A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

Guillaume Flament[1] & Valentin Patilea[2]

[1] *CREST, Ensai, Univ. Rennes, and Square Research Center ; France, guillaume.flament@square-management.com*
[2] *CREST, Ensai, Univ. Rennes ; France, valentin.patilea@ensai.fr*

Résumé. Dans le contexte de la gestion des risques, les institutions financières ont tendance à utiliser des modèles dits “réglementaires”, comme par exemple le modèle Merton-Vašíček pour estimer la perte inattendue d’un portefeuille de crédits dans le cas des stress-tests. Dans ce modèle, un facteur commun Gaussien représente l’état de l’économie. Malheureusement, le modèle “réglementaire” ne permet pas de prendre en compte de manière explicite des données macroéconomiques qui permettraient de rafiner les prédictions. Pour ce faire, nous proposons de modéliser le facteur commun à l’aide d’une régression de type “location-scale” et d’estimer ses quantiles conditionnels. Les fonctions “location” et “scale” sont considérées de type semi-paramétriques à direction révélatrice unique, et la loi du terme d’erreur est générale. Plusieurs estimateurs non paramétriques de la fonction de répartition de l’erreur sont proposés. La performance du modèle est illustrée par des simulations. Enfin, une application sur des données réelles issues d’un exercice de stress-test climatique est présentée.

Mots clefs. Modèles semi-paramétrique, Lissage à noyau, Économétrie, Finance.

Abstract. In the context of risks management, financial institutions tend to use “regulatory models”, such for example the Merton-Vašíček model which estimates the unexpected loss of a credit portfolio in the case of credit stress-testing. In this model, a Gaussian common factor represents the state of the economy. This model does not allow to explicitly account for the information from macroeconomic data. We therefore propose to model this common factor with a semiparametric single-index location-scale model for estimating conditional quantiles of the common factor. Several nonparametric estimators of the error distribution function are proposed. The finite sample performance is investigated by a simulation study. Finally, an application on real data from a climate stress-testing exercise is presented.

Keywords. Semiparametric models, Kernel smoothing, Econometrics, Finance.

---

1 Introduction

Financial institutions are required to manage their unexpected losses, which refer to financial losses that cannot be accurately forecast using conventional risk management models. Unexpected losses can result from a large range of factors, such as incidents on financial markets (e.g., the subprime crisis), natural disasters (e.g., the tidal waves following earthquakes and the Fukushima nuclear power plant failure), economic crises (e.g., the Asian crisis in the 90’s), counterparty failures, sudden regulatory changes, etc. These losses are typically challenging to predict due to their complex and often unique nature. To prevent financial institutions defaults as consequence of such unexpected losses, for each borrower[1] all banks are required to have a minimum amount of capital (see also, Roncalli, 2020).

This capital requirement depends on the *EAD* which is *Exposure at Default* and represents the amount of money a lender is exposed to, or stands to lose, if a borrower fails to fulfill their financial obligations. It also depends on the maturity of the loan, the *Losses Given Defaults* (LGD) which is the amount that could not be retrieved by the bank after all assets are liquidated. It finally depends on the probability of default over a given period (e.g., one year). Let p be the unconditional probability of default of a lender.

In the Merton, 1974 model, a company defaults when the normalized value of its assets A falls below a given threshold B. Let D be a default indicator such that :

$$\{D = 1\} \iff \{A \leq B\}. \tag{1}$$

Therefore, assuming that A is standard Gaussian random variable, the unconditional probability of default is $p = \mathbb{P}(D = 1) = \Phi(B)$ where $\Phi(\cdot)$ is a standard Gaussian distribution function. Furthermore one can assume that A depends on an idiosyncratic factor ϵ and a random common factor Y both supposed to be standard Gaussian variables. The common factor is then interpreted as the state of the global economy, the lower its value the worst the state of the economy, and higher the probability of default. This leads to the so-called Merton-Vašíček model, see Vašíček, 2002. In that model, assuming further that the correlation between A and Y is equal to $\sqrt{\xi}$, the random variable A is decomposed as

$$A = \sqrt{\xi} Y + \sqrt{1 - \xi}\, \epsilon. \tag{2}$$

Given that $B = \Phi^{-1}(p)$, the conditional probability of default given $Y = y$ can be obtained by plugin (2) in (1) which yields :

$$\pi(y) = \mathbb{P}(\sqrt{\xi} y + \sqrt{1 - \xi}\epsilon \leq \Phi^{-1}(p)) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\xi}\, y}{\sqrt{1 - \xi}}\right). \tag{3}$$

