HAL open science

Least Impulse Response Estimator for Stress Test Exercises

Christian Gourieroux, Yang Lu

▶ To cite this version:

Christian Gourieroux, Yang Lu. Least Impulse Response Estimator for Stress Test Exercises. 2019. hal-02089698

HAL Id: hal-02089698
https://hal.science/hal-02089698

Preprint submitted on 4 Apr 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.

CEPN Centre d'économie de l'Université Paris Nord
CNRS UMR n° 7234

Document de travail N° 2019-05
Axe : Macroéconomie Appliquée Finance et Mondialisation

Least Impulse Response Estimator for Stress Test Exercises

Christian Gouriéroux* and Yang Lu**,

*University of Toronto and Toulouse School of Economics
**CEPN, Université Paris 13, Sorbonne Paris Cité

March 2019

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 the stress test exercises demanded in the current prudential regulation. We show that the Least Impulse Response Estimator, which minimizes the estimated effect of a stress, leads to consistent parameter 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. The approach is illustrated by both simulation experiments and the case study of a retail P2P lending portfolio.

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

JEL Codes: C51, G21

1 Introduction

The aim of this paper is to explain how recent results on transformation groups [Gourieroux, Monfort, Zakoian (2019)] 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)] :

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

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 (E)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 (E)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[1]. 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 existing approaches 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 (E)LGD. In particular we introduce models

[1]See e.g. Lutkepohl (2008) for the definition and use of the notion of Impulse Response Function in Economics.

based on Moebius transformations. The properties of the new approach are illustrated by simulation experiments in Section 6, as well as by a real loan portfolio in Section 6. Section 7 concludes. Proofs of asymptotic results are gathered in Appendices.

2 The Transformation Model

The objective of the new modelling is to extend to variables valued in $[0, 1]$ the standard linear model. This linear model can be written as:

$$y_t = x_t'\beta + u_t \qquad \Longleftrightarrow \qquad y_t - x_t'\beta = u_t,$$

depending on if either we want to understand how the output y_t reacts to the two types of inputs, that are $x_t'\beta$, the summary score function of the explanatory variables, and the unobserved error term u_t, or, if we focus on the reconstitution of the unobserved error u_t from observed data. This extension is the following. We consider a semi-parametric transformation model for variables valued in $(0, 1)$. The model can be written as :

$$y_t = c[a(x_t, \beta), u_t], \qquad t = 1, \ldots, T, \tag{2.1}$$

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;
- $u \to c(a, u)$ is a one-to-one function[2] mapping $[0, 1]$ to itself, parametrized by $a \in \mathcal{A} \subset \mathbb{R}^J$;
- $x \to a(x, \beta)$ is a score function from $\mathcal{X}$ to $\mathcal{A}$, parametrized by $\beta \in \mathcal{B} \subset \mathbb{R}^K$.

In applications to credit risk, the variable y_t can be the observed PD on a given segment, or the observed average loss given default (ELGD) on this

[2]Function $u \to c(a, u)$ is not necessarily monotonous. However, in many examples we consider in the paper it is indeed increasing. In these cases $u \to c(a, u)$ maps $(0, 1)$ to itself and satisfies $c(a, 0) = 0$, $c(a, 1) = 1$.

segment, and x_t the macro risk factors. In this section, the model is described for a single endogenous variable for expository purpose, but the results are easily extended to several endogenous variables. Such a bivariate extension is discussed in Section 4 for a joint analysis of PD and ELGD.

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

$$u_t = c^{-1}_{a(x_t,\beta)}(y_t), \tag{2.2}$$

where the error is expressed in terms of the explanatory variables and the endogenous variable.

Next, we assume that :

Assumption 1. *The errors $u_t, t = 1, \ldots, T$ 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.*

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 LGD[3] [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,[4] as well as a continuous component on $(0, 1)$. Therefore the assumption of continuous distribution

[3]or historical LGD [Gupton, Stein (2002)].

[4]In practice the loss can be larger than 1 for instance due to legal costs, or negative due to penalties and interest on delayed payments. These data are truncated to follow the supervision rules, see e.g. Altman et al. (2005). In other words the individual LGD data are first computed without truncation, before being sent truncated to the prudential authorities.

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 rating[5]. 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."

Assumption 1 is also satisfied for market values of the recovery rates, obtained for instance by comparing the market value of the debt (i.e. corporate bonds) just before default and its value one-month after default. This methodology [see e.g. Renault, Scaillet (2004), Bruche and Gonzalez-Aguado (2010)] is used by Moody's for firms with debt traded on an organized corporate bond market. These market-valued recovery rates are more sensitive to cyclical effects than the averages computed directly from the recovery processes. Note also that we consider a semi-parametric model with a vector of parameter β measuring the effect of variables x_t depending on time, and a functional parameter, that is the unknown distribution of (u_t). Therefore our approach differs from the pure nonparametric approaches without explanatory variables [Calabrese, Zenger (2010)], or applied on time independent segments defined for instance by seniority, or industry [Renault, Scaillet (2004)]. Indeed we are interested in the dynamic analysis of the ELGD and its dependence on macro risk factors x_t. This type of modelling is the basis for stress tests and the determination of the additional capital to be hedged against stressed situations.

2.1 Transformation Groups

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

Assumption 3. *We have :*

[5]These averages are either dollar weighted, or event weighted in practice. They typically lie strictly between 0 and 1 even if the individual LGD have not been artificially truncated.

$$c[a, c(b, u)] = c(a * b, u), \qquad \forall a, b \in \mathcal{A}, u \in [0, 1],$$

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

The group operation $*$ combines two elements $a, b \in \mathcal{A}$ of the group to form 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 :

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

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 : $c(a, u) = u^a$, 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 : $c(a, u) = \dfrac{au}{1 + (a - 1)u}$, 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.

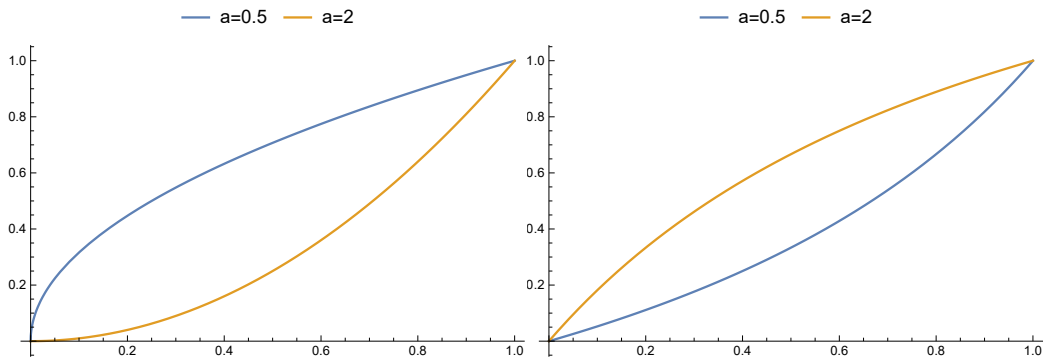

Figure 1: Plot of power transformations (left panel) and homographic transformations (right panel). In both cases, we plot the curves for two values of the parameter $a = 2$ and $a = 0.5$.

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

We have : $c(a, u) = \exp[-a_1(-\log u)^{a_2}]$, 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 parameter values.

