



Least Impulse Response Estimator for Stress Test Exercises

Christian Gouriéroux, Yang Lu

► To cite this version:

Christian Gouriéroux, Yang Lu. Least Impulse Response Estimator for Stress Test Exercises. Journal of Banking and Finance, 2019, 103, pp.62-77. 10.1016/j.jbankfin.2019.03.021 . hal-02419030

HAL Id: hal-02419030

<https://hal.science/hal-02419030>

Submitted on 20 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Least Impulse Response Estimator for Stress Test Exercises

Christian GOURIEROUX¹, Yang LU²

July 2018

We gratefully acknowledge financial support of the Chair ACPR : "Regulation and Systemic Risks", the Chair LCL: "New Challenges for New Data" and the ERC DYSMOIA.

¹University of Toronto, Toulouse School of Economics and CREST, correspondence: gouriero@ensae.fr

²University of Paris 13, correspondence: luyang000278@gmail.com

Least Impulse Response Estimator for Stress Test Exercises

Abstract

We introduce new semi-parametric models for the analysis of rates and proportions, such as proportions of default, (expected) loss-given-default and credit conversion factor encountered in credit risk analysis. These models are especially convenient for stress test exercises. We show that minimizing the estimated effect of a stress leads to consistent estimates. The new models with their associated estimation method are compared with the other approaches currently proposed in the literature such as the beta, and logistic regressions.

Keywords : Basel Regulation, Stress Test, Loss-Given-Default, Impulse Response, Pseudo-Maximum Likelihood, LIR Estimation, Beta Regression, Moebius Transformation.

JEL codes: C51, G21.

1 Introduction

The aim of this note is to explain how recent results on transformation groups [Gourieroux, Monfort, Zakoian (2018)] can be used in the framework of stress test exercises for homogeneous pools of credit lines. For such a homogeneous pool, the total expected loss is conventionally decomposed along the following formula [see e.g. BCBS (2001), EBA (2016)] :

$$\text{Expected Loss} = \text{EAD} \times \text{CCF} \times \text{PD} \times \text{LGD}, \quad (1.1)$$

where EAD denotes the exposure-at-default, that is the total level of the credit line, CCF is the credit conversion factor, that is the proportion of the credit line really used at the default time, PD is the probability of default and LGD the (expected) loss given default defined as 1 minus the (expected) recovery rate. Formula (1.1) involves variables that are value constrained. The EAD is positive, whereas the three other variables are constrained to lie between 0 and 1. When performing a stress exercise the EAD is usually crystallized, i.e. kept fixed, but the three other characteristics depend on how the borrowers react to the environment and are thus sensitive to local and/or extreme shocks. In this paper we introduce semi-parametric transformation models for variables in $[0, 1]$, such as CCF, PD, or LGD. We propose a consistent estimation method minimizing the sensitivity of the risk variables with respect to shocks, which we call the Least Impulse Response (LIR) approach. Moreover, we show that this method leads also to simple and consistent estimate of the effect of the stresses on the risk variables.

The semi-parametric transformation model with values in $[0, 1]$ is described in Section 2. We provide different examples of transformation models and explain how to introduce a notion of “intercept” parameter. The LIR estimation approach is introduced in Section 3. We discuss its interpretation as a pseudo maximum likelihood approach, and show that this approach provides consistent approximations of the effect of local as well as global stresses on the error term. We also compare the LIR approach with the approaches currently proposed in the credit risk literature such as the beta regression model, or the transformed Gaussian regression. In Section 4 the approach is extended to the joint treatment of PD and LGD. In particular we introduce models based on Moebius transformations. Section 5 concludes. Proofs of asymptotic results are gathered in Appendices.

2 The Transformation Model

We consider a semi-parametric transformation model for variables valued in $[0, 1]$. The model can be written as :

$$u_t = c[a(x_t, \beta), y_t], \quad t = 1, \dots, T, \quad (2.1)$$

where:

- the observed endogenous variable y_t and the unobserved error term u_t are valued in $[0, 1]$, x_t are observed explanatory variables that can include exogenous as well as lagged endogenous variables;
- $u \rightarrow c(a, u)$ is a one-to-one function mapping $[0, 1]$ to itself, parametrized by $a \in \mathcal{A} \subset \mathbb{R}^J$;
- $x \rightarrow a(x, \beta)$ is an index function (or score) from \mathcal{X} to \mathcal{A} , parametrized by $\beta \in \mathcal{B} \subset \mathbb{R}^K$.

Due to the invertibility of function $c_a(u) \equiv c(a, u)$, model (2.1) can be equivalently written in the standard form as :

$$y_t = c_{a(x_t, \beta)}^{-1}(u_t), \quad (2.2)$$

where the endogenous variable is expressed in terms of the explanatory variables and errors. In other words, model (2.2) is an input-output model with input (u_t, x_t) and output y_t . However we will see below that representation (2.1) is often more convenient than (2.2) for estimation purpose.

Next, we assume that :

Assumption 1. *The errors $u_t, t = 1, \dots, T$ are independent, identically distributed (i.i.d.). Their distribution is absolutely continuous with respect to the Lebesgue measure on $[0, 1]$, with probability density function (p.d.f.) denoted by f .*

Assumption 2. *The model (2.1) is well specified with true parameter value β_0 and true error distribution with p.d.f. f_0 .*

Remark 1 : Assumption 1 implies a continuous distribution on $[0, 1]$ for the endogenous variable y . This assumption has to be discussed in details for the LGD variable. Up to recent years, the supervisors have not sufficiently distinguished the loss-given-default observed ex-post, that is, at the end of the recovery process, from the expected loss-given-default measured ex-ante for a credit (corporate) not yet entered in default. The first notion is now called realized LGD¹ [see e.g. EBA (2016)], whereas the second notion is called ELGD for expected LGD in the recent academic literature. The distinction between the two notions is important. Indeed, if the observations concern the realized individual losses, the observed distributions of individual realized LGD's have significant point masses at 0 and 1, as well as a continuous component on $(0, 1)$. Therefore the assumption of continuous distribution on $(0, 1)$ is not satisfied. However, if the observed data concerns homogeneous segments of individual risks, what is observed is the ELGD, which is strictly between 0 and 1. Therefore Assumption 1 is satisfied when the observations concern average LGDs with corporate segments defined for instance by crossing the industrial sector, the country of domiciliation and the corporate rating². This corresponds to the demand of the supervisor [BCBS (2001), paragraph 336]: "A bank must estimate an LGD for each of its internal grades... Each estimate of LGD must be grounded in historical experience and empirical evidence."

Remark 2 : Assumption 1 is also satisfied for market values of the recovery rates, obtained for instance by comparing the market value of the debt just before default and its value one-month after default. This methodology is used by Moody's for firms with debt traded on an organized bond market. These market valued recovery rates are more sensitive to cyclical effects than the averages computed directly from the recovery processes.

2.1 Transformation Groups

Let us further assume that the set of transformations : $\mathcal{C} = \{c(a, \cdot), a \in \mathcal{A}\}$ is a group for the composition of functions, so that equivalently the group structure (\mathcal{C}, o) can be transferred on set \mathcal{A} :

¹or historical LGD [Gupton, Stern (2002)].

²These averages are either dollar weighted, or simple event weighted in practice.

Assumption 3. *We have :*

$$c[a, c(b, y)] = c(a * b, y),$$

where $(\mathcal{A}, *)$ is a group and $*$ denotes the group operation.

Remind that a group structure combines two elements $a, b \in \mathcal{A}$ into a third element $a * b \in \mathcal{A}$ and satisfies the axioms of associativity, identity and invertibility. We denote below by e the identity element such that: $a * e = e * a = a, \forall a \in \mathcal{A}$, and by a^{-1} the inverse of element a , such that $a * a^{-1} = a^{-1} * a = e$. Under Assumption 3, model (2.2) can be also written as :

$$y_t = c[a^{-1}(x_t, \beta), u_t]. \quad (2.3)$$

Let us provide examples of transformation groups on $[0,1]$. For each example we explicit the transformation, the group $(\mathcal{A}, *)$, the identity element and the form of the inverse.

