On the existence of maximum likelihood estimators in Poisson-gamma HGLM and negative binomial regression model

Lucien Diégane Gning[*]

LSTA, Université Pierre et Marie CURIE, Paris, France
and
LERSTAD, Université Gaston BERGER, Saint-Louis, Sénégal
e-mail: luciendesgning@yahoo.fr

and

Daniel Pierre-Loti-Viaud

LSTA, Université Pierre et Marie CURIE, Paris, France
e-mail: daniel.pierre-loti-viaud@upmc.fr

Abstract: A breakthrough is provided in the study of the existence problem for maximum likelihood estimators (MLE) in the hierarchical generalized linear model (HGLM) of Poisson-gamma type, as well as in the negative binomial regression model. Any more than the uniqueness problem associated, the existence problem of MLE for these models has not yet been studied except in the very special case of the sample. This issue is addressed here for the Poisson-gamma HGLM, and a sufficient condition is obtained to ensure the MLE existence in that case. It is also shown that this condition has the same effect in the negative binomial regression model with the index parameter considered as unknown. In the latter model, the obtained condition appears as a natural extension of the necessary and sufficient condition well known for solving the existence and uniqueness problems for the index parameter MLE in the sample case.

The Poisson-gamma HGLM are members of the hierarchical generalized linear model family (Lee and Nelder [14]), which is an extension of the generalized linear model family and of the generalized linear mixed model family. HGLM have many fields of application, and are specifically adapted for representing longitudinal data that are generally correlated (Cameron and Trivedi [4], Hilbe [12], Lee and Nelder [14], Molas and Lesaffre [19]). For instance, Poisson-gamma HGLM are used in practice to describe longitudinal count data, as the claim numbers in non-life insurance (Dionne and Vanasse [6], Boucher et al. [3]), among

---

[*]Corresponding author.

others. It should be noted that a Poisson-gamma HGLM considered at a single moment follows a negative binomial regression model (Cameron and Trivedi [4], Hilbe[12]), since a mixture of Poisson distributions by a gamma distribution gives a negative binomial distribution (Johnson et al. [13]). Therefore, without the regression part, a very special case of Poisson-gamma HGLM is the negative binomial sample. To estimate parameters in HGLM, several methods have been proposed in the literature, see, for instance, Lee and Nelder [14], Lee et al. [15], Molas and Lesaffre [19], or Liang and Zeger [18] and Zeger et al. [24]. In this work, we only focus on the classical maximum likelihood estimation approach.

The existence and uniqueness problems for maximum likelihood estimators (MLE) in Poisson-gamma HGLM are not considered in the literature, except in the case of the negative binomial sample (see Levin and Reeds [17] for a correct solution in this case). And it may be observed that this very special case has had a "long" history before being entirely resolved (see Fisher [7] and Haldane [11] for the beginning). The aim of this paper is to provide a breakthrough in the study of the existence problem for MLE in Poisson-gamma HGLM. The same issue for generalized linear models and log linear models is discussed for instance in Christensen [5], Haberman [10], Santner and Duffy [22], a special case of these types of models consisting of Poisson regression models.

However, it should be noticed that the problem addressed is only partially solved here. A condition on the observations is presented, and it is only shown that it is sufficient to solve the MLE existence problem. At present, interesting extensions of this work seem not to be easy to achieve, despite the fact that the presented condition appears to be a good candidate to be a necessary and sufficient condition for solving the existence and uniqueness problems of MLE for the HGLM Poisson-gamma. On the contrary, the transposition of the obtained result to the negative binomial regression model is trivial.

To complete this introduction, it is probably useful to discuss the importance to be given to the fact that maximum likelihood estimators do not always exist. First, it may be observed that two conditions must be introduced to ensure that these estimators exist for Poisson-gamma HGLM. The former is typical of the presence of covariables in a Poisson regression model and ensures that there is a solution to the likelihood equations associated (Haberman [10]). The second is concerned with the additional parameter which exists in the Poisson-gamma HGLM and fixes the longitudinal structure of the model. Also note that the need for such conditions is well referenced, for example in the special case of samples of discrete distributions. Moreover, in this latter case, it is known that these conditions are not satisfied on a set of observations of nonzero probability, even if the probability of this set converges to 0 when the sample size goes to infinity. Now, some additional details specifying the role of these conditions are given by separating theoretical and practical aspects.

The theoretical aspects are considered discussing a well-known example which is among the simplest possible to be described. Consider a sample of negative binomial distribution $\mathcal{NB}(r, p)$, with parameters $r \in]0, \infty[$ and $p \in]0, 1[$, and probability mass function f given by $f(y) = (\Gamma(y + r)/(y!\,\Gamma(r)))\,p^y(1 - p)^r$ for $y \in \mathbb{N}$. The empirical mean and variance are calculated for this sample (the

empirical variance has to be normalized by the sample size). The necessary and sufficient condition for solving the existence and uniqueness problems of MLE for the parameter r and p is then that the empirical variance is strictly greater than the empirical mean. This condition has long been known (Fisher [7] and Haldane [11]), but its role was fully demonstrated nearly forty years later (Levin and Reeds [17]); moreover, incorrect demonstrations have even been published since this last publication. Under the condition set out the MLE of r and p exist in $]0, \infty[\times]0, 1[$. What will happen when this condition is not satisfied? A laborious calculation shows that a solution that maximizes the likelihood function exists on the boundary of the area where the parameters are defined. This solution is such that r is equal to infinity, p is equal to 0 and rp is equal to the empirical mean (and the likelihood function is finite). Then, recalling the convergence result of the negative binomial distribution $\mathcal{NB}(r, p)$ to the Poisson distribution $\mathcal{P}(\lambda)$ when λ is positive, r goes to infinity, p goes to 0 and rp goes to λ, the "boundary" solution may be connected to the MLE of parameter λ in the Poisson sample when the empirical mean is positive. And it remains to handle the case where the sample mean is zero, again a change of statistical models can be considered for this purpose. However, from a theoretical point of view, a comfortable approach is to not change the statistical model and to seek estimators within their domain of definition. It is a choice which is usually done and this is the one retained in this paper: the conditions of existence presented below ensure that the MLE will exist in the area where the parameters are defined for the model used and not on the boundary of this area. However, it is important to note that the same boundary phenomenon as the one just described for the negative binomial sample arises for the Poisson-gamma HGLM. The Poisson regression models are then the boundary statistical models having to be considered in that case. As well, the convergence of Poisson-gamma HGLM to Poisson regression models is a known fact (Gning [8], Rodríguez-Avi et al. [21]) that is used in this paper in proving the main result presented. Indeed, this convergence result allows to obtain the behavior of the likelihood function when the parameters are at the border of their domain of definition and this is crucial to conclude on the existence of MLE in the sense that has been reminded.

From a practical point of view, several numerical methods already exist to find estimators in the context of HGLM and these algorithms address these models in a wider context than envisaged here. Some of them, however, rely on the concept of maximum likelihood (Molas and Lesaffre [19]), other methods are also used (see, for instance, Lesaffre and Spiessens [16]). What meaning should be given to the conditions of existence obtained? To the extent that it would be shown that these conditions are necessary and sufficient to ensure that MLE exist, it seems inevitable that an algorithm takes into account their function to change statistical models when they are not satisfied by the observations, and also when they are close to not be satisfied. Otherwise, errors and instabilities will appear in a numerical research of MLE. Moreover, it may be indicated that, at least for the case of a simple negative binomial sample, other estimation methods, such as the method of moments, encounter the same problem and there exist recent publications still questioning methods for estimating both parameters r and p

when the condition of existence of MLE is not satisfied (see, for instance, Al-Khasawneh [2], Robinson and Smyth [20]). Finally, the following point should be emphasized. It will be easy to see that the two conditions obtained in the case of Poisson-gamma HGLM become less easy to check when the number of covariables in the regression part of the model increases, therefore, the subset of observations satisfying these conditions is likely to become smaller in this case.

