Revue Internationale des Sciences de Gestion
ISSN: 2665-7473
Volume 6 : Numéro 1

PREDICTING THE ACT OF ENGAGEMENT OF INFORMAL PRODUCTION UNITS DURING THE COVID-19 EPIDEMIC

PRÉDICTION DE L'ACTE D'ENGAGEMENT DES UNITÉS DE PRODUCTION INFORMELLES DURANT L'ÉPIDÉMIE DE COVID-19

ROUAINE Zakaria
Research Professor
Faculty of Economics and Management
Ibn Tofail University
Laboratory of Economics, Management, and Development of Organizations
Morocco
rouainezakaria@gmail.com

TAITAI Zineb
Ph.D. student
Faculty of Economics and Management
Ibn Tofail University
Laboratory of management, finance, and accounting Morocco
taitaizineb@gmail.com

Date submitted : 20/01/2023

Date of acceptance : 20/02/2023

To cite this article :

ROUAINE Z. & TAITAI Z. (2023) «PREDICTING THE ACT OF ENGAGEMENT OF INFORMAL PRODUCTION UNITS DURING THE COVID-19 EPIDEMIC», Revue Internationale des Sciences de Gestion « Volume 6 : Numéro 1 » pp : 1142 - 1171

Abstract
The new tax reforms have been an important support triggering the process of engagement of informal units with their tax administration. The latter has put in place various corrective measures to contain and mitigate the unfavorable consequences on the companies that make up the national economic fabric. In other words, the multilateral collaboration between the formal production units and the tax administration has allowed a successful economic recovery, increasing the resilience of enterprises during the coronavirus epidemic. However, the introduction of these new tax reforms, characterized by a downward trend in taxes, has triggered the willingness of production units operating in the informal sector to submit to the tax authorities, seizing the opportunity of the new tax bases to encourage everyone to pay taxes and improve the relationship between taxpayers and the administration. Nevertheless, this article will focus on the prediction of the act of engagement of informal units with the tax administration after the new tax reforms during the Covid-19 epidemic. This study will mobilize one of the extensions of generalized linear models, such as the binary logistic regression model.
Keywords: « Tax reforms », « Engagement act », « Generalized linear models », « Binary logistic regression », « the covid-19 epidemic ».

Résumé
Les nouvelles réformes fiscales ont été un appui important pour déclencher le processus d'engagement des unités informelles avec leur administration fiscale. Cette dernière a mis en place diverses mesures correctives pour contenir et atténuer les conséquences défavorables sur les entreprises composant le tissu économique national. En d'autres termes, la collaboration multilatérale entre les unités de production formelles et l'administration fiscale a permis une relance économique réussie, augmentant la résilience des entreprises pendant l'épidémie de coronavirus. Cependant, l'introduction de ces nouvelles réformes fiscales, caractérisées par une tendance à la baisse des impôts, a déclenché la volonté des unités de production opérant dans le secteur informel de se soumettre à l'administration fiscale, saisissant l'opportunité des nouvelles bases d'imposition. Néanmoins, cet article se concentrera sur la prédiction de l'acte d'engagement des unités informelles avec l'administration fiscale après les nouvelles réformes fiscales pendant l'épidémie de Covid-19. Cette étude mobilisera une des extensions des modèles linéaires généralisés, comme le modèle de régression logistique binaire.
Mots-clés : « Réformes fiscales », « acte d'engagement », « modèles linéaires généralisés », « régression logistique binaire », « l'épidémie de covid-19 ».

1. INTRODUCTION

Unprecedented in recent history, the coronavirus has caused a health crisis and a collapse of economic activity on a global scale. In this unfortunate context, the public authorities have introduced several sanitary measures with the ultimate motive of mitigating the spread of Covid-19, reducing the number of newly infected people, limiting the pressure on health systems, and preventing a new epidemic outbreak while restarting economic activity and increasing the immunity of production units to the risks of the resurgence of the epidemic. The measures implemented had negative economic repercussions that sharply impacted the supply and demand side of the economies (Guerrieri et al., 2020). In other words, these remedial measures have hampered domestic consumption, investment, and net foreign demand, causing a monthly decline in the gross domestic product (GDP) of 2% of annual GDP according to the organization for economic cooperation and development estimates (2020).

There are many challenges for the national economy. The public authorities have tried to limit the negative consequences of containment measures affecting households and businesses and adapting to the risks presented by the epidemic. Also, to support the economic recovery by putting in place expansive plans to revive the economic activity on the national field. To build confidence and increase the resilience of production units, the tax authorities have approved several decisive measures to contain and mitigate the coronavirus epidemic's fallout from the health and economic crisis. In this context, a multilateral collaboration between public authorities and local production units would be crucial to ensure recovery and increase the resilience of the national economy to future shocks.

To adapt tax rules to the national context, policymakers have attempted to implement various cyclical corrective measures such as accelerating the allowable depreciation, income tax cuts, major corporate tax reforms, capital gains tax reductions, accelerated value-added tax refunds, etc. Nevertheless, these reforms accompanying the covid-19 health crisis have resulted in a contraction of tax revenues, thus widening the ordinary budget deficit. On the other hand, several companies operating in the informal sector have seized this opportunity to register with their tax authorities, taking direct advantage of the tax reforms characterized by a downward trend in taxes and duties with the ultimate aim of boosting economic activity and easing disbursements for national production units.

The derogatory measures undertaken by the Moroccan authorities, more precisely, by the Council of Economic Watch (CVE.), have for mission to adapt the tax system of rigor to the hazards of this heavy carried which tossed to the world economy, notably, to that Moroccan. Where the informal sector is of weight. The aforementioned measures concern the companies in a general way that the professionals. The first key measure observed is the postponement of the legal deadlines for corporate income tax and income tax (IR). Not to mention some incidental incentives such as the spreading of the deductibility of support donations to the COVID-19 fund over 5 years in order to reduce the impact on the tax result. On the other hand, eligible taxpayers must be "in difficulty", as this notion is defined by regulation and applicable exclusively to the period of the state of a health emergency. The extension of the state of emergency throughout the country until February 28, 2023, is proof of this.

In terms of IR, any additional allowance paid to employees (affiliated to the CNSS) by their employers is exempt, up to a limit of 50% of the average net monthly salary, excluding bonuses and annual premiums. This measure was extended by the 2022 Finance Act. In 2023, the legislator has conditioned the obtaining of this benefit by the following conditions:

- the employee must be hired during the year 2021;
- the employee must have benefited from the fund for loss of employment in accordance with the provisions of law n° 03-14 modifying and completing the Dahir bearing law n° 1-72-184 of 15 Joumada II 1392 (July 27, 1972) relating to the social security system;
- the employee cannot benefit twice from the above-mentioned exemption;

To this end, the CVE press release specifies that this preferential measure was aimed particularly at employees who received the 2,000 Dirham compensation for work stoppage following the consequences of the health crisis. In addition, the exceptional measures were spread over the interest rate of the credit interests applicable to the current account of credit partners which depends on the vast movement of decrease of the rates noted in particular of the short-term credits granted by the credit institutions, especially for the years post-COVID-19.

In this furrow, our prospection will analyze the impact of the tax reforms at the level of the corporate tax, the income tax, and the tax on the added value of the act of engagement of the informal units during the epidemic coronavirus covid-19. This study will focus on the use of generalized linear models, specifically the binary logistic regression model. Generalized linear models are defined as a development of the general linear model, allowing model responses

that are not normally distributed. These models have been developed in response to the shortcomings of classical linear models. In other words, these models are limited to describing the relationship between a variable to be explained and explanatory variables, to test the significance and compare the intensity of the impact of each independent variable on the variability of the dependent variable. Throughout this paper, we make explicit an attempt to model the act of engagement of informal units with the tax administration triggered by the tax reforms introduced during the covid-19 health pandemic in Morocco. In other words, **what would be the decision of informal units after the new tax bases during the crisis towards their tax authorities?** To answer this question, we will use the binary logistic regression method, which is one of the most important extensions of generalized linear models.

