldsymbol{\rho(\theta)} = \mathbf{0}$ if and only if $\boldsymbol{f_\theta} = \boldsymbol{f_{\theta_0}}$.*

The following example illustrates the Hellinger pseudo-metric function.

Example 2.1 *(Normal Squared Location Model, Continued). The Hellinger pseudo-metric function in this model is $\boldsymbol{\rho(\theta)} = \mathbf{1} - \boldsymbol{exp}\left(-\left(\boldsymbol{\theta}^{\mathbf{2}} - \boldsymbol{\theta}_{\mathbf{0}}^{\mathbf{2}}\right)/\mathbf{8}\right)$, see e.g., Pacini (2022).*

2.5 Diameter

Let $\boldsymbol{S}$ be a nonempty convex set in $\mathbb{R}^{\boldsymbol{K}}$. Let $\mathbb{S} = \{\boldsymbol{q} \in \mathbb{R}^{\boldsymbol{K}}: \|\boldsymbol{q}\| = \mathbf{1}\}$ denote the unit sphere in $\mathbb{R}^{\boldsymbol{K}}$. The support $\boldsymbol{\delta_S(q)}$ and width $\boldsymbol{\omega_S(q)}$ functions of $\boldsymbol{S}$ in the direction $\boldsymbol{q} \in \mathbb{S}$ are $\boldsymbol{\delta_S(q)} := \boldsymbol{sup_{s \in S}}\,\langle \boldsymbol{q}, \boldsymbol{s}\rangle$ and $\boldsymbol{\omega_S(q)} := \boldsymbol{\delta_S(q)} + \boldsymbol{\delta_S(-q)}$, respectively. The following example illustrates the support and with functions of different convex sets.

Example 2.2 *(Support and Width Functions). Consider first a singleton $\boldsymbol{S} = \{\boldsymbol{s}\}$. We have $\boldsymbol{\delta_{\{s\}}(q)} = \boldsymbol{q's}$ and $\boldsymbol{\omega_{\{s\}}(q)} = \boldsymbol{q's} - \boldsymbol{q's} = \mathbf{0}$. Consider now the Euclidean unit ball $\mathbb{B} := \{\boldsymbol{q} \in \mathbb{R}^{\boldsymbol{K}}: \|\boldsymbol{q}\| \leq \mathbf{1}\}$. The support function is $\boldsymbol{\delta_{\mathbb{B}}(q)} = \boldsymbol{q'q} = \|\boldsymbol{q}\|^{\mathbf{2}} = \mathbf{1}$ and the width is function is $\boldsymbol{\omega_{\mathbb{B}}(q)} = \|\boldsymbol{q}\|^{\mathbf{2}} + \|\boldsymbol{-q}\|^{\mathbf{2}} = \mathbf{2}$. Both functions are constant; the support function is the radius,* *and the width function is the diameter. Finally, consider the case when* $\boldsymbol{S}$ *is an ellipse in* $\mathbb{R}^{\mathbf{2}}$. *The support function is the signed distance from the origin to a supporting plane. The width is the length of a chord in a direction.*

For later use, we now use the width function to characterize the diameter of the set of minimizers a continuous function $\boldsymbol{f}: \mathbb{R}^{\boldsymbol{K}} \rightarrow \mathbb{R}$. We resort to the conjugate $\boldsymbol{f}^{*}$ of the lower semi-continuous regularization of the extended value extension of $\boldsymbol{f}$ defined by

$$\boldsymbol{f}^{*}(\boldsymbol{y}) = \boldsymbol{sup}_{\boldsymbol{x} \in \boldsymbol{cl}(\boldsymbol{C})}\{\langle \boldsymbol{x}, \boldsymbol{y} \rangle - \boldsymbol{f}(\boldsymbol{x})\},$$

where $\boldsymbol{cl}(\boldsymbol{C})$ denotes the closure of a convex subset $\boldsymbol{C} \subseteq \mathbb{R}^{\boldsymbol{K}}$. We have the following result.