The rest of this paper is organized as follows. In the first section, we present the Poisson-gamma HGLM and introduce notation that will be useful later. In the second section, we state our main result, and also preliminary results that will be used to prove the main result. This section concludes with a conjecture, since it seems quite likely that our condition is also a necessary and sufficient condition for the MLE existence and uniqueness in the models considered. All the results are proved in the third section. And some possible extensions of this work are discussed in the last section.

1. Poisson-gamma HGLM

Consider a sample of n individuals kept under observation over T periods. For the individual k during the period t, a random count variable Y_{kt} is observed and represents the dependent variable, while the J deterministic characteristics $\boldsymbol{x_{kt}} = (x_{kt1}, \ldots, x_{ktJ})' \in \mathbb{R}^J$ are known and represent the covariables. For individual k, it is also assumed that there is a real and positive random characteristic Θ_k, which is attached but is not observed. Finally, the regression structure will use the vector of regression parameters $\boldsymbol{\beta} = (\beta_1 \ldots \beta_J)' \in \mathbb{R}^J$, the quantities $\lambda_{kt} = \exp(\boldsymbol{x'_{kt}\beta})$, for $k = 1, \ldots, n$ and $t = 1, \ldots, T$, and the $nT \times J$ regression matrix $X = (\boldsymbol{x_{11}x_{12}} \cdots \boldsymbol{x_{nT}})'$.

In this setting, denoting by $\mathcal{P}(\lambda)$ the Poisson distribution of parameter λ, $\lambda > 0$, and, by $gamma(a, b)$ the gamma distribution of parameters a and b, $a > 0$, $b > 0$, the Poisson-gamma HGLM requires the three following assumptions:

$(H1)$ the individual k is represented by the random vector $(\Theta_k \ Y_{k1} \ldots Y_{kT})'$, and these vectors are independent for $k = 1, \ldots, n$;

$(H2)$ for each k and $\theta_k > 0$, given $\Theta_k = \theta_k$, the variables $Y_{k1}, \ldots, Y_{kT}$ are independent and for each t the conditional distribution of Y_{kt} is the Poisson distribution $\mathcal{P}(\lambda_{kt}\theta_k)$;

$(H3)$ for each k the distribution of Θ_k is the gamma distribution $gamma(a, a)$.

For the presentation of the results obtained, one additional assumption must still be made:

$(H4)$ X is a matrix of full rank, and $nT \geq J$.

From these assumptions, it follows that the model defined is identifiable and satisfies the following constraint:

$$\log \boldsymbol{\lambda} \in \mathcal{F},$$

with $\boldsymbol{\lambda} = (\lambda_{11}\lambda_{12}\cdots\lambda_{nT})' \in \mathbb{R}^{nT}$, $\log\boldsymbol{\lambda} = (\log\lambda_{11}\log\lambda_{12}\cdots\log\lambda_{nT})' \in \mathbb{R}^{nT}$ and $\mathcal{F} = \mathrm{Im}(X) \subset \mathbb{R}^{nT}$ which is the range of X. Observe also that the non-conditional distribution of Y_{kt} is the negative binomial distribution $\mathcal{NB}(a, \lambda_{kt}/(a + \lambda_{kt}))$, since, by $(H1)$ and $(H2)$, for $y_{kt} \in \mathbb{N}$:

$$
\begin{aligned}
\mathbb{P}(Y_{kt} = y_{kt}) &= \int_0^{+\infty} \frac{(\lambda_{kt}\theta_k)^{y_{kt}}}{y_{kt}!} e^{-\lambda_{kt}\theta_k} \frac{a^a}{\Gamma(a)} \theta_k^{a-1} e^{-a\theta_k} \, d\theta_k \\
&= \frac{\Gamma(a + y_{kt})}{y_{kt}!\,\Gamma(a)} \left(\frac{\lambda_{kt}}{a + \lambda_{kt}}\right)^{y_{kt}} \left(\frac{a}{a + \lambda_{kt}}\right)^a.
\end{aligned}
$$

Therefore, when $T = 1$, the Y_{k1}, $k = 1, \ldots, n$, constitute one of the main forms of negative binomial regression models, with $\boldsymbol{\beta}$ as regression parameter and a as index parameter (see Hilbe [12]).

An extension of the Poisson-gamma HGLM, useful for instance in insurance, is to assume that individuals are not necessarily observed for the same number of periods. All results presented seem to be easily extended to cover this case.

2. Existence results for MLE

Starting with the calculus of the log-likelihood expression, this section introduces the main results obtained. Several preliminary results are also stated.

2.1. The log-likelihood function

For any $k = 1, \ldots, n$, given Θ_k, the count variables $Y_{k1}, \ldots, Y_{kT}$ are assumed independent. Therefore, their joint probability function is given for $(y_{k1} \cdots y_{kT})' \in \mathbb{N}^T$ by:

$$
\begin{aligned}
\mathbb{P}(Y_{k1} = y_{k1}, \ldots, Y_{kT} = y_{kT}) &= \int_0^{+\infty} \left(\prod_{t=1}^T \frac{(\lambda_{kt}\theta_k)^{y_{kt}}}{y_{kt}!} e^{-\lambda_{kt}\theta_k}\right) \frac{a^a}{\Gamma(a)} \theta_k^{a-1} e^{-a\theta_k} \, d\theta_k \\
&= \left(\prod_{t=1}^T \frac{\lambda_{kt}^{y_{kt}}}{y_{kt}!}\right) \frac{a^a}{\Gamma(a)} \frac{\Gamma(a + s_k)}{(a + \mu_k)^{a + s_k}},
\end{aligned}
$$

where $s_k = \sum_{t=1}^T y_{kt}$ and $\mu_k = \sum_{t=1}^T \lambda_{kt}$. So that the log-likelihood function of the model differs of the following function $\ell_{\boldsymbol{y}}$ by a constant term, where $\boldsymbol{y} = (y_{11} \ldots y_{nT})' \in \mathbb{N}^{nT}$ is the vector of observation:

$$
\ell_{\boldsymbol{y}}(a, \boldsymbol{\beta}) = \sum_{k=1}^n \sum_{t=1}^T y_{kt} \boldsymbol{x}_{\boldsymbol{kt}}' \boldsymbol{\beta} - \sum_{k=1}^n (a + s_k) \log\left(a + \sum_{t=1}^T e^{\boldsymbol{x}_{\boldsymbol{kt}}'\boldsymbol{\beta}}\right) + na\log a
$$

$$
+ \sum_{k=1}^n \log\left(\frac{\Gamma(a + s_k)}{\Gamma(a)}\right)
$$

$$= \boldsymbol{y}'X\boldsymbol{\beta} - \sum_{k=1}^{n}(a+s_k)\log\left(a+\sum_{t=1}^{T}e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}\right) + na\log a + \sum_{k=1}^{n}\sum_{j=0}^{s_k-1}\log(a+j),$$

where, for $k = 1, \ldots, n$, $\sum_{j=0}^{s_k-1}\log(a+j) = 0$ whenever $s_k = 0$ (for properties of the function gamma, see Abramovitz and Stegun [1]).