By construction, we have $p = \mathbb{E}(\pi(Y))$. Moreover, $\pi(y)$ is a decreasing function of y. Using the relationship $\Phi^{-1}(1 - \alpha) = -\Phi^{-1}(\alpha)$, $\alpha \in (0, 1)$, and replacing y by minus the $\alpha-$th quantile of the standard Gaussian distribution, we can rewrite (3) and redefine under the form

$$\pi(\alpha) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\xi}\, \Phi^{-1}(\alpha)}{\sqrt{1 - \xi}}\right). \tag{4}$$

To prevent banks' failure due to extreme event, the regulator is interested by the evaluation the conditional probabilities corresponding to extreme left quantiles $\Phi^{-1}(\alpha)$. In the case of the Basel II regulation, $\alpha = 0.001$. The required capital due to credit risk for a banking

[1] The regulation is available here : https://www.bis.org/publ/bcbs107.pdf. Formulae for the Internal Rating based, or Foundation Internal Rating Based Ratings (FIRB), method are stated in part 2 section 3, paragraph 272.

institution is then depending on $\pi(\alpha)$. The higher the probability of default, the higher the required capital.

The classical Merton-Vašíček model presents several pitfalls. The most important one is that it assumes that the required capital does not directly depends on any covariates. Indeed, the regulatory formula (3) simply evaluates the conditional probability of default given $Y = y$, without any reference to observed covariates. In the short term, this is perhaps less problematic for the calculation of the required capital. Indeed, fixing this required capital under the scenario that a financial crisis will occur in the next year seems reasonable, since stress-testing exercises aim at protecting the bank from a catastrophic event. However, in the long run, for example for climate stress-testing exercises, the practitioners would prefer to take into account the dynamics of macroeconomic, financial and other types of predictors.

Therefore we construct a statistical model that links the common factor Y to a vector of covariates W (e.g., macroeconomic variables, financial variables). More precisely, our goal will be to estimate the conditional distribution function of Y given the covariates W. We will then be able to estimate the conditional quantiles of Y given W. In particular, this will allow practitioners to include covariates in the regulatory stress-testing exercise by simply replacing the unconditional (marginal) quantiles $\Phi^{-1}(\alpha)$ by their conditional versions in (4).

1.1 A brief look at the statistical model

Let $\tau \in (0,1)$ and define the conditional quantile function of order τ

$$q_\tau(w) = \inf\{y \ : \Psi_{Y|W}(y \mid w) \geq \tau\} \qquad \text{where} \qquad \Psi_{Y|W}(y \mid w) = \mathbb{P}(Y \leq y \mid W = w).$$

Here, $\Psi_{Y|W}$ is the conditional cumulative distribution function of Y given the covariate vector value, and we have

$$\Phi_Y(\cdot) = \mathbb{E}[\Psi_{Y|W}(\cdot \mid W)],$$

where Φ_Y is the unconditional (marginal) distribution of Y. Under the Gaussianity assumptions like in Merton-Vašíček's model, $\Phi_Y = \Phi$ and the conditional version of (3) given $W = w$ is

$$\pi(\tau \mid w) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\xi}\, q_\tau(w)}{\sqrt{1-\xi}}\right), \qquad \tau \in (0,1). \tag{5}$$

In our statistical analysis, τ can be any fixed value between 0 and 1. For the bank regulation rules, for example $\tau = \alpha = 0.001$.

