ECONOMETRIC MODELLING and TRANSFORMATION GROUPS

C. GOURIEROUX (with MONFORT, A., ZAKOIAN, J.M.)

E.T. Lecture, Amsterdam, December, 2017

1. INTRODUCTION

The simple linear regression model (as a starting point)

$$y_t = \alpha + \beta x_t + u_t,$$

where the $u_t' s$ are i.i.d., independent of x_t.

• The Gaussian Pseudo-Maximum Likelihood estimator of β is consistent (even if the Gaussian distribution of u is not zero mean).

• It coincides with the solution $\hat{\beta}_T$ of the first-order conditions :

$$cov_e(y_t - \hat{\beta}_T x_t, x_t) = 0.$$

• Thus we have a double interpretation of the OLS estimator :

- as a PML estimator,
- as a covariance estimator.

Question :

Can we extend these results to much more general frameworks ?

What are the respective roles of

- the sum (between α and βx_t, between βx_t and u_t) ;
- the intercept ;
- the Gaussian choice of the pseudo-distribution ?

Answer

What really matters :

- A group structure of the transformations.
- the "intercept".

What does not matter :

- the Gaussian pseudo-distribution.
- the linearity in u.

Outline

- The Semi-Parametric Model
- Consistent Estimation
- Examples
- Abelian Lie Group
- Concluding Remarks

2. THE SEMI-PARAMETRIC MODEL

2.1 The Transformation Group

A triple $(\mathcal{A}, \mathcal{Y}, \mathcal{C})$, where $(\mathcal{A}, *)$ is a group, $\mathcal{Y}$ a (Hausdorff topological) space and $\mathcal{C}$ a set of maps from $\mathcal{A} \times \mathcal{Y}$ to $\mathcal{Y}$ such that :

i) $c[a, c(b, y)] = c[a * b, y], \forall y \in \mathcal{Y}, a, b \in \mathcal{A}$,
ii) $c(e, y) = y, \forall y \in \mathcal{Y}$, where e is the identity of $(\mathcal{A}, *)$.

[see e.g. Bredon (1972) : "Introduction to Compact Transformation Groups", Vol 46, Pure and Applied Mathematics, Chapter 1].

The group structure $(\mathcal{A}, *)$ is transferred (isomorphic) to the group structure $(\mathcal{C}, o)$, where o is the composition of functions. In particular each function : $y \to c(a, y)$ is invertible.

- Additional constraints can be introduced on the group :
  - Abelian group, when $a * b = b * a, \forall a, b \in \mathcal{A}$;
  - Lie group, when a differentiable structure with respect to a.

- A group transformation is easily transferred by any one-to-one change of parameter (or change of argument y).

2.2 The Econometric Model

The semi-parametric model is deduced from the transformation group by making parameter a depend on explanatory variables and introducing an "intercept" parameter :

$$\begin{aligned} u_t &= c[\alpha * a(x_t, \beta), y_t] \\ \iff \quad y_t &= c[a^{-1}(x_t, \beta) * \alpha^{-1}, u_t], \end{aligned}$$

where • the errors (u_t) are i.i.d.,

• the error u_t is independent of the explanatory variable x_t (either exogenous, or lagged endogenous).

• the parameter : $\alpha, \beta : \alpha \in \mathcal{A}, \beta \in \mathcal{B}$.

As many intercepts α as scores (indexes) $a(x, \beta)$.

2.3 An Identification Issue

Let us assume that the econometric model is well-specified :

true parameter value : α_0, β_0,

true error distribution : f_0.

• If f_0 is known, (α, β) is identifiable iff :

$$\alpha * a(x, \beta) = \alpha_0 * a(x, \beta_0), \forall x \in \mathcal{X} \Rightarrow \alpha = \alpha_0, \beta = \beta_0.$$

• If f_0 is unknown, we cannot expect to identify separately α_0, f_0. This is the reflection problem discussed by Manski (1993) for linear models. Indeed :

$$\begin{aligned} y_t &= c[a^{-1}(x_t, \beta_0) * \alpha_0^{-1}, u_t] \\ &= c[a^{-1}(x_t, \beta_0), c(\alpha_0^{-1}, u_t)] \\ &= c[a^{-1}(x_t, \beta_0) * e, v_t]. \end{aligned}$$

(α_0, f_0) cannot be distinguished from (e, distribution of v_t).

