



Master 1 Économie Appliquée parcours
Ingénierie des Données et Évaluations
Économétrie

Les modèles à variables limitées (tronquées)

Enseignant :

M. Philippe COMPAIRE

Matière :

**Econométrie des
choix discrets**

Réalisé par :

ANDE Abra Letsime Irene

KADDAR Halima

Année universitaire : **2024/2025**

PLAN

1- INTRODUCTION

2- PROBLEMATIQUE

3- PRESENTATION DES VARIABLES EXOGENES

4- MODELE ECONOMETRIQUE

4-1 Estimation par OLS sur toutes les observations

4-2 Choix du seuil de troncature

4-3 Estimation du modèle TOBIT

5-PRESENTATION DES RESULTATS & INTERPRETATION DES EFFETS

MARGINAUX

6-CONCLUSION

7- ANNEXE

1-INTRODUCTION

Notre étude s'appuie sur une base de données des tirages de l'Euromillions, qui regroupe tous les numéros sortis à chaque tirage entre 2004 et 2023. Notre base de données contient 1685 observations et plusieurs variables que nous avons sélectionnées pour notre étude. Elle nous permet d'analyser un aspect particulier des tirages : la somme des numéros.

Dans ce travail, on s'intéresse à la somme des numéros tirés lors d'un même tirage. C'est cette somme que l'on cherche à expliquer, elle sera donc notre variable dépendante. **L'objectif principal est de comprendre quelles sont les variables qui peuvent influencer cette somme.**

Nous avons remarqué que les très petites sommes sont rarement observées dans les tirages. C'est logique, car il faut que les cinq numéros tirés soient parmi les plus petits nombres (1-50) pour obtenir une somme basse. Cette rareté des petites sommes pose un problème pour l'analyse statistique. Si on utilisait une simple régression linéaire, on risquerait d'obtenir des résultats faussés. En effet, une régression linéaire essaierait de trouver une droite qui s'ajuste au mieux à toutes les données, même là où il manque des valeurs (les petites sommes). Cela pourrait nous amener à sous-estimer l'influence de certains facteurs sur la somme.

C'est pourquoi on utilise la méthode Tobit. Elle est spécialement conçue pour gérer ce type de situation, où certaines valeurs de la variable que l'on étudie sont comme coupées ou « tronquées ». La méthode Tobit permet de tenir compte de cette particularité et d'obtenir des estimations plus fiables.

2-PROBLEMATIQUE :

Quelles sont les variables qui influencent la somme des numéros tirés lors des tirages de l'Euromillions ?

3-PRESENTATION DES VARIABLES EXOGENES :

Std : l'écart -type.

nombre consécutif : une nouvelle variable créée à partir des numéros des boules tirées.

Code 1 si existence d'au moins deux variable consécutives sinon 0.

Is_fibo : l'existence de la suite de Fibonacci dans un tirage.

Is_premier : nombre de nombre premiers dans un tirage.

intervalle_b : (max-min)/médiane.

multi7 : nombre de numéros qui sont des multiples de 7 dans un tirage.

4-MODELE ECONOMETRIQUE

L'équation que l'on désire estimer se présente comme suit :

Somme= constante +a1*nombre_consecutif +a2*std +a3*multi7 +a4*is_fibo
+a5*is_premier + a6*intervalle_b+ε

4-1 Estimation par OLS sur toutes les observations :

Estimation OLS somme toutes observations

La procédure REG
Modèle : MODEL1
Variable dépendante : somme somme

Nb d'observations lues	1685
Nb d'obs. utilisées	1685

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	6	781680	130280	315.81	<.0001
Erreur	1678	692229	412.53245		
Total sommes corrigées	1684	1473909			

Root MSE	20.31089	R carré	0.5303
Moyenne dépendante	127.20178	R car. ajust.	0.5287
Coeff Var	15.96746		