We here consider that the value p is given. In the credit risk analysis, the probability of default p can be accurately estimated from external data, such as the so-called transition matrices obtained from the reports of the rating agencies.

A flexible model for the conditional distribution of Y given W, and thus for the conditional quantile function, can be constructed as a location-scale regression model

$$Y = m(Z) + \sigma(X)\varepsilon, \tag{6}$$

where $Z \in \mathbb{R}^{d_Z}$ and $X \in \mathbb{R}^{d_X}$ are subvectors of $W \in \mathbb{R}^{d_W}$, with possibly common components in which case $d_Z + d_X < d_W$. The error term ε is independent of W, and the functions m

and σ are unknown. For identification purposes, the variance $\text{Var}(\varepsilon)$ has to be set to some value. The location-scale model was extensively studied in the statistically literature. See, for example, Akritas and Van Keilegom, 2001, Neumeyer and Van Keilegom, 2010, Racine and Van Keilegom, 2020. In the location-scale model, the conditional quantile function is

$$q_\tau(w) = \inf_y \left\{ y \ : F_\varepsilon \left(\frac{y - m(z)}{\sigma(x)} \right) \geq \tau \right\}, \tag{7}$$

where $F_\varepsilon(\cdot)$ denotes the distribution function of ε. The vectors w, z and x belong to the support of W, Z and X, respectively.

In order to avoid the curse of dimensionality due to the nonparametric estimation of the multivariate location and scale functions m and σ, we here consider a single-index modeling of these functions. Under the single-index assumptions, the model (6) becomes

$$Y = m(Z^\top \gamma) + \sigma(X^\top \beta)\varepsilon, \tag{8}$$

where γ and β are unknown vectors and $m(\cdot)$ and $\sigma(\cdot)$ are now univariate function to be estimated. See also Neumeyer and Van Keilegom, 2010.

In this paper, we study the semiparametric single-index location-scale model (8). Moreover, we consider a smooth estimator of $F_\varepsilon(\cdot)$ using the residuals of the single-index location scale model, following the lines of Azzalini, 1981, see also Neumeyer and Van Keilegom, 2010 and Racine and Van Keilegom, 2020. Keeping in mind the application to credit risk and stress-testing, we consider that the marginal distribution of the response Y is known (for example is the standard normal). This introduces some specific identification constraints in the treatment of the model (8). Knowing the marginal distribution of the response also allows us to propose a data-driven rule for the smooth estimates of the error distribution function. The paper is organized as follows. In section 2 we introduce the single-index estimators of the location and scale functions. Moreover, we consider the estimator of the error distribution function F_ε obtained as the empirical distribution function of the residuals of the semiparametric location-scale model (8). For the purpose of estimating extreme conditional quantiles, we also consider smoothed versions of the empirical distribution function of the residuals. The finite sample performance of our estimators is investigated by simulations, the results are presented in section 3.1. A real data application is presented in section 3.2.

2 Model estimation

The goal is to estimate the finite-dimensional parameters γ and β, the univariate functions $m(\cdot)$ and $\sigma(\cdot)$ and the distribution F_ε of the error term ε. Plugging into (7) these estimates yields the estimate of $q_\tau(w)$. Let (Y_i, W_i), $1 \leq i \leq n$, be an independent sample of $(Y, W) \in \mathbb{R} \times \mathbb{R}^{d_W}$. Thus Z_i and X_i are independent copies of Z and X, the subvectors of W appearing in (8).

2.1 Single-index estimators

In the first step, we estimate γ and m by semiparametric least squares. By construction, we have

$$\mathbb{E}[Y \mid W] = \mathbb{E}[Y \mid Z^\top \gamma], \qquad \text{for some } \gamma \in \mathbb{R}^{d_Z}.$$