3. CONSISTENT ESTIMATION

3.1 Two Estimation Methods

• Covariance Estimators

Based on the covariance restrictions :
$cov[\varphi(v_t), \psi(v_{t-h})] =, \forall t, h$, and a large class of functions φ, ψ,
$\iff Cov[\varphi[c(a(x_t, \beta), y_t)], \psi[c(a(x_{t-h}, \beta), y_{t-h})]\} = 0, \forall t, h, \varphi, \psi.$

Under stationarity assumption and an appropriate choice of functions φ, ψ and lag h, consistent estimators of β_0 can be derived from the empirical covariance counterparts.

Generalized Covariance (GCov) Estimators :
Gourieroux, Jasiak (2017) : "Noncausal Vector Autoregressive Process : Representation, Identification and Semi-Parametric Estimation", Journal of Econometrics, 200, 118-134, Section 4 and Appendix C.

• The covariance estimators are similar to GMM estimators, but differ from them. Indeed the restrictions involve both moments and products of moments.

This is an indirect way to adjust for the nonidentifiable intercept.

• The GCov estimators can be applied to continuous as well as discrete variables y.

- **Pseudo Maximum Likelihood (PML) Estimators**

An alternative is to apply a maximum likelihood approach on the econometric model with intercept and a given distribution f for the $u_t's$. Since in general f is different from the unknown f_0, this is a PML approach.

The optimisation is :

$$(\hat{\alpha}_T, \hat{\beta}_T) = \arg\max_{\alpha,\beta} \frac{1}{T} \sum_{t=1}^{T} \log l(\alpha * a(x_t, \beta), y_t),$$

and the asymptotic criterion is :

$$E_0 E_x \log l\{\alpha * a(x, \beta), c[a^{-1}(x; \beta_0) * \alpha_0^{-1}, u]\}.$$

The transformation group structure implies a special form of the pseudo likelihood function and the following result :

The consistency result

Let us assume that the optimisation problem :

$$\max_\alpha E_0 E_x \log l(\alpha, u),$$

has a unique solution $\tilde{e}_0$, and the identification condition for fixed f, then the limiting optimization problem has the unique solution :

$$p\lim \hat{\alpha}_T = \tilde{e}_0 * \alpha_0, p\lim \hat{\beta}_T = \beta_0.$$

In other words,
For any choice of pseudo distribution f (satisfying minimal regularity conditions), the introduction at the right place of intercept parameters adjusts for the asymptotic bias of the PML estimator of β.
This result is valid for any type of observation : discrete as well as continuous.

How to understand the consistency result ?

Consider the discrete state case : $\mathcal{Y} = \{1, \ldots, K\} (= \mathcal{U})$, say, and introduce a pseudo distribution on u, that are, fix elementary probabilities : $p(k)$ (or $p(u)$). We have :

$$\begin{aligned} \log l_t &= \log p[c(\alpha * a(x_t, \beta), y_t)] \\ &= \log p[c[\alpha * a(x_t, \beta), c[a^{-1}(x_t, \beta_0) * \alpha_0^{-1}, u_t]]] \\ &= \log p[c[\alpha * a(x_t, \beta) * a^{-1}(x_t, \beta_0) * \alpha_0^{-1}, u_t]]. \end{aligned}$$

Then

$$\begin{aligned} & E_0 E_x \log p[c[\alpha * a(x, \beta) * a^{-1}(x, \beta_0) * \alpha_0^{-1}, u_t]] \\ & \leq E_0 \log p[c(\tilde{e}_0, u)], \end{aligned}$$

and the maximum is reached for :

$$\begin{aligned} & \beta = \beta_0, \alpha * \alpha_0^{-1} = \tilde{e}_0, \\ \iff & \beta = \beta_0, \alpha = \tilde{e}_0 * \alpha_0. \end{aligned}$$