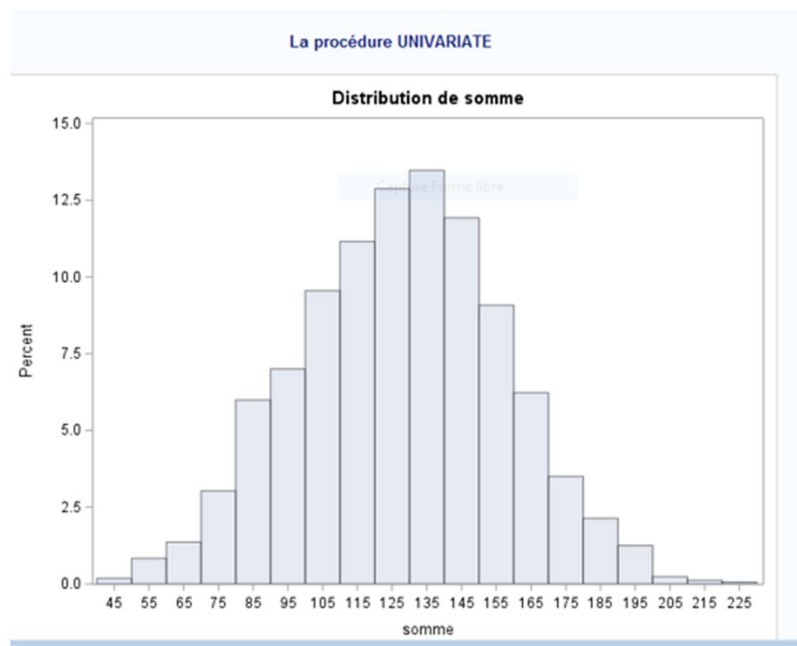
Résultats estimés des paramètres						
Variable	Libellé	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr > t
Intercept	Intercept	1	126.80521	2.17961	58.18	<.0001
std	std	1	3.06395	0.15111	20.28	<.0001
nombre_consecutif	nombre consecutif	1	2.77769	1.07758	2.58	0.0100
is_fibo	is_fibo	1	-9.32399	0.69100	-13.49	<.0001
is_premier	is_premier	1	-2.05552	0.51395	-4.00	<.0001
intervalle_b	intervalle_b	1	-21.82223	0.66112	-33.01	<.0001
multi7	multi7	1	1.54745	0.67583	2.29	0.0222

Source : logiciel SAS

4-2 Choix du seuil de troncature :

Nous avons réalisé une analyse des fréquences cumulées de notre variable dépendante (tableau 1 ANNEXE) et nous avons remarqué que seulement 2.79% des observations ont une somme inférieure à 70. Bien que l'histogramme (graphique 1) ne montre pas de discontinuité abrupte, la faible occurrence de ces valeurs suggère une forme de censure à gauche. Par conséquent, nous avons décidé d'utiliser un modèle Tobit avec un seuil de troncature à 70 pour tenir compte de cette censure potentielle et éviter un biais d'estimation.

Graphique1



Source : logiciel SAS

4-3-Estimation du modèle TOBIT

Estimations tobit							
La procédure QLIM							
Statistiques descriptives des réponses continues							
Variable	Moyenne	Erreur type	Type	Borne inférieure	Borne supérieure	Borne inférieure N obs.	Borne supérieure N obs.
somme	127.4463	29.045590	Censored	70		47	

Synthèse de l'ajustement de modèle	
Nombre de variables endogènes	1
Variable endogène	somme
Nombre d'observations	1685
Log-vraisemblance	-7281
Gradient absolu maximal	2.20795E-6
Nombre d'itérations	13
Méthode d'optimisation	Quasi-Newton
AIC	14577
Critère de stratification Schwarz	14621

L'algorithm a convergé.

Résultats estimés des paramètres					
Paramètre	DDL	Estimation	Erreur type	Valeur du test t	Approx Pr > t
Intercept	1	126.710079	2.175249	58.25	<.0001
std	1	3.251684	0.153675	21.16	<.0001
nombre_consecutif	1	2.530491	1.070818	2.36	0.0181
is_fibo	1	-9.585944	0.693969	-13.81	<.0001
is_premier	1	-2.015166	0.510704	-3.95	<.0001
multi7	1	1.556011	0.671640	2.32	0.0205
intervalle_b	1	-23.472236	0.693993	-33.82	<.0001
_Sigma	1	20.112978	0.352655	57.03	<.0001

Source : logiciel SAS

Après estimation du modèle sur sas, on obtient l'équation suivante :

$$\text{Somme} = 126.805208 + 2.77690 * \text{nombre_consecutif} + 3.063948 * \text{std} + 1.547452 * \text{multi7} - 9.323992 * \text{is_fibo} - 2.055525 * \text{is_premier} - 21.822225 * \text{intervalle_b}$$

Avec $\sigma = 20.268662$

On constate que les variables comme nombre_consecutif, std, multi7 sont corrélées positivement avec la somme des numéros des boules, cela signifie que leurs augmentations entraînent une augmentation de la sommes des boules.

Par exemple, l'existence d'au moins deux nombres consécutifs dans un tirage fait augmenter la somme des numéros tirés toute choses égale par ailleurs.