Example 1 : Power transformation (single score)

We have : $c(a, y) = y^a$, where $a \in \mathcal{A} = \mathbb{R}^{+*}$. The group operation is : $a * b = ab$, with identity element $e = 1$ and the inverse : $a^{-1} = 1/a$.

Example 2 : Homographic transformation (single score)

We have : $c(a, y) = \frac{ay}{1 + (a-1)y}$, where $a \in \mathcal{A} = \mathbb{R}^{+*}$. The group transformation is : $a * b = ab$, with identity element $e = 1$ and the inverse : $a^{-1} = 1/a$.

Figure 1 plots examples of power and homographic transformations for different values of parameters.

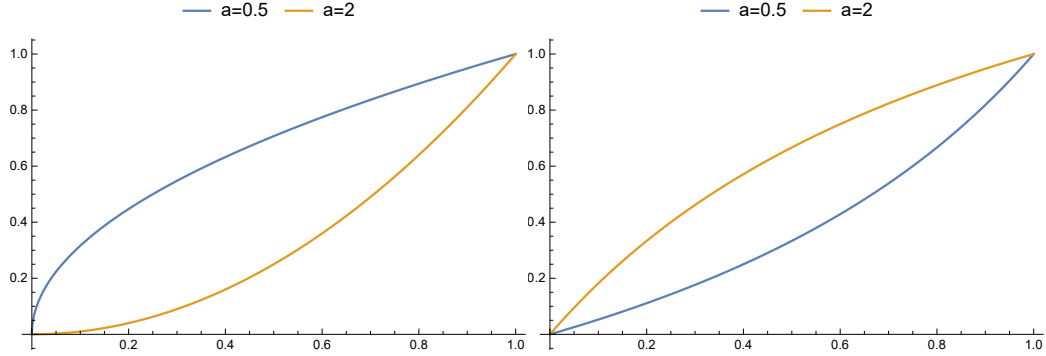


Figure 1: Plot of power transformations (left panel) and homographic transformations (right panel). In both cases, we plot the curves for two values of the parameter $a = 2$ and $a = 0.5$.

Example 3 : Exp-log power transformation (two scores)

We have : $c(a, y) = \exp[-a_1(-\log y)^{a_2}]$, where $a = (a_1, a_2)' \in \mathcal{A} = (\mathbb{R}^{+*})^2$. The group operation is : $a * b = (a_1 b_1^{a_2}, a_2 b_2)'$. The identity element is $e = (1, 1)'$ and the inverse is : $a^{-1} = (a_1^{-1/a_2}, 1/a_2)'$. Figure 2 plots examples of Exp-log power transformations for various values of the parameters.

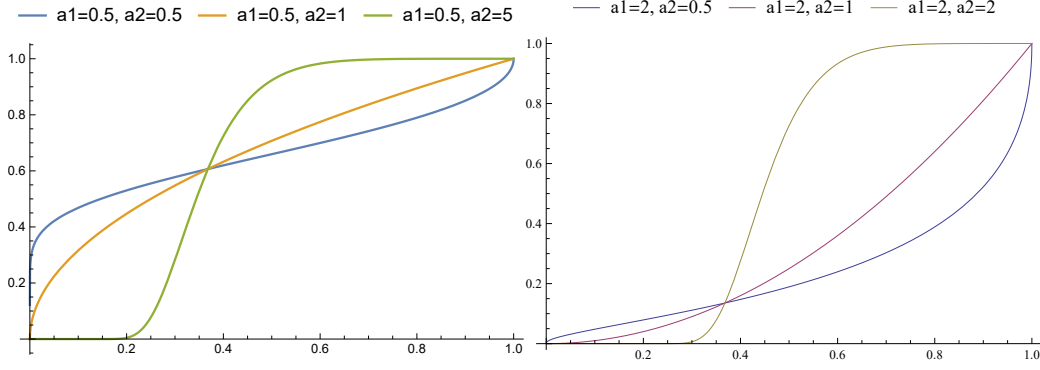


Figure 2: Examples of Exp-log power transformations. In the left (resp. right) panel, the three transformations share the same $a_1 = 0.5$ (resp. $a_2 = 2$), but have different a_2 . Thus the three curves pass through the same points $(0, 0)$, $(1, 1)$, and (e^{-1}, e^{-a_1}) .

Other groups can be derived by introducing piecewise transformations.

Example 4 : Piecewise transformation (multiscore)

The above elementary transformations can also be “combined” to increase the flexibility of the transformation. Let us for instance partition the interval $[0, 1]$ by deciles $k/10, k = 1, \dots, 10$. Then we consider 10 basic groups of transformations $c_k(a_k, y), k = 1, \dots, 10$, on $[0, 1]$, as well as ten monotonous bijections³ F_k , each mapping $[0, 1]$ to $[\frac{k-1}{10}, \frac{k}{10}]$. Then we construct the set of transformations :

$$c(a, y) = \sum_{k=1}^{10} \mathbb{1}\left(\frac{k-1}{10} \leq y \leq \frac{k}{10}\right) F_k \circ c_k(a_k, F_k^{-1}(y)). \quad (2.4)$$

It is easily checked that $c(a, \cdot)$ is one-to-one from $[0, 1]$ to itself. It defines the product group $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_{10}$ with multi-index : $a = (a'_1, \dots, a'_{10})'$, and the group operation, the identity and the inverse are all defined component-wise. This example shows how to define “splines” based on simple transformation groups. As a special case, one can set the transformations c_k to be

³For instance, we can choose the affine transformation: $F_k(y) = \frac{k-1+y}{10}, \forall y \in [0, 1]$.

identical up to bijections:

$$c_k(a_k, y) = F_k \circ c_1(a_k, F_k^{-1}(y)), \quad \forall k = 1, \dots, 10, a_k \in \mathcal{A}_1.$$

In other words $\mathcal{A}_1 = \mathcal{A}_2 = \dots = \mathcal{A}_{10}$ and the product group becomes $\mathcal{A} = (\mathcal{A}_1)^{10}$.

Example 5 : Exogenous switching regime

One shortcoming of the above spline group is that it maps an interval $[\frac{k-1}{10}, \frac{k}{10}[$ to itself. In other words, it does not allow the probability $\mathbb{P}[\frac{k-1}{10} \leq Y_t \leq \frac{k}{10} | a]$ to depend on a . Let us now construct a two-layer “hierarchical group”. For expository purpose, let us first divide $[0, 1]$ into two subintervals $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ only⁴.

First, remark that any $y \in [0, 1]$ can be alternatively represented by the couple

$$\begin{bmatrix} J(y) = E[2y] = \mathbb{1}_{2y > 1}, \\ \phi(y) = 2(y - \frac{1}{2}J(y)) \end{bmatrix} \in \{0, 1\} \times [0, 1], \quad (2.5)$$

where $E[2y]$ is the integer part of $2y$, that is the quotient of the Euclidean division of y by $\frac{1}{2}$, whereas $\phi(y)$ is twice the remainder in the division. As this division is unique, function $y \mapsto (J(y), \phi(y))$ is one-to-one from $[0, 1]$ to $\{0, 1\} \times [0, 1]$ and its inverse function is:

$$y = \frac{\phi(y) + J(y)}{2}. \quad (2.6)$$

In the following we will use alternately these two representations of a proportion value.

Let us now consider the group of permutations \mathcal{S}_2 on $\{0, 1\}$. It is composed of two elements that are the identity function Id , and the permutation p mapping 0 to 1. The group operation is the standard composition \circ of functions. These permutations can be equivalently represented by the functions:

$$y \mapsto by + (1 - b)(1 - y) := b(y).$$

If $b = 1$, we get the identity; if $b = 0$, we get permutation p . Moreover, the composition $b' \circ b$ can be represented by:

$$b' \circ b = b'b + (1 - b')(1 - b), \quad (2.7)$$

⁴The extension to a model with 10 subintervals, say, is straightforward and omitted.

where b' , b on the RHS should be regarded as elements of $\{0, 1\}$. Thus when there is no ambiguity we can write $Id = 1$, $p = 0$, that is, $\{0, 1\}$ is isomorphic to \mathcal{S}_2 .