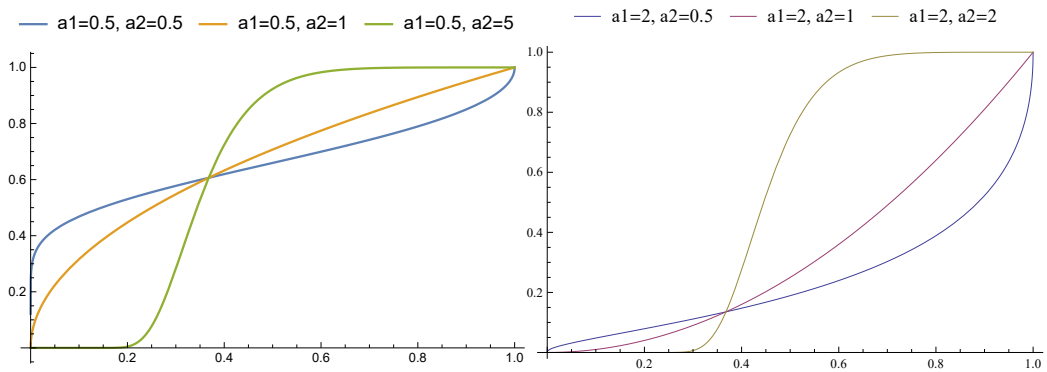

Figure 2: Examples of Exp-log power transformations. In the left (resp. right) panel, the three transformations share the same $a_1 = 0.5$ (resp. $a_2 = 2$), but have different a_2. Thus the three curves pass through the same points $(0,0)$, $(1,1)$, and (e^{-1}, e^{-a_1}).

Other groups can be derived by introducing piecewise transformations.

Example 4 : Piecewise transformation (multi-score)

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 $k/10, k = 1, \ldots, 10$. Then we consider 10 basic groups of transformations $c_k(a_k, u), k = 1, \ldots, 10$, on $[0,1]$, as well as ten monotonous bijections[6] F_k, each mapping $[0,1]$ to $[\frac{k-1}{10}, \frac{k}{10}]$. Then we construct the set of transformations :

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

It is easily checked that $c(a, \cdot)$ is one-to-one from $[0,1]$ to itself. It defines the product group $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_{10}$ with multi-score : $a = (a_1', \ldots, a_{10}')'$, 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

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

identical:

$$c_k(a_k, \cdot) = c_1(a_k, \cdot), \qquad \forall k = 1, ..., 10, a_k \in \mathcal{A}_1.$$

In other words $\mathcal{A}_1 = \mathcal{A}_2 = \cdots = \mathcal{A}_{10}$ 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 $[\frac{1}{2}, 1]$ only[7].

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

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

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

$$u = \frac{\phi(u) + J(u)}{2}. \tag{2.6}$$

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:

$$u \mapsto bu + (1 - b)(1 - u) := b(u).$$

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

$$b' \circ b = b'b + (1 - b')(1 - b), \tag{2.7}$$

[7]The 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 u equivalently reparameterized as $(J(u), \phi(u))'$:

$$\tilde{c}(\begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}, \begin{bmatrix} J(u) \\ \phi(u) \end{bmatrix}) = \begin{bmatrix} b(J(u)) \\ c(a_{b(J(u))}, \phi(u)) \end{bmatrix}, \tag{2.8}$$

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

$$\begin{aligned} \tilde{\tilde{c}}(\begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}, u) = b\Big\{ \mathbb{1}_{2u<1}\frac{1}{2}c(a_0, 2u) + \mathbb{1}_{2u>1}\big[\frac{1}{2} + \frac{1}{2}c(a_1, 2u-1)\big]\Big\} \\ + (1-b)\Big\{ \mathbb{1}_{2u<1}\big[\frac{1}{2} + \frac{1}{2}c(a_1, 2u)\big] + \mathbb{1}_{2u>1}\frac{1}{2}c(a_0, 2u-1)\big]\Big\}. \end{aligned} \tag{2.9}$$

In other words, the integer part $J(u)$ is transformed into another element of $b(J(u))$ of $\{0,1\}$, whereas the remainder $\phi(u)$ is transformed into a new "remainder", by applying the transformation $c(a_{b(J(u))})$. In Appendix 1 we prove that the family of transformations $\tilde{\tilde{c}}$, indexed by $\theta = \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix}$, defines a group for the operation:

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

Finally, in all the examples above, the econometric model is deduced by substituting a parametrized score 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 the beta distribution $\mathcal{B}(3.5,\ 2.5)$. We consider three transformations: the identity, a spline (Example 4), as well as a hierarchical transformation (Example 5).

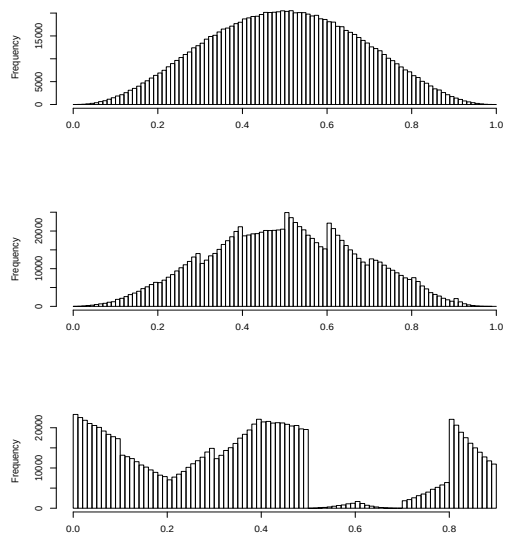

Figure 3: Histogram of simulated y_t, when u_t follows the $\mathcal{B}(3.5,\ 2.5)$ distribution. Upper panel: the transformation c is identity; middle panel: the transformation is a spline with 9 knots; lower panel: the transformation is a two-layer hierarchical transformation. In particular, in the 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 non-constant explanatory variables:

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

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

$$y_t = c[a(x_t, \theta) * \lambda, u_t], \qquad t = 1, \ldots, T, \tag{2.12}$$

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* $a(x, \theta) * \lambda$, *that is,*

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

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

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

$$v_t = c[a^{-1}(x_t, \theta_0), y_t], \qquad t = 1, \ldots, T, \tag{2.13}$$

where $v_t = c(\lambda_0, u_t)$ 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 $c(a, .)$, necessary to define the effect on y of a small shock on error u.

Assumption 5. *The function* $u \to c(a, u)$ *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 :

$$\mathrm{IR}_t = \frac{\partial y_t}{\partial u_t} = \frac{\partial}{\partial u_t} c[a(x_t, \theta_0) * \lambda_0, u_t] \tag{3.1}$$

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

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

Note that any transformation model is uniquely characterized by the form of its Impulse Response Function (IRF), since the function $c(a,.)$ 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 : $\dfrac{\partial c}{\partial u}(a,u) = au^{a-1}$, and $\log \dfrac{\partial c}{\partial u}(a,u) = \log a + (a-1)\log u$.

Example 2 : Homographic transformation (cont.)

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

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

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

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

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 displayed in Figures 1 and 2 (resp. 3).

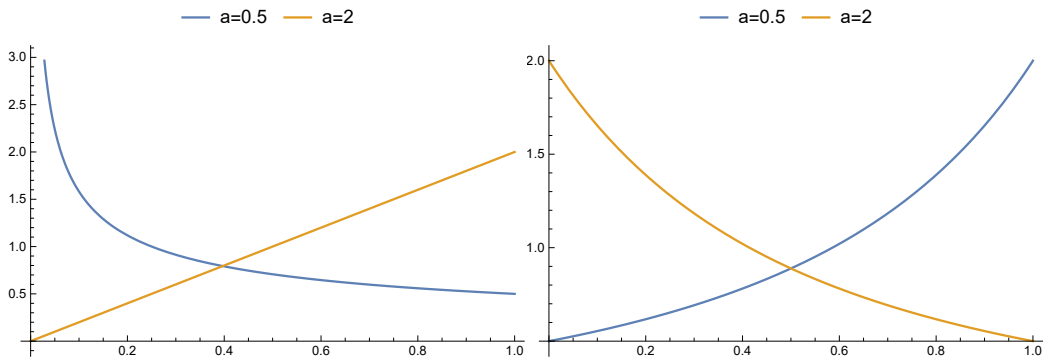

Figure 4: The IRF of the power transformations (left panel) and homographic transformations (right panel) considered in Figures 1 and 2, respectively.