For identification purposes, we set a component γ (e.g., the first one) equal to 1. Next, given a value γ, we consider the regression function

$$m(t;\gamma) = \mathbb{E}[Y \mid Z^\top \gamma = t], \qquad t \in \mathbb{R}.$$

This function can be estimated by local linear smoothing (Fan and Gijbels, 1996), that is

$$\widehat{m}(t;\gamma) = \arg_a \min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^n \left\{Y_i - a - b(Z_i^\top \gamma - t)\right\}^2 k\left(\frac{Z_i^\top \gamma - t}{h}\right), \tag{9}$$

where k is a second order symmetric kernel and h is the bandwidth. The index γ is then defined as the least squares estimator

$$\widehat{\gamma} = \arg\min_\gamma \sum_{i=1}^n \left\{Y_i - \widehat{m}(Z_i\gamma^\top;\gamma)\right\}^2, \tag{10}$$

where the optimization is considered under the constraint that the first component of γ is equal to 1. See, for example, Ichimura, 1993, Hardle et al., 1993, Carroll et al., 1997, Delecroix et al., 2006 for some references on semiparametric single index models.

To estimate the scale factor $\sigma(\cdot)$ under the single-index assumption, we set the first component of β equal to 1, and define

$$\sigma^2(t;\beta) = \frac{1}{\mathrm{Var}(\varepsilon)}\mathbb{E}\left[\left\{Y - m(Z^\top\gamma;\gamma)\right\}^2 \mid X^\top\beta = t\right], \qquad t \in \mathbb{R}.$$

In usual location scale models, the variance of ε is set equal to 1, and hence no scaling by the inverse of $\mathrm{Var}(\varepsilon)$ is required. However, for the application we have in mind for stress-testing, where there is an additional information on the unconditional distribution of Y (typically, Y is standard Gaussian), the usual identification choice $\mathrm{Var}(\varepsilon) = 1$ is no longer admissible. We therefore have to set a specific identification constraint for the scale function, such for instance

$$\mathrm{Var}(\varepsilon) = \sigma_Y^2/2 := \mathrm{Var}(Y)/2.$$

To estimate $\sigma^2(t;\beta)$ we consider the local linear smoother

$$\widehat{\sigma}^2(t;\beta) = \arg_a \min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^n \left\{(2/\sigma_Y^2)\left[Y_i - \widehat{m}(Z_i^\top\widehat{\gamma};\widehat{\gamma})\right]^2 - a - b(X_i^\top\beta - t)\right\}^2 k\left(\frac{X_i^\top\beta - t}{h}\right). \tag{11}$$

For simplicity we use the same bandwidth rate for the local linear estimators of $m(\cdot;\gamma)$ and $\sigma^2(\cdot;\beta)$. The semiparametric least squares estimator of β is then

$$\widehat{\beta} = \arg\min_\beta \sum_{i=1}^n \left[(2/\sigma_Y^2)\left\{Y_i - \widehat{m}(Z_i^\top\widehat{\gamma};\widehat{\gamma})\right\}^2 - \widehat{\sigma}^2(X_i^\top\beta;\beta)\right]^2, \tag{12}$$

the minimization being considered under the identification assumption that the first component of β is equal to 1. See Zhu et al., 2013; see also Fan and Yao, 1998, Yin et al., 2010. Let us point out that, the same $\widehat{\beta}$ can be obtained without the scaling factor $(2/\sigma_Y^2)$ in (11) and (12). However, the scaling is necessary for the construction of the residuals and the estimation of the distribution function, as explained below.

2.2 Error distribution function estimators

Given the estimators $\widehat{m}(Z_i^\top\widehat{\gamma};\widehat{\gamma})$ and $\widehat{\sigma}^2(X_i^\top\widehat{\beta};\widehat{\beta})$ obtained from (9), (10) and (11), (12), respectively, we define the residuals

$$\widehat{\varepsilon}_i = \frac{Y_t - \widehat{m}(Z_i^\top\widehat{\gamma};\widehat{\gamma})}{\widehat{\sigma}(X_i^\top\widehat{\beta};\widehat{\beta})}, \qquad 1 \leq i \leq n.$$

Based on the residuals $\widehat{\varepsilon}_i$, we study four methods to estimate the distribution function F_ε.