En revanche, les variables comme, is_fibo, is_premier, intervalle_b sont négativement corrélées avec la somme des numéros des boules.

Tobit effets marginaux moyens :

tobit effets marginaux moyens									
Obs.	Intercept	std	nombre_consecutif	is_fibo	is_premier	multi7	intervalle_b	_Sigma	t
1	126.710	3.25168	2.53049	-9.58594	-2.01517	1.55601	-23.4722	20.1130	1

Source : logiciel SAS

5-PRESENTATION DES RESULTATS & INTERPRETATION DES EFFETS MARGINAUX

- Effet marginal et probabilités pour « std » :**

Nous avons considéré la moyenne pour les variables _std et intervalle_b et nous avons choisi pour les autres variables certaines valeurs de leurs modalités pour faire des comparaisons.

	Cas 1	Cas 2
	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif =0 Multi7 =0 Is_fibo min =0 Is_premier=0	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif =1 Multi7 =4 Is_fibo min =4 Is_premier=5
Proba 0	5.351E-11	0.000000919
Proba1	0.99999999994649	0.999999081
Effet marginal	3.25168	3.25168

On remarque que les effets marginaux restent les mêmes mais les probabilités ont légèrement changé.

L'augmentation d'une unité de la variation (écart-type) des numéros des 5 boules entraîne une augmentation de 3.25168 unités de la somme des boules.

Dans le cas 1, on a 5.351E-9% de chances que la somme des numéros d'un tirage soit inférieur ou égale 70 et presque 100% de chances que la somme des boules d'un tirage soit supérieure à 70.

Dans le cas 2, on a 0.0000919% de chances que la somme des numéros d'un tirage soit inférieur ou égale 70 et presque 100% de chances que la somme des boules d'un tirage soit supérieure à 70.

- **Effet marginal et probabilités pour « nombre_consecutif » :**

	Cas 1	Cas 2
	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif =1 Multi7 =2 Is_fibo =3 Is_premier=2	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif =0 Multi7 =4 Is_fibo =4 Is_premier=5
Proba 0	5.8481E-8	0.000002477
Proba1	0.999999941519	0.999998659
Effet marginal	2.40028	2.40028

Pour tout tirage dont la somme des boules est strictement supérieure au seuil 70, la présence de nombres consécutifs dans les numéros tirés est associée, en moyenne, à une augmentation de 2.40028 unités de somme des boules.

Dans le cas 1, on a 5.8481E-6% de chance que la somme des numéros d'un tirage soit inférieur ou égale 70 et presque 100% de chance que la somme des boules d'un tirage soit supérieure à 70 avec l'existence d'au moins deux nombres consécutifs lors d'un tirage

Dans le cas 2, on a 0.0002477% de chance que la somme des numéros d'un tirage soit inférieur ou égale 70 et presque 100% de chance que la somme des boules d'un tirage soit supérieure à 70 avec l'existence de 0 nombres consécutifs lors d'un tirage

- Effet marginal et probabilités pour « is_fibo »

	Cas 1	Cas 2
	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif=1 Multi7 =2 Is_fibo =3 Is_premier=5	Std :14.1372401 Intervalle_b :1.5790370 Nombre_consecutif=0 Multi7 =4 Is_fibo =4 Is_premier=5
Proba 0	0.000008595	0.000007992
Proba1	0.99999	0.999992008
Effet marginal	-9.32398	-9.32398

Lorsqu'une suite de Fibonacci est présente dans un tirage, la somme des numéros des boules tirées diminue de 9.32398 unités.

Dans le cas 1, on a **0.0008595 %** de chances que la somme des numéros d'un tirage soit inférieur ou égale 70 et **99.999%** de chances que la somme des boules d'un tirage soit supérieure à 70.

Dans le cas 2, on a **0.0007992%** de chances que la somme des numéros d'un tirage soit inférieur ou égale 70 et **99.99%** de chances que la somme des boules d'un tirage soit supérieure à 70.

- Effet marginal et probabilités pour « is_premier »

	Cas 1	Cas 2
	Std=14.14 Intervalle_b=1.58 Nombre_consécutif :1 Is_premier :5 Is_fibo :4 Multi7 :4	Std=14.14 Intervalle_b=1.58 Nombre_consécutif :1 Is_premier :1 Is_fibo :4 Multi7 :4
Proba 0	.000004574	.000000624
Proba 1	1.0000	1.0000
Effet marginal	-2.01516	-2.01517

➤ **cas 1 :**

Il y a 0,00046% de chances pour que la somme des boules soit inférieure ou égale à 70 et 99,95% de chances pour qu'elle soit supérieure à 70 avec 14,14 pour std, 1,58 pour intervalle_b, 1 pour le nombre_consécutif, 5 pour is_premier, 4 pour is_fibo et 4 pour multi7.