Finally, the following notation is introduced. Let $r = \frac{1}{a}$ and Φ be the function defined from $]0, +\infty[\times\mathbb{R}^J$ to $\mathbb{R}$ by:

$$\Phi(r, \boldsymbol{\beta}) = \ell_{\boldsymbol{y}}\left(\frac{1}{r}, \boldsymbol{\beta}\right) = \boldsymbol{y}'X\boldsymbol{\beta} - \sum_{k=1}^{n}\left(\frac{1}{r}+s_k\right)\log\left(1+r\sum_{t=1}^{T}e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}\right) + \sum_{k=1}^{n}\sum_{j=0}^{s_k-1}\log\left(1+rj\right).$$

(2.1)

Therefore, passing to the limit as r tends to 0, the function $\Phi(0, \cdot)$ can be defined from $\mathbb{R}^J$ to $\mathbb{R}$ by:

$$\Phi(0, \boldsymbol{\beta}) = \boldsymbol{y}'X\boldsymbol{\beta} - \sum_{k=1}^{n}\sum_{t=1}^{T}e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}.$$

(2.2)

And it must be noticed that the function $\Phi(0, \cdot)$ differs of the log-likelihood function of a Poisson regression model by an additive term which is independent of the model parameter $\boldsymbol{\beta}$. Indeed, that is the model where the Y_{kt}, $k = 1, \ldots, n$, $t = 1, \ldots, T$, are independent and the distribution of Y_{kt} is the Poisson distribution $\mathcal{P}(\lambda_{kt})$. Moreover, the following condition is known to be a necessary and sufficient condition for solving the existence and uniqueness problems of MLE for this Poisson regression model (see Haberman [10]):

($C1$) there exists $\boldsymbol{\delta} \in \mathcal{F}^{\perp}$ such that $y_{kt} + \delta_{kt} > 0$, $k = 1, \ldots, n$, $t = 1, \ldots, T$.

Under ($C1$), the vector $\boldsymbol{y}$ is nonzero, and the MLE of $\boldsymbol{\beta}$ in this model is denoted in the following by $\hat{\boldsymbol{\beta}}(0)$.

The following remarks can be made on the condition (C1). Recall that all y_{kt} are natural numbers, so that condition (C1) requires that there is not too many of these observations of the dependent variable that are equal to zero. In the particular case of a Poisson sample, the subspace $\mathcal{F}$ is generated by the vector in $\mathbb{R}^{nT}$ with all coordinates equal to 1 and condition (C1) is then equivalent to at least one of the y_{kt} has to be different from zero (which is also equivalent to their sum has to be different from zero). On the other hand, when all covariates are categorical and the Poisson regression model has a multiplicative structure (Haberman [10], Santner and Dufy [22]), condition (C1) is equivalent to an elementary form. Indeed, in that case, it can be shown (Gning and Pierre-Loti-Viaud [9]) that (C1) is equivalent to the fact that for any covariate and any category of this covariate, there is an observation with this characteristic and a nonzero dependent variable (in other words, all sums on the margins of the observations of the dependent variable are different from zero). In general, in a less accurate way, the dimension of the subspace $\mathcal{F}$ gets larger when the number of covariables increases, which in turn requires an augmentation of the number of nonzero observations of the dependent variable.

2.2. *Main result*

Now, we can state the sufficient condition for the existence of MLE of the Poisson-gamma HGLM introduced in Section 1, letting, when $\hat{\boldsymbol{\beta}}(0)$ exists:

$$\hat{\lambda}_{kt}(0) = e^{\boldsymbol{x}'_{kt}\hat{\boldsymbol{\beta}}(0)}, \ \ k = 1, \ldots, n, \ t = 1, \ldots, T,$$

and considering the condition:

$$(C2) \quad \frac{1}{n}\sum_{k=1}^{n}\frac{1}{T}\Big(s_k - \sum_{t=1}^{T}\hat{\lambda}_{kt}(0)\Big)^2 - \frac{1}{nT}\sum_{k=1}^{n}\sum_{t=1}^{T}y_{kt} > 0.$$

Theorem 1. *If* $(C1)$ *and* $(C2)$ *are satisfied, then the MLE of the parameters a and* $\boldsymbol{\beta}$ *of the model satisfying* $(H1)$-$(H4)$ *exist.*

Remark 1 (Negative binomial regression model). By taking $T = 1$, Theorem 1 applies to one of the main forms of negative binomial regression models. Assuming further that all λ_{k1} are equal, so that Y_{k1}, $k = 1, \ldots, n$, is a negative binomial sample, then, the condition $(C2)$ becomes: the empirical variance is strictly greater than the empirical mean. The latter condition is also necessary and sufficient for the existence and uniqueness of MLE of the index parameter in a negative binomial sample (Levin and Reeds [17]).

Remark 2. $(C2)$ can be interpreted as the empirical variance "between class" of the Poisson regression model related to our model has to be greater than the empirical mean of the observations. Moreover, proceeding as in the first section, it is easily seen that for $k = 1, \ldots, n$ the sum $S_k = \sum_{t=1}^{T} Y_{kt}$ follows the negative binomial distribution $\mathcal{NB}(a, \mu_k/(a+\mu_k))$, where $\mu_k = \sum_{t=1}^{T}\lambda_{kt}$. Since the variance of a negative binomial distribution is strictly greater than the mean, a theoretical counterpart of (C2) is thus obtained in writing that the variance of S_k is strictly greater than its mean and summing in $k = 1, \ldots, n$. Indeed, the variance of S_k can be rewritten as $\mathbb{E}((S_k - \mu_k)^2) = \mathbb{E}((S_k - \sum_{t=1}^{T}\lambda_{kt})^2)$, so that, the Poisson-gamma HGLM introduced in Section 1 satisfies the inequality:

$$\sum_{k=1}^{n}\mathbb{E}\Big(\Big(S_k - \sum_{t=1}^{T}\lambda_{kt}\Big)^2\Big) > \sum_{k=1}^{n}\mathbb{E}S_k = \sum_{k=1}^{n}\sum_{t=1}^{T}\mathbb{E}Y_{kT}\,.$$

2.3. *Preliminary results*

The proof of Theorem 1 is based on the following results, where $\mathbb{1}_{nT} = (1\ldots1)' \in \mathbb{R}^{nT}$, $\boldsymbol{s} = (s_{11}\cdots s_{nT})' \in \mathbb{R}^{nT}$ with $s_{kt} = s_k$, $k = 1, \ldots, n$, $t = 1, \ldots, T$, and $\mathrm{diag}(u_{kt})$ denote the $nT \times nT$ diagonal matrix in which the diagonal entries are $u_{11}, \ldots, u_{nT}$.

Lemma 1. *For each compact subset* K *of* $\mathbb{R}^J$, *when* $r \searrow 0$ *the function* $\Phi(r, \cdot)$ *converges uniformly on* K *to the function* $\Phi(0, \cdot)$.