The first method is using the empirical distribution function of the residuals :

$$\widehat{F}_{\varepsilon,\mathrm{emp}}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\widehat{\varepsilon}_i \leq t\}}, \qquad t \in \mathbb{R}. \tag{13}$$

See Koul et al., 2017 for a similar estimator in the case of a single-index model for the location function, and without scale function. See also Neumeyer and Van Keilegom, 2010 for a general setup of location scale regression.

This empirical distribution may be a simple and quite effective estimator of $F_\varepsilon(\cdot)$ if the interest lies in quantiles $q_\tau(\cdot)$ with τ away from 0 and 1. It is however expected to behave poorly for extreme quantiles. We therefore consider smoothed versions of the empirical distribution. More precisely, for some kernel function $k(\cdot)$, which may not identical to that in (9), (11) or (12), we consider

$$\widehat{F}_{\epsilon,1}(u) = \frac{1}{n}\sum_{i=1}^{n} K\left((\widehat{\varepsilon}_i - u)/h_\varepsilon\right) \qquad \text{with} \qquad K(u) = \int_{-\infty}^{u} k(v)dv, \tag{14}$$

and h_ε some bandwidth devoted to the estimation of F_ε. In the case of the Epanechnikov kernel, that is with $k(u) = (3/4)(1-u^2)\mathbb{1}_{\{|u|\leq 1\}}$, we get

$$K(u) = (1/4)\left(3u - u^3 + 2\right)\mathbb{1}_{\{|u|\leq 1\}} + \mathbb{1}_{\{u \geq 1\}}.$$

The smooth empirical distribution is the distribution function obtained by integrating the Parzen-Rosenblatt density estimator. See Azzalini, 1981 for the properties of the smooth empirical distribution estimator in the case where the sample of the variable (here ε) is observed. See also Racine and Van Keilegom, 2020 for the case where $\widehat{\varepsilon}_i$ are obtained after fitting a nonparametric location scale model.

Finally, keeping in mind the application to stress-testing where the interest focuses on the accurate estimation of the left tail of the distribution, we also investigate a version of the smoothed empirical distribution with asymmetric kernel. More precisely, we replace $k(u)$

by $k_l(u) = 2k(u)\mathbb{1}_{\{u\leq 0\}}$, and thus $K(u)$ by $K_l(u) = 2\int_{-\infty}^{u} k(v)dv$, and construct a third estimator under the form

$$\widehat{F}_{\epsilon,2}(u) = \frac{1}{n}\sum_{i=1}^{n}\left\{2K\left((\widehat{\varepsilon}_i - u)/h_\varepsilon\right)\mathbb{1}_{\{\widehat{\varepsilon}_i\leq u\}} + \mathbb{1}_{\{\widehat{\varepsilon}_i\geq u\}}\right\}. \tag{15}$$

In our empirical study, we consider this idea with $k(\cdot)$ the Epanechnikov kernel.

Finally, in order to diminish the bias of the smoothed empirical distribution estimator, we consider it with a higher order kernel. For simplicity, we only investigate the case of the asymmetric kernel. More precisely, we define

$$\widehat{F}_{\epsilon,3}(u) = \frac{1}{n}\sum_{i=1}^{n}\left\{2K_6\left((\widehat{\varepsilon}_i - u)/h_\varepsilon\right)\mathbb{1}_{\{\widehat{\varepsilon}_i\leq u\}} + \mathbb{1}_{\{\widehat{\varepsilon}_i\geq u\}}\right\}, \tag{16}$$

with

$$K_6(u) = C_1\left[u - 7u^3/3 + 63u^5/25 - 33u^7/25 - C_2\right]\mathbb{1}_{\{|u|\leq 1\}} + \mathbb{1}_{\{u>1\}},$$

where $C_1 = 525/256$ and $C_2 = 128/525$.