• **Effet marginal et probabilités pour « multi7 » :**

	Cas 1	Cas 2
	Std=14.14 Intervalle_b=1.58 Nombre_consécutif :1 Is_premier :4 Is_fibo :4 Multi7 :3	Std=14.14 Intervalle_b=1.58 Nombre_consécutif :1 Is_premier :4 Is_fibo :4 Multi7 :1
Proba 0	.000004101	.000008507
Proba 1	0.9999	0.99999
Effet marginal	1.55601	1.55601

➤ **cas 2 :**

Il y a 0,00085% de chances pour que la somme des boules soit inférieure ou égale à 70 et 99,99% de chances pour qu'elle soit supérieure à 70 avec 14,14 pour std, 1,58 pour intervalle_b, 1 pour le nombre_consécutif, 4 pour is_premier, 4 pour is_fibo et 1 pour multi7.

• **Effet marginal et probabilités pour « intervalle_b »**

	Cas 1	Cas 2
	Std=14.14 Intervalle_b=1.58 (moyenne) Nombre_consécutif :1 Is_premier :3 Is_fibo :1 Multi7 :2	Std=14.14 Intervalle_b=8.2 (max) Nombre_consécutif :1 Is_premier :3 Is_fibo :1 Multi7 :2
Proba 0	1.3522E-9	0.13338
Proba 1	1.00000	0.86662
Effet marginal	-23.4722	-1.62903

➤ **cas 2 :**

Il y a 13,38 %de chances pour que la somme des boules soit inférieure ou égale à 70 et 86,67 % de chances pour qu'elle soit supérieure à 70 avec 14,14 pour std, 8.2 pour intervalle_b, 1 pour le nombre_consécutif, 3 pour is_premier, 1 pour is_fibo et 2 pour multi7.

6-CONCLUSION

Notre étude a pour but d'analyser l'influence de plusieurs variables sur la somme des numéros tirés lors des tirages de l'Euromillions entre l'année 2004 et 2023. Face à la rareté des petites sommes, induisant une troncature à gauche des données, l'utilisation d'un modèle Tobit avec un seuil de troncature à 70 s'est avérée cruciale pour obtenir des estimations non biaisées.

Nos résultats montrent que plusieurs variables exercent une influence significative sur la somme des numéros tirés. L'écart-type des numéros tirés (std), la présence de nombres consécutifs (nombre_consecutif) et le nombre de multiples de 7 (multi7) sont positivement corrélées avec la somme. En d'autres termes, une plus grande dispersion des numéros, la présence de nombres consécutifs et un plus grand nombre de multiples de 7 présents dans un tirage, tendent à augmenter la somme des numéros tirés. Inversement, la présence de nombres de la suite de Fibonacci (is_fibo), le nombre de nombres premiers (is_premier) et l'amplitude de l'intervalle entre le plus petit et le plus grand numéro tiré (intervalle_b) sont négativement corrélées avec la somme. Ainsi, la présence de nombres de Fibonacci, un nombre élevé de nombres premiers et un grand écart entre les numéros extrêmes tendent à diminuer la somme.

L'analyse des effets marginaux a permis de quantifier l'impact de chaque variable sur la somme, tout en tenant compte de la troncature. Par exemple, nous avons observé qu'une augmentation d'une unité de l'écart-type entraîne une augmentation d'environ 3.25 unités de la somme, et que la présence d'une suite de Fibonacci diminue la somme d'environ 9.32 unités. L'analyse des probabilités associées aux différents scénarios (cas 1 et cas 2) confirme la faible probabilité d'observer des sommes inférieures ou égales à 70, renforçant la pertinence de l'utilisation du modèle Tobit.

Points forts de notre étude :

- **Prise en compte de la troncature :** L'utilisation du modèle Tobit a permis de corriger le biais potentiel lié à la distribution non normale des données et à la rareté des petites sommes.
- **Analyse des effets marginaux :** L'interprétation des effets marginaux a fourni une mesure concrète de l'impact de chaque variable sur la somme.
- **Variables explicatives pertinentes :** Les variables choisies, basées sur des propriétés mathématiques des nombres, se sont avérées pertinentes pour expliquer les variations de la somme.

