SUPPLEMENT TO “CONSISTENT PSEUDO-MAXIMUM LIKELIHOOD ESTIMATORS AND GROUPS OF TRANSFORMATIONS”
(Econometrica, Vol. 87, No. 1, January 2019, 327–345)

C. GOURIÉROUX
University of Toronto, TSE, and PSL, Dauphine

A. MONFORT
CREST, ENSAE

J-M. ZAKOÏAN
CREST, ENSAE and University of Lille

THIS DOCUMENT consists of three sections of additional results: (i) regularity conditions for Proposition 3; (ii) sketch of proof of Proposition 3; (iii) derivatives of functions based on exponential of matrices.

8. REGULARITY CONDITIONS FOR PROPOSITION 3

Let \( \vartheta = (\theta', \lambda') = (\partial_j)_{1 \leq j \leq K+J} \) and \( l_i(\vartheta) = l[a(x_i; \theta) + \lambda, y_i] \).
Together with Assumptions A.1–A.7, Proposition 3 requires the following assumptions:

ASSUMPTION A.9: \( \vartheta^* = (\theta_0', \lambda_0') \) belongs to the interior of \( \theta(B) \times A \).

ASSUMPTION A.10: For any \( x \), the function \( \theta \to a(x; \theta) \) has continuous third-order derivatives. The pseudo-density function \( g \) is three times continuously differentiable.

ASSUMPTION A.11: The matrices \( K \) and \( L \) defined in the proof are positive definite.

ASSUMPTION A.12: For at least one \( j \in \{1, \ldots, J\} \), the matrix \( V_0(\frac{\partial a_j}{\partial \theta}(x_i, \theta_0)) \) is positive definite.

ASSUMPTION A.13: There exists a neighborhood \( V(\vartheta^*_0) \) of \( \vartheta^*_0 \) such that, for \( i, j = 1, \ldots, r \), for all \( \vartheta \in V(\vartheta^*_0) \), the process \( \left\{ \frac{\partial}{\partial \vartheta} \left( \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \ell_t(\vartheta) \right) \right\} \) is strictly stationary and ergodic, and,

\[
E_0 \sup_{\vartheta \in V(\vartheta^*_0)} \left\| \frac{\partial}{\partial \vartheta} \left( \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \ell_t(\vartheta) \right) \right\| < \infty.
\]

9. PROOF OF PROPOSITION 3

In this section, we will explain how to use an appropriate central limit theorem (CLT), and we will derive the asymptotic covariance matrix.

---

C. Gouriéroux: Christian.Gourieroux@ensae.fr
A. Monfort: Alain.Monfort@ensae.fr
J-M. Zakoïan: Jean-Michel.Zakoian@ensae.fr

© 2019 The Econometric Society

https://doi.org/10.3982/ECTA14727
9.1. The Pseudo-Score

For model (6.1), the pseudo log-likelihood for one observation takes the form

\[
l_t(\theta, \lambda) = \log g \left[ \exp \left\{ \sum_{j=1}^{J} [a_j(x_t; \theta) + \lambda_j] C_j \right\} y_t \right] + \sum_{j=1}^{J} [a_j(x_t; \theta) + \lambda_j] \text{Tr}(C_j).
\]

Let

\[
z_t(\theta, \lambda) = \exp \left\{ \sum_{j=1}^{J} [a_j(x_t; \theta) + \lambda_j] C_j \right\} y_t.
\]

For \(\gamma = (\text{Tr}(C_1), \ldots, \text{Tr}(C_J))'\), we have

\[
l_t(\theta, \lambda) = \log g \{ z_t(\theta, \lambda) \} + \gamma' \{ a(x_t; \theta) + \lambda \}.
\]

Using the computations of Section 10.3, it follows that

\[
\frac{\partial}{\partial \theta'} l_t(\theta, \lambda) = \frac{\partial}{\partial u} \log g \{ z_t(\theta, \lambda) \} \frac{\partial z_t(\theta, \lambda)}{\partial \theta'} + \gamma' \frac{\partial a(x_t; \theta)}{\partial \theta'}
\]

\[
= \left( \frac{\partial}{\partial u} \log g \{ z_t(\theta, \lambda) \} C [I_J \otimes z_t(\theta, \lambda)] + \gamma' \right) \frac{\partial a(x_t; \theta)}{\partial \theta'}
\]

\[
:= h'_t(\theta, \lambda) \frac{\partial a(x_t; \theta)}{\partial \theta'}.
\]

Proceeding similarly with parameter \(\lambda\), we find that

\[
\frac{\partial}{\partial \lambda} l_t(\theta, \lambda) = \left( \frac{\partial}{\partial \theta} l_t(\theta, \lambda) \right) \left( \frac{\partial}{\partial \lambda} l_t(\theta, \lambda) \right) = \left( \frac{\partial a(x_t; \theta)}{\partial \theta} \right) h_t(\theta, \lambda).
\]

