Statistical models and likelihood functions *

Jean-Marie Dufour †
Université de Montréal

First version: February 1998
This version: January 12, 2021, 5:06pm

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec).

† Canada Research Chair Holder (Econometrics). Centre de recherche et développement en économique (C.R.D.E.), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

Contents

List of Definitions, Propositions and Theorems ii

1. Statistical models 1

2. Identification 3

3. Likelihood and score functions 4

4. Efficiency bounds 8

List of Definitions, Assumptions, Propositions and Theorems

1.1 **Definition :** Statistical model 1
1.2 **Definition :** Dominated model 1
1.3 **Definition :** Homogeneous model 1
1.4 **Definition :** Parametric model 1
1.5 **Definition :** Functional parameter 2
2.1 **Definition :** Identification of a parameter value 3
2.2 **Definition :** Identification of a model 3
2.3 **Definition :** Identification of a parametric function 3
2.4 **Definition :** Local identification 3
3.1 **Definition :** Likelihood function 4
3.2 **Definition :** Score function 6
3.3 **Proposition :** Mean of a score 6
3.4 **Proposition :** Information matrix 6
3.5 **Proposition :** Information matrix identity 7
4.1 **Definition :** Regular estimator 8
4.2 **Theorem :** Fréchet-Darmois-Cramer-Rao bound 8

1. Statistical models

1.1 Definition STATISTICAL MODEL. *A* statistical model *is a pair* $(\mathcal{Z},\ \mathcal{P})$ *where* $\mathcal{Z}$ *is a set of possible observations and* $\mathcal{P}$ *a nonempty family of probability measures which assign probabilities to subsets of* $\mathcal{Z}$*. When the probability measures in* $\mathcal{P}$ *are all defined on the same* σ*-algebra of events* $\mathcal{A}_{\mathcal{Z}}$ *in* $\mathcal{Z}$*, we shall also refer to the triplet* $(\mathcal{Z},\ \mathcal{A}_{\mathcal{Z}},\ \mathcal{P})$ *as a statistical model.*

1.2 Definition DOMINATED MODEL. *A statistical model* $(\mathcal{Z},\ \mathcal{A}_{\mathcal{Z}},\ \mathcal{P})$ *is* dominated *if all the probability measures in* $\mathcal{P}$ *have a density with respect to the same measure* μ *on* $\mathcal{Z}$*.* μ *is called the dominating measure* and we say that $(\mathcal{Z},\ \mathcal{P})$ *is* μ*-dominated.*

1.3 Definition HOMOGENEOUS MODEL. *A statistical model* $(\mathcal{Z},\ \mathcal{A}_{\mathcal{Z}},\ \mathcal{P})$ *is* homogeneous *if it is dominated and the dominating measure* μ *can be chosen so that the densities are all strictly positive.*

1.4 Definition PARAMETRIC MODEL. *A statistical model* $(\mathcal{Z},\ \mathcal{P})$ *is said to be parametrized* by the elements of a nonempty set Θ *if the set* $\mathcal{P}$ *of probability* *measures has the form*

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\} \ .$$

If the set Θ is a subset of $\mathbb{R}^p$ or we can define a one-to-one transformation between Θ and the elements of a subset of $\mathbb{R}^p$, we say that $(\mathcal{Z}, \mathcal{P})$ is a parametric model*. Otherwise, the model $(\mathcal{Z}, \mathcal{P})$ is said to be* nonparametric*.*

1.5 Definition FUNCTIONAL PARAMETER. *A* functional parameter *on a statistical model $(\mathcal{Z}, \mathcal{P})$ is an application*

$$g : \mathcal{P} \to \Theta$$

which assigns to each element $P \in \mathcal{P}$ a parameter $\theta = g(P) \in \Theta$, where Θ is a nonempty set (the parameter space).

Functional parameters allow one to associate parameters with the distributions of parametric or nonparametric models. The mean, variance, median, etc., of a probability distribution may all be interpreted as functional parameters.

