**Projet Econométrie**

1. Introduction (sujet, problématique, modèle, base de données)
2. Théorie du modèle (modèle de comptage)
3. Explication base de données (variables exogènes gardées et construction variable endogène)
4. Codes et interprétations (sorties sas + interprétations des coefficients et des probas)
5. Conclusion (réponse à la problématique)
6. Introduction

Dans le cadre du cours *Econométrie des choix discrets* nous sommes amenés à travailler sur une base de données contenant diverses informations sur les tirages du jeu Euromillions, notamment les boules et étoiles gagnantes. Nous avons choisi de mettre en application les méthodes vues au chapitre 6 concernant les modèles de comptage. Nous avons ensuite choisi de nous focaliser sur les 2 étoiles. C’est alors que nous est naturellement venue notre problématique d’étude : Qu’elle est la probabilité que les deux étoiles soient supérieures ou égales à 10 ?

Pour cela, il va nous falloir construire notre variable endogène Y, qui peut prendre les modalités 0,1 ou 2. Ensuite, nous déterminerons par une méthode de réduction de dimension les variables exogènes les plus significatives pour ne garder que celles-ci dans l’étude.

Enfin, nous conclurons sur les différentes probabilités obtenues et essaierons de déterminer les variables les plus influentes sur la valeur des étoiles.

1. Théorie du modèle

Comme énoncé précédemment, nous avons choisi de travailler avec les modèles de comptage. Nous rappelons ci-dessous la théorie derrière ces modèles.

Dans un tel modèle, la variable endogène Y est une variable qualitative ayant plusieurs modalités.

La particularité de ce modèle est que les modalités, codées par des chiffres, représentent le nombre d’occurrence d’un évènement. Par exemple, on peut s’intéresser en économétrie au nombre d’enfant d’un couple, un nombre d’élève par classe, un nombre de voiture par foyer etc…

Les variables exogènes quant à elles peuvent être quantitatives comme qualitatives.

Ceci étant fixé, nous détaillons dans la suite les mathématiques sous-jacentes au modèle.

Il s’avère que la meilleure loi de probabilité pour modéliser un nombre d’occurrence est la loi de Poisson.

Donc dans les modèles de comptage, nous considérerons toujours que la variable endogène Y suit une loi de Poisson.

On rappelle la formule de la densité de la loi de Poisson : Soit Y~P(λ), alors P(Y=k) =

Où k appartient à N et λ>0 représente le nombre moyen d’occurrence d’un évènement.

Pour estimer = P(Y=k), il faut donc réussir à estimer λ.

Pour cela, on utilise la fonction génératrice des moments de la loi de Poisson et on obtient :

**E(Y) = λ = exp( + βX)**, où :

* X est le tableau de données constitué des individus en ligne et des variables exogènes en colonne (de taille n\*p)
* est la constante du modèle,
* β = ( est le vecteur des coefficients de la régression

Or pour estimer λ, on voit qu’il faut connaitre pour tout i>=0. Mais ces coefficients ne sont pas connus par définition. On va donc les estimer par la méthode du maximum de vraisemblance.

Une fois et déterminés, on peut donc calculer et par les formules suivantes :

= exp( + X)

=

*Remarque* : dans ce modèle, les coefficients peuvent s’interpréter.