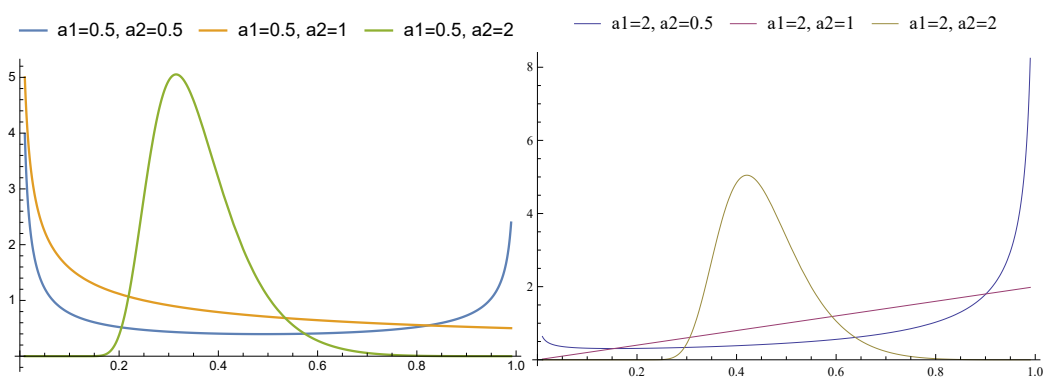

Figure 5: The IRF of the Exp-log power transformations considered in Figure 3.

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 given x_t is:

$$\frac{\partial u_t}{\partial y_t} = c'[a(x_t, \beta)^{-1}, y_t],$$

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

$$L_T(\lambda, \theta) = \sum_{t=1}^{T} \log \frac{\partial}{\partial u}[(a(x_t, \theta) * \lambda)^{-1}, y_t] = \sum_{t=1}^{T} \log \frac{\partial}{\partial u}[\lambda^{-1} * a(x_t, \theta)^{-1}, y_t]. \tag{3.3}$$

By (3.2), this objective function can also be written as :

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

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 : $\tilde{L}_\infty(\lambda, \theta) = \lim_{T\to\infty} L_T(\lambda, \theta)/T$. These RC are standard and not explicitly written here [see e.g. Basawa et al. (1976), White (1994), Gourieroux, Monfort, Zakoian (2019)]. Then we also need additional assumptions ensuring a simple form of $\tilde{L}_\infty$ 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 :*

$$\lim_{T\to\infty} \hat{\lambda}_T = \lambda_0 * (\tilde{e}_0)^{-1}, \qquad \textit{and} \qquad \lim_{T\to\infty} \hat{\theta}_T = \theta_0.$$

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 assumption[8] 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 $\hat{\lambda}_T$ 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 $\tilde{e}_0$, which is generically unrecoverable. 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 during the estimation to ensure the consistency of $\hat{\theta}_T$.

Example 4 (cont.) : spline group.
As a special case, consider the spline transformation as in Example 4:

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

[8]Only the independence between variables x_t and u_t is required, but not the independence between processes (x_t) and (u_t). This is a weak exogeneity assumption.

with $c_k^*(a_k, u) = \dfrac{k-1}{10} + \dfrac{1}{10} c_1(a_k, 10u - (k-1))$. Then we introduce the covariates, and make the following econometric specification:

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

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 $\beta_1, ..., \beta_{10}$. 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. At first sight this is rendered difficult by the fact that $u_t = c(\lambda_0^{-1} * a(x_t, \theta_0)^{-1}, y_t)$ 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:

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

This suggests to approximate v_t by its empirical counterpart:

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

Proposition 2. *Under Assumptions 1-6, the empirical c.d.f. of the $\hat{v}_{t,T}, t = 1, \ldots, T$, 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 increasing[9]. Then the c.d.f. G_0 of v is equal to:

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

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 $G_0(\cdot) = F_0[c(\lambda_0^{-1}, \cdot)]$

[9]This 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 exercises. In particular it is possible to estimate the quantiles of the ELGD distribution before and after stress, and then to deduce the additional capital to be hedged against stressed situation (see below).

Remark 1. *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 :*

$$y_t = c[a(x_t, \theta_0) * \lambda_0, u_t], \qquad u_t \sim F_0,$$

$$\textit{and} \quad y_t = c[a(x, \theta_0), v_t], \qquad v_t \sim G_0,$$

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 (i.e. ELGD) 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:

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

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 2. *Another consistent estimation method of θ is based on the covariance restrictions :*

$$Cov(\alpha(x_t), \gamma\{c[a(x_t, \theta)^{-1}, y_t]\}) = 0,$$

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 restrictions[10].

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 parametric 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), Calabrese (2014a), Huang, Oosterlee (2011), Hartman-Wendels et al. (2014) for its application to ELGD data[11]]. 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 :

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

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.

[10] see Gourieroux, Jasiak (2017), Section 4 and Appendix C, for the analysis of the class of generalized covariance estimators.

[11] So also Gupten, Stein (2002) for the LossCalc procedure developed by Moody's KMV.

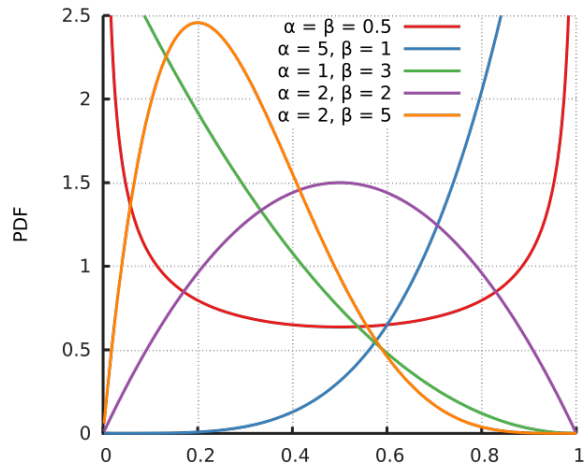

Figure 6: Density of a beta distribution.

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 [see e.g. Bruche, Gonzalez-Aguado (2010)], or another link function [see e.g. Calabrese (2014a), Section 3.1]. This parametric model is usually estimated by either a moment estimation method[12], 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 of 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

[12]Usually 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 : $\mu = \dfrac{\alpha}{\alpha+\beta}, \sigma^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

x_t. Then the transformation model is :

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

where $\tilde{F}^{-1}(.|\alpha, \beta)$ is the quantile function of the beta distribution $\mathcal{B}(\alpha, \beta)$. 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 leads to a semi-parametric model. The variable with value in $(0, 1)$ is first transformed[13] into a variable with domain $(-\infty, \infty)$. Then a linear regression model is applied [see e.g. Atkinson (1985), p. 60]. A typical example is the logit regression :

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

where parameter θ is estimated by OLS. This approach involves only one 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 scores :

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

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

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

The above equation defines a group of transformations on $[0, 1]$:

$$y_t = c(a, u_t) = \frac{1}{1 + \exp\big(- a_1 - a_2 \log \dfrac{u_t}{1 - u_t}\big)}, \tag{3.9}$$

[13] Sometimes the regression is defined from the recovery rate or the log-recovery rate without really taking into account the domain restrictions [see e.g. Altman et al. (2005b)].

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 distribution[14] of $\frac{1}{1+\exp(-\epsilon)}$, where $\epsilon \in \mathcal{N}(0,1)$, we assume a uniform pseudo-distribution, or equivalently replace the logit regression model by a probit regression model.

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 ELGD 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.[15]

4 Joint Analysis of PD and LGD

In practice the stresses impact both the PD's and the ELGD's. These impacts cannot be evaluated separately, since the PD's and ELGD's are linked [see e.g. the discussions in Altman et al. (2015b)]. They can depend on common macro-factors, but also on the more or less severe definition of default followed by the financial institution. Typically, a severe institution will declare in default some borrowers, that just have minor delinquencies, and for such borrowers the LGD will be small (often equal to zero). This can create a negative link between PD and LGD, even after the aggregation on homogeneous segments.