For this article, we are content to undertake a quantitative study to assess the impact of each tax reform on the engagement decision of informal economic groups. For this, we start with an introduction explaining the interest, the context and the tax reforms implemented, the methodology used, the results obtained, and finally, a general conclusion synthesizing the entire study.

2. METHODOLOGY

Being a set of statistical models used to analyze the relationship of a variable to one or more others, the Generalized Linear Models (GLM), usually known by their English initials, operate as adequate tools to estimate the parameters of the model used in the most impartial way possible. These models are understood as a development of the general linear model, where the dependent variable or variable to be explained is linearly related to the independent variables via a precise link function. They cover statistical models such as linear regression for normally distributed responses, logistic models for binary or dichotomous data, log-linear models for headcount data, complementary log-log models for interval-censored survival data, etc. However, they have been used to address the shortcomings of linear models. In other words, the latter is limited to describing the relationship between a variable to be explained and explanatory variables, to test the significance and compare the intensity of the impact of each independent variable on the variability of the dependent variable.

The Generalized Linear Model (GLM), is a more flexible device compared to the linear model, agreeing to cross the four assumptions mentioned above, in a process of treatment of the observations, to realize a relevant estimation of the parameters of the model and to test the

hypotheses conceived in a motive of exploring the quality of the latter. These models were introduced and defined by (Nelder John Ashworth, and Robert Wedderburn 1972) stating that they "allow us to model responses that are not normally distributed, using methods closely analogous to linear methods for normal data. However, (Anderson Duncan, Sholom Feldblum, Claudine Modlin, Doris Schirmacher, Ernesto Schirmacher, and Neeza Thandi 2004), present in detail, the concepts of the density function, the exponential distribution function, the form of the moment-generating function, and the specific types of the family of exponential distribution functions such as Gamma, Poisson, Bernoulli, Dirichlet, Exponential, Normal, Chi-square, Beta, and so on. We explain throughout this article, the generalized linear models, and a brief application of one of its extensions namely, the binary logistic regression.

A generalized linear model is an extension of the classical general linear model, so linear models are a suitable starting point for the introduction of generalized linear models. The linear regression model is characterized by four essential elements such as the column vector of dimension (*n)* of the dependent random variables (*Y*), a systematic component defined as a matrix of size ($n \times p$), and rank (*p*), called the design matrix $X = X_1, X_2, \ldots, X_p$, grouping together the column vectors of the explanatory variables, also known as the control variables endogenous, or independent, where (x_i) is the row vector of these explanatory variables associated with the observation (*i*) such that, $i = 1, 2, \ldots n$, (β) the column vector of dimension *p* of the unknown parameters of the model, i.e. the unknown regression coefficients associated with the column vector of the matrix (*X*), and finally, the vector dimension *n* of the errors (ε). The data are assumed to be drawn from observations of a statistical sample of size $n \in \mathcal{H}^{(p+1)}$(where $n > p+1$). However, linear models seem to be based on a set of assumptions such as (i) ε_i are error terms, of a variable *E*, unobserved, independent, and identically distributed, noting that $E(\varepsilon_i)=0$, (ii) the $V(\varepsilon_i)=\sigma^2$. I, about the character of homoscedasticity, referring to a constant stochastic error variance of the regression, i.e., identical dispersion for each *i*, (iii) the normality of the distribution of the error random variable ε noting: $\varepsilon_i \sim N_n\ (0, \sigma^2 I_n)$ We can also consider that ε_i is an observation of the random variable *E*, also distributed according to a normal distribution, noting that $\varepsilon_i \sim N\ (0, \sigma^2$, (iv) the *n* real random variables ε_i are considered independent, i.e. ε_i is independent of ε_j for $i \neq j$, (v) y_i is an observation of *Y* of normal distribution, such that, $Y \sim N_n\ (\beta X, \sigma^2 I_n)$. The linear regression model is defined by an equation of the form:

$$Y = \beta X + \varepsilon \quad \text{with} \quad \varepsilon \sim N_n\ (0, \sigma^2 I_n)$$

Where:

- $\mathbf{Y} \in \mathcal{R}^{(n)}$
- $\mathbf{X} \in M_{(n,p)}$ known, deterministic, with rank p
- $\boldsymbol{\beta} \in \mathcal{R}^{(p)}$, unknown
- $\sigma^2 \in \mathcal{R}^{(*+)}$, unknown

In statistics, generalized linear models is an extraordinarily flexible generalization of ordinary linear regression, which takes into account dependent variables, called responses, that have distribution patterns other than the normal distribution. GLM generalizes linear regression by allowing the linear model to be related to these response variables by a (g) link function. This mechanism was founded by (John Nelder and Robert Wedderburn 1972), who were able to formulate generalized linear models to unify various other statistical models, including linear regression, logistic regression, Poisson regression, etc. However, the model's linear predictor or deterministic component is a quantity with the skill and ability to incorporate information about the independent variables into the model. It is linked to the expected value of the data thanks to the linking function (g). This linear predictor noted η is expressed in the form of linear combinations of the unknown parameters β and the matrix of column vectors of the explanatory variables X (see the works of Denuit M. and Charpentier A. (2005), J-J. Droesbeke, Lejeune M., and Saporta G. (2005)). η can thus be expressed as:

$$\eta = \beta X$$

The normality of the response variable Y, such that, $Y \sim N_n\ (\beta X, \sigma^2 I_n)$, for any observation i, allows us to write, $E(Y) = \beta X$, and to note $E(Y) = \mu$ for simplification reasons. Thanks to the link function (g), it is possible to establish a non-linear relationship between the expectation of the response variable $E(Y)$ and the explanatory variable(s) and to apprehend observations and responses of diversified natures, such as the example of binary data of failures/successes, frequencies of successes of the treatments, lifetimes, etc., by noting that:

$$g\ (E(Y)) = g(\mu) = \eta = \beta X$$

As mentioned in the work of (Esbjörn Ohlsson, and Björn Johansson 2010), we can also write that:

$$E(Y) = \mu = g^{-1}(\eta)$$

The linkage function (g) states the relationship between the linear predictor η and the mean of the distribution function μ. There are many commonly used link functions, and their choice is based on several considerations. There is always a well-defined canonical link function that is derived from the exponential response density function (Y). However, in some cases, it makes sense to try to match the domain of the link function to the range of the mean of the distribution function. A linkage function transforms the probabilities of a category response variable into a continuous unbounded scale. Once the transformation is complete, the relationship between the η predictors and the response can be modeled using linear regression. For example, a dichotomous response variable may have two unique values. Converting these values to probabilities causes the response variable to vary between 0 and 1. When an appropriate linkage function is chosen to be applied to the probabilities, the resulting numbers are between $-\infty$ and $+\infty$. However, any probability law of the random component Y has associated with it a specific function of the expectation called the canonical parameter. For the normal distribution, it is the expectation itself. For the Poisson distribution, the canonical parameter is the logarithm of the expectation. For the binomial distribution, the canonical parameter is the logit of the probability of success. In the family of generalized linear models, the functions using these canonical parameters are called canonical link functions. In most cases, generalized linear models are built using these link functions. Below is a table of several commonly used exponential family distributions, the data for which they are commonly used, and the canonical link functions and their means.