The PML estimator of the intercept is not consistent of α_0, but tends to the pseudo-true value :

$$\alpha_0^* = \tilde{e}_0 * \alpha_0 \Leftrightarrow \alpha_0^* * \alpha_0^{-1} = \tilde{e}_0.$$

When the group operation is the addition : $* = +$, we get :

$$\alpha_0^* - \alpha_0 = \tilde{e}_0,$$

that is the bias is constant, i.e. independent of α_0.
This result of constant bias is general with another definition of the bias on α_0 as : $\alpha_0^* * \alpha_0^{-1}$. Indeed in general the right interpretation of an intercept is not to be a mean.

3.2 Consistent Estimation of Standardized Errors

Once parameter β is estimated consistently, we deduce :

• consistent approximations of the standardized errors v_t by computing the residuals :

$$\hat{v}_{tT} = c[a(x_t, \hat{\beta}_T), y_t];$$

• consistent functional estimators of the common distribution of the $v_t' s$:
by an appropriate (kernel) smoothing of the empirical distribution of the residuals ;

• then, the possibility to develop adaptive estimation method for parameter β.

3.3 Unobserved Heterogeneity

The PML approach can also be applied assuming that the (misspecified) distribution of errors belongs to a parametric family : $f(u;\gamma)$, say.

Then we get PML estimators for α, β, γ. They converge to values : $\hat{\alpha}_T \to \alpha_0^*, \hat{\beta}_T \to \beta_0, \hat{\gamma}_T \to \gamma_0^*$.

Thus we still get the consistency of the estimator of parameter β.

This result has an interesting interpretation for group transformation model with unobserved heterogeneity.

The model is defined by :

$$\begin{aligned} u_t &= c[\eta_t * \alpha * a(x_t, \beta), y_t] \\ \iff \quad y_t &= c[a^{-1}(x_t, \beta) * \alpha^{-1} * \eta_t^{-1}, u_t] \end{aligned}$$

where η_t denotes the unobserved heterogeneity : (η_t, u_t) i.i.d.

Let us assume that the DGP is parametric : $u_t \sim f(u, \gamma_0)$, $\eta_t \sim h(\eta, \delta_0), \beta = \beta_0, \alpha = \alpha_0$, say.

Proposition : The ML estimator of α, β, γ computed without assuming heterogeneity is such that $\hat{\beta}_T \to \beta_0$.

This is a consequence of the alternative form of the model :

$$y_t = c[a^{-1}(x_t, \beta), v_t], \text{ where } v_t = c[\alpha^{-1} * \eta_t^{-1}, u_t].$$

Message : be aware about the way of introducing unobserved heterogeneity.

4. EXAMPLES

4.1 Affine Transformation Model

$$\begin{aligned} u_t &= a_0(x_t,\beta) + A_1(x_t,\beta)y_t. \\ \iff \quad y_t &= -A_1^{-1}(x_t,\beta)a_0(x_t,\beta) + A_1^{-1}(x_t,\beta)u_t. \end{aligned}$$

$c(a,y) = a_0 + A_1 y$, with $a_0 \in \boldsymbol{R}^n, A_1$ an invertible (n,n) matrix.

• The group operation :

$$(b_0, B_1) * (a_0, A_1) = (b_0 + B_1 a_0, B_1 A_1)$$

• The econometric model with intercept :

$$y_t = b_0 + B_1 a_0(x_t,\beta) + B_1 A_1(x_t,\beta)y_t,$$

where : $\alpha = (b_0, B_1)$ is the intercept.

• This is the multivariate extension of :
Newey, Steigerwald (1997) : "Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroscedastic Models', Econometrica, 65, 587-599.

i) Multivariate ARCH-Model

When $u_t = A_1(x_t, \beta)y_t, \alpha = B_1$ and $x_t = \underline{y_{t-1}}$, the model includes the multivariate ARCH models.