2.3 Semiparametric quantile function estimation

Given the semiparametric estimators of the single-index location and scale functions, as well as an estimator $\widehat{F}_\varepsilon(\cdot)$ of the distribution function $F_\varepsilon(\cdot)$ of ε, we can build our semi-parametric estimator of the conditional quantile function. More precisely, we define

$$\widehat{q}_\tau(w) = \widehat{q}_\tau(z,x) = \inf\left\{y \ : \widehat{F}_\varepsilon\left(\frac{y - \widehat{m}(z^\top\widehat{\gamma})}{\widehat{\sigma}(x^\top\widehat{\beta})}\right) \geq \tau\right\}. \tag{17}$$

where w, z and x belong to the support of W, Z and X, respectively. Our estimator requires several tuning parameters. For the estimators $\widehat{m}(Z_i^\top\widehat{\gamma};\widehat{\gamma})$ and $\widehat{\sigma}^2(X_i^\top\widehat{\beta};\widehat{\beta})$ we impose bandwidths in the range required for deriving the $\sqrt{n}-$asymptotic normality of $\widehat{F}_\varepsilon(\cdot)$. When considering the smoothed versions of the empirical distribution of the residuals $\widehat{\varepsilon}_i$, we propose a choice of the bandwidth which use the unconditional distribution of Y; see section 3.1 below.

3 Empirical study

In the following, we illustrate our semi-parametric approach for the estimation of the conditional quantiles by means of simulations and a real data application.

3.1 Simulation study

In this section we study the estimator proposed in (17) using the location scale model

$$Y = \left(Z^\top\gamma\right)^3 + \exp(X^\top\beta)\varepsilon, \tag{18}$$

where $X, Z \in \mathbb{R}^3$ have three components, and they don't have common components. The variable Y has a standard Gaussian distribution. The parameters are fixed as $\gamma = (1, -1.2, 0.4)$ and $\beta = (1, 0.8, -0.2)$. The covariate vectors are decomposed like $X = (X_1, \widetilde{X})$ and $Z = (Z_1, \widetilde{Z})$, and $\widetilde{X}$ and $\widetilde{Z}$ are simulated as standard Gaussian bivariate random vectors. The variables ε, X_1 and Z_1 are generated such that $Y \sim \mathcal{N}(0, 1)$. This is achieved using the Box-Müller method, see Box and Muller, 1958. We generate iid samples from (18), for sample sizes from $n \in \{50, 100, 150, 250\}$.

The values $\tau \in \{0.01, 0.05, 0.2\}$ are considered. In both local linear problem (9) and (11), we assumed a bandwidth $h = n^{-1/3.5}$ which matches the assumptions needed for the theory. We consider four different methods to estimate the distribution function $F_\varepsilon(\cdot)$. The first one is $\widehat{F}_{\varepsilon,\mathrm{emp}}(\cdot)$ from (13), the second is the smoothed version (14), denoted by $\widehat{F}_{\varepsilon,1}(\cdot)$, and the last ones are asymmetric smoothed estimators as in (15) and (16), denoted by $\widehat{F}_{\varepsilon,2}(\cdot)$ and $\widehat{F}_{\varepsilon,3}(\cdot)$, respectively. For each smooth estimator $\widehat{F}_{\epsilon,j}(\cdot)$ of $F_\varepsilon(\cdot)$, the bandwidth is set as the solution to the following minimization problem :

$$\min_h \sum_{i=1}^{n} \left(\widehat{F}_{\epsilon,j} \left(\frac{q_\tau - \widehat{m}(\widehat{\gamma}^\top Z_i)\}}{\widehat{\sigma}(\widehat{\beta}^\top X_i)} \right) - \tau \right)^2, \qquad j \in \{1, 2, 3\},$$