Lemma 2. *For $r \geq 0$, $\Phi(r, .)$ is a strictly concave function, and $(C1)$ is satisfied if and only if there exists a unique $\hat{\boldsymbol{\beta}}(r) \in \mathbb{R}^J$ such that:*

$$\Phi(r, \hat{\boldsymbol{\beta}}(r)) = \max_{\boldsymbol{\beta} \in \mathbb{R}^J} \Phi(r, \boldsymbol{\beta}).$$

Lemma 3. *If $(C1)$ is satisfied, then, for all neighborhood $\mathcal{V}$ of $\hat{\boldsymbol{\beta}}(0)$ and for $r > 0$ small enough, $\Phi(r, .)$ has a local maximum on $\mathcal{V}$ which is $\hat{\boldsymbol{\beta}}(r)$.*

Lemma 4. *If $(C1)$ is satisfied, then $\lim_{r \searrow 0} \hat{\boldsymbol{\beta}}(r) = \hat{\boldsymbol{\beta}}(0)$.*

Lemma 5. *If $(C1)$ is satisfied, then, for $r > 0$ in a neighborhood of 0:*

$$\hat{\boldsymbol{\beta}}(r) = \hat{\boldsymbol{\beta}}(0) + rD\hat{\boldsymbol{\beta}}(0) + o(r),$$

with $D\hat{\boldsymbol{\beta}}(0)$ being equal to:

$$\left(X^{'} \operatorname{diag}(e^{\boldsymbol{x}_{\boldsymbol{kt}}^{'}\hat{\boldsymbol{\beta}}(0)})X\right)^{-1} \left(X^{'} \operatorname{diag}\left(e^{\boldsymbol{x}_{\boldsymbol{kt}}^{'}\hat{\boldsymbol{\beta}}(0)} \sum_{u=1}^{T} e^{\boldsymbol{x}_{\boldsymbol{ku}}^{'}\hat{\boldsymbol{\beta}}(0)}\right) \mathbb{1}_{nT} - X^{'} \operatorname{diag}\left(e^{\boldsymbol{x}_{\boldsymbol{kt}}^{'}\hat{\boldsymbol{\beta}}(0)}\right)\boldsymbol{s}\right).$$

2.4. Conjecture

The foregoing subsections, and especially Remark 1, lead us to introduce the following conjecture.

Conjecture The MLE of the parameters of Poisson-gamma HGLM, and therefore that of the related negative binomial regression models, exist and are unique if and only if $(C1)$ and $(C2)$ are satisfied.

3. Proofs

The proofs of the results stated previously are established, starting with the proofs of the preliminary results of Section 2.3.

3.1. Proof of the preliminary results

Proof of Lemma 1. Fix a compact subset K of $\mathbb{R}^J$. With $A_k(\boldsymbol{\beta}) = \sum_{t=1}^{T} e^{\boldsymbol{x}_{\boldsymbol{kt}}^{'}\boldsymbol{\beta}}$ and $B(r) = \sum_{k=1}^{n} \sum_{j=0}^{s_k-1} \log\left(1 + rj\right)$, it follows from (2.1) and (2.2) that, for all $r > 0$ and all $\boldsymbol{\beta} \in \mathbb{R}^J$:

$$|\Phi(r, \boldsymbol{\beta}) - \Phi(0, \boldsymbol{\beta})| = \left| \sum_{k=1}^{n} A_k(\boldsymbol{\beta}) - \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log\left(1 + rA_k(\boldsymbol{\beta})\right) + B(r) \right|$$

$$\leq \sum_{k=1}^{n} \left(A_k(\boldsymbol{\beta}) \left| 1 - \frac{\log\left(1 + rA_k(\boldsymbol{\beta})\right)}{rA_k(\boldsymbol{\beta})} \right| + s_k \log\left(1 + rA_k(\boldsymbol{\beta})\right) \right) + B(r).$$

Then, since we have $0 < 1 - \frac{\log(1+u)}{u} \leq \frac{u}{2}$ for $u > 0$, we obtain, for all $r > 0$ and all $\boldsymbol{\beta} \in \mathbb{R}^J$:

$$|\Phi(r, \boldsymbol{\beta}) - \Phi(0, \boldsymbol{\beta})| \leq \sum_{k=1}^{n} \left(\frac{r}{2} A_k^2(\boldsymbol{\beta}) + s_k \log \left(1 + r A_k(\boldsymbol{\beta}) \right) \right) + B(r),$$

Finally, observing that $A_k(\boldsymbol{\beta}) \leq \sum_{t=1}^{T} e^{\|\boldsymbol{x}_{kt}\| \|\boldsymbol{\beta}\|}$, where $\|.\|$ denotes the euclidean norm on $\mathbb{R}^J$, clearly, the last inequality leads to the identity:

$$\lim_{r \searrow 0} \left(\sup_{\boldsymbol{\beta} \in K} |\Phi(r, \boldsymbol{\beta}) - \Phi(0, \boldsymbol{\beta})| \right) = 0. \qquad \square$$

Proof of Lemma 2. See Haberman [10] for the case $r = 0$ and Gning [8] or Gning and Pierre-Loti-Viaud [9] for the case $r > 0$. $\qquad \square$

Proof of Lemma 3. Let $\gamma > 0$, and let $\bar{B}_\gamma$ be the closed ball in $\mathbb{R}^J$ of radius γ, centered at $\hat{\boldsymbol{\beta}}(0)$. Moreover, for $0 < \gamma_1 < \gamma$, let B_{γ_1} be the open ball of radius γ_1, centered at $\hat{\boldsymbol{\beta}}(0)$. Under $(C1)$, we have by Lemma 2:

$$\varepsilon = \Phi(0, \hat{\boldsymbol{\beta}}(0)) - \max_{\substack{\boldsymbol{\beta} \in \bar{B}_\gamma \ \boldsymbol{\beta} \notin B_{\gamma_1}}} \Phi(0, \boldsymbol{\beta}) > 0.$$

Then, according to Lemma 1, there exists $r_0 > 0$ small enough such that:

$$\sup_{\boldsymbol{\beta} \in \bar{B}_\gamma} |\Phi(r, \boldsymbol{\beta}) - \Phi(0, \boldsymbol{\beta})| < \frac{\varepsilon}{3}, \ \ 0 \leq r \leq r_0,$$

from which we deduce the following inequalities:

$$\Phi(r, \hat{\boldsymbol{\beta}}(0)) > \Phi(0, \hat{\boldsymbol{\beta}}(0)) - \frac{\varepsilon}{3}, \ \ 0 \leq r \leq r_0,$$

and:

$$\Phi(r, \boldsymbol{\beta}) < \Phi(0, \hat{\boldsymbol{\beta}}(0)) - \frac{2\varepsilon}{3}, \quad \boldsymbol{\beta} \in \bar{B}_\gamma - B_{\gamma_1}, \ \ 0 \leq r \leq r_0.$$

Hence, for $0 < r < r_0$, the smooth function $\Phi(r, .)$ has a local maximum in B_{γ_1} and this local maximum is necessarily the global maximum $\hat{\boldsymbol{\beta}}(r)$ because of the strict concavity of $\Phi(r, .)$. $\qquad \square$

Proof of Lemma 4. The proof is a direct consequence of Lemmas 2 and 3. $\quad \square$

Proof of Lemma 5. We have to prove that the function $\hat{\boldsymbol{\beta}}$ is of class $\mathcal{C}^1$ in a right neighborhood of 0, and we must establish a first order Taylor expansion at this point.

Consider the function f defined from $[0, +\infty[\times \mathbb{R}^J$ to $\mathbb{R}^J$ by:

$$f(r, \boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} \Phi(r, \boldsymbol{\beta}).$$

If $(C1)$ is satisfied, then Lemma 2 yields that for $r \geq 0$, $\hat{\boldsymbol{\beta}}(r)$ exists and is the unique solution of the equation $f(r, \boldsymbol{\beta}) = 0$, with:

$$
\begin{aligned}
f(r, \boldsymbol{\beta}) &= X'\boldsymbol{y} - \sum_{k=1}^{n}(1 + rs_k)\sum_{t=1}^{T}\frac{e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}}{1 + r\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}}}\boldsymbol{x}_{kt} \\
&= \sum_{k=1}^{n}\sum_{t=1}^{T}\boldsymbol{x}_{kt}\left(y_{kt} - (1 + rs_k)\frac{e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}}{1 + r\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}}}\right).
\end{aligned}
$$

This will show that the function $\hat{\boldsymbol{\beta}}$ is of class $\mathcal{C}^1$ by using the Implicit Function Theorem (see, for example, Spivak [23]). This theorem is applied at 0, in a right neighborhood of 0. The hypotheses of the Implicit Function Theorem are now checked.