Table N°1: Laws of the exponential family and their canonical links

Y distribution	Canonical links	Means
Normal distribution $N(\mu, \sigma^2)$	Identity: $\eta = \mu$	$\mu = \beta X$
Bernoulli distribution $B(\mu)$	Logit: $\eta = \ln(\mu/(1-\mu))$	$\mu = 1/(1+\exp^{-(\beta X)})$
Poisson distribution $P(\mu)$	Log: $\eta = \ln(\mu)$	$\mu = \exp^{(\beta X)}$
Gamma distribution $G(\mu, v)$	Inverse: $\eta = 1/(\mu)$	$\mu = (\beta X)^{-1}$
Gaussian Inverse distribution $I.G\ (\mu, \lambda)$	Inverse carré : $\eta = 1/(\mu^2)$	$\mu = (\beta X)^{-2}$

Source: Author

2.1.Probability law of the response variable *Y*

The inadequacy of the so-called classical general linear model, of the laws that it associates with the response variables, leads us to use generalized linear models (GLM), which allow us to connect other laws than the normal law, such as Bernoulli's law, the binomial law, Poisson's law, Gamma law, etc. These laws are part of the exponential family, offering a common framework for estimation and modeling. These laws are part of the exponential family, offering a common framework for estimation and modeling. This natural exponential family has laws that are written in exponential form, which allows us to unify the presentation of results. Let f_Y be the probability density of the response variable *Y*. We can admit that f_Y belongs to the natural exponential family if it is written in the form:

$$f(Y/\theta,\varphi,\omega) = \exp\left(\frac{Y\theta - b(\theta)}{a(\phi)}\omega + c(Y,\varphi,\omega)\right),\ Y \in S$$

With:

- a(.), b(.), c(.): Functions specified according to the type of the exponential family considered.
- θ: Natural parameter, also called canonical parameter or mean parameter.
- φ: Parameter of dispersion. This parameter may not exist for some laws of the exponential family, in particular when the law of Y depends only on one parameter (in these cases $\varphi = 1$). Otherwise, it is a nuisance parameter that must be estimated. As its name indicates, this parameter is related to the variance of the law. It is also a very important parameter in that it

controls the variance and therefore the risk. In some cases, a weighting is necessary to grant relative importance to the different observations and the parameter φ is replaced by φ/ω, where ω designates a weight known as a priori.

- S: Subset of R or N
- ω: The weights of the observations.

Moreover, if f_Y , belongs to the natural exponential family, we can deduce the following properties:

- $E[Y] = \mu = b'(\theta) = \partial b(\theta)/\partial(\theta)$
- $V[Y] = a(\varphi) \times b''(\theta) = a(\varphi) \times \partial^2 b(\theta)/\partial(\theta)^2$
- $g(\mu) = g(b'(\theta)) = \beta X$

With: $b'(\theta) = g^{-1}(\beta X)$ and $\theta = \eta = \beta X$

For a probability law to belong to the natural exponential family, it is sufficient to write it as an exponential function and determine its terms. We try below to propose some examples of commonly used probability laws, and explain all their components (See the works of Michel Denuit, and Arthur Charpentier (2005).

- **The Gaussian distribution**, with mean μ and variance σ^2. $Y \sim N(\mu, \sigma^2)$ belongs to the exponential family, with $\theta = \mu$, $\varphi = \sigma^2$, $a(\varphi) = \varphi$, $b(\theta) = \theta^2/2$, and $c(Y,\varphi,\omega) = -1/2\ (y^2/\sigma^2 + \ln(2\pi\sigma^2))$, where $Y \in R$.
- **The Bernoulli distribution**, with mean π, and variance π(1-π). Y ~ B(π) is catalogued among the exponential family, with $\theta = \ln\{\pi/(1-\pi)\}$, $\varphi = 1$, $a(\varphi) = 1$, $b(\theta) = \ln(1+\exp(\theta))$, and $c(Y,\varphi,\omega) = 0$ where $Y \in N$.
- **The Poisson distribution**, with mean λ, and variance λ. Y ~ P(λ), is part of the exponential family, with $\theta = \ln(\lambda)$, $\varphi = 1$, $a(\varphi) = 1$, $b(\theta) = \exp(\theta) = \lambda$, et $c(Y,\varphi,\omega) = -\ln(\lambda!)$ with $Y \in N$.
- **The Gamma distribution**, with mean μ and variance v^{-1}. Y ~ G(μ, v), also joins the exponential family, with $\theta = -1/\mu$, $\varphi = v^{-1}$, $a(\varphi) = \varphi$, $b(\theta) = -\ln(-\theta)$, and $c(Y,\varphi,\omega) = ((1/\varphi)-1)\ln(Y) - \ln(\Gamma(1/\varphi))$ where $Y \in R^+$.

Table Nº 2: Components of the exponential family of usual probability laws

Y distribution	θ(μ)	φ	a(φ)	b(θ)	c (Y,φ,ω)
Normal distribution *N* (μ, σ^2)	μ	σ^2	φ	$\theta^2/2$	$-\frac{1}{2}(y^2/\sigma^2 + \ln(2\pi\sigma^2))$
Bernoulli distribution *B*(μ)	$\ln\{\mu/(1-\mu)\}$	1	1	$\ln(1+\exp(\theta))$	0
Poisson distribution *P*(μ)	ln(μ)	1	1	exp(θ)	− ln (Y !)
Gamma distribution *G*(μ, *v*)	−1/μ	1/v	φ	−ln(−θ)	$((1/\varphi) - 1)\ln(Y) - \ln(\Gamma(1/\varphi))$
Gaussian Inverse distribution *I.G* (μ, λ)	$-1/2\mu^2$	σ^2	φ	$-(-2\theta)^{1/2}$	$-1/2(\ln(2\pi\varphi Y^3) + 1/\varphi Y)$

Source: Author

Table Nº 3: Expectation and variance of usual probability laws

Y distribution	μ = E(Y) =b'(θ)	V(Y) = a(φ)b"(θ)
Normal distribution N(μ, σ2)	θ	σ^2
Bernoulli distribution B(μ)	exp(θ)/((1+exp(θ)	$\mu(1-\mu)$
Poisson distribution P(μ)	exp(θ)	μ
Gamma distribution G(μ, v)	−1/θ	μ^2/v
Gaussian Inverse distribution I.G (μ, λ)	μ	μ^3/λ

Source: Author

The two tables above summarize respectively, the different components of the exponential family for usual probability laws, as well as their expectation and variance, assuming that the weight ω = 1.

2.2. Parameters estimation

At this stage, it is a question of estimating the column vector $\beta = (\beta_0, \beta_1, \cdots, \beta_p)$ noted $(\widehat{\beta_0}, \widehat{\beta_1}, \cdots, \widehat{\beta_p})$ of dimension p of the unknown parameters of the model, i.e. the unknown regression coefficients associated with the column vectors of the matrix (X) representing a set of explanatory variables, by maximizing the natural log-likelihood of the generalized linear model. This estimation applies to all laws with a distribution belonging to the exponential family of the form:

$$f(Y/\theta, \varphi, \omega) = \exp\left(\frac{Y\theta - b(\theta)}{a(\phi)}\omega + c(Y, \varphi, \omega)\right), \quad Y \in S$$

The main idea of the maximum likelihood method is to look for the parameters' value that maximizes the probability of having observed what we observed. Moreover, the standard approach to finding the maximum of any function of several variables consists in canceling its gradient (first derivative) and checking that it's hessian (second derivative) is negative. However, to obtain the maximum likelihood estimator (L), we solve the following system of p unknowns β:

$$\frac{\partial \ln L(\beta)}{\partial \beta_1} = 0$$
$$\vdots$$
$$\frac{\partial \ln L(\beta)}{\partial \beta_p} = 0$$