But with a new modelling perspective hidden in the identification condition of A_1 in the (pseudo) generic model. Let us assume that the components $u_{it}, i = 1, \ldots, n$ of u_t are independent, a condition needed for analyzing the consequences of shocks in such an ARCH model :

$$y_t = A_1(x_t, \beta)^{-1} u_t.$$

If either the true or the pseudo distributions are Gaussian, say, the square $A_1 A_1'$ is identifiable, but not necessarily A_1. In other words, there exists a multiplicity of square roots of the volatility matrix that could be chosen for deriving the impulse response functions.

If both the true and pseudo distributions admit at most one Gaussian component, we can identify and estimate consistently all β parameters, not only the subparameters of β identifiable from $A_1(x_t, \beta)A_1(x_t, \beta)'$.

ii) Peer Effects

The standard modelling of peer effects in a linear framework is :

$$y_t = By_t + \alpha(x_t, \theta) + u_t,$$

or equivalently :

$$\begin{aligned} y_t &= (Id - B)^+ \alpha(x_t, \beta) + (Id - B)^+ w_t \\ &\equiv a(x_t, \beta) + Aw_t, \end{aligned}$$

(where i the student, $t = j$ the school, say).
This model features the standard identification issue, i.e. the reflection problem.

The model can be extended to an explanatory modelling of peer effects $A(\Longleftrightarrow B)$, i.e. to :

$$y_t \equiv a(x_t, \beta) + A(x_t, \beta)u_t.$$

Difficulty : the rather large dimension $n = \dim y_t$ and n^2 for the number of additional intercepts.

Solution : Constrain the explained peer effects in order A to belong to a subgroup of a much smaller dimension.

3.2 The Least Impulse Response (LIR) Estimator

The stress test exercices demanded by the supervisors for credit risk are based on a decomposition of the total expected loss for an homogenous segment of loans as :

$$\text{Expected Loss} = \text{EL} = \text{EAD. CCF. PD. (E)LGD},$$

where : EAD = Exposure-at-Default,
CCF = Credit Conversion Factor,
PD = Probability of Default,
(E)LGD = (Expected) Loss-Given-Default.

The variables PD, (E)LGD, CCF are valued in $[0, 1]$ and there is a lack of flexible models and of interpretable estimation methods for their analysis.

Instead of the standard parametric beta model, we suggest the use of a transformation group and a PML approach based on the uniform distribution on [0,1].

$$\text{DGP} : u_t = c[\alpha_0 * a(x_t, \beta_0), y_t], u_t \text{ i.i.d.}$$

pseudo log-likelihood :

$$\begin{aligned} L_T(\alpha, \beta) &= \sum_{t=1}^{T} \log \frac{\partial c}{\partial u}[\alpha * a(x_t; \beta), y_t] \\ &= -\sum_{t=1}^{T} \log \frac{\partial c}{\partial u}[a^{-1}(x_t; \beta) * \alpha^{-1}, u_t] \\ &= -\sum_{t=1}^{T} \log IR_t(\alpha, \beta), \end{aligned}$$

where $IR_t(\alpha, \beta)$ is the effect on y_t of a local shock on u_t.

This justifies the name :

Least Impulse Response (LIR) estimator.

An unfair estimation approach

• The idea is to select the estimates $\hat{\alpha}_T, \hat{\beta}_T$ in order to minimize the estimated consequences of (local) shocks on u; in other words, to minimize the amount of estimated reserves introduced to be hedged against such shocks.

• As unfair as the OLS approach in the standard regression model, where the OLS approach selects the estimate to give the impression that the estimated model is accurate : minimize the sum of squared residuals.

Nevertheless :

• the unfair approach provides consistent results ;
• is also appropriate for analyzing the effect of nonlocal shocks (just the distribution of v_t matters, not separately f_0 and α_0).