9.2. The Martingale Difference Property

Replacing \(y_t\) by \(\exp\{- \sum_{j=1}^{J} [a_j(x_t; \theta_0) + \lambda_{0j}] C_j\} u_t\), we find that

\[
z_t(\theta_0, \lambda_0^*) = \exp \left\{ \sum_{j=1}^{J} (\lambda_{0j}^* - \lambda_{0j}) C_j \right\} u_t := \Gamma(\lambda_0^* - \lambda_0) u_t.
\]

Thus,

\[
\frac{\partial}{\partial \theta} l_t(\theta_0, \lambda_0^*) = \left( \frac{\partial}{\partial \theta} l_t(\theta_0, \lambda_0^*) \right) \left( \frac{\partial}{\partial \lambda} l_t(\theta_0, \lambda_0^*) \right) = \left( \frac{\partial a(x_t; \theta_0)}{\partial \theta} \right) k(u_t),
\]

where

\[
k(u_t) = \left[ I_J \otimes (\Gamma(\lambda_0^* - \lambda_0) u_t) \right] C \frac{\partial}{\partial u} \{ \Gamma(\lambda_0^* - \lambda_0) u_t \} + \gamma.
\]
Noting that
\[ k(u_t) = [I_J \otimes (\Gamma(a^*_0)u_t)]C \frac{\partial \log g}{\partial u} \{\Gamma(a^*_0)u_t\} + \gamma, \]
where \( a^*_0 \) is defined in Assumption A.5, we have \( E\{k(u_t)\} = 0 \) from the first-order conditions in the generic model. Thus, using Assumption A.8, \( (\frac{\partial}{\partial \theta} l_t(\theta_0, \lambda_0^*)\}_{\cdot \cdot} ) \) is a martingale difference sequence, where \( F_t = \sigma\{(u_i, x_i), i \leq t\} \). The asymptotic normality follows from applying a CLT for the square integrable, ergodic, and stationary martingale difference (see Billingsley (1961)). We get
\[ V_{as} \left[ \sqrt{T} (\hat{\theta}_T - \theta_0) \right] = A^{-1} BA^{-1}, \]
where
\[ A = E_0 \left[ - \frac{\partial^2}{\partial \theta \partial \theta'} l_t(\theta_0, \lambda_0^*) \right], \quad B = V_0 \left[ \frac{\partial}{\partial \theta} l_t(\theta_0, \lambda_0) \right]. \]

9.3. Computation of the Asymptotic Covariance Matrix

We have
\[ B = V_0 \left[ \frac{\partial}{\partial \theta} l_t(\theta_0, \lambda_0^*) \right] = E_0 \left[ \left( \frac{\partial a' (x_i; \theta_0)}{\partial \theta} I_J \right) K \left( \frac{\partial a' (x_i; \theta_0)}{\partial \theta} I_J \right) \right], \quad K = V_0 \left[ k(u_t) \right]. \]

Now,
\[ \frac{\partial^2}{\partial \theta \partial \theta'} l_t(\theta, \lambda) = \frac{\partial h'_t(\theta, \lambda)}{\partial \theta} \frac{\partial a(x_i; \theta)}{\partial \theta'} + \left\{ h'_t(\theta, \lambda) \otimes I_p \right\} A(x_i; \theta), \]
where \( A(x_i; \theta) \) is the \( Jp \times p \) matrix:
\[ A(x_i; \theta) = \begin{pmatrix} \frac{\partial^2 a_1 (x_i; \theta)}{\partial \theta \partial \theta'} \\ \vdots \\ \frac{\partial^2 a_j (x_i; \theta)}{\partial \theta \partial \theta'} \end{pmatrix}. \]
Noting that
\[ h'_t(\theta, \lambda) = \left[ \frac{\partial \log g}{\partial u'} \{z_i(\theta, \lambda)\} C_1 z_i(\theta, \lambda), \ldots, \frac{\partial \log g}{\partial u'} \{z_i(\theta, \lambda)\} C_j z_i(\theta, \lambda) \right] + \gamma', \]
we compute
\[ \frac{\partial}{\partial \theta} \left\{ \frac{\partial \log g}{\partial u'} \{z_i(\theta, \lambda)\} C_j z_i(\theta, \lambda) \right\} \]
\[ = \frac{\partial}{\partial \theta} \left\{ \frac{\partial \log g}{\partial u'} \{z_i(\theta, \lambda)\} \right\} C_j z_i(\theta, \lambda) + \left[ \frac{\partial}{\partial \theta} \left\{ C_j z_i(\theta, \lambda) \right\} \right]' \left\{ \frac{\partial \log g}{\partial u'} \{z_i(\theta, \lambda)\} \right\}'. \]
\[
\begin{align*}
= \frac{\partial z_i(\theta, \lambda)}{\partial \theta} \frac{\partial^2 \log g}{\partial u \partial u'} \left\{ z_i(\theta, \lambda) \right\} C_j z_i(\theta, \lambda) \\
+ \frac{\partial a'(x_i; \theta)}{\partial \theta} \left[ I_J \otimes z_i(\theta, \lambda) \right] C'C_j \frac{\partial \log g}{\partial u} \left\{ z_i(\theta, \lambda) \right\} \\
= \frac{\partial a'(x_i; \theta)}{\partial \theta} \left[ I_J \otimes z_i(\theta, \lambda) \right] C' \\
\times \left\{ \frac{\partial^2 \log g}{\partial u \partial u'} \left\{ z_i(\theta, \lambda) \right\} C_j z_i(\theta, \lambda) + C_j' \frac{\partial \log g}{\partial u} \left\{ z_i(\theta, \lambda) \right\} \right\}.
\end{align*}
\]