2. Identification

Let $(\mathcal{Z},\ \mathcal{A}_{\mathcal{Z}},\ \mathcal{P})$ a statistical model such that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

2.1 Definition IDENTIFICATION OF A PARAMETER VALUE. *We say that a parameter value $\theta_1 \in \Theta$ is* identifiable *if there is no other value $\theta_2 \in \Theta$ such that $P_{\theta_1} = P_{\theta_2}$.*

2.2 Definition IDENTIFICATION OF A MODEL. *We say that the model $(\mathcal{Z},\ \mathcal{A}_{\mathcal{Z}},\ \mathcal{P})$ is identifiable if all the elements of Θ are identifiable.*

2.3 Definition IDENTIFICATION OF A PARAMETRIC FUNCTION. *Let $\psi : \theta \longrightarrow \Psi$ be a function of θ. We say that the function $\psi(\theta)$ is* identifiable *if*

$$\psi(\theta_1) \neq \psi(\theta_2) \Rightarrow P_{\theta_1} \neq P_{\theta_2}\ , \forall \theta_1, \theta_2 \in \Theta$$

or, equivalently,

$$P_{\theta_1} = P_{\theta_2} \Rightarrow \psi(\theta_1) = \psi(\theta_2)\ \ , \forall \theta_1, \theta_2 \in \Theta\ .$$

2.4 Definition LOCAL IDENTIFICATION. *Suppose the set Θ has a set of neighborhoods defined on it (a topology). Then we say that a parameter value $\theta_1 \in$*

Θ *is* locally identifiable *if there is a neighborhood* $V(\theta_1)$ *of* θ_1 *such that*

$$\theta_2 \in V(\theta_1) \text{ and } \theta_2 \neq \theta_1 \Rightarrow P_{\theta_1} \neq P_{\theta_2} \, .$$

3. Likelihood and score functions

3.1 Definition LIKELIHOOD FUNCTION. *Let* $(\mathcal{Z}, \mathcal{P})$ *be a statistical model which satisfies the following assumptions:*

(A1) $(\mathcal{Z}, \mathcal{P})$ *is a* μ*-dominated model;*

(A2) $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$;

(A3) $L(z\,;\theta)$, $z \in \mathcal{Z}$, *is the density function* $($*with respect to* $\mu)$ *associated with* P_θ.

The density function $L(z;\theta)$ *viewed as a function of* θ *is called the* likelihood function *of model* $(\mathcal{Z}, \mathcal{P})$. *The symbol* $\underset{\theta}{E}(\cdot)$ *refers to the expected value with respect to* θ *(provided it exists) :*

$$\underset{\theta}{E}[h(Z)] = \int_{\mathcal{Z}} h(z)\, dP_\theta(z) = \int_{\mathcal{Z}} h(z)\, L(z\,;\theta)\, d\mu(z) \, .$$

The vector Z often has the form

$$Z = (Y_1', Y_2', \ldots, Y_n')'$$

where $Y_t \in \mathbb{R}^m$ is an "individual" observation vector and $\theta = (\theta_1, \theta_2, \ldots, \theta_p)' \in \Theta$. Usually, the density $L(z;\theta)$ is written in the form

$$L(z;\theta) = \prod_{t=1}^{n} f_t(z\,;\theta) \equiv L_n(z;\theta) \tag{3.1}$$

where $f_t(z;\theta)$ is a density for an "individual observation". $f_t(z;\theta)$ usually has one of the following forms :

$$f_t(z\,;\theta) = f(y_t\,;\theta)\ ,\ \ y_t \in \mathbb{R}^m \tag{3.2}$$

$$f_t(z\,;\theta) = f(y_t \mid x_t\,;\theta) \tag{3.3}$$

where x_t is a $k \times 1$ vector of conditioning variables ("explanatory variables") and $f(y_t\,;.)$ is the density function of y_t (given x_t) as a function of the parameter vector θ, or

$$L_t(z\,;\theta) = f(y_t \mid \bar{y}_{t-1}, x_t\,;\theta) \tag{3.4}$$

where $\bar{y}_{t-1} = (\bar{y}_0, y_1, \ldots, y_{t-1})'$ is a vector of past values of y and $\bar{y}_0$ is a vector of "initial conditions".

3.2 Definition SCORE FUNCTION. *Under the assumption* $(A1)$ *to* $(A3)$, *suppose also that:*

(A4) Θ *is an open set in* $\mathbb{R}^p$;

(A5) $\partial L(z\,;\theta)\,/\partial\theta$ *exists,* $\forall z \in \mathcal{Z}\ ,\forall \theta \in \Theta$;

(A6) $L(z\,;\theta) > 0\ ,\forall z \in \mathcal{Z}\ ,\forall \theta \in \Theta$;

(A7) $\int_{\mathcal{Z}} \frac{\partial}{\partial\theta}\left[L(z\,;\theta)\right] d\mu(z) = \frac{\partial}{\partial\theta}\left[\int_{\mathcal{Z}} L(z\,;\theta)\, d\mu(z)\right]$.