First, note that f is of class $\mathcal{C}^1$, and let $D_{\boldsymbol{\beta}}f$ denote its partial derivative with respect to $\boldsymbol{\beta}$, and $D_r f$ denote its partial derivative with respect to r. Then, it is easily obtained that, for $r \geq 0$:

$$
D_{\boldsymbol{\beta}}f(r, \boldsymbol{\beta}) = -X'\text{diag}\left(\frac{(1 + rs_k)\,e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}}{(1 + r\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}})^2}\right)X,
$$

$$
D_r f(r, \boldsymbol{\beta}) =
$$

$$
-X'\text{diag}\left(\frac{e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}}{1 + r\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}}}\right)\boldsymbol{s} + X'\text{diag}\left(\frac{e^{\boldsymbol{x}'_{kt}\boldsymbol{\beta}}\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}}}{(1 + r\sum_{u=1}^{T}e^{\boldsymbol{x}'_{ku}\boldsymbol{\beta}})^2}\right)(\mathbb{1} + r\boldsymbol{s}).
$$

Now, we have:

$$
f(0, \hat{\boldsymbol{\beta}}(0)) = \nabla_{\boldsymbol{\beta}}\Phi(0, \hat{\boldsymbol{\beta}}(0)) = \mathbf{0},
$$

and:

$$
D_{\boldsymbol{\beta}}f(0, \hat{\boldsymbol{\beta}}(0)) = -X'\text{diag}\left(e^{\boldsymbol{x}'_{kt}\hat{\boldsymbol{\beta}}(0)}\right)X,
$$

where $D_{\boldsymbol{\beta}}f(0, \hat{\boldsymbol{\beta}}(0))$ is invertible since X is a full-rank matrix and $nT \geq J$ by $(H4)$.

It thus follows by using the Implicit Function Theorem that there exist $r_1 > r_0$, $\mathcal{V}$ an open neighborhood of $\hat{\boldsymbol{\beta}}(0)$ and h a function of class $\mathcal{C}^1$ defined from $[0, r_1[$ to $\mathcal{V}$, such that, for all $0 \leq r < r_1$:

$$
f(r, \boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}}\Phi(r, \boldsymbol{\beta}) = \mathbf{0} \iff \boldsymbol{\beta} = h(r),
$$

moreover, for all $0 \leq r < r_1$:

$$
Dh(r) = -\left(D_{\boldsymbol{\beta}}f(r, h(r))\right)^{-1}D_r f(r, h(r)).
$$

Finally, since there exists a unique solution for the equation $f(r, \boldsymbol{\beta}) = \mathbf{0}$ for $r \geq 0$, we have $\hat{\boldsymbol{\beta}}(r) = h(r)$ for $0 \leq r < r_1$. Hence, the function $\hat{\boldsymbol{\beta}} = h$ is of class $\mathcal{C}^1$ in $[0, r_1[$, and its first order Taylor expansion at 0 is given by:

$$
\hat{\boldsymbol{\beta}}(r) = \hat{\boldsymbol{\beta}}(0) + rD\hat{\boldsymbol{\beta}}(0) + o(r),
$$

where some straightforward calculations lead to the expression:

$$
D\hat{\boldsymbol{\beta}}(0) = \left(X^{'}\mathrm{diag}(e^{\boldsymbol{x}_{kt}^{'}\hat{\boldsymbol{\beta}}(0)})X\right)^{-1}\left(X^{'}\mathrm{diag}\left(e^{\boldsymbol{x}_{kt}^{'}\hat{\boldsymbol{\beta}}(0)}\sum_{u=1}^{T}e^{\boldsymbol{x}_{ku}^{'}\hat{\boldsymbol{\beta}}(0)}\right)\mathbb{1}_{nT}\right.
$$
$$
\left.-X^{'}\mathrm{diag}\left(e^{\boldsymbol{x}_{kt}^{'}\hat{\boldsymbol{\beta}}(0)}\right)\boldsymbol{s}\right). \quad \square
$$

3.2. Proof of Theorem 1

First, under $(C1)$, and for each $r \geq 0$, the problem $\underset{\boldsymbol{\beta}\in\mathbb{R}^{J}}{\mathrm{argmax}}\,\Phi(r,\boldsymbol{\beta})$ is solved according to Lemma 2 and its unique solution is denoted by $\hat{\boldsymbol{\beta}}(r)$. In addition, it must be recalled that $\hat{\boldsymbol{\beta}}(r)$ is also the unique solution of $\nabla_{\boldsymbol{\beta}}\Phi(r,\boldsymbol{\beta}) = \boldsymbol{0}$.

The completion of the proof of Theorem 1 then consists in showing that $(C2)$ is a sufficient condition for the existence of solutions of the problem $\underset{r\in]0,+\infty[}{\mathrm{argmax}}\,\Phi(r,\hat{\boldsymbol{\beta}}(r))$, given that, for each $r > 0$, $\nabla_{\boldsymbol{\beta}}\Phi(r,\hat{\boldsymbol{\beta}}(r)) = \boldsymbol{0}$.

To achieve this goal, let $\Psi(r) = \Phi(r,\hat{\boldsymbol{\beta}}(r))$, $r \geq 0$, and observe that, using $(C1)$:

$$
\Psi(r) = (\boldsymbol{y}+\boldsymbol{\delta})'X\hat{\boldsymbol{\beta}}(r) - \sum_{k=1}^{n}(\frac{1}{r}+s_k)\log\left(1+r\sum_{t=1}^{T}e^{\boldsymbol{x}_{kt}^{'}\hat{\boldsymbol{\beta}}(r)}\right) + \sum_{k=1}^{n}\sum_{j=0}^{s_k-1}\log\left(1+rj\right),
$$
$$(3.1)$$

where $\boldsymbol{\delta} = (\delta_{11}\ldots\delta_{nT})^{'} \in \mathcal{F}^{\perp}$ (so that $\boldsymbol{\delta}'X = 0$), and is such that $y_{kt} + \delta_{kt} > 0$ for $k = 1,\ldots,n$ and $t = 1,\ldots,T$. Then, the rest of the proof will be completed by studying:

- the limit of Ψ at $+\infty$, and we will show that $\lim_{r\to+\infty}\Psi(r) = -\infty$;
- the behavior of Ψ in a neighborhood of 0, and we will recall that $\lim_{r\to 0}\Psi(r) = \Psi(0)$, specifying in addition the position of Ψ with respect to its limit.

In fact, thanks to these properties, if the function Ψ is strictly above its limit in a neighborhood of 0, and since it is a continuous function, then it reaches its maximum in $]0,+\infty[$.