Let n be independent variables Y_i, with $i = 1\cdots, n$ of law belonging to the exponential family, X the design matrix, where are arranged the observations of p column vectors representing the explanatory variables, β the column vector of p parameters of the model, η the linear predictor with n components noted $\eta = \beta X$, g the link function, is supposed to be monotonic and differentiable such that, $\eta = g(\mu)$, as well as the canonical link function, is expressed by $g(\mu) = \theta$. For n observations assumed to be independent, and taking into account the link between θ and β, the likelihood (L) and the natural logarithm of the likelihood (ℓ) are written as follows:

$$L(Y, \theta, \phi, \omega) = \prod_{i=1}^{n} f(Y_i, \theta_i, \phi, \omega)$$

$$\ell(Y, \theta, \phi, \omega) = \ln(L(Y, \theta, \phi, \omega)) = \ln\left(\prod_{i=1}^{n} f(Y_i, \theta_i, \phi, \omega)\right)$$

$$= \sum_{i=1}^{n} \ln(f(Y_i, \theta_i, \phi, \omega)) = \sum_{i=1}^{n} \ell_i(Y_i, \theta_i, \phi, \omega)$$

With:

$$\ell_i = \frac{Y_i \theta_i - b(\theta_i)}{a_i(\phi)}\omega + c(Y_i, \phi, \omega), \text{ and } \theta_i = \beta_j . x_i^T$$

Indeed, we try this method to reach the maximum likelihood. The logarithm function is strictly increasing, and the likelihood and the natural logarithm of the likelihood reach their maximum at the same point. Moreover, the search for the maximum likelihood generally requires the calculation of the first derivative of the likelihood, and this is much simpler than the natural log-

likelihood, in the case of multiple independent observations, since the logarithm of the product of the likelihoods is written as the sum of the logarithms of the likelihoods, and it is easier to derive a sum of terms than a product. However, the derivative of the natural log-likelihood can be realized by solving the following equality:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j}$$

From the above equality, we try to give the meaning of each term of the latter as follows:

- $\frac{\partial \ell_i}{\partial \theta_i} = \frac{Y_i - b\prime(\theta_i)}{a_i(\phi)} = \frac{Y_i - (\mu_i)}{a_i(\phi)}$
- $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{V(Y_i)}{a_i(\phi)}$
- $\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial (\beta X_i)}{\partial \beta_j} = X_{ij}$

And $\frac{\partial \mu_i}{\partial \eta_i}$ depends on the link function $\eta_i = g(\mu_i)$ with $\eta_i = \beta_j \cdot X_{ij}$

The partial differential equations are therefore written in the following form:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{Y_i - (\mu_i)}{a_i(\phi)} \times \frac{a_i(\phi)}{V(Y_i)} \times \frac{\partial \mu_i}{\partial \eta_i} \times X_{ij}$$

$$\frac{\partial \ell(Y,\theta(\beta),\phi)}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{Y_i - (\mu_i) \times X_{ij}}{V(Y_i)} \times \frac{\partial \mu_i}{\partial \eta_i} \right) = 0, \forall j = 1, \cdots p$$

In the case where the link function used coincides with the canonical link function ($\eta_i = \theta_i$), these equations are simplified as follows:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \times \frac{\partial \eta_i}{\partial \beta_j}$$

Thus, the partial differential equations can take the following form:

$$\frac{\partial \ell(Y,\theta(\beta),\phi)}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{Y_i - (\mu_i)}{a_i(\phi)} \times X_{ij} \right) = 0, \forall j = 1, \cdots p$$

However, μ_i is unknown, so it is impossible to obtain an analytical expression of the maximum likelihood estimator of β by canceling the first derivative (gradient): these equations are called transcendental. In other words, they are non-linear β equations whose solution requires iterative

optimization methods, such as the Newton-Raphson algorithm referring to the Hessian matrix and the Fisher-scoring algorithm referring to the information matrix, whose approach can be summarized as follows:

a. Choose a starting point β^0
b. Put down $\beta^{k+1} = \beta^k + A_k \times \nabla L(\beta^k)$
c. Shutdown condition : $\beta^{k+1} \approx \beta^k$

Or :

$\nabla L(\beta^{k+1}) \approx \nabla L(\beta^k)$
$A_k = -[\nabla^2 L(\beta^k)]^{-1}$ For Newton-Raphson algorithm
$A_k = -(E[\nabla^2 L(\beta^k)])^{-1}$ For the iterative Reweighted Least Squares

2.3. Properties of the maximum likelihood estimator and confidence interval

In general, it is insufficient for a statistician to stop in the estimation phase of the value of the regression parameters. However, given that the value of the regression estimator depends closely on the sample on which the modeling is done, it is more legitimate to look at the confidence interval in which it lies, by setting a confidence level beforehand. Thus, the smaller the interval, the more robust the estimate. Let us note $\widehat{\beta_n}$ the maximum likelihood estimator (MLE). This estimator verifies certain properties, under certain classical assumptions of the regularity of the probability density, such as:

- $\widehat{\beta_n}$: Converges in probability to β, which implies that $\widehat{\beta_n}$ is asymptotically unbiased.
- $\widehat{\beta_n}$: Converges to a normal distribution.

Indeed, it is possible to write:

$$\sqrt{n}(\widehat{\beta_n} - \beta) \sim \mathcal{N}(0, \mathbb{I}_n^{-1}(\beta))$$

- $\widehat{\beta_n}$: Estimator of the maximum log-likelihood of $\beta = (\beta_0, \beta_1, \cdots, \beta_p)$
- $\mathbb{I}_n^{-1}(\beta)$: $-(E[\partial \ell^2(Y, (\beta), \phi)/\, \partial^2 \beta])$ is the Fisher information matrix evaluated in β and ϕ on a sample of size *n*.

Let $\widehat{\beta_n}$ be the estimator of the parameter β such that $\widehat{\beta_n}$verifies a central limit theorem, i.e., when

n tends to infinity, the random variable of centered reduced Gaussian distribution *z* tends to the value below:

$$\frac{\widehat{\beta_n}-\beta}{\sqrt{V(\widehat{\beta_n})}} \sim z$$

As a way of determining the confidence interval at risk α for $\widehat{\beta_n}$: from the bounds $(z_{1-\alpha/2})$ and $(-z_{1-\alpha/2})$ such that:

$$P\left(-z_{1-\alpha/2} < \frac{\widehat{\beta_n}-\beta}{\sqrt{V(\widehat{\beta_n})}} < z_{1-\alpha/2}\right) = 1-\alpha$$

If n is large enough, we can suppose that $\frac{\widehat{\beta_n}-\beta}{\sqrt{V(\widehat{\beta_n})}}$ follows approximately a Gaussian distribution and *F* the distribution function of the centered reduced Gaussian distribution, so we can write that:

$$P\left(-z_{1-\alpha/2} < \frac{\widehat{\beta_n}-\beta}{\sqrt{V(\widehat{\beta_n})}} < z_{1-\alpha/2}\right)$$

$$= F(z_{1-\alpha/2}) - F(-z_{1-\alpha/2})$$

$$= 2\,F(z_{1-\alpha/2}) - 1$$

With:

$$F(-z_{1-\alpha/2}) = 1 - F(z_{1-\alpha/2})$$

We can then deduce that:

$$2\,F(z_{1-\alpha/2}) - 1 = 1-\alpha$$

$$z_{1-\alpha/2} = F^{-1}(1-\alpha/2)$$

So, the bounds of the confidence interval for $\widehat{\beta_n}$ are written as follows:

$$B^- = \widehat{\beta_n} - F^{-1}((1-\alpha/2) \times \sqrt{V(\widehat{\beta_n})}$$

$$B^+ = \widehat{\beta_n} + F^{-1}((1-\alpha/2) \times \sqrt{V(\widehat{\beta_n})}$$

However, an asymptotic confidence interval at the level of 100 ×(1−α) % of the regression coefficients β can be designed as follows:

$$I.C_{\beta_n} = [\ \widehat{\beta_n} \pm (z_{1-\alpha/2}) \times \sqrt{V(\widehat{\beta_n})}\ ]$$

With:

$z_{1-\alpha/2}$ is the quantile at (1 − α/2) of the standard normal distribution, N (0, 1), and $V(\widehat{\beta_n})$ is the diagonal term of the inverse of the Fisher information matrix.