We also assume a “baseline” group of transformations $c(a, \cdot)$ on $[0, 1]$, where a belongs to a certain group $(\mathcal{A}, *)$. Then we construct a transformation \tilde{c} parameterized by $\mathcal{S}_2 \times \mathcal{A}^2$ as follows: for any $b \in \mathcal{S}_2$, $a_0, a_1 \in \mathcal{A}$, we have, for any y equivalently reparameterized as $(J(y), \phi(y))'$:

$$\tilde{c}\left(\begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}, \begin{bmatrix} J(y) \\ \phi(y) \end{bmatrix}\right) = \begin{bmatrix} b(J(y)) \\ c(a_{b(J(y))}, \phi(y)) \end{bmatrix}, \quad (2.8)$$

or equivalently, by applying (2.6), (2.8), in terms of y , the transformation defined on $[0, 1]$ can be written as:

$$\begin{aligned} \tilde{c}\left(\begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}, y\right) &= b\left\{\mathbb{1}_{2y < 1} \frac{1}{2}c(a_0, 2y) + \mathbb{1}_{2y > 1} \left[\frac{1}{2} + \frac{1}{2}c(a_1, 2y - 1)\right]\right\} \\ &\quad + (1 - b)\left\{\mathbb{1}_{2y < 1} \left[\frac{1}{2} + \frac{1}{2}c(a_1, 2y)\right] + \mathbb{1}_{2y > 1} \frac{1}{2}c(a_0, 2y - 1)\right\}. \end{aligned} \quad (2.9)$$

In other words, the integer part $J(y)$ is transformed into another element of $b(J(y))$ of $\{0, 1\}$, whereas the remainder $\phi(y)$ is transformed into a new “remainder”, by applying the transformation $c(a_{b(J(y))})$. In Appendix 1 we

prove that the family of transformations \tilde{c} , indexed by $\theta = \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}$, defines a group for the operation:

$$\begin{bmatrix} b' \\ a'_0 \\ a'_1 \end{bmatrix} \tilde{*} \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} bb' + (1 - b)(1 - b)' \\ a'_0 * [b'a_0 + (1 - b')a_1] \\ a'_1 * [b'a_1 + (1 - b')a_0] \end{bmatrix} = \begin{bmatrix} b' \circ b \\ a_{b'(0)} \\ a_{b'(1)} \end{bmatrix}. \quad (2.10)$$

Finally, in all the examples above, the econometric model is deduced by substituting a parametrized index function $a(x_t, \beta)$ to parameter a indexing the set of transformations.

As an illustration, we plot the histogram of y_t , in a transformation model where U_t follows a beta distribution $B(3.5, 2.5)$. We consider three transformations: the identity, a spline (Example 4), as well as a hierarchical transformation (Example 5).

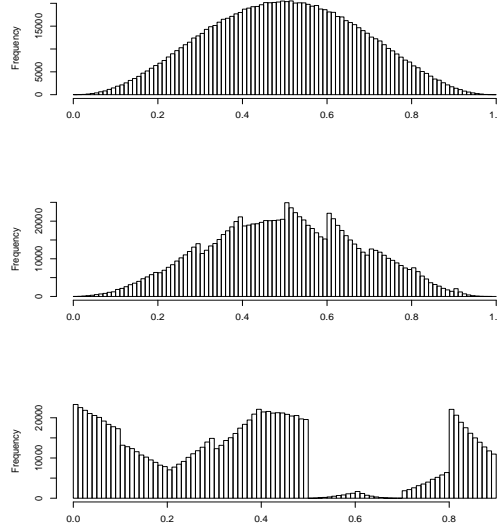


Figure 3: Histogram of simulated Y_t , when U_t follows a $B(3.5, 2.5)$ distribution. Upper panel: the transformation c is identity; middle panel: the transformation is a spline with 9 knots; lower panel: the transformation is a two-layer hierarchical transformation. In particular, in the latter case, the simulated histogram has three modes, two near the boundaries, one near 0.4.

2.2 Model with intercept

It is usual in linear models to consider a regression with intercept, that is to separate the intercept λ from the sensitivity parameters θ corresponding to the nonconstant explanatory variables:

$$y_t = x_t' \theta + \lambda + u_t = \lambda * x_t' \theta + u_t, \text{ say,} \quad (2.11)$$

where the group operation $* = +$ is the addition. This notion of intercept can be extended to any transformation group model by considering a specification :

$$u_t = c[\lambda * a(x_t, \theta), y_t], \quad t = 1, \dots, T, \quad (2.12)$$

where the intercept $\lambda \in \mathcal{A}$ has the same dimension as the index, is not constrained, and θ belongs to some appropriate parameter space Θ . From now on we assume that:

Assumption 4. *i) The data generating process (DGP) is a model with intercept, the true value of (λ, θ) being (λ_0, θ_0) .*

*ii) The parameter (λ, θ) is identifiable from $\lambda * a(x, \theta)$, that is,*

$$\lambda * a(x, \theta) = \lambda_0 * a(x, \theta_0), \quad \forall x \in \mathcal{X},$$

implies $\lambda = \lambda_0, \theta = \theta_0$, if $\lambda \in \mathcal{A}, \theta \in \Theta$.

Note that due to the group structure, the true model with intercept can also be written as a model without intercept :

$$v_t = c[a(x_t, \theta_0), y_t], \quad t = 1, \dots, T, \quad (2.13)$$

where $v_t = c(\lambda_0^{-1}, u_t)$ is another i.i.d. error term with values in $[0, 1]$. This transformation will be shown to be more useful for stress testing, whereas the representation (2.12) is more convenient for estimation purpose.

3 Least Impulse Response Estimator

Let us now introduce a smoothness condition on function $c(a, \cdot)$, necessary to define the effect on y of a small shock on error u .

Assumption 5. *The function $u \rightarrow c(a, u)$ is continuous and differentiable except at a finite number of points.*

For a model with intercept, the local impact on y_t of a small shock on u_t is measured by :

$$IR_t = \frac{\partial}{\partial u_t} y_t = \frac{\partial}{\partial u_t} c[(\lambda_0 * a(x_t, \theta_0))^{-1}, u_t] \quad (3.1)$$

$$= 1 / \frac{\partial c}{\partial u} [\lambda_0 * a(x_t; \theta_0), y_t], \quad (3.2)$$

where we have used the derivative formula of an inverse function to pass from equation (3.1) to (3.2).

Any transformation model is uniquely characterized by the form of its Impulse Response Function (IRF), since the function $c(a, \cdot)$ satisfies the boundary conditions $c(a, 0) = 0, c(a, 1) = 1$. Let us explicit the form of impulse response functions.

Example 1 : Power transformation (cont.)

We have : $\frac{\partial c}{\partial u}(a, y) = ay^{a-1}$, and $\log \frac{\partial c}{\partial u}(a, y) = \log a + (a - 1) \log y$.

Example 2 : Homographic transformation (cont.)

We have : $\frac{\partial c}{\partial u}(a, y) = \frac{a}{[1 + (a - 1)y]^2}$ and $\log \frac{\partial c}{\partial u}(a, y) = \log a - 2 \log(1 + (a - 1)y)$.

Example 3 : Exp-log power transformation (cont.)

We have : $\frac{\partial c}{\partial u}(a, y) = \frac{a_1 a_2}{y} (-\log y)^{a_2-1} c(a, y)$, and

$$\log \frac{\partial c}{\partial u}(a, y) = \log(a_1 a_2) - \log y + (a_2 - 1) \log(-\log y) - a_1 (-\log y)^{a_2}.$$

Thus, we get different patterns of the IRF, and the possibility to interpret the components of a as IRF or log IRF shape parameters. Figure 4 (resp. 5) plots the IRF of the transformations plotted in Figures 1 and 2 (resp. 3).