Lemma 2.3 *Let* $\boldsymbol{f}: \mathbb{R}^{\boldsymbol{K}} \rightarrow [\mathbf{0}, \mathbf{1}]$ *be a continuous function that is convex relative to the non-empty open convex set* $\boldsymbol{C} \subseteq \mathbb{R}^{\boldsymbol{K}}$ *with* $\boldsymbol{inf}\ \boldsymbol{f} := \boldsymbol{inf}_{\boldsymbol{x} \in \boldsymbol{C}}\ \boldsymbol{f}(\boldsymbol{x})$. *Then, the set of minimizers* $\boldsymbol{arg}\ \boldsymbol{inf}_{\boldsymbol{x} \in \boldsymbol{C}}\ \boldsymbol{f}(\boldsymbol{x})$ *is a non-empty convex set with diameter*

$$\boldsymbol{diam}(\boldsymbol{arg}\ \boldsymbol{inf}_{\boldsymbol{x} \in \boldsymbol{C}}\ \boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{sup}_{\boldsymbol{q} \in \mathbb{S}} \boldsymbol{\omega}_{\boldsymbol{\partial f}^{*}(\mathbf{0})}(\boldsymbol{q})$$

where $\boldsymbol{\partial f}^{*}(\mathbf{0})$ *is the subdifferential of* $\boldsymbol{f}^{*}$ *evaluated at 0.*

2.6 Equivalence Class

Let s and s_0 be two points in a set $\mathcal{S}$ A binary relationship $s \sim \mathrm{s}_0$ is an equivalence if and only if it is reflexive ($s \sim s\ for\ any\ s$), symmetric ($s \sim \mathrm{s}_0$ if and only if $\mathrm{s}_0 \sim s$ for any $s, \mathrm{s}_0\ in\ \mathcal{S}$) and transitive (if $s \sim \tilde{s}$ and $\tilde{s} \sim \mathrm{s}_0$, then s$\sim \mathrm{s}_0$ for any $s, \tilde{s}, \mathrm{s}_0\ in\ \mathcal{S}$). We define the equivalence class of s_0 under $\sim$ as $[\mathrm{s}_0]_{\mathcal{S}} = \{s \in \mathcal{S}: s \sim \mathrm{s}_0\}$. The following example illustrates this notion.

Example 2.3 *(Equivalence class) Two parameter points* θ *and* θ_0 *are observationally equivalent if* $f_\theta = f_{\theta_0}$. *Let denote this binary relationship by* $\theta \sim \theta_0$. *It is reflexive, symmetric, and transitive. The equivalence class* $[\theta_0]_\Theta = \{\theta \in \Theta: \theta \sim \theta_0\}$ *is known as the identified set of* θ_0.

3. Main Results

We are now able to quantify the identifying strength of a model for a point θ_0 in the parameter space.

3.1 Identifying Negentropy

We begin by relating the concept of local identifiability in Definition 2.1 to the smallest eigenvalues of the Fisher matrix. This matrix, being a variance-covariance matrix, is positive semi-definite. Since a positive semi-definite matrix is non-singular if and only if its smaller eigenvalue is positive, one can re-state Lemma 2.1 as follows.

Lemma 3.1 *Let θ_0 be a regular point of the Fisher matrix. This point is locally identifiable if and only if $ln\left(1 + e_K\big(\mathcal{I}(\theta_0)\big)\right) > 0$, where $e_K\big(\mathcal{I}(\theta_0)\big)$ denotes the smallest eigenvalue of the matrix $\mathcal{I}(\theta_0)$.*

The identifying negentropy of the model $\mathcal{F}_\Theta$ at θ_0 is

$$ln\left(1 + e_K\big(\mathcal{I}(\theta_0)\big)\right).$$

It is a measure, certainly not the only one, associated to the ability of the statistical model to discern nearby points in the parameter space. It misses, however, the inability of the model to discern nearby points in the parameter space. We next introduce a measure of this inability.

3.2 Identifying Entropy

Let Θ_0 denote an open convex subset of Θ such that $\theta_0 \in \Theta_0$ and the Hellinger pseudo-metric function is a convex function when defined on it. Θ_0 is non-empty. It is not a singleton because $\rho\colon \Theta \to [0,1]$ is locally convex around θ_0 (see Lemma 3.3 below). We have the following result concerning the convexity of the set of observational equivalent values to θ_0

Lemma 3.2 *The equivalent class $[\theta_0]_{\Theta_0} := \{\theta \in \Theta_0 \colon \theta \sim \theta_0\}$ is a non-empty convex subset of Θ.*

We now define the identifying entropy in terms of the diameter of $[\theta_0]_{\Theta_0}$ as

$$ln\left(1 + diam\big([\theta_0]_{\Theta_0}\big)\right),$$

where we set $ln(\infty) = \infty$. The identifying entropy is a measure, certainly not the unique one, of the inability of a statistical model to distinguish nearby values in the parameter space.