2.4 Binary logistic regression, an extension of generalized linear models

The essays of (Hosmer D. W., and Lemeshow S. 2000) as well as the work of (King G., and Zeng L. 2001), underline that logistic regression is understood as a relevant statistical choice, for situations in which the occurrence of a binary outcome must be predicted. In addition, (Burns R. B., Burns R., Burns, R. P. 2008), and (Muijs D., 2010) have offered clarifications of the steps necessary to perform such an analysis using a variety of statistical packages, such as SPSS, R, etc. While the explanation of the phases of performing such analysis in different particular contexts has also been mentioned on many websites, as highlighted in the works of (Greenhouse J. B., Bromber, J. A., and Fromm D. A. 1995) as well as the writings of (Wuensch D. 2009).

2.4.1 Logit transformation

We consider a population P subdivided into two groups of individuals G_1, and G_2 identifiable by an assortment of quantitative or qualitative explanatory variables $X_1, X_2, \dots X_p$ and let Y be a dichotomous qualitative variable to be predicted (explained variable), worth (1) if the individual belongs to the group G_1, and (0) if he/she comes from the group G_2. In this context, we wish to explain the binary variable Y from the variables $X_1, X_2, \dots X_p$.

We have a sample of n independent observations of y_i, with i = 1, 2, ..., n. y_i denotes a dependent random variable presented as a column vector such that, $y_i = (y_1, y_2, \dots y_n)$ expressing the value of a qualitative variable known as a dichotomous outcome response, which means that the outcome variable y_i can take on two values 0 or 1, evoking respectively the absence or the presence of the studied characteristic. We also consider a set of p explanatory variables denoted by the design matrix (X) = $(X_1, X_2, \dots X_p)$ grouping the column vectors of the independent variables, of size ($n \times p$) and rank (p), where (x_i) is the row vector of these explanatory variables associated with the observation (i) such that, i = 1, 2, …, n, and the column vector (β) of

dimension p of the unknown parameters of the model, i.e. the unknown regression coefficients associated with the column vectors of the matrix (X). We consider in this paper that y_i (response variable) is a realization of a random variable y_i that can take the values 1 in the case that corresponds to the probability of tourism companies succeeding in overcoming the health crisis or 0 in the case of the probability of failing to overcome this crisis with probabilities of (π) and (1-π) respectively.

The distribution of the response variable y_i is called Bernoulli distribution with parameter (π). And we can write y_i ~ B(1, π). Let the conditional probability that the outcome is absent be expressed by P(y_i = 0|X) = 1 − π and present, denoted P(y_i = 1|X) = π, where X is the matrix of explanatory variables with *p* column vectors. The modeling of response variables that have only two possible outcomes, which are the "presence" and "absence" of the event under study, is usually done by logistic regression (Agresti, 1996), which belongs to the large class of generalized linear models introduced by (John Nelder and Robert Wedderburn 1972). The Logit of the logistic regression model is given by the equation:

$$\text{Logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \sum_{k=0}^{p} \beta_k x_{ik}, \text{ with } i = 1, \ldots, n \quad (1)$$

By the Logit transformation, we obtain from equation (1) the equation (2):

$$\left(\frac{\pi}{1-\pi}\right) = \exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right) \quad (2)$$

We evaluate equation (2) to obtain π et 1 − π as:

$$\pi = \exp\left(\sum_{k=}^{p} \beta_k x_{ik,}\right) - \pi \exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right) \quad (3)$$

$$\pi + \pi \exp\left(\sum_{k=0}^{p} \beta_k x_{ik,}\right) = \exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right) \quad (4)$$

$$\pi \left(1 + \exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)\right) = \exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right) \quad (5)$$

$$\pi = \left(\frac{\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}{1+\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}\right) \quad (6)$$

$$\pi = \left(\frac{1}{1+\exp\left(-\sum_{k=0}^{p}\beta_k x_{ik}\right)}\right) (7)$$

In the same way, we obtain $(1 - \pi)$:

$$1 - \pi = 1 - \left(\frac{1}{1+\exp\left(-\sum_{k=0}^{p}\beta_k x_{ik}\right)}\right)$$

$$1 - \pi = \left(\frac{1}{1+\exp\left(\sum_{k=0}^{p}\beta_k x_{ik}\right)}\right)$$

$$1 - \pi = \frac{\exp\left(-\sum_{k=0}^{p}\beta_k x_{ik}\right)}{1+\exp\left(-\sum_{k=0}^{p}\beta_k x_{ik}\right)} \quad (8)$$

2.4.2 Estimation of the β parameters of the nonlinear equations of the Bernoulli distribution using the maximum likelihood estimator (MLE).

If y_i takes strictly two values 0 or 1, the expression for π given in equation (7) provides the conditional probability that y_i is equal to 1 given X, and will be reported as P(y_i = 0|X). And the quantity 1-π gives the conditional probability that y_i is equal to 0 given X, and this will be reported as P(y_i = 0|X). Thus, for y_i =1, the contribution to the likelihood function is π, but when y_i =0, the contribution to this function is $1 - \pi$. This contribution to the likelihood function will be expressed as follows:

$$\pi^{y_i}\ (1-\pi)^{1-y_i}$$

At this stage, we will estimate the P+1 unknown parameters β, using the maximum likelihood estimator (MLE) as follows:

$$L\ (y_1, y_2, \ldots y_n,\ \pi) = \prod_{i=1}^{n} \pi^{y_i}\ (1-\pi)^{1-y_i}$$

Maximum likelihood is one of the most widely used estimation procedures for determining the values of the unknown β parameters that maximize the probability of obtaining an observed data set. In other words, the maximum likelihood function explains the probability of the observed data based on unknown regression parameters β. This method was developed by the British

statistician Ronald Aylmer Fisher between (1912 - 1922) as it was assigned in John Aldrich's book "*R. A. Fisher and the making of maximum likelihood 1912-1922* " published in (1997). This method aims to find estimates of the p explanatory variables to maximize the probability of observation of the response variable *Y*.

$$L\ (y_1, y_2, \dots y_n,\ \pi) = \prod_{i=1}^{n} \pi^{y_i}\ (1-\pi)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left(\frac{\pi}{1-\pi}\right)^{y_i}\ (1-\pi)$$

Substituting equation (2) for the first term and equation (8) for the second term, we obtain:

$$L\ (y_1, y_2, \dots y_n,\ \beta_1, \beta_2, \dots \beta_p,\ ) = \prod_{i=1}^{n} \left(\exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik},\right)\right)^{y_i} \left(1 - \frac{\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}{1+\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}\right)$$

So,

$$L\ (y_1, y_2, \dots y_n,\ \beta_1, \beta_2, \dots \beta_p) = \prod_{i=1}^{n} \left(\exp\left(y_i\ \sum_{k=0}^{p} \beta_k\, x_{ik},\right)\right)\ \left(1 + \exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik},\right)\right)^{-1}$$