It follows that

\[
\frac{\partial^2}{\partial \theta \partial \theta} l_i(\theta, \lambda)
= \frac{\partial a'(x_i; \theta)}{\partial \theta} \left[ I_J \otimes z_i(\theta, \lambda) \right] C' \\
\times \sum_j \left\{ \frac{\partial^2 \log g}{\partial u \partial u'} \left\{ z_i(\theta, \lambda) \right\} C_j z_i(\theta, \lambda) + C_j' \frac{\partial \log g}{\partial u} \left\{ z_i(\theta, \lambda) \right\} \right\} \left\{ \frac{\partial a_j(x_i; \theta)}{\partial \theta'} \right\} \\
+ \left\{ h'_i(\theta, \lambda) \otimes I_p \right\} A(x_i; \theta)
= \frac{\partial a'(x_i; \theta)}{\partial \theta} \left[ I_J \otimes z_i(\theta, \lambda) \right] C' \frac{\partial^2 \log g}{\partial u \partial u'} \left\{ z_i(\theta, \lambda) \right\} C \left[ I_J \otimes z_i(\theta, \lambda) \right] \frac{\partial a(x_i; \theta)}{\partial \theta'} \\
+ \left\{ h'_i(\theta, \lambda) \otimes I_p \right\} A(x_i; \theta),
\]

where \( G(u) \) is the \( n \times J \) matrix:

\[
G(u) = \begin{bmatrix}
C_1' \frac{\partial \log g}{\partial u} (u) & C_2' \frac{\partial \log g}{\partial u} (u) & \ldots & C_J' \frac{\partial \log g}{\partial u} (u)
\end{bmatrix}.
\]

Note that the first-order conditions imply \( Eh'_i(\theta_0, \lambda^*_0) = 0 \). Therefore,

\[
A = E_0 \left[ - \frac{\partial^2}{\partial \theta \partial \theta} l_i(\theta_0, \lambda^*_0) \right] = E_0 \left[ \left( \frac{\partial a'(x_i; \theta_0)}{\partial \theta} \right) I_J \left( \frac{\partial a'(x_i; \theta_0)}{\partial \theta} \right)' \right],
\]

where

\[
L = -E_0 \left[ \left[ I_J \otimes \Gamma(\lambda^*_0 - \lambda_0) u_i \right] C \frac{\partial^2 \log g}{\partial u \partial u'} \left\{ \Gamma(\lambda^*_0 - \lambda_0) u_i \right\} C \left[ I_J \otimes \Gamma(\lambda^*_0 - \lambda_0) u_i \right] \\
+ \left[ I_J \otimes \Gamma(\lambda^*_0 - \lambda_0) u_i \right] C' G(\Gamma(\lambda^*_0 - \lambda_0) u_i) \right].
\]
10. Derivatives of Functions Based on Exponential of Matrices

10.1. Derivatives of $a \rightarrow \log(g(e^{a}C) y)$

For $a \in \mathbb{R}$, $y \in \mathbb{R}^{n}$, $C$ an $n \times n$ matrix, $g : \mathbb{R}^{n} \rightarrow \mathbb{R}^{+}$ a function,

$$\frac{\partial}{\partial a}(e^{a}y) = Ce^{a}y,$$

$$\frac{\partial^{2}}{\partial a^{2}}(e^{a}y) = C^{2}e^{a}y,$$

$$\frac{\partial}{\partial a} \log(g(e^{a}y)) = \left[ \frac{\partial \log g}{\partial u}(e^{a}y) \right] Ce^{a}y,$$

$$\frac{\partial^{2}}{\partial a^{2}} \log(g(e^{a}y)) = (Ce^{a}y) \left[ \frac{\partial^{2} \log g}{\partial u^{2}}(e^{a}y) \right] Ce^{a}y + (C^{2}e^{a}y) \left[ \frac{\partial \log g}{\partial u}(e^{a}y) \right].$$