3.3 Goodness-of-Identifiability Criterion

Define the local goodness-of-identifiability criterion $\mathcal{g}(\theta_0)$ evaluated at θ_0 as the difference between the identifying negentropy and the identifying entropy

$$\mathcal{g}(\theta_0) := ln\left(1 + e_K\big(\mathcal{I}(\theta_0)\big)\right) - ln\left(1 + diam\big([\theta_0]_{\Theta_0}\big)\right).$$

The function $\theta \to \mathcal{g}(\theta)$ takes values on the extended real line. It quantifies the variation in the identifying strength of a parametric statistical model. We have the following result.

Lemma 3.3 *The point* θ_0 *is locally identifiable if and only if the identifying negentropy is greater or equal than the identifying entropy:* $\mathcal{g}(\theta_0) \geq 0$.

This result characterizes local identifiability for any point in the parameter space. It generalizes the main result in Rothenberg (1971, Theorem 1) by considering irregular points of the Fisher matrix, c.f., Lemma 3.3, and Lemma 2.1.

We now characterize the local goodness-of-identifiability criterion in terms of the Hellinger pseudo-metric function. This characterization relates the identifying entropy and negentropy to convex optimization problems. We proceed by relating the Fisher matrix to the Hellinger pseudo-metric function.

Lemma 3.4 *Assume that* $\theta \to f_\theta^{1/2}$ *is continuously differentiable* $\mu\, a.e.$. *Then,* $\mathcal{I}(\theta_0) = 4\nabla^2\rho(\theta_0)$, *where* $\nabla^2\rho(\theta_0)$ *denotes the matrix of second partial derivates evaluated at* θ_0 *of the Hellinger pseudo-metric function.*

The assumption on the differentiability of $\theta \to f_\theta^{1/2}$ is mild given that we have already assumed $\theta \to ln\, f_\theta$ that is continuously differentiable. Since the Fisher matrix is a variance-covariance matrix, it is positive semi-definite. It follows from Lemma 3.4, by the characterization of a convex function in Rockafellar and Wets (1998, Theorem 2.14), that the Hellinger pseudo-metric function is a locally convex function. Since Hellinger pseudo-metric function is also bounded between zero and one, one is then justified to use the characterization of the minimizers of convex functions in Lemma 2.3 to obtain the following characterization of the goodness-of-identifiability criterion.

Theorem 3.1 *Let Assumption 2.1 and the assumption in Lemma 3.4. Then,*

$$\mathscr{g}(\theta_0) = inf_{q\in\mathbb{S}}\langle q, \nabla^2\rho(\theta_0)q\rangle - sup_{q\in\mathbb{S}}\omega_{\partial\rho^*(0)}(q)$$

and the point θ_0 *is locally identifiable if and only if*

$$inf_{q\in\mathbb{S}}\langle q, \nabla^2\rho(\theta_0)q\rangle \geq sup_{q\in\mathbb{S}}\omega_{\partial\rho^*(0)}(q).$$

Two remarks are in order. First, when the Hellinger pseudo-metric function has a unique minimizer, one has that the identifying entropy is zero and θ_0 is locally identifiable even if the Fisher matrix is singular. This case was not covered by the criterion based on the Fisher matrix. Second, the objective functions in the optimization problems are both convex functions. One could use this result, for instance, for constructing a test for checking that a point θ_0 is locally identifiable. This construction, which is of practical importance in applications where identifiability is costly to deduce by analytical calculations, is out of the scope of this note and it is left for future research.

4. Conclusion

This note provides a novel criterion quantifying the notion of identifying strength of a parametric statistical model. This criterion, unlike the qualitative identifying criteria based on the Fisher matrix, applies to regular and irregular points in the parameter space. It also offers a characterization of the set of observational equivalent values in the parameter space in terms of the Hellinger pseudo-metric function. These are novel theoretical advances towards the quantification of the identifying strength of econometric models.

ACKNOWLEDGEMENTS. I would like to thank Santiago Aceranza and seminar participants at ORT Uruguay for useful comments on a previous version of this note.