For simplicity, we incorporate the neperian logarithm into the above equation. Since the logarithm is a monotonic function, any maximum in the likelihood function will also be a maximum in the log-likelihood function and vice versa. Thus, considering the natural logarithm of this equation, we obtain the log-likelihood function ℓ expressed as follows:

$$\ln\left(L\ (y_1, y_2, \dots y_n,\ \beta_1, \beta_2, \dots \beta_p)\right) =$$

$$\ln\left(\prod_{i=1}^{n} \left(\exp\left(y_i\ \sum_{k=0}^{p} \beta_k\, x_{ik}\right)\right)\ \left(+\exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik}\right)\right)^{-1}\ \right)$$

$$\ell\ (y_1, y_2, \dots y_n,\ \beta_1, \beta_2, \dots \beta_p) = \sum_{i=1}^{n} y_i\ \left(\sum_{k=0}^{p} \beta_k\, x_{ik}\right) - \ln\left(1 + \exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik}\right)\right)$$

Deriving the last natural logarithm equation of the likelihood function above, we should write:

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} y_i\ x_{ik} - \frac{1}{1+\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}\ \times \frac{\partial}{\partial \beta_k}\left(1 + \exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik}\right)\right) \quad (9)$$

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} y_i\ x_{ik} - \frac{1}{1+\exp\left(\sum_{k=0}^{p} \beta_k x_{ik}\right)}\ \times \exp\left(\sum_{k=0}^{p} \beta_k\, x_{ik}\right) \times \frac{\partial}{\partial \beta_k} \sum_{k=0}^{p} \beta_k\, x_{ik} \quad (10)$$

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} y_i \; x_{ik} - \frac{x_{ik}}{1+\exp(\sum_{k=0}^{p} \beta_k x_{ik})} \times \exp\left(\sum_{k=0}^{p} \beta_k \, x_{ik}\right) (11)$$

Knowing that:

$$\frac{\partial}{\partial \beta_k} \sum_{k=0}^{p} \beta_k \, x_{ik} = x_{ik}$$

So,

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \ell'_{\beta_k} = \sum_{i=1}^{n} y_i \; x_{ik} - \; \pi . x_{ik} \quad (12)$$

Therefore, the estimation of the parameters $\hat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1}, \; \cdots, \widehat{\beta_p})$that maximize the log-likelihood function l can be determined by canceling each of the *P +1* equations of ℓ' (gradient of ℓ) as mentioned in equation (12), and verify that its Hessian matrix (second derivative) is negative definite, i.e. that each element of the diagonal of this matrix is less than zero (Gene H. Golub and Charles F. Van Loan 1996). The Hessian matrix consists of the second derivative of equation (12). The general form of the second partial derivative matrix (Hessian matrix) can be written as follows:

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_k \, \partial \beta_{k'}} = \frac{\partial}{\partial \beta_{k'}} \; \sum_{i=1}^{n} y_i \; x_{ik} - \; \pi . x_{ik} \; (13)$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_k \, \partial \beta_{k'}} = \frac{\partial}{\partial \beta_{k\prime}} (- \; \pi . x_{ik}) \quad (14)$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_k \, \partial \beta_{k'}} = - \; x_{ik} \frac{\partial}{\partial \beta_{k\prime}} \left(\frac{\exp(\sum_{k=0}^{p} \beta_k x_{ik} \,)}{1+\exp \, (\sum_{k=0}^{p} \beta_k x_{ik})} \right)$$

$$\ell''_{\beta_{k\beta\prime_k}} = - \; x_{ik} \, \pi(1-\pi) \, x_{ik} \quad (15)$$

To solve the (*P +1*) nonlinear β equations (12), we use the Newton-Raphson iterative optimization method, referring to the Hessian matrix. Using this method, the estimation of the β parameters starts with the first step of choosing a starting point $\beta^{\,0}$ or $\beta^{\,old}$. The second step consists in mentioning the way the method works by posing: $\beta^{\,k+1} = \beta^{\,k} + A_k \times \nabla \, L(\beta^{\,k})$, and finally stop when the condition $\beta^{\,k+1} \approx \beta^{\,k}$ or $\nabla \; \mathrm{L}\beta^{\,k+1}) \approx \nabla \; \mathrm{L}(\beta^{\,k})$ is realized. The result of

this algorithm in matrix notation is:

$$\beta^{new} = \beta^{old} + [-\ell''(\beta^{old})]^{-1} \times \ell'(\beta^{old})$$

By putting $\hat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1}, \cdots, \widehat{\beta_p})^t$ we have:

$$V(\hat{\beta}) = \left(-\frac{\partial^2}{\partial \beta^2} \ln L(\beta, Y)\right)^{-1} \|_{\beta=\hat{\beta}} = (X^t W X)^{-1}$$

To simplify this equation above, we substitute the value of $\ell'(\beta)$, and $\ell''(\beta)$ with another matrix form in the following way:

$$\beta^{new} = \beta^{old} + (X^t W X)^{-1} \times X^t (Y - \mu) \quad (16)$$

$$\beta^{new} = (X^t W X)^{-1} \times X^t W (X\beta^{old} + W^{-1}(Y - \mu)) \; \beta^{new} = (X^t W X)^{-1} X^t W Z \quad (17)$$

Where $Z = (X\beta^{old} + W^{-1}(Y - \mu))$ is a vector, and W is the vector of weights of the values of the diagonal of the inputs $\widehat{\pi_i}(1 - \widehat{\pi_i})$. We can also write:

$$\beta^{new} = \beta^{old} + (X^t W X)^{-1} \times X^t (Y - \mu) \quad (18)$$

With:

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}$$

$$\hat{V} = \begin{pmatrix} \widehat{\pi_1}(1 - \widehat{\pi_1}) & 0 & \cdots & 0 \\ 0 & \widehat{\pi_2}(1 - \widehat{\pi_2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \widehat{\pi_n}(1 - \widehat{\pi_n}) \end{pmatrix}$$

And:

$$W = \text{Diag } \widehat{\pi_1}(1 - \widehat{\pi_1}), \cdots, \widehat{\pi_n}(1 - \widehat{\pi_n})$$

2.4.3. Odds and Odds-ratios

The odds ratio (OR) is a statistical procedure used to evaluate the association between two qualitative random variables. This procedure is often used in logistic regression to measure a relative effect. Knowing that we are in a case of a dichotomous response variable y_i (binary logistic regression), the probability of having Y=1 knowing that X= x is noted π_i. We determine the chance (odds) of having (Y =1|X = x) rather than having (Y = 0|X = x) by the ratio $\{\pi\}/\{1-\pi\}$. The odds ratio can be expressed as follows:

$$OR = \frac{\pi(x+1)/[1-\pi(x+1)]}{\pi(x)/[1-\pi(x)]}$$

3. RESULTS AND DISCUSSION

In a context particularly marked by the worsening of the Covid-19 pandemic, the public authorities have legislated various measures and tax reforms to alleviate the tensions caused by the health crisis, to stimulate considerably the economic recovery and the productive fabric and specifically to attract the intention of agents operating in the informal sector to widen the base of the State's tax revenues. Nevertheless, we will attempt through this study to predict the act of commitment of informal units, sensitized by the new tax reforms in terms of corporate tax (C.T), income tax (I.T), and value-added tax (VAT), using binary logistic regression.

We consider a sample of n = 1000 informal production units taken from the Rabat-Salé-Kénitra area. The collection of responses is carried out through direct interviews, using a questionnaire composed of 20 closed questions. These answers declined 34.9% of the informal units from Rabat, 31.1% from Salé, and 34% from Kenitra. However, 3.6% of them carry out industrial activities, 50.3% construction activities, 28.7% commercial activities, and 17.4% service activities. However, 18% of these informal units have 1 year of seniority, 50.3% have between 2 and 4 years, 28.7% have between 5 and 6 years, and 3% have more than 6 years of experience in the informal sector.