10.2. Derivatives of $\theta \rightarrow e^{a(\theta)C} y$ and $\theta \rightarrow \log(g(e^{a(\theta)C} y))$

For $a : \mathbb{R}^{p} \rightarrow \mathbb{R}$ and $\theta \in \mathbb{R}^{p}$,

$$\frac{\partial}{\partial \theta} \{e^{a(\theta)C} y\} = Ce^{a(\theta)C} y \frac{\partial a(\theta)}{\partial \theta},$$

$$\frac{\partial}{\partial \theta} \log \{e^{a(\theta)C} y\} = \frac{\partial}{\partial a} \log \{e^{a(\theta)C} y\} \cdot \frac{\partial a(\theta)}{\partial \theta} = \left\{ \left[ \frac{\partial \log g}{\partial u}(e^{a(\theta)C} y) \right] Ce^{a(\theta)C} y \right\} \cdot \frac{\partial a(\theta)}{\partial \theta},$$

$$\frac{\partial^{2}}{\partial \theta^{2}} \log \{e^{a(\theta)C} y\} = \frac{\partial}{\partial a} \log \{e^{a(\theta)C} y\} \cdot \frac{\partial^{2} a(\theta)}{\partial \theta^{2}} + \frac{\partial^{2}}{\partial a^{2}} \log(g(e^{aC}y)) \cdot \frac{\partial a(\theta)}{\partial \theta} \frac{\partial a(\theta)}{\partial \theta} = \left\{ \left[ \frac{\partial \log g}{\partial u}(e^{a(\theta)C} y) \right] Ce^{a(\theta)C} y \right\} \cdot \frac{\partial^{2} a(\theta)}{\partial \theta^{2}}$$

$$+ \left\{ (Ce^{a(\theta)C} y) \left[ \frac{\partial^{2} \log g}{\partial u^{2}}(e^{a(\theta)C} y) \right] Ce^{a(\theta)C} y \right\} \cdot \frac{\partial a(\theta)}{\partial \theta} \frac{\partial a(\theta)}{\partial \theta}.$$

In these equalities, “.” indicates the multiplication of a matrix by a scalar.

10.3. Derivatives of $\theta \rightarrow \exp(\sum_{j=1}^{J} [a_{j}(x_{i}; \theta) + \lambda_{j}]C_{j})y_{i}$

Let $z_{i}(\lambda, \theta) = \exp(\sum_{j=1}^{J} [a_{j}(x_{i}; \theta) + \lambda_{j}]C_{j})y_{i}$, where $\lambda = (\lambda_{1}, \ldots, \lambda_{J}) \in \mathbb{R}^{J}$, $a_{j}(\cdot)$ are real valued functions with $\theta \in \mathbb{R}^{p}$, $y_{i} \in \mathbb{R}^{n}$, $C_{j}$ are $n \times n$ matrices. Let $a(x_{i}; \theta) = (a_{1}(x_{i}; \theta), \ldots, a_{J}(x_{i}; \theta))^{\top}$. For $i = 1, \ldots, J$, let the $n \times 1$ vectors

$$z_{i}^{(j)}(\theta, \lambda) = \exp \left\{ \sum_{j=1}^{i-1} [a_{j}(x_{i}; \theta) + \lambda_{j}]C_{j} \right\} C_{i} \exp \left\{ [a_{i}(x_{i}; \theta) + \lambda_{i}]C_{i} \right\} \times \exp \left\{ \sum_{j=i+1}^{J} [a_{j}(x_{i}; \theta) + \lambda_{j}]C_{j} \right\} y_{i},$$

where
where the first and last sums are replaced by 0 when $i = 1$ and $i = J$, respectively. Let the $n \times J$ block-matrix

$$Z_i(\theta, \lambda) = [z_i^{(1)}(\theta, \lambda) | \ldots | z_i^{(J)}(\theta, \lambda)].$$

We have

$$\frac{\partial z_i(\theta, \lambda)}{\partial \theta} = Z_i(\theta, \lambda) \frac{\partial a(x_i; \theta)}{\partial \theta}.$$

When the matrices $C_j$ commute, we have $z_i^{(i)}(\theta, \lambda) = C_i z_i(\theta, \lambda)$ and

$$Z_i(\theta, \lambda) = [C_1 z_i(\theta, \lambda) | \ldots | C_J z_i(\theta, \lambda)] = C[I_J \otimes z_i(\theta, \lambda)],$$

where $C = [C_1 | \ldots | C_J]$. Thus, when the $C_j$ commute,

$$\frac{\partial z_i(\theta, \lambda)}{\partial \theta} = C[I_J \otimes z_i(\theta, \lambda)] \frac{\partial a(x_i; \theta)}{\partial \theta}.$$

REFERENCES