References

[1] Beran, R. (1977). Minimum Hellinger Distance Estimates of Parametric Models, Annals of Statistics, 5(3), pp. 445-463.
[2] Bowden, R. (1973). The Theory of Parametric Identification, Econometrica, 41(6), pp. 1069-1074.
[3] Hallin, M. and C. Ley (2012) Skew Symmetric Distributions and Fisher Information: The Double Sin of the Skew-Normal. Bernoulli, 20(3), 1432-1453.
[4] Hausman, J. (1974). Full Information Instrumental Variable Estimation of Simultaneous Equation Systems. Annals of Economic and Social Measurement, 4(3), pp. 641-652.
[5] Lee, L. and A. Chesher (1986). Specification Testing when Score Test Statistics are Identically Zero. Journal of Econometrics, 31(2), pp. 121-149.
[6] Lewbel, A. (2019). The Identification Zoo: Meanings of Identification in Econometrics. Journal of Economic Literature, 57(4), pp. 835-903.
[7] Pacini, D. (2022). Identification in Parametric Models: The Hellinger Distance Criterion, 10(1).
[8] Rockafellar, T. and R. Wets (1998). Variational Analysis. Springer, Berlin.
[9] Rothenberg, T. (1971). Identification in Parametric Models. Econometrica, 39(3), pp. 577-591.
[10] Sargan, D. (1983). Identification and Lack of Identification. Econometrica, 51(6), pp. 1605-1633.
[11] Stoica, P. and T. Soderstrom (1982). On Non-singular Information Matrices and Local Identifiability. International Journal of Control, 36(2) pp. 323-329.
[12] Tamer, E., X. Chen and M. Ponomareva (2014). Likelihood Inference in Some Finite Mixture Models. Journal of Econometrics, 182(1) pp. 87-99.
[13] Vijverberg, W. (1993). Measuring the Unidentified Parameter of the Extended Roy Model. Journal of Econometrics, 57(1), pp. 69-89.

Appendix: Proofs

Proof of Lemma 2.1. This result was established by Rothenberg (1971, Theorem 1). For the sake of completeness, we replicate the proof in Rothenberg (1971).

We first show that if θ_0 is not locally identifiable, then the Fisher matrix is singular. Assume that θ_0 is not locally identifiable. By the Mean Value Theorem, there is a point θ_* between θ_0 *and* θ such that

$$\ell(\theta) - \ell(\theta_0) = \nabla\ell(\theta_*)'(\theta - \theta_0).$$

Then, there is a sequence $\{\theta_j \in \Theta\}_{j\in\mathbb{N}}$ converging to θ_0 such that $\ell(\theta_j) - \ell(\theta_0) = 0$. After dividing both sides of the last display by $\|\theta_j - \theta_0\|$, this implies $\nabla\ell(\theta_*)'q_j = 0, where\ q_j := (\theta_j - \theta_0)/\|\theta_j - \theta_0\|$. The sequence $\{q_j \in \mathbb{S}\}_{j\in\mathbb{N}}$ converges to a limit $q_0 \in \mathbb{S}$ (passing to a subsequence if necessary) because $\mathbb{S}$ is compact. As θ_j approaches θ_0, q_j approaches q_0 and $\nabla\ell(\theta_*)'q_j$ approaches $\nabla\ell(\theta_0)'q_0 = 0$. But this implies

$$0 = q_0{}'\mathbb{E}[\nabla\ell(\theta_0)\nabla\ell(\theta_0)']\, q_0 = q_0{}'\mathcal{I}(\theta_0)\, q_0,$$

Where the second equality follows from observing that $\mathbb{E}[\nabla\ell(\theta)] = 0$. Hence, $\mathcal{I}(\theta_0)$ must be singular.