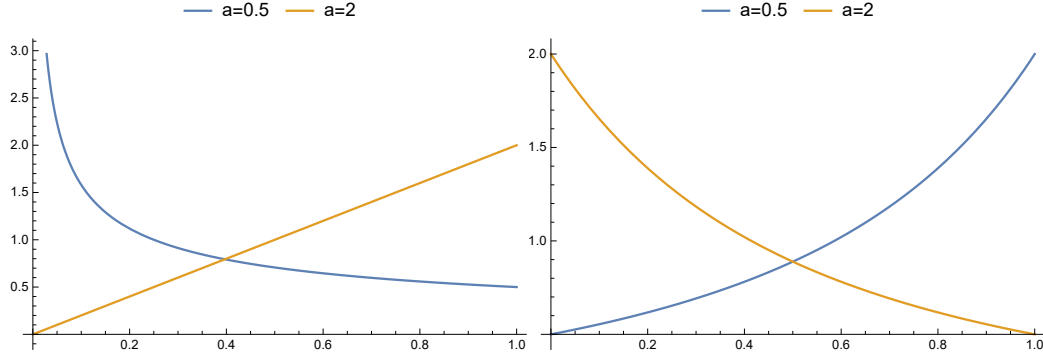


Figure 4: The corresponding IRF of the power transformations (left panel) and homographic transformations (right panel) considered in Figures 1 and 2.

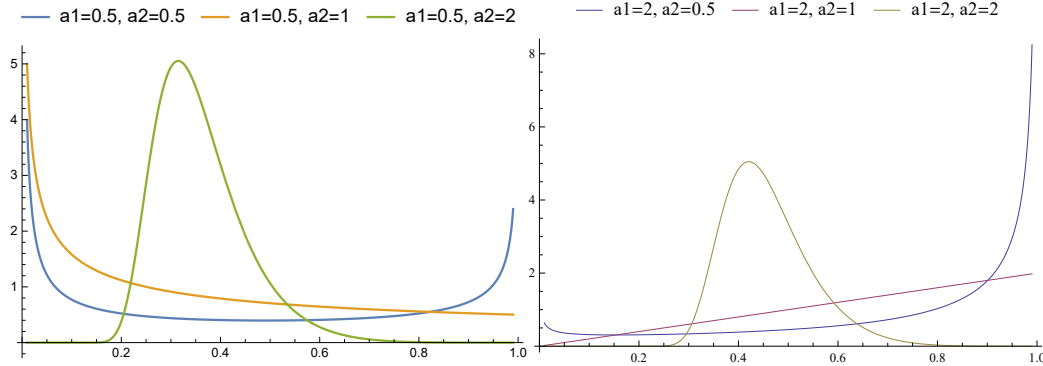


Figure 5: The corresponding IRF of the Exp-log power transformations considered in Figure 3.

3.1 The LIR estimator

We can estimate the parameter (λ, θ) by a pseudo maximum likelihood (PML) method, in which we assume a (misspecified) distribution for the error u_t . In this paper we suggest the uniform distribution, which has the advantage of leading to the interpretable Least Impulse Response estimator

(LIR). More precisely, under this distributional assumption, the conditional density of $y_t|x_t$ is,

$$\frac{\partial u_t}{\partial y_t} = c'(a(x_t, \beta), y_t),$$

by (2.2) and the Jacobian formula. Thus the pseudo log-likelihood function is :

$$L_T(\lambda, \theta) = \sum_{t=1}^T \log c'[\lambda * a(x_t, \theta), y_t]. \quad (3.3)$$

This objective function can also be written as :

$$L_T(\lambda, \theta) = - \sum_{t=1}^T \log IR_t(\lambda, \theta). \quad (3.4)$$

Thus the PML estimator minimizes the historical geometric average of the impulse responses. This motivates the terminology of LIR estimator.

3.2 Consistency

To analyze the consistency of the LIR estimator, we need regularity conditions (RC) to ensure the existence of a limiting criterion : $\tilde{L}_\infty(\lambda, \theta) = \lim_{T \rightarrow \infty} L_T(\lambda, \theta)/T$. These RC are standard and not explicitly written here [see e.g. Basawa et al. (1976), White (1994), Gouriéroux, Monfort, Zakoian (2018)]. Then we also need additional assumptions ensuring a simple form of \tilde{L}_∞ and that it admits a unique maximum.

Assumption 6. *i) The process of explanatory variables (x_t) is strongly stationary.*

ii) The variables u_t and x_t are independent.

iii) The function :

$$\tilde{l}_\infty(\lambda) = E_0[\log \frac{\partial c}{\partial u}(\lambda, u)] = \int_0^1 \log \frac{\partial c}{\partial u}(\lambda, u) f_0(u) du,$$

has a unique maximum on \mathcal{A} , attained at $\lambda = \tilde{e}_0$, say, where f_0 is the true p.d.f. of U .

Then we get the following consistency result :

Proposition 1. *Under Assumptions 1-6, the LIR estimator $\hat{\lambda}_T, \hat{\theta}_T$ is such that :*

$$\lim_{T \rightarrow \infty} \hat{\lambda}_T = \tilde{e}_0 * \lambda_0, \quad \text{and} \quad \lim_{T \rightarrow \infty} \hat{\theta}_T = \theta_0.$$

Proof. See Appendix 2. □

In other words, consistent estimators are obtained by minimizing the consequences of local stress on error u , *i.e.* the sensitivity of the reserves with respect to (local) shocks on u . This surprising result is due to the independence assumption⁵ between u_t and x_t , that allows to treat separately the local stress on x , measured via θ , and the local stress on u .

This result is the analogue of what is usually noted for Ordinary Least Squares (OLS) in a linear regression model $y_t = x_t' \theta + u_t$, say. When an estimator of θ is selected, the quality of the estimated model is usually measured by the sum of square residuals (SSR) : $\Sigma_t \hat{u}_t^2 = \Sigma_t (y_t - x_t' \hat{\theta}_T)^2$. The OLS approach selects an estimate minimizing the SSR to give the impression of a good accuracy of the estimated model. Nevertheless, this a priori "unfair" strategy provides a consistent estimator of the regression coefficient.

The LIR estimator of the intercept is in general not consistent. Since \tilde{e}_0 is independent of the true value of the intercept, we know that the theoretical asymptotic bias is equal to $(\tilde{e}_0 * \lambda_0)^{-1} * \lambda_0$. In fact the intercept estimator captures all the necessary adjustments needed to balance the misspecification of the error distribution. In particular, if the initial econometric model contains no intercept parameters, they have to be artificially introduced to ensure consistency.

Example 4 (cont.) : spline group.

As a special case, consider the spline transformation as in Example 4:

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

with $c_k^*(a_k, y) = \frac{k-1}{10} + \frac{1}{10} c_1(a_k, 10y - (k-1))$. Then we introduce the

⁵Only the independence between variables x_t and u_t is required, not that between processes (x_t) and (u_t) . This is a weak exogeneity assumption.

covariates, and make the following econometric specification:

$$a_k(x_t) = a(x_t, \beta_k) * \lambda_{0,k}, \quad \forall k = 1, \dots, 10.$$

In other words, the different score functions on different subintervals $[\frac{k-1}{10}, \frac{k}{10}[$ belong to the same parametric family. Then, by the above proposition, we can consistently estimate $\beta_1, \dots, \beta_{10}$. This suggests that it is also possible to construct statistical tests of the equality of these coefficients.

3.3 Nonlocal stress

The transformation model is also convenient to perform nonlocal stress test on error u . This is rendered difficult by the fact that $u_t = c(\lambda_0^{-1} * a(x_t, \theta_T)^{-1}, y_t)$ depends on the intercept λ_0 , for which we do not have a consistent estimator. The solution is to use equation (2.7), which rewrites the initial model with intercept and error u into an equivalent model without intercept and with transformed error:

$$v_t = c(\lambda_0^{-1}, u_t) = c[a(x_t, \theta_T)^{-1}, y_t].$$

This suggests to approximate v_t by its empirical counterpart:

$$\hat{v}_{t,T} = c[a(x_t, \hat{\theta}_T)^{-1}, y_t], t = 1, \dots, T. \quad (3.5)$$

Proposition 2. *Under Assumptions 1-6, the empirical c.d.f. of the $\hat{v}_{t,T}, t = 1, \dots, T$, converges pointwise to the true c.d.f. G_0 of v .*

From now on we assume, for expository purpose, that the transformation $c(a, \cdot)$ is increasing⁶. Then the c.d.f. G_0 of v is equal to:

$$\begin{aligned} G_0(v) &= P(v_t < v) = P_0[c(\lambda_0^{-1}, u_t) < v] \\ &= P_0(u_t < c(\lambda_0, v)) = F_0[c(\lambda_0, v)], \quad v \in [0, 1], \end{aligned}$$

where F_0 denotes the true c.d.f. of u_t . Thus although λ_0 and the distribution F_0 of (u_t) are not identified, the composite function $G_0(\cdot) = F_0[c(\lambda_0, \cdot)]$

⁶This assumption is satisfied by all the examples given in Section 2, except Example 5. The treatment of non-monotonous transformation follows the same principle and is omitted.

is identified. This is sufficient for a large variety of exercises, such as the simulation of y_t given x_t , as well as nonlocal stress analysis.