The observations are fragmented into two groups of units G1 and G2 identifiable by a set of independent variables X_1 , X_2 , X_3 . More precisely X_1 represents the corporate tax reform, X_2 is the income tax reform, and X_3 . Value-added tax reform. Let Y be the dichotomous qualitative variable to be predicted (response variable) expressing the decision of the informal units to engage with the general tax department. Y has the value (1) if the unit belongs to the G_1 group

and (0) if it comes from the G_2 . group. Noting also that G_1 is dedicated to units choosing to commit to the tax administration, and G_2 is dedicated to those deciding to remain uncommitted. Hence, we can write:

- **Y = 1**: Engagement of informal units with the general directorate of taxes
- **Y = 0**: No engagement of informal units with the general directorate of taxes
- **X_1 :** Corporate tax reform (CT)
- **X_2 :** Income tax reform (IT)
- **X_3 :** Value-added tax reform (VAT)

In this study, we have tried to measure the consequences of each of the explanatory variables on the engagement decision of informal units with their tax department. In other words, we will attempt to quantify the impact of tax reforms implemented by the public authorities in the areas of corporate income tax, income tax, and value-added tax on the decision of informal units to engage with the general tax directorate or to give up this option.

Table N° 4: Reliability test

Cronbach's Alpha	Cronbach's Alpha based on standardized elements	Number of elements
0.842	**0,841**	**3**

Source: Author

According to the reliability test, we note that the value of the coefficient $\hat{\alpha}$ = 0.842 largely exceeds the conventional minimum threshold of α = 0.70 (Nunnally J. C. 1978), (Darren and Mallery 2008) revealing that we obtain, for this assortment composed of three explanatory elements of the dependent variable, a satisfactory internal consistency.

Table N° 5 : R^2 ajusted

-2 Log of Likelihood	R^2 of Cox and Snell	R^2 of Nagelkerke	R^2 of the sum of squares	R^2 (Adjusted) of the sum of squares
396.009	0.839	0.811	0.848	0.846

Source: Author

The model summary provides the values of (-2LL), Cox and Snell, and Nagelkerke for the full model. The value of (-2LL) for this model reaches 396.009. This value was compared to that

of the base model using the chi-square test to reveal a highly significant decrease between the two ($p = 0.000 < 0.05$). This degradation justifies that the new model is significantly better fitted than the null model. Furthermore, the values tell us approximately how much variation in the outcome is explained by the model used. The Cox and Snell of the full model are 0.839 indicating that there is an 83.9% probability that an economic unit operating in the informal sector will engage with the tax administration and formally declare its activity. Furthermore, the Nagelkerke, which is an adjusted version of the Cox-Snell and therefore closer to reality, is 0.811. Thus, we can say that the explanatory variables contribute to explaining 81.1% of the variation in the probability that an informally operating firm can report its activity to the tax authorities after the implemented post-Covid-19 reforms, covering corporate taxes, income taxes, and value-added tax. On the other hand, a high value of the fitted or interpolated coefficient of determination refers to a better fit of the model to the data used. In our case the adjusted coefficient of determination (adjusted) = 0.846, i.e. 84.6% of the dispersion is explained by the binary logistic regression model.

Table Nº 6: Interelements correlation matrix

	Corporate tax reform (CT)	**Income tax reform (IT)**	**Value-added tax reform (VAT)**
Corporate tax reform (CT)	1	0,781	0,871
Income tax reform (IT)	0,781	1	0,895
Value-added tax reform (VAT)	0,871	0,895	1

Source: Author

The matrix of inter-element correlations is a matrix of statistical correlation coefficients calculated based on several variables taken two by two. It allows for quick detect the existing links between the introduced variables by foreseeing several studies and statistical explanations beforehand. However, the correlation matrix is symmetrical, and its diagonal is made up of 1's since the correlation of a variable with itself is perfect. The correlation matrix based on our study's answers shows that all the variables used are sufficiently correlated, with a correlation

coefficient varying between r = 0.781 and r = 0.895 noting that: $0.781 \leq r \leq 0.895$, confirming moreover the result of Cronbach's Alpha reliability coefficient.

Table Nº 7: Chi-square test

	Corporate tax reform (CT)	**Income tax reform (IT)**	**Value-added tax reform (VAT)**
Chi-square value of pearson	19.512 ddl = 1	15.376 ddl = 1	16.876 ddl = 1
Asymptotic significance (bilateral)	0.000 <0.05	0.000 <0.05	0.000 <0.05

Source: Author

The Chi-square test shows the relationship between the explanatory variables X_1: Corporate tax reform (CT), X_2: Income tax reform (IT), X_3: Value − added tax reform (VAT) and the response variable "the engagement deed of the informal units with the Directorate General of Taxes" is highly significant, and an asymptotic significance (two-sided) of $p = 0,000 < 0.05$. These results refer to rejecting the null hypothesis H_0. In other words, the explanatory variables selected in this study have a significant relationship with the dependent variable, the engagement act of economic units operating in the informal sector with the tax administration.

Table Nº 8: Cramer test

		Value	**Approximate significance**
Cramer's V	X_1: **Corporate tax reform (CT)**	0,416	0,000
	X_2 : **Income tax reform (IT)**	0,581	0,000
	X_3: **Value-added tax reform (VAT)**	0,526	0,000

Source: Author

The value of Cramer's V varies in the interval [0,1]. In our case, we notice that the three explanatory variables "X_1: **Corporate tax reform (CT)**", "X_2 : **Income tax reform (IT)**" and "X_3: **Value-added tax reform (VAT)**" have a strong link with the response variable, "engagement of informal units" (Louis M. Rea and Richard A. Parker (1992)). According to the work of Louis M. Rea and Richard A. Parker, if the value of Cramer's V is between 0.4 and

0.6 the association between the dependent variable and the independent variables is relatively strong. As mentioned in the table of the Cramer's V test, the set of values is bounded between the values 0.4 and 0.6.

Table Nº 9: Area under curve

	AUC	Standard error	Asymptotic Sig.	Asymptotic confidence interval for 95%	
				Inferior	superior
X_1: **Corporate tax reform (CT)**	0.618	0.029	0.001	0.562	0.674
X_2 : **Income tax reform (IT)**	0.640	0.028	0.007	0.601	0.676
X_3: **Value-added tax reform (VAT)**	0.667	0.028	0.003	0.654	0.712

Source: Author

The AUC (area-under-curve) expresses the probability of placing a positive element in front of a negative element. However, this technique proposes an AUC = 0.5 as a baseline situation that our classifier needs to improve. At first glance, all results are highly significant with a p = 0.000 ≤ 0.05. On the other hand, the table also reports AUCs that exceed the baseline situation (AUC = 0.5), which means that the explanatory variables used in the model all significantly impact the response variable. However, it can be predicted that informal units are 61.8 % ($IC^{5\%}$= [0.562, 0.674]) more likely to engage with their tax authorities than to operate informally if they are exposed to the new Corporate tax reform (CT). Similarly, the new Income tax reform (IT) is likely to generate 64%. ($IC^{5\%}$ = [0.601, 0.676]) chance of engagement. In addition, the Value-added tax reform (VAT) is capable of generating 66.7% engagement act ($IC^{5\%}$ = [0.654, 0712]).