Examples of transformation groups on [0,1]

i) Power transformation

$$c(a,y) = y^a, a \in \boldsymbol{R}^{+*}, a * b = ab;$$

ii) Homographic transformation

$$c(a,y) = \frac{ay}{1+(a-1)y}, a \in \boldsymbol{R}^{+*}, a * b = ab;$$

iii) Piecewise transformation

$$c(a,y) = \sum_{k=1}^{10} c_k^*(a_k,y)\mathbf{1}_{\frac{k-1}{10} \le c \le \frac{k}{10}},$$

with $\quad c_k^*(a_k,y) = \frac{k-1}{10} + \frac{1}{10}c(a_k, 10y-(k-1)), k = 1, 10.$

iv) Moebius transformations

For a joint analysis of two variables valued in (0,1), such as PD and (E)LGD.

3.3 Boolean Group and Adjusted Maximum Score

An (exotic ?) example with discrete variables : $\mathcal{Y} = \mathcal{U} = \{0, 1\}$.

- The group of permutations :

$$\begin{aligned} u &= y = c(1, y), \text{the identity Id}, \\ u &= 1 - y = c(0, y), \text{the other permutation } P. \end{aligned}$$

- The group operation :

$a \in \mathcal{A} = \{0, 1\}, a * b = ab + (1 - a)(1 - b)$.

$$\begin{cases} Id \ o \ Id = Id \\ Id \ o \ P = P \\ P \ o \ Id = P \\ P \ o \ P = Id \end{cases} \Longrightarrow \begin{cases} 1 * 1 = 1 \\ 1 * 0 = 0 \\ 0 * 1 = 0 \\ 0 * 0 = 1. \end{cases}$$

- Simplified form :

$$c(a, y) = ay + (1 - a)(1 - y).$$

- The econometric model (without intercept).

$$\log \Pi \; S_T(\beta) + \log(1 - \Pi)(T - S_T(\beta)],$$

where $\Pi = P[u_t = 1]$ and

$$S_T(\beta) = \sum_{t=1}^{T} [\mathbb{1}_{\tilde{a}(x_t,\beta)>0} y_t + \mathbb{1}_{\tilde{a}(x_t,\beta)<0}(1 - y_t)],$$

is the score (here $t = i$ individual index).

If $\Pi > 0.5$, the PML is simply the maximum score estimator.

Manski, C. (1975) : "Maximum Score Estimation of the Stochastic Utility Model of Choice", Journal of Econometrics, 3, 205-228.

- The econometric model (with intercept) :

$$u_t = a(x_t, \beta) * \alpha * y_t.$$

The associated concentrated pseudo log-likelihood is the max of :

$$\log \Pi S_T(\beta) + \log(1-\Pi)[T - S_T(\beta)], \log \Pi[T - S_T(\beta)] + \log(1-\Pi) S_T(\beta).$$

The PML estimator of β maximizes :

$$\hat{\beta}_T = \arg\max_{\beta} \max(S_T(\beta), T - S_T(\beta)).$$

The standard maximum score has to be adjusted for "orientation". Indeed we do not know a priori if $\Pi_0 > 0.5$.

5. ABELIAN LIE GROUP AND COVARIANCE ESTIMATORS

It is well known that a maximum likelihood estimator can be interpreted as a moment estimator, where the moment restrictions correspond to the expected score being zero.

Do we have a similar result for the consistent PML estimator taking into account the adjustment for not identifiable intercept ?

First we need a differentiability assumption on the group to be able to derive first-order conditions :

- Lie group ;

We also need the commutativity property of the operation :

- Abelian Lie group.

5.1 A Characterization of an Abelian Lie Group

Theorem [Bredon (1972), Th 5-4] :

A (connected) Abelian Lie Group $(\mathcal{A}, *)$ is isomorphic to $T^K \times R^{p-K}$ for some p, K, where T^K is the K-dimensional torus, i.e. the product of K copies of the circle group $S^1 = R/2\Pi\mathbb{Z}, S^1$ and R with their standard group operations.