Thus there is a need for a joint modelling capable of capturing the multin-

[14]This distribution is called logit normal distribution.

[15]It 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)].

odality of the joint distribution, and its relationship with the modes that can be observed on the univariate distribution of PD (resp. ELGD). This can be easily achieved by extending the above transformation model. More precisely, 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, componentwise, either a probit, or a logit function, or any other given quantile function. Examples of basic bidimensional transformation groups include affine functions, rotations and the 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 3 and 4. Such an enlarged number of scores is desirable to account for the dependencies between PD and ELGD variables.

Example 5 : Bivariate affine model (6 scores)

For any $u \in \mathbb{R}^2$, we define: $c(a, A, u) = a + Au$, where $a \in \mathbb{R}^2$, A belongs to the group 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 :

$$\tilde{c}(a, B, u) = \begin{bmatrix} \psi\left(a_1 + A_1 \begin{bmatrix} \psi^{-1}(u_1) \\ \psi^{-1}(u_2) \end{bmatrix}\right) \\ \\ \psi\left(a_2 + A_2 \begin{bmatrix} \psi^{-1}(u_1) \\ \psi^{-1}(u_2) \end{bmatrix}\right) \end{bmatrix},$$

where a_1, a_2 (resp. $u_1, u_2; A_1, A_2$) are the components of a (resp. components of u; 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 transformation[16]

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. $u = u_1 + iu_2$), where

[16]See e.g. Arnold, Rogness (2008).

$i = \sqrt{-1}$. Moebius transformations are transformations in $\mathbb{C} \cup \{\infty\}$, 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. More precisely, the Moebius transform is :

$$c(a, b, c; u) = \frac{au + b}{cu + 1}, \tag{4.1}$$

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

$$\begin{aligned} c(a, b, c; \infty) &= a/c, & (4.2) \\ c(a, b, c; -1/c) &= \infty. & (4.3) \end{aligned}$$

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 :

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

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 (u_1, u_2) into (y_1, y_2). Simple, but tedious, computations lead to the formulas :

$$\begin{aligned} y_1 &= \frac{(a_1u_1 - a_2u_2 + b_1)(c_1u_1 - c_2u_2 + 1) + (a_2u_1 + a_1u_2 + b_2)(c_2u_1 + c_1u_2)}{(c_1u_1 - c_2u_2 + 1)^2 + (c_2u_1 + c_1u_2)^2}, \\ y_2 &= \frac{(a_2u_1 + a_1u_2 + b_2)(c_1u_1 - c_2u_2 + 1) - (a_1u_1 - a_2u_2 + b_1)(c_2u_1 + c_1u_2)}{(c_1u_1 - c_2u_2 + 1)^2 + (c_2u_1 + c_1u_2)^2}. \end{aligned}$$

Thus the transformation from (u_1, u_2) to (y_1, y_2), as well as 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 transformation group, we simulate i.i.d. samples of (y_1, y_2), using three different Moebius transformations and independent, uniformly distributed u_{1t}, u_{2t}. Figure 7 plots the joint isocontours of these three samples along with the marginal densities, obtained using a kernel estimator. The parameter values are set to:

$$a_1 = -0.1, a_2 = 0.3, b_1 = 10, b_2 = 0.3, c_1 = 0.5, c_2 = 2, \qquad \text{in the upper panel,}$$
$$a_1 = -10, a_2 = 10, b_1 = -3, b_2 = 3, c_1 = -2, c_2 = -5, \qquad \text{in the middle panel,}$$
$$a_1 = 1, a_2 = -3, b_1 = -10, b_2 = -3, c_1 = -5, c_2 = 2, \qquad \text{in the lower panel.}$$

We can observe that these samples are such that y_{1t} and y_{2t} are not independent. Indeed, we observe several modes in each marginal distribution whereas the joint distribution admits only one mode, so that the Cartesian product of a pair of modes of the two marginal distributions is generically not a mode of the joint distribution. In particular, the first simulation shows that two modes close to 0 and 1, which are typically observed on real LGD data, can exist on the marginal distributions, whereas the joint distribution has a single mode away from 0 and 1.

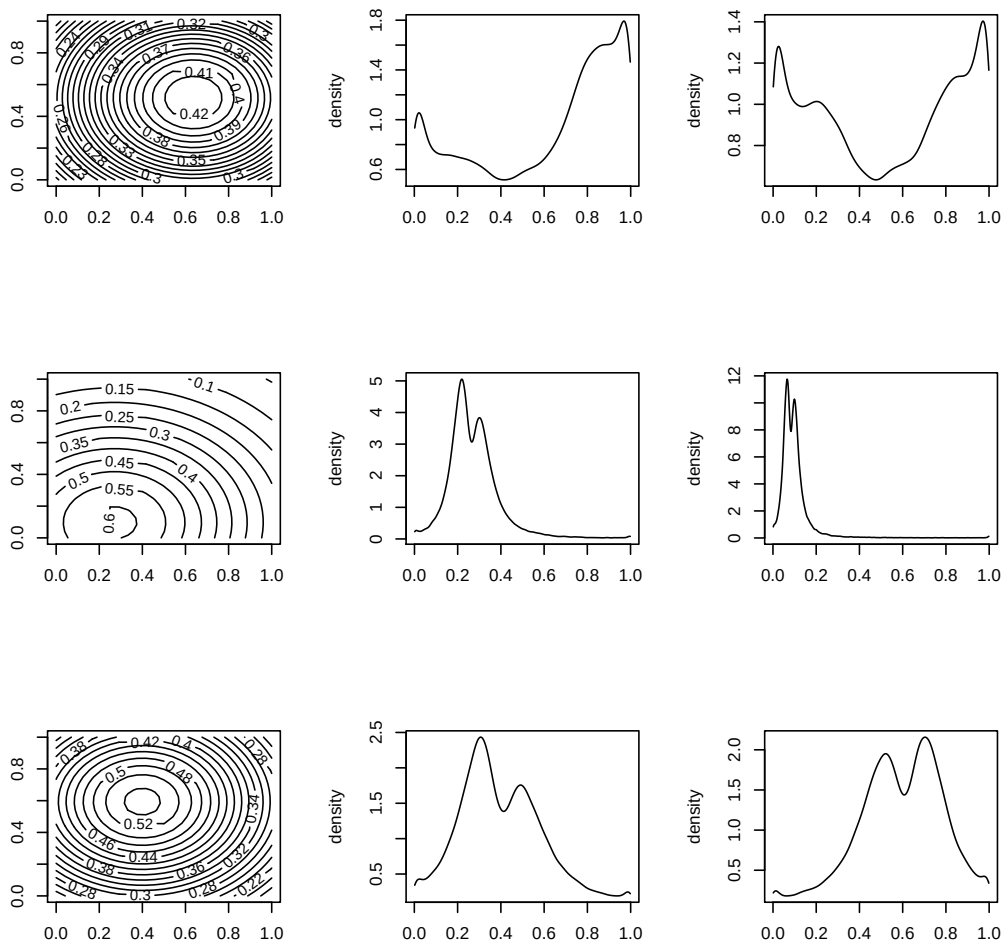

Figure 7: Joint isocontours (first column) and marginal histograms (second and third columns) for simulated couples of (y_{1t}, y_{2t}), obtained from Moebius transformations of (u_{1t}, u_{2t}) which are independent and uniformly distributed. Figures of the same model are plotted in the same row.

Example 7 : A subgroup of Moebius transformations (4 scores)
We define a transformation[17] on $\mathbb{C} \cup \{\infty\}$ by :

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

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

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 :

$$C(\theta;u) = \frac{\max(|u_1|;|u_2|)}{\max(|u_1\cos\theta - u_2\sin\theta|, |u_1\sin\theta + u_2\cos\theta|)} R(\theta)u.$$

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

$$y_t = c[a(x_t;\theta)*\lambda, u_t], \qquad t = 1,\ldots,T. \tag{4.5}$$

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

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

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 Gourieroux, Monfort, Zakoian (2019), the consistency property of these PML estimators remains valid in this extended case.