Remark 3 : The difficulty of identifying separately λ_0 and F_0 is not due to the LIR estimation approach, but is intrinsic to the transformation model. Indeed the two transformation models :

$$y_t = c[\lambda_0^{-1} * a^{-1}(x_t, \theta_0), u_t], u_t \sim F_0,$$

$$\text{and } y_t = c[e * a^{-1}(x, \theta_0), v_t], v_t \sim G_0,$$

lead to the same conditional distribution of y_t given x_t . Therefore we cannot disentangle the triples $(\lambda_0, \theta_0, F_0)$ and (e, θ_0, G_0) . This is the reflection problem mentioned by Manski (1993) for linear models.

Let us now explain how to derive the effect of a nonlocal stress on u . These stresses are usually defined by comparing the values of y corresponding to two quantiles of the distribution of u for given x , such as a quantile at $p_1 = 95\%$, say, and the median quantile at $p_2 = 50\%$. However since v is obtained from u by an increasing transformation, the two values of y to compare are also associated with the p_1 and p_2 quantiles of v . More precisely, these two values of y are:

$$y_j^s = c[a^{-1}(x, \theta_0), G_0^{-1}(p_j)], \quad j = 1, 2.$$

These values can be consistently approximated by replacing $G_0^{-1}(p_1), G_0^{-1}(p_2)$ and θ_0 with their empirical counterparts $\hat{G}_T^{-1}(p_1), \hat{G}_T^{-1}(p_2)$ and $\hat{\theta}_T$.

Remark 4 : Another consistent estimation method of θ is based on the covariance restrictions :

$$Cov(\alpha(x_t), \gamma\{c[a(x_t, \theta), y_t]\}) = 0,$$

valid for any pairs of (square integrable) functions α, γ . This set of covariance restrictions, where the $\alpha(x_t)$ plays the role of instrumental variables, is equivalent to the independence between x_t and u_t when $\theta = \theta_0$. We explain in Appendix 3 that the LIR approach asymptotically selects an appropriate subset of such covariance restrictions⁷.

⁷see Gouriéroux, Jasiak (2017), Section 4 and Appendix C, for the analysis of the class of generalized covariance estimators.

3.4 Comparison with the credit risk literature

As seen in Section 2.1, the LIR estimation approach can be applied to a class of semi-parametric econometric models with a number of underlying scores equal to the dimension of parameter a . As a consequence this modelling is much more flexible than the models currently considered in either the theoretical, or applied credit risk literature developed for the analysis of ELGD [see e.g. Qi, Zhao (2011), Yashkir, Yashkir (2013), Li et al. (2016), for the comparison of parametric modeling approaches for ELGD].

3.4.1 Beta Regression

The benchmark distribution for continuous variables with value in $(0,1)$ is the beta family [see e.g. Ferrari, Cribari-Neto (2008), Huang, Oosterlee (2011), Hartman-Wendels et al. (2014) for the its application to ELGD data⁸]. This is a two parameter family with density :

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)},$$

where the shape parameters α, β are positive and the beta function $B(\alpha, \beta)$ is defined by :

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad \text{with } \Gamma(\alpha) = \int_0^\infty \exp(-y)y^{\alpha-1}dy, \quad \forall \alpha > 0.$$

This family allows for densities with bell shape, U-shape, J-shape, or inverted J-shape. Figure 6 plots examples of beta densities for different values of the parameters α, β .

⁸So also Gupten, Stern (2002) for the LossCalc procedure developed by Moody's KMV.

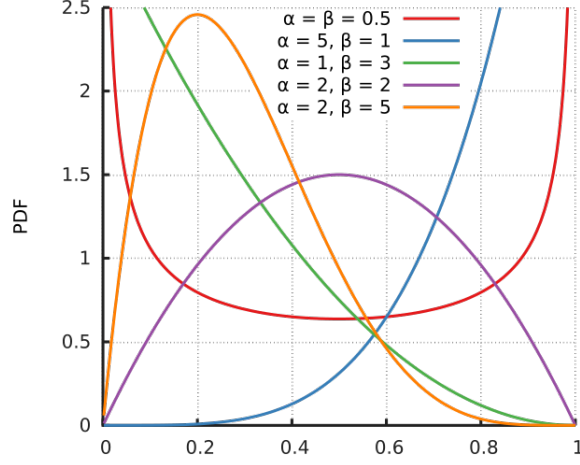


Figure 6: Density of a beta distribution.

The associated econometric model is deduced by letting parameters α, β depend on explanatory variables, such as : $\alpha = \exp(x'\theta_1), \beta = \exp(x'\theta_2)$, say. This parametric model is usually estimated by either a moment estimation method⁹, or by maximum likelihood. Compared to the new approach developed in the present paper, the beta modelling has the following drawbacks :

i) It is less flexible than a semi-parametric model, and the estimation approach can lead to inconsistent estimator if the true underlying conditional distribution of y_t given x_t is not beta.

ii) The derivation of the estimates can be computationally demanding since it requires the numerical computation of the gamma function and of its derivative (the digamma function).

iii) Finally, in order to derive the effect on a shock on the “error”, it is necessary to rewrite the model as a transformation model. This is usually done by defining $u_t = F(y_t|x_t)$, where F is the conditional c.d.f. of y_t given x_t . Then the transformation model is :

⁹Usually the calibration is based on the conditional moments of order 1 and 2, using the relation between the shape parameters and the mean μ and variance σ^2 of the beta distribution : $\mu = \frac{\alpha}{\alpha + \beta}, \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

$$y_t = F^{-1}(u_t|x_t) = \tilde{F}^{-1}(u_t|\alpha(x_t), \beta(x_t)), \quad (3.6)$$

where $\tilde{F}^{-1}(\cdot|\alpha, \beta)$ is the quantile function of the beta distribution with parameters α, β . However this quantile function has no closed-form expression, which makes it numerically difficult to analyze the effect of shocks on u , even by simulation.

3.4.2 Transformed Gaussian regression

An alternative modelling also leads to a semi-parametric model. The variable with value in $(0, 1)$ is first transformed into a variable with domain $(-\infty, \infty)$. Then a linear regression model is applied [see e.g. Atkinson (1985), p60]. A typical example is the logit regression :

$$\log \frac{y_t}{1 - y_t} = x_t' \theta + \epsilon_t, \quad (3.7)$$

where parameter θ is estimated by OLS. This approach involves only one index function (score) and is not flexible enough for the analysis of the ELGD. It can be extended by considering a regression model including conditional heteroskedasticity, leading to an analysis with two index functions (two scores) :

$$\text{logit} y_t = \log \frac{y_t}{1 - y_t} = a_1(x_t, \theta) + a_2(x_t, \theta) \epsilon_t, \quad (3.8)$$

and is usually estimated by Gaussian PML, that is, by applying a maximum likelihood approach based on the misspecified Gaussian assumption of v_t .

Now we remark that the inverse function of the logit function is the logistic function: $x \mapsto \frac{1}{1+\exp(-x)}$ mapping \mathbb{R} to $]0, 1[$. Then we define the transformed error u_t by $u_t = \frac{1}{1+\exp(-\epsilon_t)}$ so that equation (3.8) becomes:

$$\text{logit} y_t = a_1 + a_2 \text{logit} u_t.$$

It is easily seen that the above equation defines a group of transformations on $[0, 1]$:

$$y_t = c(a, u_t) = \frac{1}{1 + \exp \left(-a_1 - a_2 \log \frac{u_t}{1 - u_t} \right)}, \quad (3.9)$$

with group operation: $(b_1, b_2) * (a_1, a_2) = (b_1 + b_2 a_1, a_2 b_2)$, for all $(b_1, b_2), (a_1, a_2)$ belonging to \mathbb{R}^2 .