We now verify the converse. Let θ_0 be a regular point of the Fisher matrix. Suppose that $\mathcal{I}(\theta_0)$ has constant rank $r < K$ in a neighborhood of θ_0. Consider the eigenvalue v_{θ_0} associated to one of the zero eigenvalues of $\mathcal{I}(\theta_0)$. Since $v_{\theta_0}{}'\mathcal{I}(\theta_0)v_{\theta_0} = 0$, we have that, for all θ in a neighborhood of θ_0,

$$v_\theta{}'\mathcal{I}(\theta) = 0.$$

Since $\mathcal{I}(\theta)$ has constant rank and $\theta \to \mathcal{I}(\theta)$ is continuous, the function $\theta \to v_\theta$ is continuous in a neighborhood of θ_0. Consider now the curve $\gamma: [0, t_1] \to \Theta$ defined as the solution to the differential equation $\frac{d\gamma(t)}{dt} = v_{\gamma(t)}$ with initial condition $\gamma(0) = \theta_0$. The log density function is differentiable in t with

$$\frac{d\ell(\gamma(t))}{dt} = v_{\gamma(t)}\nabla\ell(\gamma(t))$$

By the preceding display this is zero for all $0 \leq t \leq t_1$. Thus, $t \to \ell(\gamma(t))$ is constant and θ_0 is not locally identifiable.

Proof of Lemma 2.2. Write

$$\boldsymbol{\rho}(\boldsymbol{\theta}) = \frac{1}{2}\left\|\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2} - \boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\right\|_{L_2(\mu)}^2 = \frac{1}{2}\int\left(\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2} - \boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\right)^2 \boldsymbol{d\mu} = 1 - \int \boldsymbol{f}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{d\mu}.$$

From the last expression, notice that $\boldsymbol{\rho}(\boldsymbol{\theta}) = \mathbf{0}$ if and only if $\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2} = \boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}$ and $\boldsymbol{\rho}(\boldsymbol{\theta}) = \mathbf{1}$ if and only if $\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2} = \mathbf{0}$. To verify that $\boldsymbol{\rho}(\boldsymbol{\theta})$ does not depend on the choice of the dominating measure, we need additional notation. Let $\boldsymbol{g}_{\boldsymbol{\theta}}$ and $\boldsymbol{g}_{\boldsymbol{\theta}_0}$ denote, respectively, the densities of $\boldsymbol{P}_{\boldsymbol{\theta}}$ and $\boldsymbol{P}_{\boldsymbol{\theta}_0}$ relative to a dominating measure $\boldsymbol{v}$ different from $\boldsymbol{\mu}$. Let $\boldsymbol{h}_{\boldsymbol{\theta}}$ and $\boldsymbol{h}_{\boldsymbol{\theta}_0}$ denote, respectively, the densities of $\boldsymbol{P}_{\boldsymbol{\theta}}$ and $\boldsymbol{P}_{\boldsymbol{\theta}_0}$ relative to the measure $\boldsymbol{\mu} + \boldsymbol{v}$. Let $\boldsymbol{m}$ and v denote, respectively, the densities of $\boldsymbol{\mu}$ and $\boldsymbol{v}$ relative to $\boldsymbol{\mu} + \boldsymbol{v}$. We have $\boldsymbol{h}_{\boldsymbol{\theta}} = \boldsymbol{f}_{\boldsymbol{\theta}}\boldsymbol{m} = \boldsymbol{g}_{\boldsymbol{\theta}}\boldsymbol{v}$ and $\boldsymbol{h}_{\boldsymbol{\theta}_0} = \boldsymbol{f}_{\boldsymbol{\theta}_0}\boldsymbol{m} = \boldsymbol{g}_{\boldsymbol{\theta}_0}\boldsymbol{v}$. Hence, and $\left(\boldsymbol{f}_{\boldsymbol{\theta}}\boldsymbol{f}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{m} = \left(\boldsymbol{g}_{\boldsymbol{\theta}}\boldsymbol{g}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{v}$ and

$$\int\left(\boldsymbol{g}_{\boldsymbol{\theta}}\boldsymbol{g}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{dv} = \int\left(\boldsymbol{g}_{\boldsymbol{\theta}}\boldsymbol{g}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{vd}(\boldsymbol{\mu} + \boldsymbol{v}) = \int\left(\boldsymbol{f}_{\boldsymbol{\theta}}\boldsymbol{f}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{md}(\boldsymbol{\mu} + \boldsymbol{v}) = \int\left(\boldsymbol{f}_{\boldsymbol{\theta}}\boldsymbol{f}_{\boldsymbol{\theta}_0}\right)^{1/2}\boldsymbol{d}\,\boldsymbol{\mu},$$

which completes the proof.