5 Simulation experiments

In this section we conduct two simulation experiments to illustrate the finite sample properties of the LIR estimator.[18] We first consider a model in which the distribution of u_t is multi-modal, and then propose a model in which the distribution of y_t is multi-modal. The common point of these two models is that the distribution of u_t (and hence that of y_t) is a mixture of simpler distributions. While mixture distributions have already been used in the LGD literature [see Calabrese (2014b)], existing models usually do not allow the mixture distribution to depend on the covariates.

5.1 A model with multi-modal distribution for u_t

The design of the first simulation is the following. For each replication, we simulate $T = 5000$ i.i.d. observations of $(x_t, y_t), t = 1, ..., T$, where x_t has a dimension of 8 and has independent components. Its first four components are Gaussian $\mathcal{N}(0, 0.5)$ distributed, the next four components are Bernoulli distributed with probability 0.8. Then y_t is defined by:

$$y_t = c[a(x_t, \theta), u_t],$$

where:

- c is the homographic transformation (see Example 2).
- (u_t) is an i.i.d. sequence, and has a mixture distribution with three components. The first component is the beta distribution $\mathcal{B}(12, 1)$, with weight 0.2, the second component is $\mathcal{B}(15, 20)$, with weight 0.56, and the third component is $\mathcal{B}(1, 12)$, with weight 0.24. Figure 8 plots the density function of u_t. This distribution has three modes and is very

[18] An R program allowing to replicate these experiments is available from the authors upon request.

different from the uniform distribution, which is the pseudo distribution used in the LIR estimation approach.

- the score function $a(x_t, \theta) = \exp(\theta' x_t)$, where parameter $\theta = (\theta_1, ..., \theta_8)$ is set to be: $\theta = (-1, -2, -1, +2, -1, +1, -2, +2)$.

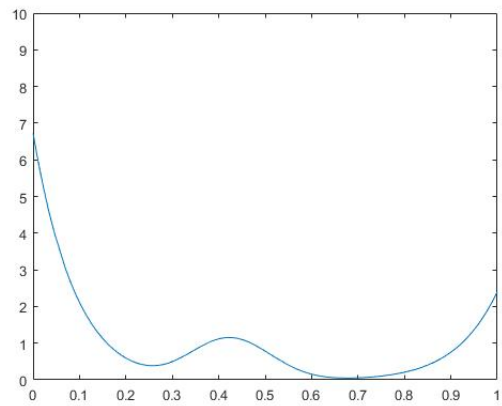

Figure 8: Density function of u_t.

We also plot, in Figure 9, the histogram of 5000 observations of y_t.

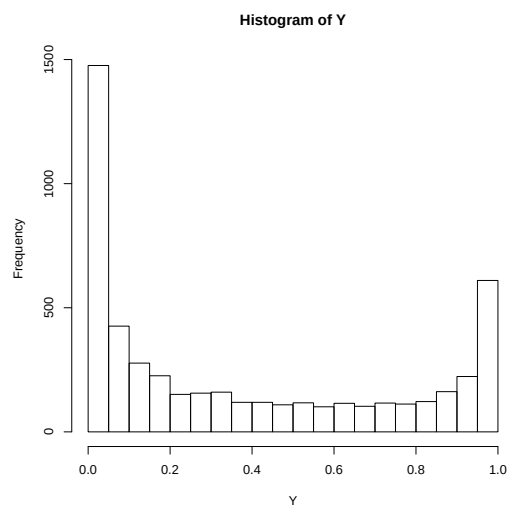

Figure 9: Histogram of y_t.

We see that the nonlinear transformation from u_t to y_t significantly dwarfs the intermediate mode of the distribution of u_t.

Then we use the LIR approach to estimate the parameter θ on the simulated data $(x_t, y_t)_{t=1,...,5000}$. Note that the DGP corresponds to an econometric model without intercept, but as explained in Section 3 below Proposition 1, in order to ensure the consistency of the LIR estimator, an intercept λ_0 has to be artificially added to the set of parameters to estimate. In other words, the LIR approach consists in maximizing the pseudo-likelihood of the model: $y_t = c[a(x_t, \theta)a_0, u_t]$, where θ and $a_0 = \exp(\lambda_0)$ are to be estimated jointly, and the error (u_t) is mis-specified to follow the uniformly distribution on $[0, 1]$.

This simulation exercise is repeated for a total of 400 replications, leading to 400 estimates of parameter $\hat{\theta}_T^{(i)}, i = 1, ..., 400$. Figure 10 plots the histogram of each of the eight components of $\hat{\theta}_T^{(i)}$, for $i = 1, ..., 400$.

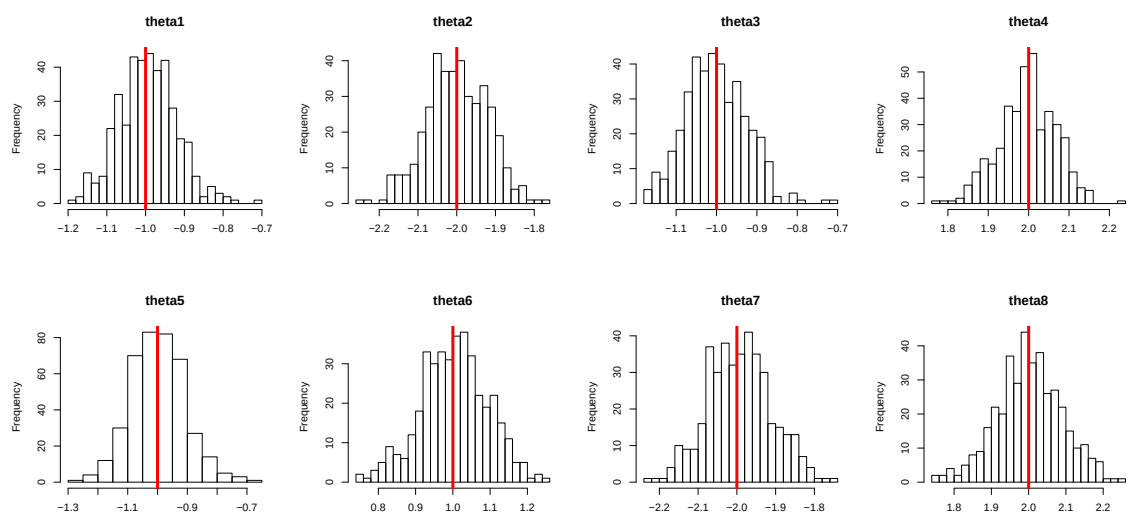

Figure 10: Histogram of the empirical distribution of the LIR estimator of θ, with a sample size of 400, where each $\hat{\theta}_T$ is estimated using a hypothetical dataset of $T = 5000$ observations. To facilitate comparison, in each of the eight histograms, we have plotted in red vertical line the theoretical values of the corresponding component of θ.

We see that the empirical distribution of each component of $\hat{\theta}_T$ is rather

concentrated around its theoretical value. This echoes the consistency property of the LIR estimator when T goes to infinity.

Once the value of θ estimated, we can use equation:

$$\hat{u}_{t,T} = c(a(x_t, \hat{\theta}_T)^{-1}, y_t) \tag{4.7}$$

to recover the empirical distribution of the error term u_t. Note that in equation (4.7), we have not used the estimate of the intercept $\hat{a_0}$. This latter is generically biased by Proposition 1, but by Proposition 2, the sample distribution of $(\hat{u}_{t,T})$ converges to the true distribution of $v_t = c(a(x_t, \theta)^{-1}, y_t)$ regardless of the inconsistency of $\hat{a_0}$. Moreover since the initial econometric model does not contain any intercept, we have $u_t = v_t$. In Figure 11, we plot the histogram of $\hat{u}_{t,T}$ for eight randomly selected Monte-Carlo replications, and check that as expected, in all these replications, the recovered histogram has a similar three-mode form as the true density displayed in Figure 8.