Limit of Ψ at $+\infty$

First, observe that if $\boldsymbol{\delta}$ satisfies $(C1)$, then for $0 < \alpha < 1$ any $\alpha\boldsymbol{\delta}$ satisfies too $(C1)$ since all y_{kt} are natural numbers. We thus consider a $\boldsymbol{\delta}$ of the form $\boldsymbol{\delta} = \frac{\boldsymbol{\Delta}}{r}$, where $\boldsymbol{\Delta} = (\Delta_{11}\ldots\Delta_{nT})^{'} \in \mathbb{R}^{nT}$ and $\Delta_{kt} < 1/T$ for $k = 1,\ldots,n$ and $t = 1,\ldots,T$. Then, Ψ may be rewritten as follows:

$$
\Psi(r) = (\boldsymbol{y}+\boldsymbol{\delta})'X\hat{\boldsymbol{\beta}}(r) - \sum_{k=1}^{n}(\frac{1}{r}+s_k)\log\left(\frac{1}{r}+\sum_{t=1}^{T}e^{\boldsymbol{x}_{kt}^{'}\hat{\boldsymbol{\beta}}(r)}\right) + \sum_{k=1}^{n}\sum_{j=0}^{s_k-1}\log\left(\frac{1}{r}+j\right)
$$
$$
+ \frac{n}{r}\log(\frac{1}{r}). \tag{3.2}
$$

We start by observing that obviously:

$$\lim_{r \to +\infty} \frac{n}{r} \log(\frac{1}{r}) = 0. \tag{3.3}$$

In addition, it is easy to show that:

$$\lim_{r \to +\infty} \sum_{k=1}^{n} \sum_{j=0}^{s_k-1} \log(\frac{1}{r} + j) = -\infty. \tag{3.4}$$

Indeed, if $(C1)$ is satisfied, the observation that $\boldsymbol{y}$ is nonzero is crucial, so that the set $A = \{k \in \{1, \ldots, n\} : s_k > 0\}$ is not empty. Now, since for $k \in A$, the presence of the term $j = 0$ yields:

$$\lim_{r \to +\infty} \sum_{j=0}^{s_k-1} \log(\frac{1}{r} + j) = -\infty,$$

while for the other k, by definition we have:

$$\sum_{j=0}^{s_k-1} \log(\frac{1}{r} + j) = 0,$$

the proof of (3.4) is therefore complete.

Then, the final step consists in showing that the remaining term defining Ψ in (3.2) is bounded above when $r \to +\infty$. For that purpose, let $\hat{\boldsymbol{\eta}}(r) = (\hat{\eta}_{11}(r) \cdots \hat{\eta}_{nT}(r)) \in \mathbb{R}^{nT}$, where $\hat{\eta}_{kt}(r) = \boldsymbol{x'_{kt}}\hat{\boldsymbol{\beta}}(r)$, $k = 1, \ldots, n$, $t = 1, \ldots, T$, and let, for $\boldsymbol{\eta} = (\eta_{11} \ldots \eta_{nT})' \in \mathbb{R}^{nT}$ and $r > 0$:

$$\varphi(\boldsymbol{\eta}, r) = \sum_{k=1}^{n} \sum_{t=1}^{T} (y_{kt} + \frac{\Delta_{kt}}{r})\eta_{kt} - \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log \left(\frac{1}{r} + \sum_{t=1}^{T} e^{\eta_{kt}} \right). \tag{3.5}$$

We observe that the function φ is not explicitly related to the log-likelihood of some statistical model (because of the presence of the Δ_{kt}), but this function retains the properties summarized in the following lemma, whose proof is deferred to the next sub-section.

Lemma 6. *Assuming that $\Delta_{kt} < 1/T$, $k = 1, \ldots, n$, $t = 1, \ldots, T$, then, for each $r > 0$, the function $\varphi(\cdot, r)$ is of class C^{∞} and strictly concave in $\mathbb{R}^{nT}$. Moreover, its maximum is reached at a unique point $\tilde{\boldsymbol{\eta}}(r) \in \mathbb{R}^{nT}$, which is the unique solution of $\nabla_{\boldsymbol{\eta}}\varphi(\boldsymbol{\eta}, r) = 0$ and is given by $\tilde{\boldsymbol{\eta}}(r) = (\tilde{\eta}_{11} \cdots \tilde{\eta}_{nT})'$ with:*

$$\tilde{\eta}_{kt} = \log \left(\frac{y_{kt} + \Delta_{kt}/r}{1 - \sum_{u=1}^{T} \Delta_{ku}} \right), \quad k = 1, \ldots, n, \ k = 1, \ldots, T.$$

From this, the following upper bound is obtained:

$$(\boldsymbol{y} + \boldsymbol{\delta})'X\hat{\boldsymbol{\beta}}(r) - \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log \left(\frac{1}{r} + \sum_{t=1}^{T} e^{\boldsymbol{x'_{kt}}\hat{\boldsymbol{\beta}}(r)} \right) = \varphi(\hat{\boldsymbol{\eta}}(r), r)$$

$$\leq \sup_{\boldsymbol{\eta} \in \mathbb{R}^{nT}} \varphi(\boldsymbol{\eta}, r) = \varphi(\tilde{\boldsymbol{\eta}}(r), r),$$

where:

$$\varphi(\tilde{\boldsymbol{\eta}}(r), r) = \sum_{k=1}^{n} \sum_{t=1}^{T} (y_{kt} + \frac{\Delta_{kt}}{r}) \log \left(\frac{y_{kt} + \frac{\Delta_{kt}}{r}}{1 - \sum_{u=1}^{T} \Delta_{ku}} \right)$$

$$- \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log \left(\frac{1}{r} + \sum_{t=1}^{T} \frac{y_{kt} + \frac{\Delta_{kt}}{r}}{1 - \sum_{u=1}^{T} \Delta_{ku}} \right)$$

$$= \sum_{k=1}^{n} \sum_{t=1}^{T} (y_{kt} + \frac{\Delta_{kt}}{r}) \log \left(y_{kt} + \frac{\Delta_{kt}}{r} \right) - \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log \left(\frac{1}{r} + s_k \right)$$

$$+ \sum_{k=1}^{n} \frac{1 - \sum_{t=1}^{T} \Delta_{kt}}{r} \log \left(1 - \sum_{u=1}^{T} \Delta_{ku} \right),$$

Moreover, making use of the convention $z \log z = 0$ when $z = 0$, it is easily seen that:

$$\lim_{r \to +\infty} \varphi(\tilde{\boldsymbol{\eta}}(r), r) = \sum_{k=1}^{n} \sum_{t=1}^{T} y_{kt} \log y_{kt} - \sum_{k=1}^{n} s_k \log s_k < +\infty.$$

In particular, these results imply that:

$$\limsup_{r \to +\infty} \left((\boldsymbol{y} + \boldsymbol{\delta})' X \hat{\boldsymbol{\beta}}(r) - \sum_{k=1}^{n} (\frac{1}{r} + s_k) \log \left(\frac{1}{r} + \sum_{t=1}^{T} e^{\boldsymbol{x}'_{kt} \hat{\boldsymbol{\beta}}(r)} \right) \right) < +\infty,$$

which, when combined with (3.2), (3.3) and (3.4), completes the proof that:

$$\lim_{r \to +\infty} \Psi(r) = -\infty.$$

Behavior of Ψ in a neighborhood of 0

In this step, we take $\boldsymbol{\delta} = \boldsymbol{0}$ in the expression (3.1) of Ψ. A Taylor expansion at order 1 will be established for Ψ in a right neighborhood of 0. And it should be noted that all the expansions in what follows are given for r in a right neighborhood of 0.