Thus the LIR approach developed in the present paper extends the regression approach on a transformed y . However, our estimation approach differs. Instead of applying a PML approach in which u_t has the distribution¹⁰ of $\frac{1}{1 + \exp(-\epsilon)}$, where $\epsilon \in \mathcal{N}(0, 1)$, we assume a uniform pseudo-distribution.¹¹

A common drawback of the beta regression and the transformed Gaussian regression is that the number of scores (two) is insufficient to capture the multimodalities of the (E)LGD distribution. Different extensions of the beta regression model have been introduced to solve this issue, but are rather ad-hoc as they require a preliminary treatment of the data to replace some modes by point masses. Typically, the observed ELGDs close to 0 (resp. to 1) are assigned to 0 (resp. 1), in order to create artificial point masses at 0 (resp. 1). Then the model is defined in two steps : first the modelling of the position of the ELGD, that is 0, 1, or value strictly between 0 and 1. Second a beta regression, when the observations are strictly between 0 and 1. Such an approach proposed in Ospina, Ferrari (2010), (2012) leads to the inflated beta regression.¹²

4 Joint Analysis of PD and LGD

The transformation model is easily extended to the joint analysis of two variables on $[0, 1]$, such as PD and ELGD. The transformation groups are first defined for endogenous variables and errors with values in \mathbb{R}^2 ; then both of them are transformed into variables in $(0, 1)$ by applying either a probit, or a logit function¹³ componentwise. Examples of basic bidimensional transformation groups include affine functions, rotations and Moebius transform, some of which admit up to 6 scores. Moreover, further flexibility can be accommodated for by considering bivariate piecewise transformations, in a similar way as their univariate counterparts introduced in Examples 4 and 5. Such an enlarged number of scores is desirable to account for the dependence between PD and ELGD variables.

¹⁰This distribution is called logit normal distribution.

¹¹Or equivalently replace the logit regression model by a probit regression model.

¹²It has also been proposed to apply Tobit type models with two underlying latent variables to such preliminarily treated data [see e.g. Sigrist, Stahel (2011)].

¹³Or any other given quantile function.

Example 5 : Bivariate affine model (6 scores)

For any $y \in \mathbb{R}^2$, we define: $c(a, A, y) = a + Ay$, where $a \in \mathbb{R}^2$, A belongs to $GL(2)$, i.e., the set of $(2, 2)$ invertible matrices. The group operation is :

$$(a, A) * (b, B) = (a + Ab, AB).$$

The identity element is : $e = (0, Id)$ and the inverse is : $(a, A)^{-1} = (-A^{-1}a, A^{-1})$.

After a logit transformation this model becomes :

$$\tilde{c}(a, B, y) = \begin{bmatrix} \psi \left(a_1 + A_1 \begin{bmatrix} \psi^{-1}(y_1) \\ \psi^{-1}(y_2) \end{bmatrix} \right) \\ \psi \left(a_2 + A_2 \begin{bmatrix} \psi^{-1}(y_1) \\ \psi^{-1}(y_2) \end{bmatrix} \right) \end{bmatrix},$$

where a_1, a_2 (resp. $y_1, y_2; A_1, A_2$) are the components of a (resp. components of y ; rows of A) and ψ denotes the c.d.f. of the logistic distribution. This is the extension of the model discussed in 3.4.2.

Example 6 : Moebius transformation¹⁴

A bidimensional real vector $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ [resp. $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$] can be equivalently represented by a complex number $y = y_1 + iy_2$ (resp. $u = u_1 + iu_2$). The Moebius transformations define a transformation group on $\mathbb{C} \cup \{\infty\}$, that is on the set of complex numbers augmented by a point at infinity. For our application this additional infinite element has no real impact, since it has zero mass both for the errors and endogenous variables. The Moebius transform is :

$$c(a, b, c; y) = \frac{ay + b}{cy + 1}, \quad (4.1)$$

where a, b, c are complex numbers, and with the following standard conventions apply :

$$c(a, b, c; \infty) = a/c, \quad (4.2)$$

$$c(a, b, c; -1/c) = \infty. \quad (4.3)$$

¹⁴See e.g. Arnold, Rogness (2008).

This is a model with 6 scores corresponding to the real and imaginary parts of a, b, c . The Moebius transformations form a group for the composition of functions, that can be transferred into a group on $[\mathbb{C} \cup \{\infty\}]^3 \sim (\mathbb{R}^2 \cup \{\infty\})^3$. The associated operation is :

$$(a, b, c) * (\tilde{a}, \tilde{b}, \tilde{c}) = \left(\frac{a\tilde{a} + b\tilde{c}}{c\tilde{b} + 1}, \frac{a\tilde{b} + b}{c\tilde{b} + 1}, \frac{c\tilde{a} + \tilde{c}}{c\tilde{b} + 1} \right). \quad (4.4)$$

The identity element is : $e = (1, 0, 0)$ and the inverse element is :

$$(a, b, c)^{-1} = (1/a, -b/a, -c/a).$$

The complex Moebius transformation (4.1) can be rewritten to highlight the transformation of real arguments (y_1, y_2) into (u_1, u_2) . Simple, but tedious, computations lead to the formulas :

$$\begin{aligned} u_1 &= \frac{(a_1 y_1 - a_2 y_2 + b_1)(c_1 y_1 - c_2 y_2 + 1) + (a_2 y_1 + a_1 y_2 + b_2)(c_2 y_1 + c_1 y_2)}{(c_1 y_1 - c_2 y_2 + 1)^2 + (c_2 y_1 + c_1 y_2)^2}, \\ u_2 &= \frac{(a_2 y_1 + a_1 y_2 + b_2)(c_1 y_1 - c_2 y_2 + 1) - (a_1 y_1 - a_2 y_2 + b_1)(c_2 y_1 + c_1 y_2)}{(c_1 y_1 - c_2 y_2 + 1)^2 + (c_2 y_1 + c_1 y_2)^2}. \end{aligned}$$

Thus the transformation from (y_1, y_2) to (u_1, u_2) , and its inverse are rational functions with quadratic numerator and denominator.

Other models can be derived by considering subgroups of the group of affine models, or of the group of Moebius transformations.

To illustrate the flexibility of this family of transformations, we simulate i.i.d. samples of (Y_1, Y_2) , using three different Moebius transformations. Figure 7 plots the joint isocontours of these three samples along with the marginal densities, obtained using a kernel estimator. We observe several modes in each marginal distribution whereas the joint distribution admits only one mode.

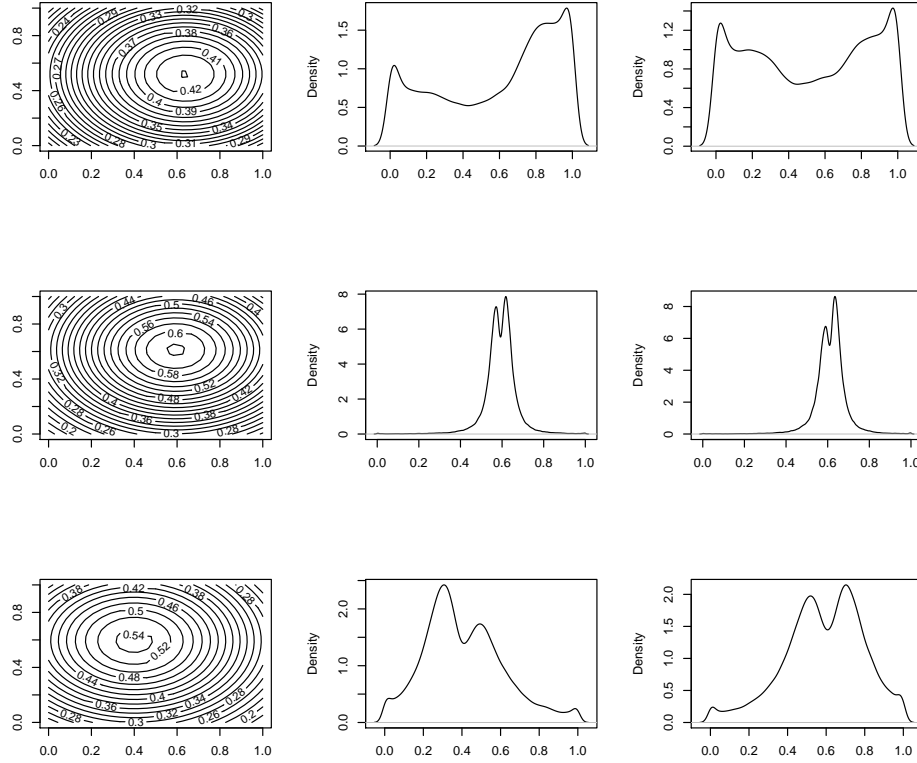


Figure 7: Joint isocontours (first column) and marginal histograms (second and third columns) for simulated couples (Y_{1t}, Y_{2t}) , when the distribution of (U_{1t}, U_{2t}) is independent uniform distribution. Figures of the same model are plotted in the same row.