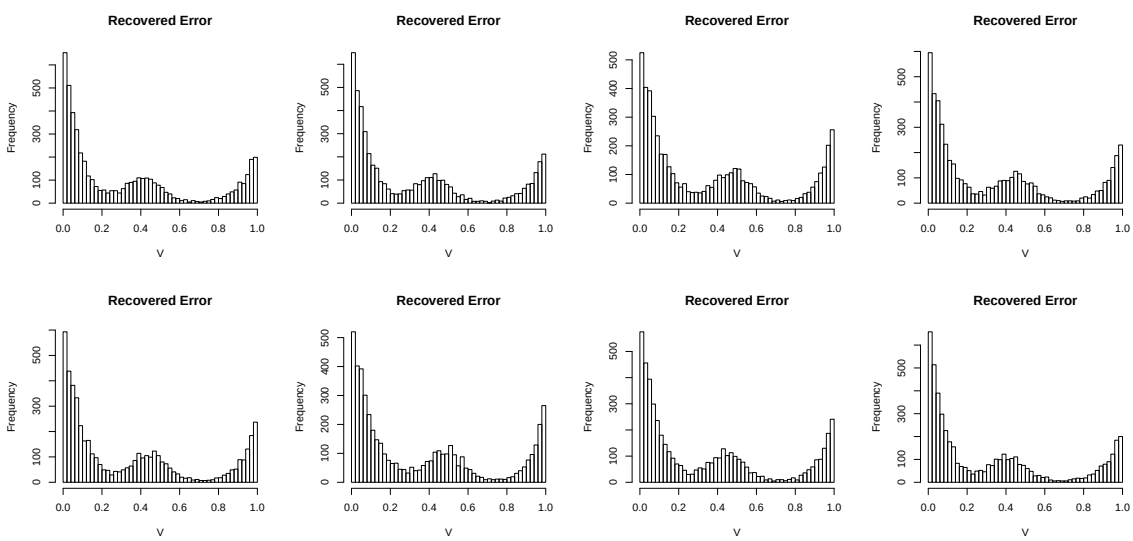

Figure 11: Histogram of the recovered errors in eight different Monte-Carlo replications.

5.2 A model with multi-modal distribution for LGD y_t

In the DGP of subsection 5.1, we have intentionally considered a multi-modal distribution for u_t, to demonstrate that even for this complicated model, the

LIR approach allows to recover consistently the parameter θ, as well as the distribution of v_t (or equivalently that of u_t if original model does not have intercept parameter). In credit risk applications, it is not uncommon to observe a multi-modal distribution for y_t, with additional modes besides the two modes close to 0 and 1. To this end we consider $T = 5000$ i.i.d. samples of (u_t, x_t), where u_t is drawn from a mixture of beta distribution. The two components of the mixture are $\mathcal{B}(6,1)$ and $\mathcal{B}(1,6)$ distributions, respectively, with equal weights. Then we assume that $x_t = (x_{1t}, x_{2t})$ is independent from u_t, and both components are mutually independent, Gaussian $\mathcal{N}(0,1)$ distributed. The transformation model is set to $y_t = c[a(x_t,\theta) * a_0, u_t]$, where c is a piecewise spline transformation (see Example 4) with scores:

$$c(a,u) = \mathbb{1}_{u<\frac{1}{3}} \frac{c_0(a_1, 3u)}{3} + \mathbb{1}_{\frac{1}{3}<u<\frac{2}{3}} \frac{c_0(a_2, 3u-1)+1}{3} + \mathbb{1}_{\frac{2}{3}<u} \frac{c_0(a_3, 3u-2)+2}{3},$$

where transformation c_0 is defined by:

$$c_0(a,u) = \frac{1}{1+\exp(-a \log \frac{u}{1-u})},$$

that is a constraint logistic transformation obtained by omitted the term a_1 in equation (3.9). Finally, the score functions are specified as follows:

$$\begin{aligned} a(x_t,\theta) &= (a_1(x_t,\theta), a_2(x_t,\theta), a_3(x_t,\theta)) \in \mathbb{R}^3, \\ \text{where } a_1(x_t,\theta) &= \exp(\theta_1 x_1 + \theta_2 x_2 + \lambda_1), \\ a_2(x_t,\theta) &= \exp(\theta_1 x_1 + \theta_2 x_2 + \lambda_2), \\ a_3(x_t,\theta) &= \exp(\theta_5 x_1 + \theta_6 x_2 + \lambda_3), \end{aligned}$$

and parameter value is set to:

$$\theta = (-0.1, -0.2, -0.1, 0.1, -0.1, 0.1).$$

Figure 12 plots the histogram of both u_t and y_t. While the distribution of u_t possesses two modes close to 0 and 1, respectively, that of y_t has one extra mode at around 0.5.

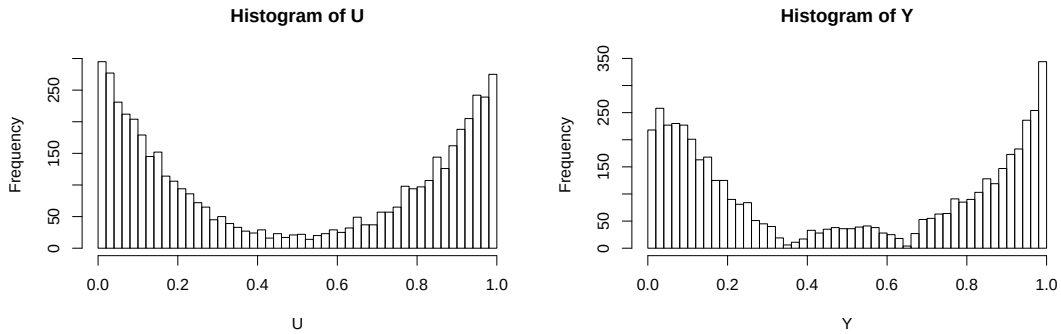

Figure 12: Left panel: histogram of u_t. Right panel: histogram of y_t.

Again, to evaluate the finite sample properties of the LIR estimator, we repeat the above simulation experiments for a total of 400 times. Figure 13 is the analogue of Figure 11, in which we plot the histogram of the 400 LIR estimators of $\theta_1, \theta_2, ..., \theta_6$, componentwise.

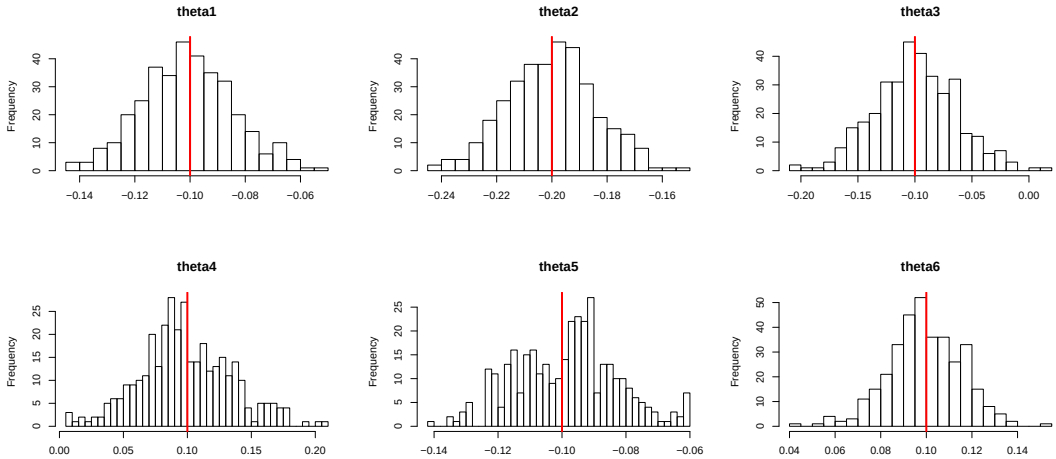

Figure 13: Histogram of the empirical distribution of the LIR estimator of θ, with a sample size of 400, where each $\hat{\theta}_T$ is estimated using a hypothetical dataset of $T = 5000$ observations. To facilitate comparison, in each of the eight histograms, we have plotted in red vertical line the theoretical values of the corresponding component of θ.