Thus, up to a one-to-one change of the group elements, we can choose for the components of a :

K "angles", $w_k, k = 1, \ldots, K$, with $w_k * w_l = w_k + w_l$ (mod 2π) ;

p reals : $a_j, j = K + 1, \ldots, p$, with $a_j * a_l = a_j + a_l$.

It is also isomorphic to the group of $(p+K, p+K)$ block-diagonal matrices, where the K first $(2,2)$ blocks are matrices of rotations : $\begin{pmatrix} \cos w_k & \sin w_k \\ -\sin w_k & \cos w_k \end{pmatrix}$, and the next $(1,1)$ blocks have diagonal elements $\exp a_j, j = K+1, \dots, p$.

It is easily checked that such matrices can be written as :

$$\exp\left(\sum_{k=1}^{K} w_k C_k + \sum_{j=K+1}^{p} a_j C_j\right),$$

where the $(p+K, p+K)$ matrices C_j are known, commute, and the exponential of a matrix is defined as :

$$\exp(aC) = \sum_{i=0}^{\infty} \frac{a^i C^i}{i!}.$$

To summarize it is equivalent to analyse an Abelian Lie group or a group of matrices of the form :

$$\exp(\sum_{j=1}^{p} a_j C_j),$$

where the matrices $C_j, j = 1, \ldots, p$ commute.

This is useful from a theoretical point of view.

From a practical point of view, the difficulty is that the change of parameter to reach this simplified expression of the transformation, or equivalently to transform the operation $*$ into an addition $+$, is not easy to find.

5.2 First-Order Conditions and Covariance Estimators

Proposition 1 : Let us consider a model :

$$\begin{aligned} u_t &= \Pi_{j=1}^{p} \exp[(a_j(x_t, \beta) + \alpha_j) C_j] y_t \\ &= \exp[\sum_{j=1}^{p} (a_j(x_t; \beta) + \alpha_j) C_j] y_t, \end{aligned}$$

where the $C_j, j = 1, \ldots, p$ commute, then the FOC for parameter β are equivalent to empirical covariance restrictions between appropriate functions of x_t (the underlying instruments) and appropriate functions of v_t.

The asymptotic variance-covariance matrix of the PML estimator has also a specific form. For instance for a single score $p = 1$, we have :

$$V_{as}[\sqrt{T}(\hat{\beta}_T - \beta_0)] \equiv \frac{i}{j^2}\Omega^{-1},$$

where i/j^2 is the asymptotic variance of the PML estimator of α_1 in the generic model $u_t = \exp(\alpha_1 C_1)y_t$ with intercept only, which depends on both the true and pseudo-distributions of the error, but not on the distribution of explanatory variables ; $\Omega = V_x \left[\frac{\partial a(x_t;\beta_0)}{\partial \beta}\right]$ is a variance matrix independent of the pseudo and true distributions of the errors.

In particular, the asymptotic variance-covariance matrices of all PML estimators are proportional.

5.3 Additional Examples

Before introducing these examples, note that the exponential of a matrix C can be written as :

$$\exp(aC) = (\exp C)^a \equiv B^a, \text{with } B = \exp C.$$

Any matrix B which is diagonalizable, with complex, or positive real eigenvalues can be written under an exponential form.
C is called the generator of this group.

5.3.1. Dynamic Model for Stochastic Measures

Let us consider a (n, n) transition matrix P, say, with complex, or positive real eigenvalues :

$$P = \exp \wedge,$$

where the generator $\wedge$ is the associated matrix of intensities.

The econometric model :

$$y_t = P^{a(x_t,\beta)} u_t = \exp(a(x_t, \beta)\wedge) u_t,$$

is appropriate for defining a dynamic model for observed probability distributions if :
$u_{it} > 0, \Sigma_i u_{it} = 1 [y_{it} > 0, \forall i, \Sigma_i y_{it} = 1],$ and $a(x_t, \beta) > 0, x_t = \underline{y_{t-1}}.$

$$a(x_t, \beta) \text{ is a time deformation.}$$

The extended model with intercept is :

$$y_t = P^{a(x_t,\beta)+\alpha} u_t,$$

where α is not constrained to be positive. α is a time origin.