Table Nº 10: Table of variables in the equation

	$\hat{\beta}$	E. S	Wald	ddl	Sig	Exp(β)	Confidence interval for Exp (β) (95%)	
							Inferior	superior
X_1	1.615	0.460	11.98	1	0.001	5.030	3.015	7.552
X_2	1.871	0.617	9.212	1	0.002	6.497	4.941	9.753
X_3	1.558	0.571	7.341	1	0.007	4.751	2.542	6.642

Source: Author

This table provides the regression coefficients $\hat{\beta}$, the Wald statistic for testing statistical

significance, the odds ratio exp ($\hat{\beta}$) for each predictor variable, and finally the confidence interval for each odds ratio (OR). However, it is easy to interpret the p-meanings, but the question that arises at this point is how to interpret the regression coefficients $\hat{\beta}$. What does this coefficient correspond to, and how can it be interpreted? Nevertheless, the regression coefficient $\hat{\beta}$ can only explain the direction of fluctuation between the explanatory variable and the response variable. That is, a positive sign of the coefficient $\hat{\beta}$ refers to a change in the same direction between the predictor variable and the dependent variable, whereas a negative sign refers to a change in two opposite directions of the two variables. Apart from the coefficient $\hat{\beta}$ is not interpretable. However, the exponential of $\hat{\beta}$ "(exp ($\hat{\beta}$))" has a meaning that is easily interpreted by statisticians. The "exp ($\hat{\beta}$)" also called odds-ratio (OR), odds ratio, or also a close relative risk, designates a relationship to the response variable.

Looking at the results, we find a highly significant effect of all predictor variables on the response variable "Act of engagement of informal units with the tax administration". However, the p (X_1: **Corporate tax reform (CT)**) = 0.001 < 0.05, p (X_2 : **Income tax reform (IT)**) = 0.002 < 0.05, and p (X_3: **Value-added tax reform (VAT)**) = 0.007 < 0.05. The column exp($\hat{\beta}$) (Odds Ratio) tells us that the different explanatory variables each influence the variable to be predicted distinctly. In line with our case, we can claim that the **Corporate tax reform (CT)** can generate a fivefold increase in the chance (OR(X_1) = 5.0 30, $IC^{5\%}$ = [2.015, 12.552]) that informal units are likely to engage with the tax department. In the same manner, an **Income tax reform (IT)** is also six times more likely (OR(X_2) = 6.497, $IC^{5\%}$= [1.941, 21.753]) that they will choose to engage than to operate outside of tax regulations. Also, a **Value-added tax reform (VAT)** is four times more likely (OR(X_3) = 4.751, $IC^{5\%}$= [1.542, 14.642]) that they will fully comply with the tax authorities. We note that the new reforms and tax bases for corporate income tax, income tax, and value-added tax have contributed significantly to the engagement of informal economic units, draining additional tax revenues from the general state budget.

4. CONCLUSION

In a fearful context marked by the coronavirus epidemic, the Moroccan State has put in place several preventive measures to ensure economic agents' sanitary, economic and social security. However, Morocco has mobilized various private and public actors to involve them in the process of the absorption of this health crisis. The measures implemented have included easing

various tax administration and tax litigation deadlines, as well as a reform marked by a downward tax trend. In addition, Morocco has used its tax system as an effective means to provide economic support to various sectors of the economy affected by the pandemic. These actions have not only contributed to the resilience of formally operating production units but also encouraged production units operating in the informal sector to engage with the tax authorities for the first time. In capturing the importance and benefit of this engagement decision by informal units on the state budget balance and gross domestic product, our study simply predicted and explained the act of engagement by the reforms implemented in terms of corporate income tax, income tax, and value-added tax during the covid-19 period. This study has shown that these new tax bases have had a positive impact on the engagement of informal units to register with their tax authorities. Nevertheless, the reform of the corporate income tax made them five times more likely to engage than to remain uncommitted. Also, the income tax reform contributed to six times more chances to engage than to operate outside the tax regulations. Also, value-added tax reform was four times more likely to engage than to operate in the informal sector. Using generalized linear models, and specifically binary logistic regression, we were able to model the binary response variable "engagement of informal units" and explain how effective the tax actions implemented were in attracting these units and converting them to taxpayers of the General Tax Directorate.

Apart from this, classical linear models are predominant in the arsenal of statistical models used by the research and academic community. However, these models include common methods such as linear regression, analysis of variance, and analysis of covariance. These models are based on the assumption of a linear relationship between the expectation of the response variable and the explanatory variables. However, this condition of linearity cannot always be present between the explanatory variables and the variable to be explained. In these circumstances, generalized linear models (GLM) are used to analyze non-linear relationships, such as the binary logistic regression model, fish regression, probit regression, etc. These generalized linear models are likely to be used in the analysis of the response variable. These generalized linear models can transform a non-linear relationship between the dependent and independent variables into a linear relationship using an interpolated link function. This function can transform categorical responses to a continuous scale without limit. This article presents an application test by choosing one of the most important extensions of generalized linear models such as binary logistic regression. This application confers to the prediction of

the engagement activities of informal production units sensitized by the new tax reform implemented during the covid-19 health crisis.

5. REFERENCES

Ayer T., Chhatwal J., Alagoz O., Kahn C. E., Woods R. W., Et Burnside E. S., (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radio Graphics, 30, pp.13–22.

Adam Duhachek, Anne T. Cough- Lan, Et Dawn Iacobucci, (2005). Results on the Standard Error of the Coefficient Alpha Index of Reliability. Marketing Science, Vol. 24, No. 2, Spring, pp. 294-301.

Anderson, Duncan, Sholom Feldblum, Claudine Modlin, Doris Schirmacher, Ernesto Schirmacher, Et Neeza Thandi, (2004). A Practitioners Guide to Generalized Linear Models, 3rd ed. Arlington: Casualty Actuarial Society, Discussion Paper Program, pp. 1–116.

Agresti A., (1996). An introduction to categorical data analysis. NY: John Wiley. An excellent, accessible introduction.

Burns, R. B., Burns, R., & Burns, R. P. (2008). Business research methods and statistics using SPSS. London: Sage.

Code Général des Impôts (2022).

https://www.finances.gov.ma/Publication/dgi/2022/CGI2022Fr.pdf. Maroc.

Christensen R., Johnson W., Branscu A. Et Hanson T. E., (2011). Bayesian Ideas and Data Analysis. Boca Raton, FL: Chapman Hall/CRC.

David W. HOSMER, and Stanley LEMESHOW (2000), "applied logistic regression", June 2000.

Denuit M., Charpentier A., (2005). Mathematics of non-life insurance, Volumes I and II, Economica, Paris.

Droesbeke J. J., Lejeune M., Et Saporta G., (2005). Statistical models for qualitative data. Paris.

Gary KING and Langche ZENG (2001). "logistic Regression in Rare Events Data". Political Analysis, pages 137–163.

Greenhouse, J. B., Bromberg, J. A., & Fromm, D. A. (1995). An introduction to logistic regression with an application to the analysis of language recovery following a stroke. Journal of Communication Disorders, 28, 229–246.

Guerrieri, V. et al. (2020), Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?

Hosmer, David Et Stanley Lemeshow (2000). Applied Logistic Regression. 2nd ed. NY: Wiley Sons. A much-cited treatment utilized in SPSS routines.

Johansson, Björn and Ohlsson, Esbjörn (2010). Gradient Boosting Machines and Non-Life Insurance Pricing. Edition: Heidelberg : Springer. SSRN:

https://ssrn.com/abstract=4294965 or http://dx.doi.org/10.2139/ssrn.4294965

Muijs D. (2010). Doing quantitative research in education with SPSS. London: Sage.

Nelder, John; Wedderburn, Robert (1972). Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General). 135 (3): 370–384.

OCDE (2020), « Evaluating the initial impact of COVID containment measures on economic activity », https://read.oecd-ilibrary.org/view/?nref=126126448 − kcrc0cs6ia.

Paul-Delvaux, L., Crépon, B., Guernaoui, K. E., Hanna, R., & Sekkarie, S. (2020). Impact of the covid-19 pandemic on the Moroccan labor market and public response to the crisis.

Wuensch D. (2009). Binary logistic regression with PASW/SPSS.