We can observe that these samples are such that Y_{1t} and Y_{2t} are not independent. In particular, the Cartesian product of a pair of modes of the two marginal distributions is generically not a mode of the joint distribution (see for instance the upper panel).

Example 7 : A subgroup of Moebius transformations (4 scores)

We define a transformation¹⁵ on $\mathbb{C} \cup \{\infty\}$ by :

¹⁵It is easily checked the set $(\text{Im} y \geq 0) \cup \{\infty\}$ is invariant by these transformations. Therefore they also define a group on $(\text{Im} y \geq 0) \cup \{\infty\}$, or equivalently on $\mathbb{R} \times \mathbb{R}^+ \cup \{\infty\}$.

$$c(a, b; y) = \frac{ay - b}{\bar{b}y + \bar{a}},$$

where \bar{a} (resp. \bar{b}) denotes the complex conjugate of a (resp. b). The group operation is : $(a, b) * (a^*, b^*) = (aa^* - b\bar{b}^*, -ab^* - \bar{a}^*b)$. The identity element is : $e = (1, 0)$ and the inverse is :

$$(a, b)^{-1} = (-\bar{a}, b).$$

Example 8 : Rotation on the square $[0, 1]^2$

The rotation on \mathbb{R}^2 can be represented by orthogonal matrices :

$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$. They can be used to define rotations on the square by considering the transformation :

$$C(\theta; y) = R(\theta)y \frac{\max(|y_1|; |y_2|)}{\max(|y_1 \cos \theta - y_2 \sin \theta|, |y_1 \sin \theta + y_2 \cos \theta|)}.$$

This is a one-parameter group of transformations on $[0, 1]^2$, that can be combined with other transformations to increase the number of underlying scores.

After the transformations, the introduction of the explanatory variables and the intercept, we get a model analogue to (2.5), except that u_t, y_t are now with values in $[0, 1]^2$:

$$u_t = c[\lambda * a(x_t; \theta), y_t], \quad t = 1, \dots, T. \quad (4.5)$$

The pseudo log-likelihood function with pseudo uniform distribution on $[0, 1]^2$ for the errors is:

$$\begin{aligned} L_T(\lambda, \theta) &= \sum_{t=1}^T \log \left| \det \frac{\partial c}{\partial u} [\lambda * a(x_t; \theta), y_t] \right| \\ &= - \sum_{t=1}^T \log IR_t(\lambda, \theta), \end{aligned} \quad (4.6)$$

where now IR_t is a bivariate measure of the effects on both components of y of local shocks on both components of u . Formula (4.6) extends formula

(3.4). From Gouriéroux, Monfort, Zakoian (2018), the consistency property of these PML estimators remains valid in this extended case.

5 Concluding Remarks

The standard models, based on beta or logistic regressions, are not able to capture some features of the distribution of PD and ELGD, especially their (multiple) modes and how they depend on covariates. In this paper we have introduced a new class of semi-parametric models, which is rather flexible to analyze the (joint) behaviour of variables with values in $(0,1)$. We have also introduced consistent semi-parametric estimators, the Least Impulse Response (LIR) estimators, with interpretations in terms of local and/or global stresses.

References

- [1] Arnold, D., and S., Rogness (2008) : "Moebius Transformation Revealed", Notices of the AMS, 55, 1226-1231.
- [2] Atkinson, A. (1985) : "Plots, Transformations and Regressions : An Introduction to Graphical Methods of Diagnostic Regression Analysis", New-York, Oxford Univ Press.
- [3] Basawa, I., Feignin, P., and C., Heyde (1976) : "Asymptotic Properties of Maximum Likelihood Estimators for Stochastic Processes", Sankhya, The Indian Journal of Statistics, 38, 259-270.
- [4] Basel Committee on Banking Supervision (2001) : "The New Basel Capital Accord", Bank of International Settlements.
- [5] European Banking Authority (2016) : "Guidelines on PD Estimation, LGD Estimation and the Treatment of Defaulted Exposures", European Banking Authority, Consultation Paper 2016/21, 14 November.
- [6] Ferrari, S., and F., Cribari-Neto (2008) : "Beta Regression for Modelling Rates and Proportions", Journal of Applied Statistics, 31, 799-815.
- [7] Gouriéroux, C., and J., Jasiak (2017) : "Noncausal Vector Autoregressive Process : Representation, Identification and Semi-Parametric Estimation", Journal of Econometrics, 200, 118-134.
- [8] Gouriéroux, C., Monfort, A., and J.M., Zakoian (2018) : "Consistent Pseudo Maximum Likelihood Estimators and Transformation Groups", forthcoming Econometrica.
- [9] Gupton, G., and R., Stern (2002) : "LossCalc : Model for Predicting Loss-Given-Default", Moody's KMV.
- [10] Hartmann-Wendels, T., Miller, P., and E., Tows (2014) : "Loss-Given-Default for Leasing : Parametric and Nonparametric Estimations", Journal of Banking and Finance, 40, 364-375.
- [11] Hlawatsch, S, and S., Ostrowski (2011) : "Simulation and Estimation of Loss-Given-Default", Journal of Credit Risk, 7, 39-73.

- [12] Huang, X., and C., Oosterlee (2011) : "Generalized Beta Regression Models for Random Loss-Given-Default", *Journal of Credit Risk*, 7, 45-70.
- [13] Iwaniec, T., and G., Martin (1998) : "The Liouville Theorem, Analysis and Topology", 339-351, World Sci. Pub. River Edge, New-Jersey.
- [14] Li, P., Qi, M., Zhang, X., and X., Zhao (2016) : "Further Investigation of Parametric Loss-Given-Default Modelling", *Journal of Credit Risk*, 12, 17-47.
- [15] Manski, C. (1993) : "Identification of Endogenous Social Effects : The Reflection Problem", *Review of Economic Studies*, 60, 531-542.
- [16] Ospina, R., and L., Ferrari (2010) : "Inflated Beta Distributions", *Statistical Papers*, 51, 111-126.
- [17] Ospina, R., and S., Ferrari (2012) : "A General Class of Zero-or-One Inflated Beta Regression Models", *Computational Statistics and Data Analysis*, 56, 1609-1623.
- [18] Qi, M., and X., Zhao (2011) : "Comparison of Modeling Methods for Loss-Given-Default", *Journal of Banking and Finance*, 35, 2842-2859.
- [19] Sigrist, F., and W., Stahel (2011) : "Using the Censored Gamma Distribution for Modeling Fractional Response Variables with an Application to Loss-Given-Default", *ASTIN Bulletin*, 41, 673-710.
- [20] White, H. (1994) : "Estimation, Inference and Specification Analysis", Cambridge Univ. Press.
- [21] Yashkir, O., and Y., Yashkir (2013) : "Loss-Given-Default Modelling : A Comparative Analysis", *Journal of Risk Model Validation*, 7, 25-69.