Proof of Lemma 2.3. The set of minimizers $\operatorname{argmin}_{x\epsilon\mathcal{C}} f(x)$ is non-empty because we have assumed that $inf\ f = \min_{x\epsilon\mathcal{C}} f(x)$. Define the lower semi-continuous (lsc) regularization of $f(x)$ as

$$\check{f}(x) := liminf_{\tilde{x}\to x}\ \mathrm{f(x)},\ \ x\epsilon\mathcal{C}.$$

Define the extended value extension of $f(x)$ as

$$\tilde{f}(\mathrm{x}) = \begin{cases} f(x)\ if\ x\epsilon\mathcal{C} \\ \infty\ \ if\ \ x \notin \mathcal{C} \end{cases}.$$

Since f is convex relative to $\mathcal{C}$, $\tilde{f}$ is a proper lsc convex function. It follows, from Rockafellar and Wets (1998, Theorem 2.6), that $\operatorname{argmin}_{x\epsilon\mathcal{C}} f(x) = \operatorname{argmin}_{x\epsilon\mathbb{R}^K} \tilde{f}(x)$ is convex.

We now characterize the diameter of $\operatorname{argmin}_{x\epsilon\mathcal{C}} f(x)$ in terms of its support function. To avoid clutter in the notation, denote $X_* := \operatorname{argmin}_{x\epsilon\mathcal{C}} f(x)$. For any two points x and $\tilde{x}$, let $d(x,\tilde{x}) = \|x - \tilde{x}\|_2$ denote their Euclidean distance. The diameter of X_* is defined as

$$diam(X_*) := sup_{x,\tilde{x}\in X_*}\, d(x,\tilde{x}).$$

For the case when X_* is unbounded, it suffices to note that $diam(X_*) = \infty$ and $sup_{q\in\mathbb{S}}\omega_{X_*}(q) = \infty$. Consider now the case when X_* is bounded. We first derive a lower bound on $diam(X_*)$ in terms of the width function. Denote the upper bound of the width function by $\bar{\omega} := sup_{q\in\mathbb{S}}\omega_{X_*}(q)$. Let $q_* \in \mathbb{S}$ be a direction such that $\omega_{X_*}(q_*) = \bar{\omega}$. On each of the supporting hyperplanes perpendicular to q_*, there is a point in $cl(X_*)$. Hence, $\bar{\omega} \leq diam(X_*)$. We now derive an upper bound on $diam(X_*)$ in terms of the width function. There are points x_* and x_0 in X_* such that $diam(X_*) = \|x_* - x_0\|_2$. Consider now two hyperplanes passing each through x_* and x_0 such that they are perpendicular to x_* and x_0. These hyperplanes are supporting hyperplanes of X_*, for otherwise we could find two points in X_* at a distance apart greater than $diam(X_*)$. Hence, $\bar{\omega} \geq diam(X_*)$. We then deduce that

$$diam(X_*) = \bar{\omega} = sup_{q\in\mathbb{S}}\, \omega_{X_*}(q) = sup_{q\in\mathbb{S}}\, [\delta_{X_*}(q) + \delta_{X_*}(-q)].$$

We finally characterize the support function $q \to \delta_{X_*}(q)$. Fix $q \in \mathbb{S}$. Since $\tilde{f}: \mathbb{R}^K \to \overline{\mathbb{R}}$ is a proper lsc continuous function, it follows, from Rockafellar and Wets (1998, Theorem 11.8) that $X_* = \partial f^*(0)$, whence $\delta_{X_*}(q) = \delta_{\boldsymbol{\partial f^*(0)}}(q)$.

Proof of Lemma 3.1. In the text.

Proof of Lemma 3.2. Since the matrix of second derivates of the Hellinger pseudo-metric function is a positive semi-definite matrix (see Lemma 3.4), it follows from Rockafellar and Wets (1998, Theorem 2.14) that Θ_0 is a non-empty convex set. Moreover, since $\rho(\theta) = 0$ if and only if $\theta \sim \theta_0$ and otherwise $\rho(\theta) > 0$, one has $[\theta_0]_{\Theta_0} = argmin_{\theta\in\Theta_0}\, \rho(\theta)$. Since $\rho: \Theta_0 \to [0,1]$ is a convex function, one is justified to claim that $[\theta_0]_{\Theta_0}$ is a non-empty convex subset of Θ.