We start from the result of Lemma 5, by combining this result with the Taylor expansions:

$$\log(1 + rj) = rj + o(r), \quad j \in \mathbb{N},$$

and, since $\hat{\boldsymbol{\beta}}(r)$ is bounded in a right neighborhood of 0, for $k = 1, \ldots, n$:

$$\log \left(1 + r \sum_{t=1}^{T} e^{\boldsymbol{x}'_{kt} \hat{\boldsymbol{\beta}}(r)} \right) = r \sum_{t=1}^{T} e^{\boldsymbol{x}'_{kt} \hat{\boldsymbol{\beta}}(r)} + \frac{r^2}{2} \left(\sum_{t=1}^{T} e^{\boldsymbol{x}'_{kt} \hat{\boldsymbol{\beta}}(r)} \right)^2 + o(r^2),$$

So that, we also have, for $k = 1, \ldots, n$:

$$\log \left(1 + r \sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(r)}\right) = r \sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)} e^{r\boldsymbol{x}'_{\boldsymbol{kt}}D\hat{\boldsymbol{\beta}}(0)} + \frac{r^2}{2} \left(\sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}\right)^2 + o(r^2)$$

$$= r \sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)} + r^2 \left(\sum_{t=1}^{T} \boldsymbol{x}'_{\boldsymbol{kt}} D\hat{\boldsymbol{\beta}}(0)\, e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)} + \frac{1}{2}\left(\sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}\right)^2\right) + o(r^2),$$

where the Taylor expansion $e^u = 1 + u + o(u)$ for u in a neighborhood of 0, has been used to establish the last equality.

Combining these Taylor expansions leads to the following expansion for the function Ψ:

$$\Psi(r) = \boldsymbol{y}'X\hat{\boldsymbol{\beta}}(0) - \sum_{k=1}^{n}\sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)} + r\left(\boldsymbol{y}'XD\hat{\boldsymbol{\beta}}(0) - \sum_{k=1}^{n}\sum_{t=1}^{T} \boldsymbol{x}'_{\boldsymbol{kt}} D\hat{\boldsymbol{\beta}}(0)\, e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}\right.$$

$$\left. + \frac{1}{2}\sum_{k=1}^{n}\left(\sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}\right)^2 - \sum_{k=1}^{n} s_k\left(\sum_{t=1}^{T} e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}\right) + \sum_{k=1}^{n}\sum_{j=0}^{s_k-1} j\right) + o(r)$$

$$= \Phi(0, \hat{\boldsymbol{\beta}}(0)) + r\sum_{k=1}^{n}\left(\frac{1}{2}\left(\sum_{t=1}^{T} \hat{\lambda}_{kt}(0)\right)^2 - s_k\sum_{t=1}^{T} \hat{\lambda}_{kt}(0) + \frac{s_k(s_k-1)}{2}\right) + o(r)$$

$$= \Phi(0, \hat{\boldsymbol{\beta}}(0)) + \frac{r}{2}\sum_{k=1}^{n}\left(\left(s_k - \sum_{t=1}^{T} \hat{\lambda}_{kt}(0)\right)^2 - s_k\right) + o(r),$$

where we recall that $\hat{\lambda}_{kt}(0) = e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)}$, $k = 1, \ldots, n$, $t = 1, \ldots, T$, and that:

$$\boldsymbol{y}'X = \sum_{k=1}^{n}\sum_{t=1}^{T} \boldsymbol{x}'_{\boldsymbol{kt}}\, e^{\boldsymbol{x}'_{\boldsymbol{kt}}\hat{\boldsymbol{\beta}}(0)} \quad \Leftrightarrow \quad \nabla_{\boldsymbol{\beta}}\Phi(0, \hat{\boldsymbol{\beta}}(0)) = 0.$$

Thus, the position of Ψ with respect to its limit $\Phi(0, \hat{\boldsymbol{\beta}}(0)) = \Psi(0)$ is given by the sign of C, where:

$$C = \sum_{k=1}^{n}\left(\left(s_k - \sum_{t=1}^{T} \hat{\lambda}_{kt}(0)\right)^2 - s_k\right) = \sum_{k=1}^{n}\left(s_k - \sum_{t=1}^{T} \hat{\lambda}_{kt}(0)\right)^2 - \sum_{k=1}^{n}\sum_{t=1}^{T} y_{kt}.$$

So that, if $C > 0$, then in a right neighborhood of 0, Ψ is strictly above its limit in 0. Hence, there exists a maximum of Ψ that is attained in $]0, \infty[$ when $(C2)$ is satisfied. $\qquad\square$

3.3. Proof of Lemma 6

First, fix $r > 0$. With the function $\varphi(., r)$ as defined by (3.5), it is easily verified that it is $\mathcal{C}^\infty$ on $\mathbb{R}^{nT}$, and satisfies too:

$$
\nabla_{\boldsymbol{\eta}} \varphi(\boldsymbol{\eta}, r) = \boldsymbol{y} + \boldsymbol{\delta} - \operatorname{diag}\left(\frac{e^{\eta_{kt}}}{\frac{1}{r} + \sum_{u=1}^{T} e^{\eta_{ku}}} \right) (\frac{1}{r} \mathbb{1}_{nT} + \boldsymbol{s}),
$$

and, for k, k_1 and k_2 in $\{1, \ldots, n\}$, and t, t_1 and t_2 in $\{1, \ldots, T\}$:

$$
\frac{\partial^2 \varphi}{\partial \eta_{kt}^2}(\boldsymbol{\eta}, r) = \left(-\frac{e^{\eta_{kt}}}{\frac{1}{r} + \sum_{u=1}^{T} e^{\eta_{ku}}} + \frac{e^{2\eta_{kt}}}{(\frac{1}{r} + \sum_{u=1}^{T} e^{\eta_{ku}})^2} \right) (\frac{1}{r} + s_k);
$$

$$
\frac{\partial^2 \varphi}{\partial \eta_{kt_1} \partial \eta_{kt_2}}(\boldsymbol{\eta}, r) = \frac{e^{\eta_{kt_1} + \eta_{kt_2}}}{(\frac{1}{r} + \sum_{u=1}^{T} e^{\eta_{ku}})^2} (\frac{1}{r} + s_k), \quad t_1 \neq t_2;
$$

$$
\frac{\partial^2 \varphi}{\partial \eta_{k_1 t_1} \partial \eta_{k_2 t_2}}(\boldsymbol{\eta}, r) = 0, \quad k_1 \neq k_2.
$$

Thus, any solution $\boldsymbol{\eta}$ of the equation $\nabla_{\boldsymbol{\eta}} \varphi(\boldsymbol{\eta}, r) = 0$ verifies:

$$
0 < y_{kt} + \frac{\Delta_{kt}}{r} = \frac{(\frac{1}{r} + s_k)}{\frac{1}{r} + \sum_{u=1}^{T} e^{\eta_{ku}}} e^{\eta_{kt}}, \quad k = 1, \ldots, n, \ t = 1, \ldots, T.
$$

That is, for each $k = 1, \ldots, n$, the quantities $e^{\eta_{kt}}$ are proportional to $y_{kt} + \frac{\Delta_{kt}}{r}$ when $t = 1, \ldots, T$, which implies, after an easy calculation, that $\boldsymbol{\eta}$ is necessarily equal to $\tilde{\boldsymbol{\eta}}(r) = (\tilde{\eta}_{11} \cdots \eta_{nT})'$ defined by:

$$
\tilde{\eta}_{kt} = \log\left(\frac{y_{kt} + \Delta_{kt}/r}{1 - \sum_{u=1}^{T} \Delta_{ku}} \right), \quad k = 1, \ldots, n, \ k = 1, \ldots, T.
$$

It should be noted that under our assumptions, we have $\sum_{u=1}^{T} \Delta_{ku} < 1$ for $k = 1 \ldots, n$.