5.3.2 Stratified Models

Let us consider a decomposition of $\boldsymbol{R}^n$ into J orthogonal vector spaces and denote $P_j, j = 1, \ldots, J$, the associated orthogonal projectors. We have :

$$\exp(\sum_{j=1}^{J} a_j P_j) = \sum_{j=1}^{J} \exp(a_j) P_j.$$

The associated econometric model is of the type :

$$y_t = \exp(\sum_{j=1}^{J} a_j(x_t, \beta) P_j) u_t = \sum_{j=1}^{J} \exp[a_j(x_t, \beta)] P_j u_t.$$

We will illustrate such econometric models with the simple case : $J = 2$, P_n the orthogonal projector on the vector space generated by the unitary vector $(1, 1, \ldots, 1)'$.

i) Model with Equi-Individual Interactions (peer effects)

The model :

$$y_t = \begin{pmatrix} \alpha_1(x_t,\beta) & \alpha_2(x_t,\beta) & \dots & \alpha_2(x_t,\beta) \\ \alpha_2(x_t,\beta) & \alpha_1(x_t,\beta) & & \\ \vdots & & \ddots & \\ \alpha_2(x_t,\beta) & & & \alpha_1(x_t,\beta) \end{pmatrix} u_t, \quad (*)$$

can be rewritten as :

$$\begin{aligned} &(\alpha_1(x_t,\beta) - \alpha_2(x_t,\beta)]Id + n\alpha_2(x_t,\beta)P_n \\ &= [\alpha_1(x_t,\beta) - \alpha_2(x_t,\beta)](Id - P_n) + (\alpha_1(x_t,\beta) + (n-1)\alpha_2(x_t,\beta)]P_n \\ &= \exp[a_1(x_t,\beta)P_n + a_2(x_t,\beta)(Id - P_n)], \end{aligned}$$

with : $a_1(x_t,\beta) = \alpha_1(x_t,\beta) + (n-1)\alpha_2(x_t,\beta), a_2(x_t,\beta) = \alpha_1(x_t,\beta) - \alpha_2(x_t,\beta).$

After the change of parameter, the introduction of the additional intercept is additive. This introduction is much more complicated when it is written from the initial parameters α_1, α_2 in $(*)$.

Interpretation for academic achievement

y_{ic} achievement for student i and class c

$$y_{ic} = a_0(x_{ic}, \beta) + a_1(x_{ic}, \beta)\bar{u}_c + u_{ic}$$

• The reflection problem is not solved : we cannot disentangle the constant part of the peer effect and the dependence between the components $u_{i,c}, i = 1, \dots, n$.

• But we can estimate consistently the "exogenous" (contextual) component of the peer effect, even if the $u_{ic}, i = 1, \dots, n$ are linked.

This extends Graham (2008) : "Identifying Social Interactions Through Conditional Variance Restrictions", Econometrica, 76, where the effect of the size of the class (large vs small) was introduced.

ii) Within and Between Equi-Interactions (peer effects)

This modelling is easily extended to several segments (strata) to distinguish the interaction between and within segments. For instance for two segments of a same size n, we can write :

$$\begin{aligned} y_t &= \exp[a_1^B(x_t,\beta)P_2 + a_2^B(x_t,\beta)(Id_2 - P_2)] \otimes \exp[a_1^w(x_t,\beta)P_n + a_2^w(x_t,\beta)(Id_n - P_n)] \\ &= \exp\{a_1^B(x_t,\beta)a_1^w(x_t,\beta)P_2 \otimes P_n + a_1^B(x_t,\beta)a_2^w(x_t,\beta)P_2 \otimes (Id_n - P_n) \\ &+ a_2^B(x_t,\beta)a_1^w(x_t,\beta)(Id_2 - P_2) \otimes P_n + a_2^B(x_t,\beta)a_2^w(x_t,\beta)(Id_2 - P_2) \otimes (Id_n - P_n)] \end{aligned}$$

where $P_2 \otimes P_n, P_2 \otimes (Id - P_n)$ are orthogonal projectors and $\otimes$ denotes the Kronecker product.