Proof of Lemma 3.3. (If) Assume that $g(\theta_0) \geq 0$. Consider first the case $g(\theta_0) = 0$. We have $ln\left(1 + e_K\left(\mathcal{J}(\theta_0)\right)\right) = ln\left(1 + diam\left([\theta_0]_{\Theta_0}\right)\right)$ so $e_K\left(\mathcal{J}(\theta_0)\right) = diam\left([\theta_0]_{\Theta_0}\right)$. Since any non-identifiable point θ_0 has $diam\left([\theta_0]_{\Theta_0}\right) > 0$, the last equality implies that $e_K\left(\mathcal{J}(\theta_0)\right) > 0$. This is in contradiction with our previous observation (see the proof of Lemma 2.1) that any non-identifiable point has singular Fisher matrix. We deduce then that $g(\theta_0) = 0$ implies that θ_0 is identifiable. Consider now the case $g(\theta_0) > 0$. We have $e_K\left(\mathcal{J}(\theta_0)\right) > diam\left([\theta_0]_{\Theta_0}\right) \geq 0$, whence $e_K\left(\mathcal{J}(\theta_0)\right) > 0$ and θ_0 is locally identifiable. (Only if) Assume now that θ_0 is locally identifiable. We have $diam\left([\theta_0]_{\Theta_0}\right) = 0$ and $g(\theta_0) = ln\left(1 + e_K\left(\mathcal{J}(\theta_0)\right)\right) \geq 0$, where the last equality follows from observing that the Fisher matrix is positive semi-definite.

Proof of Lemma 3.4. We now follow Pacini (2022, Lemma 4). Consider first the case when θ is a scalar, i.e., $K = 1$. Write

$$\boldsymbol{\rho}(\boldsymbol{\theta}) = \frac{1}{2}\left\|\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2} - \boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\right\|_{L_2(\boldsymbol{\mu})}^2 = \frac{1}{2}\int\left(\boldsymbol{f}_{\boldsymbol{\theta}}^{1/2} - \boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\right)^2 \boldsymbol{d\mu} = 1 - \int \boldsymbol{f}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{f}_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{d\mu}.$$

Differentiating the Hellinger pseudo-metric function

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

Since the Hellinger pseudo-metric function reaches a minimum at θ_0, one has $\nabla\rho(\theta_0) = 0$ and so

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

By the Lebesgue Dominated Convergence Theorem, the limit

$$\nabla^2\rho(\theta_0) := lim_{\theta\to\theta_0} \frac{\nabla\rho(\theta) - \nabla\rho(\theta_0)}{(\theta - \theta_0)}$$

satisfies $\nabla^2\rho(\theta_0) = \frac{1}{4}\,\mathcal{J}(\theta_0)$ b ecause the integrand in the preceding display converges pointwise

$$\frac{\left(f_\theta^{1/2} - f_{\theta_0}^{1/2}\right)\nabla f_\theta}{(\theta-\theta_0) f_\theta^{1/2}} \to \frac{1}{4}\nabla \ln f_{\theta_0} \nabla \ln f_{\theta_0} f_{\theta_0}$$

And it is dominated by a sum of integrable functions

$$\left|\frac{\left(f_\theta^{1/2} - f_{\theta_0}^{1/2}\right)\nabla f_\theta}{(\theta-\theta_0) f_\theta^{1/2}}\right| \le \frac{\left(f_\theta^{1/2} - f_{\theta_0}^{1/2}\right)^2}{(\theta-\theta_0)^2} + \frac{\nabla f_\theta \nabla f_{\theta'}}{f_\theta} \ .$$

To extend this argument to the case when θ is a vector, one applies the argument above elementwise to the components of $\nabla^2 \rho(\theta_0)$.

Proof of Theorem 3.1. By Lemma 3.4,

$$e_K\big(\mathcal{J}(\theta_0)\big) = e_K\big(4\nabla^2\rho(\theta_0)\big) = min_{q\in\mathbb{S}} \langle q, 4\nabla^2\rho(\theta_0) q\rangle,$$

Where the last equality follows from the Courant-Fisher Theorem characterization the smallest eigenvalue of a squared matrix. The claim then follows from Lemma 2.3 applied to $f = \rho$, $\mathcal{C} = \Theta_0$ after noticing, from Lemma 2.2 and 3.4, that $\rho: \Theta_0 \to [0,1]$ is a continuous convex function.