Then the function

$$S(z\,;\theta) = \frac{\partial}{\partial\theta}\left[\ln L(z\,;\theta)\right]\ ,\theta \in \Theta\ ,\ z \in \mathcal{Z}\ ,$$

is called the score function *associated with the likelihood* $L(z\,;\theta)$.

3.3 Proposition MEAN OF A SCORE. *Under the assumptions* $(A1)$ *to* $(A7)$, *we have :*

$$\underset{\theta}{E}\left[S(Z\,;\theta)\right] = \int_{\mathcal{Z}} S(z\,;\theta)\, L(z\,;\theta)\, d\mu(z) = 0\ .$$

3.4 Definition INFORMATION MATRIX. *In addition to* $(A1)$ *to* $(A7)$, *suppose also that:*

(A8) $S(Z\,;\theta)$ *has finite second moments with respect to* $P_\theta, \forall\theta \in \Theta$.

Then, the covariance matrix of $S(Z\,;\theta)$,

$$I(\theta) = V_\theta[S(Z\,;\theta)] = \underset{\theta}{E}\left[S(Z\,;\theta)\,S(Z\,;\theta)'\right]$$
$$= \int_{\mathcal{Z}} S(z\,;\theta)\,S(z\,;\theta)'\,L(z\,;\theta)\,d\mu(z)$$

is called the Fisher information matrix *associated with* $L(z\,;\theta)$.

3.5 Proposition INFORMATION MATRIX IDENTITY. *Under the assumptions* (*A1*) *to* (*A8*), *suppose also that:*

(A9) $\frac{\partial^2 L(z\,;\theta)}{\partial\theta\partial\theta'}$ *exists*, $\forall z \in \mathcal{Z}, \forall\theta \in \Theta$;

(A10) $\forall\theta \in \Theta$,

$$\int_{\mathcal{Z}} \frac{\partial^2 L(z;\theta)}{\partial\theta_i \partial\theta_j} d\mu(z) = \frac{\partial^2}{\partial\theta_i \partial\theta_j}\left[\int_{\mathcal{Z}} L(z;\theta)\,d\mu(z)\right] .$$

Then

$$I(\theta) = \underset{\theta}{E}\left[-\frac{\partial^2 \ln L(Z;\theta)}{\partial\theta\partial\theta'}\right] , \forall\theta \in \Theta .$$

4. Efficiency bounds

4.1 Definition REGULAR ESTIMATOR. *Under the assumptions* $(A1)$ *to* $(A5)$, *an estimator* $T(Z)$ *of some function* $\psi(\theta) \in \mathbb{R}^q$ *is* regular *if it satisfies the following properties:*

(a) $T(Z)$ *has finite second moments;*

(b) $\int_{\mathcal{Z}} T(z)\, L(z\,;\theta)\, d\mu(z)$ *is differentiable with respect to* θ;

(c) $\frac{\partial}{\partial\theta} \int_{\mathcal{Z}} T(z)\, L(z\,;\theta)\, d\mu(z) = \int_{\mathcal{Z}} T(z) \frac{\partial}{\partial\theta} [L(z\,;\theta)]\, d\mu(z)$, *for all* $\theta \in \Theta$.

4.2 Theorem FRÉCHET-DARMOIS-CRAMER-RAO BOUND. *Let the assumptions* $(A1)$ *to* $(A8)$ *hold, let* $\psi(\theta) \in \mathbb{R}^q$ *be a differentiable function of* θ, *and suppose that*

(A11) *the information matrix* $I(\theta)$ *is positive definite,* $\forall\theta \in \Theta$.

If $\underset{\theta}{E}[T(Z)] = \psi(\theta)\,, \forall\theta \in \Theta$, *then the difference*

$$V_\theta[T(Z)] - P(\theta)\, I(\theta)^{-1}\, P(\theta)'$$

is positive semi-definite for all $\theta \in \Theta$, *where* $P(\theta) = \partial\psi(\theta)/\partial\theta'$.

4.3 Remark If $\psi(\theta) = \theta$, this means that $V_\theta [T(Z)] - I(\theta)^{-1}$ is positive semi-definite.

References

GOURIÉROUX, C., AND A. MONFORT (1989): *Statistique et modèles économétriques, Volumes 1 et 2*. Economica, Paris.

LEHMANN, E. L. (1983): *Theory of Point Estimation*. John Wiley & Sons, New York.

——— (1986): *Testing Statistical Hypotheses*. John Wiley & Sons, New York, 2nd edn.