The additional intercepts (4 intercepts) have to be added to the products

$a_1^B a_1^w, a_1^B a_2^w \ldots$, not separately on $a_1^B, a_2^B \ldots$.

iii) Seasonal Adjustment

A seasonal model with two seasons (semesters) :

y_{1t} semester 1, year t,
y_{2t} semester 2, year t.

The two parametrizations :

$$\begin{cases} y_{1t} & = & \alpha_1(x_t,\beta)u_{1t}+\alpha_2(x_t,\beta)u_{2t}, \\ y_{2t} & = & \alpha_2(x_t,\beta)u_{1t}+\alpha_1(x_t,\beta)u_{2t}, \end{cases}$$

or

$$\begin{cases} \dfrac{y_{1t}+y_{2t}}{2} & = & y_{.t}=a_1(x_t,\beta)u_{.t}, \text{ with } a_1=\alpha_1+\alpha_2, \\ \dfrac{y_{1t}-y_{2t}}{2} & = & y_{1t}-y_{.t}=a_2(x_t,\beta)(u_{1t}-u_{.t}) \text{ with } a_2=\alpha_1-\alpha_2. \end{cases}$$

The second parametrization is more convenient for the interpretation and the modelling :

• $u_{.t}$ = yearly innovation, $u_{1t}-u_{.t}$ = seasonal specific innovation.

• Dynamics for the year and seasonal components can be directly introduced on the second specification.

5.3.3 Evolutionary Trees (Hierarchical Models)

Let us denote $C(k)$ the matrix with unitary values on the upper k^{th} diagonal and zero anywhere else.

These matrices commute since :

$$C(k)C(l) = C(k+l) = C(1)^{k+l},$$

and we have $C(k) = 0, \forall k \geq n$.

Then, after an appropriate change of parameter, we get :

$$\exp[\sum_{j=0}^{n-1} a_j C(j)] \equiv \sum_{j=0}^{n-1} b_j(a) C(j).$$

This explains how to introduce the additional intercepts in a model of the type :

$$y_t = \begin{pmatrix} b_0(x_t,\beta) & b_1(x_t,\beta) & \dots & b_{n-1}(x_t,\beta) \\ 0 & \ddots & & \vdots \\ \vdots & & & b_1(x_t,\beta) \\ 0 & \dots & 0 & b_0(x_t,\beta) \end{pmatrix} u_t.$$

6. CONCLUDING REMARKS

The ML approach usually leads to unconsistent estimators if the distribution of the errors is misspecified. We have seen that for models based on transformation groups, this lack of consistency is easily adjusted for by introducing an appropriate number of intercept parameters at appropriate places.

Moreover, for Abelian Lie Group, the PML estimators are interpretable as covariance estimators, and their asymptotic variance-covariance matrix has a simple form, that disentangles the effect of explanatory variables and the effect of pseudo and true distributions of the errors.

However the main message seems more important since these transformation group models with their associated PML estimator approaches can be used as an alternative modelling strategy. Typically this point has been illustrated :

• for models with unobserved heterogeneity

• for credit risk, with the interpretation of some estimation approaches in terms of reserve,

• for ARCH type models, when it is possible to estimate the nonlinear sensitivity to random shocks, instead of the standard volatility which is usually less informative.

• to account for the explanatory component in peer effects.

Up to now transformation groups were not explicitely used in econometric modelling, but they are intensively used in other fields, mainly without introducing explanatory variables and/or dynamics :

not only for the Rubik's cube,

but for facial recognition, video analysis or neuroscience,

and of course for modelling the effect of time on space (relativity).

We hope that our results will help in introducing dynamic and/or explanatory models in these other fields.