6 A real data application

In this section we analyse a personal loan portfolio. The data is downloaded freely from the website of Lending Club, which is an on-line, peer-to-peer (P2P) lending platform,[19] and concerns a total of $T = 5600$ loans that are issued on this platform between 2007 and 2011 and have already been charged off. These loans are typically of a small amount (up to 40,000 USD), and could be issued for various purposes such as house improvement, medical expenses, or credit card refinancing. Borrowers with low credit score have difficulty securing loans from usual banks and P2P lending is an alternative solution, with likely a rather high interest rate. From the lenders' perspective, it allows a large number of retail investors to earn a higher return than bank saving, without being a credit expert[20]. It is quickly gaining popularity, generating a greater than 100% year over year growth rate in the US in recent years. As the largest P2P lending company in the US, Lending Club has issued a total of 1.6 billion USD of loans from its inception in 2006 till 2015. Its dataset has previously also been used by Guo et al. (2016).

Lending Club's dataset does not provide LGD, but allows to compute another relevant proportion variable which we call the total loss ratio (TLR) of each defaulted loan. It is defined as one minus the repayment rate[21] of the balance (principal) at issuing. This ratio is important for the lender, which are retail investors. It differs from the LGD of the loan, but is linked to the latter through:

$$TLR = LGD \times \tau,$$

where τ is the ratio of the remaining balance at default event by the balance at issuing. We have checked that among the 5600 defaulted loans, the proportion of loans whose TLR is equal to 0 is negligible (about 1 percent), and they are discarded in the analysis. Moreover, there are no loans whose TLR is equal to or larger than 1. Figure 14 plots the histogram of the TLR variable, in which we see several local modes and in particular two modes near 1.

[19] See their website: www.lendingclub.com/info/download-data.action.

[20] For instance, on Lending Club, the typical interest rate of the loans lies between 6 % and 30%, with a loan period of 3 or 5 years.

[21] That is the ratio between the principal received today from the borrower and the total amount committed to that loan.

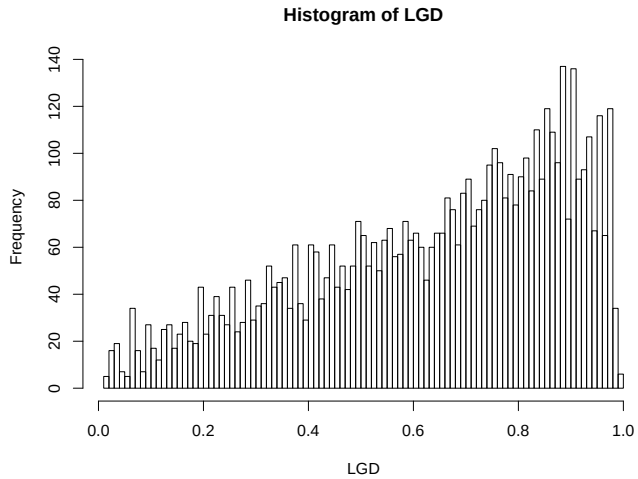

Figure 14: Histogram of the TLR.

Table 1 provides a summary of the explanatory variables we will use in the analysis.

Name of variable	type of variable	domain
Year of issuance of loan	categorical	2011, 2010, 2009, or before 2009
funded loan amount	continuous	$[0, \infty[$
Term of the loan	categorical	3 years or 5 years
Length of employment	categorical	$\{0, 1, ..., 10\}$
Delinquency	binary	0 or 1
Home ownership	categorical	"Rent", "mortgage", or "own"
Annual income	continuous	$[0, \infty[$
Debt ratio	continuous	$[0, 1]$

Table 1: Description of the covariates. The variable "delinquency" records whether the borrower has failed to make a regular debt payment during the past two years, with 0 corresponding to absence of delinquency, whereas the variable "Debt ratio" computes the ratio between the annual debt payment and the annual income.

Thus the regressors we use in the estimation includes the funded loan amount, the term of the loan, the length of the employment, the annual income and the debt ratio; two dummy variables characterizing the home

ownership, corresponding to the alternatives "mortgage" and "own"; three dummy variables defining the year of issuance, corresponding to years 2009, 2010 and 2011. Table 2 provides a breakdown of the defaulted loans by the values of the categorical explanatory variables as well as the corresponding average TLR.

Variable =Year of Issuance		
Value of variable	Number of observations	Average TLR
Before 2009	291	0.604
2009	582	0.627
2010	1464	0.620
2011	3259	0.647
Variable =Home ownership		
Value of variable	Number of observations	Average TLR
Mortgage	2335	0.634
Own	441	0.644
Rent	2820	0.636
Variable =Presence of delinquency in the past two years		
Value of variable	Number of observations	Average TLR
0	4910	0.635
1	686	0.817
Variable =Term of the loan		
Value of variable	Number of observations	Average TLR
3 years	3172	0.595
5 years	2424	0.689

Table 2: Breakdown of the defaulted loans by different values of the categorical variables

The pattern of the distribution of the TLR on the P2P platform clearly differs from patterns discussed in the previous sections. This is a consequence of the definition of variable of interest TLR, which is smaller than LGD, but also of the creditworthiness of the borrowers, with almost 50 % paying a huge interest rate of between 18 % and 25 %. In other words, these loans are similar to "junk bonds", the expected return being based on the payment of interest rates, when alive and on the recovery strategy.

Let us denote by x_t the set of all regressors (including the constant 1),

Year 2009	0.14
Year 2010	-0.19
Year 2011	-0.007
short term	-0.61
Employment duration	-0.016
House ownership	0.09
No Delinquency	-0.05
log income	-0.05

Table 3: Least Impulse Response parameter estimates

where t is the index of the borrower. Then we consider the homographic transformation model $y_t = c[a(x_t, \theta), u_t]$ where $a(x_t, \theta) = \exp(\theta' x_t + \lambda)$. Then we estimate (θ, λ) using the LIR Estimator.

Let us comment on the sign of these estimates. From the right panel of Figure 1, we know that when parameter a increases, function $c(a, u)$ is increasing. On the other hand, if a covariate X_i whose corresponding regression coefficient θ_i is positive (resp. negative), then $\exp(\theta_i X_i)$ is increasing in X_i. Thus, for instance, loans issued in 2010 tend to have a lower TLR than those issued in 2011 since $-0.05 < 0.23$, *ceteris paribus*. This phenomenon can be interpreted by the fact that borrowers are increasingly riskier in the aftermath of the financial crisis. Similarly, a shorter term or the absence of repayment delinquency are associated with a lower TLR. Note also that the sign in Table 3 are roughly consistent with the average TLR computed in Table 2. The only exception seems to be the dummy variable of Year 2011. Indeed, Table 2 indicates that loans issued in 2011 have a significantly higher TLR, whereas its corresponding estimate in Table 3 is close to zero. This can be explained by the fact that the correlation coefficient between this dummy variable and the short term dummy is -0.27, which is significantly negative.[22] In other words, in 2011, Lending Club has issued much more longer term loans than in previous years.

Finally, using the estimated parameter value, we recover the distribution of v_t. Figure 15 plots the histogram of $\hat{v}_{tT} = c[a(x_t, \hat{\theta}_T)^{-1}, y_t]$.

[22]This correlation coefficient becomes 0 and 0.27, respectively, if we replace the year dummy of 2011 by that of 2009 and 2010, respectively.