Appendix 1

The group operation in Example 5

First, function \tilde{c} is clearly one-to-one on $[0, 1]$. Let us now check that we have a group of transformations. For $\theta = \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}$ and $\theta' = \begin{bmatrix} b' \\ a'_0 \\ a'_1 \end{bmatrix}$, we have:

$$\begin{aligned} \tilde{c}(\theta', \tilde{c}(\theta, y)) &= \frac{b'}{2} \mathbb{1}_{2\tilde{c}(\theta, y) > 1} + \frac{1-b'}{2} \mathbb{1}_{2\tilde{c}(\theta, y) < 1} \\ &+ \frac{b'}{2} c(a'_0, 2\tilde{c}(\theta, y)) \mathbb{1}_{2\tilde{c}(\theta, y) < 1} + \frac{b'}{2} c(a'_1, 2\tilde{c}(\theta, y) - 1) \mathbb{1}_{2\tilde{c}(\theta, y) > 1} \\ &+ \frac{1-b'}{2} c(a'_1, 2\tilde{c}(\theta, y)) \mathbb{1}_{2\tilde{c}(\theta, y) < 1} + \frac{1-b'}{2} c(a'_0, 2\tilde{c}(\theta, y) - 1) \mathbb{1}_{2\tilde{c}(\theta, y) > 1}. \end{aligned}$$

Combining with equation (2.9), we get:

$$\begin{aligned} &\tilde{c}(\theta', \tilde{c}(\theta, y)) \\ &= \frac{1}{2} [bb' + (1-b)(1-b)'] \mathbb{1}_{2y < 1} c(a'_0 * [b'a_0 + (1-b')a_1], 2y) \\ &+ \frac{1}{2} [bb' + (1-b)(1-b)'] \mathbb{1}_{2y > 1} \left\{ c(a'_1 * [b'a_1 + (1-b')a_0], 2y - 1) + 1 \right\} \\ &+ \frac{1}{2} [b'(1-b) + b(1-b)'] \mathbb{1}_{2y < 1} \left\{ c(a'_1 * [b'a_1 + (1-b')a_0], 2y) + 1 \right\} \\ &+ \frac{1}{2} [b'(1-b) + b(1-b)'] \mathbb{1}_{2y > 1} c(a'_0 * [b'a_0 + (1-b')a_1], 2y - 1, 2y) \end{aligned}$$

Then we remark that $bb' + (1-b)(1-b)'$ and $b'(1-b) + b(1-b)'$ are equal to $b' \circ b$ and $1 - b' \circ b$, respectively, and $b'a_0 + (1-b')a_1$, $b'a_1 + (1-b')a_0$ can be equivalently written as $a_{b'(0)}$, $a_{b'(1)}$, respectively. Thus the group operation on $\mathcal{S}_2 \times \mathcal{A}^2$ is:

$$\begin{bmatrix} b' \\ a'_0 \\ a'_1 \end{bmatrix} \tilde{*} \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} bb' + (1-b)(1-b)' \\ a'_0 * [b'a_0 + (1-b')a_1] \\ a'_1 * [b'a_1 + (1-b')a_0] \end{bmatrix} = \begin{bmatrix} b' \circ b \\ a_{b'(0)} \\ a_{b'(1)} \end{bmatrix}. \quad (4.7)$$

Appendix 2

Proof of Proposition 1

Let us follow the proof in [Gourieroux, Monfort, Zakoian \(2017\)](#). Under the assumptions of Proposition 1, the limiting objective function is :

$$\begin{aligned}\tilde{L}_\infty(\lambda, \theta) &= E_x E_0 \log \frac{\partial c}{\partial u}[\lambda * a(x, \theta), y] \\ &= E_x E_0 \log \frac{\partial c}{\partial u}[\lambda * a(x, \theta), c[a^{-1}(x, \theta_0) * \lambda_0^{-1}, u]],\end{aligned}$$

where E_x denotes the expectation with respect to the stationary distribution of x_t and E_0 with respect to the true p.d.f. f_0 for the error.

By the group structure, we get :

$$\begin{aligned}\tilde{L}_\infty(\lambda, \theta) &= E_x E_0 \log \frac{\partial c}{\partial u}[\lambda * a(x, \theta) * a^{-1}(x, \theta_0) * \lambda_0^{-1}, u] \\ &= E_x \tilde{l}_\infty[\lambda * a(x, \theta) * a^{-1}(x, \theta_0) * \lambda_0^{-1}, u],\end{aligned}$$

which is smaller than $\max_\lambda \tilde{l}_\infty(\lambda) = \tilde{l}_\infty(\tilde{\lambda}_0)$. Moreover this upper bound is reached for :

$$\theta_0^* = \theta_0, \lambda_0^* = \tilde{e}_0 * \lambda_0.$$

The consistency result follows by using the identification assumption A.4 ii).

Appendix 3

An Interpretation of Asymptotic First-Order Conditions

To get this interpretation, we first rewrite the limiting objective function $\tilde{L}_\infty(\lambda, \theta)$ with a change of index function : $A(x, \theta) = a(x, \theta) * a^{-1}(x, \theta_0)$, and of error term : $v = c(\lambda_0^{-1}, u)$. Then we have :

$$\tilde{L}_\infty(\lambda, \theta) = E_x E_0 \log \frac{\partial c}{\partial u}[\lambda * A(x, \theta), v],$$

where E_0 denotes the expectation with respect to the true distribution of v . Note that, for $\theta = \theta_0$, we get $A(x, \theta_0) = e$, independent of x . The group operation can be equivalently written as a function :

$$a * b \equiv h(a, b),$$

and later on we denote $\frac{\partial h}{\partial a}$ (resp. $\frac{\partial h}{\partial b}$) the partial derivative of h with respect to the first (resp. second) component of function h .

Let us now differentiate $\tilde{L}_\infty(\lambda, \theta) = E_x E_0 \log \frac{\partial c}{\partial u}\{h[\lambda, A(x, \theta)], v\}$ with respect to parameters λ, θ . We get :

$$\begin{cases} E_x E_0 \left(\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}[h(\lambda, A(x, \theta)), v] \frac{\partial h}{\partial a'}[\lambda, A(x, \theta)] \right) = 0, \\ E_x E_0 \left(\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}[h(\lambda, A(x, \theta)), v] \frac{\partial h}{\partial b'}[\lambda, A(x, \theta)] \frac{\partial A}{\partial \theta'}(x, \theta) \right) = 0. \end{cases}$$

When $\theta = \theta_0$, these First-Order Conditions (FOC) become :

$$\begin{cases} E_x E_0 \left(\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}(\lambda, v) \frac{\partial h}{\partial a'}(\lambda, e) \right) = 0, \\ E_x E_0 \left(\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}(\lambda, v) \frac{\partial h}{\partial b'}(\lambda, e) \frac{\partial A}{\partial \theta'}(x, \theta_0) \right) = 0, \end{cases}$$

since : $h(\lambda, A(x, \theta_0)) = h(\lambda, e) = \lambda * e = \lambda$.

Moreover, since : $\frac{\partial h}{\partial a'}(\lambda, e) = \frac{\partial}{\partial \lambda'}(\lambda * e) = \frac{\partial \lambda}{\partial \lambda'} = Id$, we get :

$$\left\{ \begin{array}{l} E_0 \left[\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}(\lambda, v) \right] = 0, \\ E_0 \left[\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}(\lambda, v) \right] E_x \left[\frac{\partial h}{\partial b'}(\lambda, e) \frac{\partial A}{\partial \theta'}(x, \theta_0) \right] = 0, \end{array} \right.$$

or : $E_0 \left[\frac{\partial}{\partial a'} \log \frac{\partial c}{\partial u}(\lambda, v) \right] = 0,$

since the second subset of conditions is automatically satisfied.

In fact the consistency result is a consequence of the covariance restriction :

$$Cov_{x,0} \left(\frac{\partial}{\partial a} \log \frac{\partial c}{\partial u}(\lambda, v), vec \left[\frac{\partial h}{\partial b'}(\lambda, e) \frac{\partial A}{\partial \theta'}(x, \theta_0) \right] \right) = 0,$$

that is satisfied due to the independence between x and v .

This interpretation in terms of covariance restrictions is asymptotic. In finite sample the first-order conditions admit also interpretations in terms of empirical autocovariance restrictions for commutative (Abelian) groups.