On the other hand, for each $k = 1, \ldots, n$, let $\boldsymbol{a}_k = (a_{k1} \cdots a_{kT})' \in \mathbb{R}^T$ be such that $a_{kt} = e^{\eta_{kt}} / (1/r + \sum_{u=1}^{T} e^{\eta_{ku}})$, $t = 1, \ldots, T$, and let D_k be the $T \times T$ diagonal matrix in which the diagonal entries are the a_{kt}, $t = 1, \ldots, T$. Then, denoting by $H_{\boldsymbol{\eta}} \varphi(\boldsymbol{\eta}, r)$ the Hessian matrix with respect to the vector variable $\boldsymbol{\eta}$, this matrix is block diagonal and equal to:

$$
H_{\boldsymbol{\eta}} \varphi(\boldsymbol{\eta}, r) = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & H_n \end{pmatrix},
$$

where, for each $k = 1, \ldots, n$, H_k is the $T \times T$ symmetric matrix satisfying:

$$
H_k = (\frac{1}{r} + s_k)(\boldsymbol{a}_k \boldsymbol{a}_k' - D_k).
$$

Moreover, for all $k = 1, \ldots, n$, we have $a_{kt} > 0$, $t = 1, \ldots, T$, and $\sum_{u=1}^{T} a_{ku} < 1$, so that it can easily be shown that $\boldsymbol{a}_k \boldsymbol{a}_k' - D_k$, and therefore H_k, is a negative-definite matrix. Which obviously implies that the Hessian matrix is negative-definite and that $\varphi(., r)$ is a strictly concave function on $\mathbb{R}^J$.

Finally, the function $\varphi(., r)$ being strictly concave, and having a value $\tilde{\boldsymbol{\eta}}(r)$ that cancels its gradient, the function $\varphi(., r)$ reaches its maximum at the same value. □

4. Conclusion

When a parametric statistical model is used in the case of count observations, in general there is a nonzero probability subset of the observations on which the MLE are not found in the area where the parameters are defined. The existence of this problem is well referenced, at least for the samples, and for the GLM or log linear models. It is also known when other methods of estimation are used, as is the case when the moment method is applied to a sample of negative binomial distribution. The study of this problem in the context of HGLM is addressed in this work for the estimators obtained by the maximum likelihood method, but considering only the case of HGLM Poisson-gamma having regard to the difficulty encountered in its solving. In contrast, and although extensions are still required, the result obtained allow to evaluate the importance of this phenomenon. In particular, an augmentation of the importance of the regression component, with its own constraint, seems to increase the influence of the second constraint obtained. And one can reasonably expect that such a phenomenon exists for extensions of the model presented, always for count data, but also for other methods of research of estimators. The present study begin to measure its importance in the HGLM setting.

It may then be noted that among the HGLM that deserve to be studied, the ones where the Poisson distribution is replaced by the negative binomial distribution can be cited because another complication arises in this case. The work done here has greatly benefited of the fact that the multidimensional distribution of the count data has an explicit form, but this fact is not true for most extensions that can be considered and notably in the ones just mentioned. It can also be observed that, although numerical procedures currently exist for determining the MLE in the case of Poisson-gamma HGLM, it is important to be informed of the conditions under which such MLE exist. When these conditions are not satisfied by the data, errors and numerical instabilities will have a negative impact on the quality of estimator evaluation to be obtained.

At that time several extensions of the work done can be considered. The main objective being to complete the proof of the conjecture that the (C1) and (C2) conditions are necessary and sufficient for the existence and uniqueness of the MLE in the Poisson-gamma HGLM. In addition, by referring to the full case study of a negative binomial sample, it should be observed that the work done is the easiest part among those necessary to achieve this objective. The study of others HGLM is also an important aim. On the other hand, small extensions of

the results obtained can be easily considered. For example, if the n individuals are not observed on the same number of periods, an extension of Theorem 1 appears clearly as possible following the same pattern of demonstration. For the particular case where $\boldsymbol{x}_{kt} = \boldsymbol{x}_k$ for $k = 1, \ldots, n$, $t = 1, \ldots, T$, a detailed study is given in Gning [8].

Acknowledgements

References

[1] ABRAMOWITZ, M. AND STEGUN, I. (1972). *Handbook of Mathematical Functions*. Dover Publications, New York. MR1225604

[2] AL-KHASAWNEH, M.F. (2010). Estimating the negative binomial dispersion parameter. *Asian Journal of Mathematics and Statistics* **3**(1), 1-15. MR2725773

[3] BOUCHER, J., DENUIT, M. AND GUILLÉN, M. (2008). Models of Insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Casuality Actuarial Society* **2**(1), 136-161.

[4] CAMERON, C.A. AND TRIVEDI, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge. MR1648274

[5] CHRISTENSEN, R. (1990). *Log Linear Models*. Springer-Verlag, New York. MR1075412

[6] DIONNE, G. AND VANASSE, CH. (1989). A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *Astin Bulletin* **19**(2), 199-212.

[7] FISHER, R.A. (1941). The negative binomial distribution. *Annals of Eugenics* **11**, 182-187. MR0006689

[8] GNING, L. (2011). *Utilisation des mélanges de lois de Poisson et de lois binomiales négatives pour établir des tarifs a priori et a posteriori en assurance non vie.* Doctoral thesis, LSTA, Université Pierre et Marie CURIE, Paris, and LERSTAD, Université Gaston BERGER, Saint-Louis.

[9] GNING, L. AND PIERRE-LOTI-VIAUD, D. (2010). Sur les estimateurs du maximum de vraisemblance dans les modèles multiplicatifs de Poisson et binomial négatif. *Afrika Statistika* **5**(13), 297-305. MR2920307

[10] HABERMAN, S. (1973). Log-linear models for frequency data: sufficient statistics and likelihood equations. *Annals of Statistics* **1**, 617-632. MR0690292

[11] HALDANE, J.B.S. (1941). The fitting of the binomial distributions. *Annals of Eugenics* **11**, 179-181.

[12] HILBE, J.M. (2007). *Negative Binomial Regression*. Cambridge University Press, Cambridge. MR2359854

[13] JOHNSON, N., KOTZ, S. AND BALAKRISHNAN, N. (1996). *Discrete Multivariate Distributions*. Wiley, New York. MR1429617

[14] LEE, Y. AND NELDER, J. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)* . **58**, 619-678. MR1410182

[15] LEE, Y., NELDER, J. AND PAWITAN, Y. (2006). *Generalized Linear Models with Random Effects, Unified Analysis via H-likelihood.* Chapman & Hall/CRC, Boca Raton. MR2259540

[16] LESSAFRE, E. AND SPIESSENS, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics* **50**, 325-335. MR1856328

[17] LEVIN, B. AND REEDS, J. (1977). Compound multinomial likelihood functions are unimodal: proof of a conjecture of I.J. Good. *Annals of Statistics* **5**, 79-87. MR0451503

[18] LIANG, K. AND ZEGER, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22. MR0836430

[19] MOLAS, M. AND LESAFFRE, E. (2011). Hierarchical generalized linear models: the R package HGLMMM. *Journal of Statistical Software* **39**(13), 1-20.

[20] ROBINSON, M.D. AND SMYTH, G.K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**(2), 321-332.

[21] RODRÍGUEZ-AVI, J., CONDE-SÁNCHEZ, A., SÁEZ-CASTILLO, A.J., OLMO-JIMÉNEZ, M.J. AND MARTÍNEZ-RODRÍGUEZ, A.M. (2009). A generalized Waring regression model for count data. *Computational Statistics and Data Analysis* **53**, 3717-3725. MR2751960

[22] SANTNER, J. AND DUFFY, E. (1989). *The Statistical Analysis of Discrete Data.* Springer-Verlag, New York. MR1019836

[23] SPIVAK, M. (1979). *A comprehensive introduction to differential geometry.* Publish or Perish, Houston. MR0532830

[24] ZEGER, S., LIANG, K. AND ALBERT, P.S. (1988). Models for longitudinal data: A general equation estimating approach. *Biometrics* **44**(4), 1049-1060.