Limites de l'étude et perspectives :

- **Seuil de troncature :** Le choix du seuil de troncature à 70, bien que justifié par l'analyse des fréquences cumulées, pourrait être affiné par des méthodes plus formelles.
- **Autres variables :** D'autres variables pourraient être envisagées, comme la présence de suites arithmétiques ou d'autres motifs numériques.

En résumé, notre étude a permis de mettre en évidence les facteurs qui influencent la somme des numéros tirés à l'Euromillions, en utilisant la méthode de Tobin.

Nos résultats peuvent intéresser ceux qui aiment jouer à la loterie. Mais il est toujours bon de rappeler que le tirage reste aléatoire et que les résultats passés ne garantissent pas ceux du futur.

7-ANNEXE

Tableau 1

Estimation OLS somme des boules				
La procédure FREQ				
somme				
somme	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
44	1	0.06	1	0.06
45	1	0.06	2	0.12
46	1	0.06	3	0.18
51	1	0.06	4	0.24
52	2	0.12	6	0.36
53	1	0.06	7	0.42
54	1	0.06	8	0.47
55	1	0.06	9	0.53
56	3	0.18	12	0.71
57	3	0.18	15	0.89
59	2	0.12	17	1.01
60	4	0.24	21	1.25
61	3	0.18	24	1.42
62	1	0.06	25	1.48
63	2	0.12	27	1.60

63	2	0.12	27	1.60
64	2	0.12	29	1.72
65	2	0.12	31	1.84
66	1	0.06	32	1.90
67	3	0.18	35	2.08
68	2	0.12	37	2.20
69	3	0.18	40	2.37
70	7	0.42	47	2.79
71	3	0.18	50	2.97
72	3	0.18	53	3.15
73	2	0.12	55	3.26
74	10	0.59	65	3.86
75	9	0.53	74	4.39
76	4	0.24	78	4.63
77	6	0.36	84	4.99
78	1	0.06	85	5.04
79	6	0.36	91	5.40

80	6	0.36	97	5.76
81	9	0.53	106	6.29
82	10	0.59	116	6.88
83	9	0.53	125	7.42
84	12	0.71	137	8.13
85	10	0.59	147	8.72
86	10	0.59	157	9.32
87	11	0.65	168	9.97
88	13	0.77	181	10.74
89	11	0.65	192	11.39
90	7	0.42	199	11.81
91	11	0.65	210	12.46
92	12	0.71	222	13.18
93	13	0.77	235	13.95
94	9	0.53	244	14.48

Tableau des moyennes :

La procédure MEANS						
Variable	Libellé	N	Moyenne	Ec-type	Minimum	Maximum
nb	nb	1685	843.0000000	486.5619180	1.0000000	1685.00
B1	B1	1685	8.4718101	6.6288575	1.0000000	38.0000000
B2	B2	1685	16.8385757	8.1733044	2.0000000	44.0000000
B3	B3	1685	25.2741840	8.6666055	5.0000000	47.0000000
B4	B4	1685	34.0421365	8.3803585	9.0000000	49.0000000
B5	B5	1685	42.5750742	6.6621807	16.0000000	50.0000000
somme	somme	1685	127.2017804	29.5845079	44.0000000	222.0000000
nombre_consecutif	nombre consecutif	1685	0.3590504	0.4798644	0	1.0000000
is_fibo	is_fibo	1685	0.8011869	0.7852552	0	4.0000000
is_premier	is_premier	1685	1.5145401	0.9859144	0	5.0000000
multi7	multi7	1685	0.7270030	0.7630190	0	4.0000000
intervalle_b	intervalle_b	1685	1.5790370	0.9051460	0.2444444	8.2000000
std	std	1685	14.1372401	3.7239892	3.5355339	23.2228336
produit	produit	1685	9410314.77	15441973.54	6300.00	165816000
Meff_std	Effet marginal de std sur somme	1685	3.1138879	0.4653175	1.0403387E-7	3.2516835
Meff_nombre_consecutif	Effet marginal de nombre_consecutif sur somme	1685	2.4232565	0.3621144	8.0960124E-8	2.5304904
Meff_is_fibo	Effet marginal de is_fibo sur somme	1685	-9.1797217	1.3717529	-9.5859424	-3.066912E-7
Meff_is_premier	Effet marginal de is_premier sur somme	1685	-1.9297692	0.2883711	-2.0151653	-6.447289E-8
Meff_multi7	Effet marginal de multi7 sur somme	1685	1.4900717	0.2226658	4.9782758E-8	1.5560103
Meff_intervalle_b	Effet marginal de intervalle_b sur somme	1685	-22.4775558	3.3588876	-23.4722318	-7.50967E-7