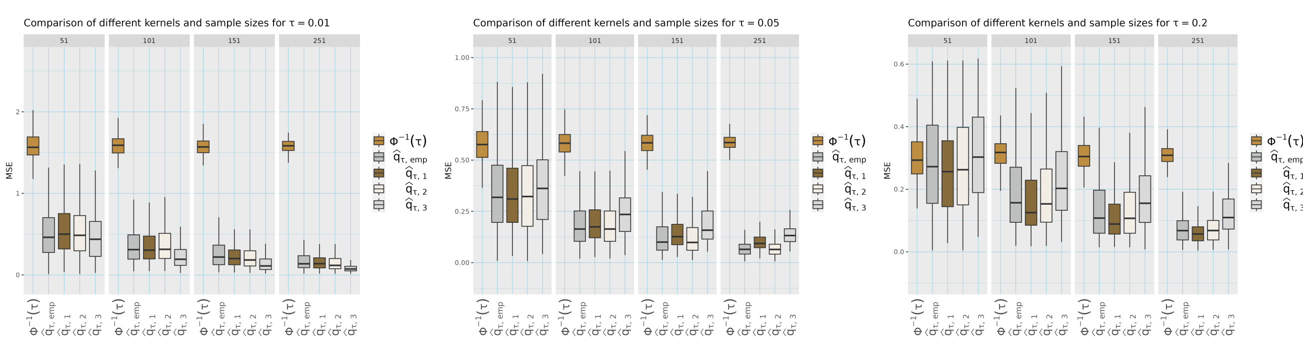
where q_τ is the marginal quantile of Y, here $q_\tau = \Phi^{-1}(\tau)$.

Let us write $\widehat{q}_\tau(Z_i, X_i)$ instead of $\widehat{q}_\tau(W_i)$ when the estimator of the $\tau-$th order conditional quantile function is computed at the observed values of the covariate vector. To measure the accuracy of our estimates of the conditional quantile function, we use the mean square error

$$\frac{1}{n} \sum_{i=1}^{n} \{\widehat{q}_\tau(Z_i, X_i) - q_\tau(Z_i, X_i)\}^2, \tag{19}$$

with $\widehat{q}_\tau(Z_i, X_i)$ computed according to (17). We also compute the mean squared error with $\Phi^{-1}(\tau)$ replacing $\widehat{q}_\tau(Z_i, X_i)$. For each simulation setup (kernel choices, τ, sample sizes) we report the performance based on $R = 250$ replications.

Figures 1a, 1b, 1c illustrates the mean squared error (19).



(a) (b) (c)

Figure 1: Mean squared error (19) for all proposed kernels, panel (a), (b) and (c) correspond to $\tau = 0.01, 0.05, 0.2$, respectively.

Across all quantiles below a certain threshold , the estimates obtained with $\widehat{F}_{\varepsilon,2}(\cdot)$ and $\widehat{F}_{\varepsilon,3}(\cdot)$ (using asymmetrical kernels) exhibit slightly better performance compared to those obtained from $\widehat{F}_{\varepsilon,\text{emp}}(\cdot)$ and $\widehat{F}_{\varepsilon,1}(\cdot)$ (see Figures 1a and 1b). This is especially true for smaller samples. However, at the highest quantile level considered ($\tau = 0.2$), using the smooth estimator $\widehat{F}_{\varepsilon,1}(\cdot)$ (constructed with a symmetric kernel) for estimating $F_\varepsilon(\cdot)$ yields superior performance (see Figure 1c). It is worth noting that all the alternative methods we propose drastically outperforms the regulatory framework.

3.2 Use case : Banking climate stress-testing

Let us apply the location-scale model with real data. The goal is to use in (4) the conditional quantiles of the common factor given covariates given by macroeconomic variables. In particular, the goal is to use macroeconomic scenarios that integrates climate risks to obtain forward looking probabilities of default that include this risk. Therefore, we will first estimate β, γ, m and σ using past data and following the method proposed in section 2. Then, conditionally given a scenario from the NGFS[2], we compute $\widehat{q}_\tau(z, x)$ for covariate values z and x corresponding to each year between 2023 and 2050 available in the macroeconomic scenario published by the NGFS.

Our historic data consists of corporate default rates from the S&P report and economic variables from U.S. Bureau of Economic Analysis, 2023; World Bank, 2023 from 1981 to 2021. Our vector of predictors W include GDP growth and inflation rate, in this case $X = Z = W$ and $x = z = w$ for the NGFS's scenario values. We assume that ξ, the correlation between the normalized assets' value and the common factor, is given by the regulator's formula, and p is also given. The Figure 2 represents the estimated conditional distribution of Y given $W = w$, while the Figure 3 represents $\pi(\tau \mid w)$ computed as in (5).