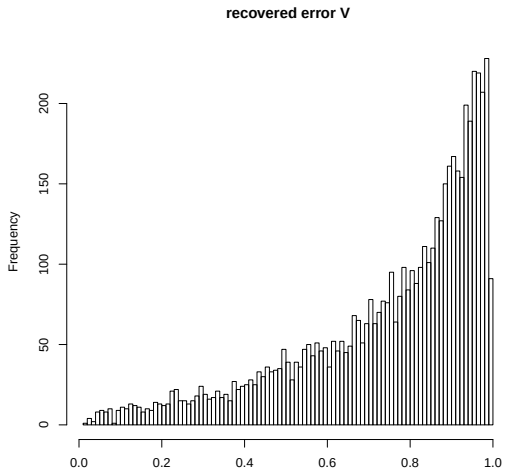

Figure 15: Histogram of the recovered error variable V.

We see that the recovered distribution of error variable y_t resembles more that of a beta distribution than the distribution of the TLR y_t. This echoes Remark 5.

7 Concluding Remarks

The standard models, based on beta or logistic regressions, are not sufficiently flexible 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 called group transformation models. This new family has several advantages. First, it is rather flexible, and includes several existing models such as logistic regression. Second, the group transformation models are associated with a consistent semi-parametric estimator called the Least Impulse Response (LIR) estimator. The LIR estimator is robust to mis-specification of the distribution of the error u_t, and has a nice interpretation in terms of local and/or global stresses. Thirdly, the model is especially convenient for stress testing exercises. Fourthly, while the existing reduced form models usually do not take into account the dependencies between PD and ELGD, our transformation model is easily extended to the analysis of the (joint) behaviour of variables with values in (0,1).

Finally, our approach has been illustrated through simulation studies and total loss ratio in P2P lending portfolio. However, both of these two exercises have focused on one response variable only, and the versatility of our approach in modelling bivariate rate/proportion variables remains empirically unexplored. In the future, it would be interesting to investigate the empirical performance of our framework, for a joint analysis of PD, ELGD, or PD, ELGD, CCF, as needed for the supervision of credit risk.

References

[1] Altman, E. I., Brady, B., Resti, A., and A., Sironi (2005a) : "The Link Between Default and Recovery Rates: Theory, Empirical Evidence, and Implications", Journal of Business, 78(6), 2203-2228.

[2] Altman, E., Resti, A., and A. Sironi (2005b) : "Loss Given Default: A Review of the Literature in Recovery Risk", NYU DP.

[3] Arnold, D., and S., Rogness (2008) : "Moebius Transformation Revealed', Notices of the AMS, 55, 1226-1231.

[4] Atkinson, A. (1985) : "Plots, Transformations and Regressions : An Introduction to Graphical Methods of Diagnostic Regression Analysis", New-York, Oxford Univ Press.

[5] 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.

[6] Basel Committee on Banking Supervision (2001) : "The New Basel Capital Accord", Bank of International Settlements.

[7] Bellotti, T., and J., Crook (2012) "Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards," International Journal of Forecasting 28(1), 171-182.

[8] Bruche, M., and C., Gonzalez-Aguado (2010) : "Recovery Rates, Default Probabilities, and the Credit Cycle," Journal of Banking and Finance 34(4), 754-764.

[9] Calabrese, R. (2014a) :"Predicting Bank Loan Recovery Rates with a Mixed Continuous-Discrete Model," Applied Stochastic Models in Business and Industry 30(2), 99-114.

[10] Calabrese, R. (2014b) :" Downturn Loss Given Default: Mixture Distribution Estimation," European Journal of Operational Research, 237, 271-277.

[11] Calabrese, R., and M., Zenga (2010) : "Bank Loan Recovery Rates: Measuring and Nonparametric Density Estimation," Journal of Banking and Finance 34(5), 903-911.

[12] European Banking Authority (2016) : "Guidelines on PD Estimation, LGD Estimation and the Treatment of Defaulted Exposures", European Banking Autority, Consultation Paper 2016/21, November 14th.

[13] Ferrari, S., and F., Cribari-Neto (2008) : "Beta Regression for Modelling Rates and Proportions", Journal of Applied Statistics, 31, 799-815.

[14] Gourieroux, C., and J., Jasiak (2017) : "Noncausal Vector Autoregressive Process : Representation, Identification and Semi-Parametric Estimation", Journal of Econometrics, 200, 118-134.

[15] Gourieroux, C., Monfort, A., and J.M., Zakoian (2019) : "Consistent Pseudo Maximum Likelihood Estimators and Transformation Groups", Econometrica, 87(1), 327-345.

[16] Guo, Y., Zhou, W., Luo, C., Liu, C., and H., Xiong (2016) : "Instance-based Credit Risk Assessment for Investment Decisions in P2P Lending", European Journal of Operational Research, 249(2), 417-426.

[17] Gupton, G., and R., Stein (2002) : "LossCalc : Model for Predicting Loss-Given-Default", Moody's KMV, New York.

[18] 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.

[19] Hlawatsch, S, and S., Ostrowski (2011) : "Simulation and Estimation of Loss-Given-Default", Journal of Credit Risk, 7, 39-73.

[20] Huang, X., and C., Oosterlee (2011) : "Generalized Beta Regression Models for Random Loss-Given-Default", Journal of Credit Risk, 7, 45-70.

[21] 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.

[22] Lutkepohl, H. (2008): "Impulse Response Function." The New Palgrave Dictionary of Economics, 2nd Edition, Palgrave Macmillan, London, 2008. 6141-6145.

[23] Manski, C. (1993) : "Identification of Endogenous Social Effects : The Reflection Problem", Review of Economic Studies, 60, 531-542.

[24] Ospina, R., and L., Ferrari (2010) : "Inflated Beta Distributions", Statistical Papers, 51, 111-126.

[25] 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.

[26] Qi, M., and X., Zhao (2011) : "Comparison of Modeling Methods for Loss-Given-Default", Journal of Banking and Finance, 35, 2842-2859.

[27] Renault, O., and O., Scaillet (2004) : "On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities," Journal of Banking and Finance 28(12), 2915-2931.

[28] 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.

[29] White, H. (1994) : "Estimation, Inference and Specification Analysis", Cambridge Univ. Press.

[30] 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 $\tilde{\tilde{c}}$ 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:

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

Combining with equation (2.9), we get:

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

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:

$$\begin{bmatrix} b' \\ a'_0 \\ a'_1 \end{bmatrix} \tilde{*} \begin{bmatrix} b \\ a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} bb' + (1-b)(1-b)' \\ a'_0 * [b'a_0 + (1-b')a_1] \\ a'_1 * [b'a_1 + (1-b')a_0] \end{bmatrix} = \begin{bmatrix} b' \circ b \\ a_{b'(0)} \\ a_{b'(1)} \end{bmatrix}. \tag{4.8}$$

Appendix 2

Proof of Proposition 1

Let us follow the proof in Gourieroux, Monfort, Zakoian (2019). Under the assumptions of Proposition 1, the limiting objective function is :

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

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 :

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

which is smaller than $\max_\lambda \tilde{l}_\infty(\lambda) = \tilde{l}_\infty(\tilde{\lambda}_0)$ [see Gourieroux et al. (2019)]. Moreover this upper bound is reached for :

$$\theta_0^* = \theta_0, \lambda_0^* = \lambda_0 * (\tilde{e}_0)^{-1}.$$

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 : $A(x,\theta) = a(x,\theta)^{-1} * a(x,\theta_0)$, and of error term : $v = c(\lambda_0^{-1}, u)$. Then we have :

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

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 $\dfrac{\partial h}{\partial a}$ (resp. $\dfrac{\partial h}{\partial b}$) the partial derivative of h with respect to the first (resp. second) component of function h.

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

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

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

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

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

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

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

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

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.