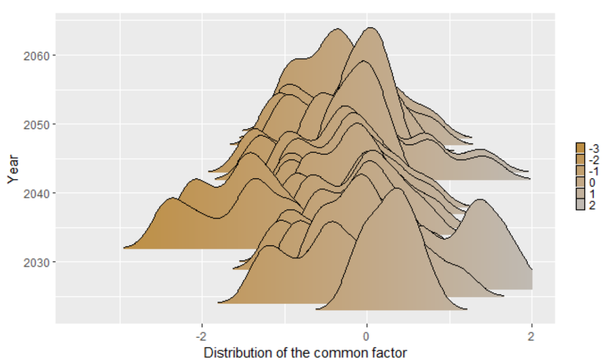


Figure 2: Distribution of the common factor conditionally to the Divergent Net Zero scenarios from the NGFS used during stress-testing exercises.

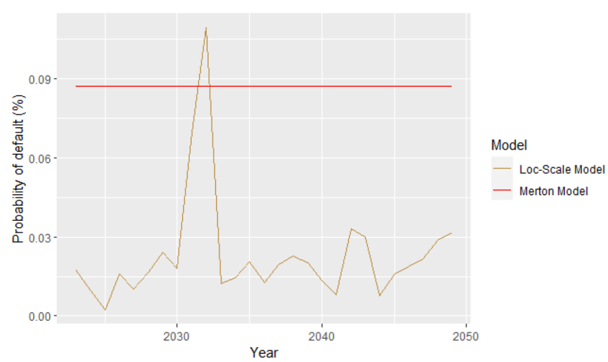


Figure 3: Conditional probability of default (under model (5)) at quantile 1% conditionally to the 'Below 2°C' scenario from the NGFS used during stress-testing exercises.

[2] See https://www.ngfs.net/ngfs-scenarios-portal/. The Divergent Net Zero scenario is used.

References

Akritas, M. G., & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, *28*(3), 549–567.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, *68*(1), 326–328.

Box, G. E., & Muller, M. E. (1958). A note on the generation of random normal deviates. *The annals of Mathematical Statistics*, *29*(2), 610–611.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Delecroix, M., Hristache, M., & Patilea, V. (2006). On semiparametric m-estimation in single-index regression. *Journal of Statistical Planning and Inference*, *136*(3), 730–769.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications* (Vol. 66). Chapman & Hall, London.

Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*(3), 645–660.

Hardle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *The annals of Statistics*, *21*(1), 157–178.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of econometrics*, *58*(1-2), 71–120.

Koul, H. L., Müller, U. U., & Schick, A. (2017). Estimating the error distribution in a single-index model. *From Statistics to Mathematical Finance: Festschrift in Honour of Winfried Stute*, 209–233.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Neumeyer, N., & Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, *101*(5), 1067–1078.

Racine, J. S., & Van Keilegom, I. (2020). A smooth nonparametric, multivariate, mixed-data location-scale test. *Journal of Business & Economic Statistics*, *38*(4), 784–795.

Roncalli, T. (2020). *Handbook of financial risk management*. CRC Press.

U.S. Bureau of Economic Analysis. (2023). Gross domestic product [GDPA] [February 5, 2024.].

Vašíček, O. (2002). The distribution of loan portfolio value. *Risk*, *15*(12), 160–162.

World Bank. (2023). Inflation, consumer prices for the united states [fpcpitotlzgusa] [Accessed February 5, 2024]. https://fred.stlouisfed.org/series/FPCPITOTLZGUSA

Yin, J., Geng, Z., Li, R., & Wang, H. (2010). Nonparametric covariance model. *Statist. Sinica*, *20*(1), 469–479.

Zhu, L., Dong, Y., & Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling. *Statist. Sinica*, *23*(3), 1235–1255.