DÉTERMINANTS DE L'ÉDUCATION

GAINS LIÉS À L'UTILISATION D'UN ESTIMATEUR NON-PARAMÉTRIQUE DU MAXIMUM DE VRAISEMBLANCE

François Laliberté-Auger, M.Sc.

Question de recherche

- Dans un modèle d'estimation des déterminants socio-économiques du parcours scolaire, l'utilisation d'un estimateur du maximum de vraisemblance semi-paramétrique permet-elle de corriger le biais introduit par l'hétérogénéité

Modèle de transition

- On regarde l'effet des déterminants socio-économiques sur la probabilité d'effectuer une transition scolaire

$$Pr(D_s = 1 | X_s = x_s, D_{s-1} = 1) = P_{s-1,s}(x_s)$$

- Régression logistique

$$P_{s-1,s}(x_s) = \frac{exp(x_s \beta_s)}{1 + exp(x_s \beta_s)}$$

- Où s est la transition observée

Hétérogénéité inobservée

- Présence de caractéristiques observables et inobservables

$$X_s = (X_{so}, X_{si})$$

- Hypothèses sur les caractéristiques inobservées
  - Les caractéristiques observées et inobservées sont indépendantes à la première période
  - L'effet des facteurs inobservés (l'« habileté » ou la « motivation ») est constant dans le temps

Biais lié à l'hétérogénéité inobservée

- Une sélection s'effectue sur les inobservables
  - Deux groupes : les «forts» réussissent mieux que les «faibles»
    - Dans les transitions subséquentes les «forts» sont sur représentés. Nous n'avons plus une population représentative
  - Il se créé une corrélation entre les caractéristiques observables et inobservables
    - Seuls ceux avec des caractéristiques observables élevées réussissent s'il sont dans le groupe «faible»
    - Dans le groupe des «forts», même ceux avec des caractéristiques observables basses peuvent réussir
    - Les enfants de familles aisées réussissent même s'ils ne sont pas doués pour l'école et les enfants doués réussissent peu importe leur milieu socio-économique

Estimateur semi-paramétrique

- Un modèle logistique

$$L = \prod_{t=1}^{T} F(x_i\beta)^{y_t}(1 - F(x_i\beta))^{1-y_t}$$

- Nous utilisons un estimateur décrit par Heckman et Cameron (1998)

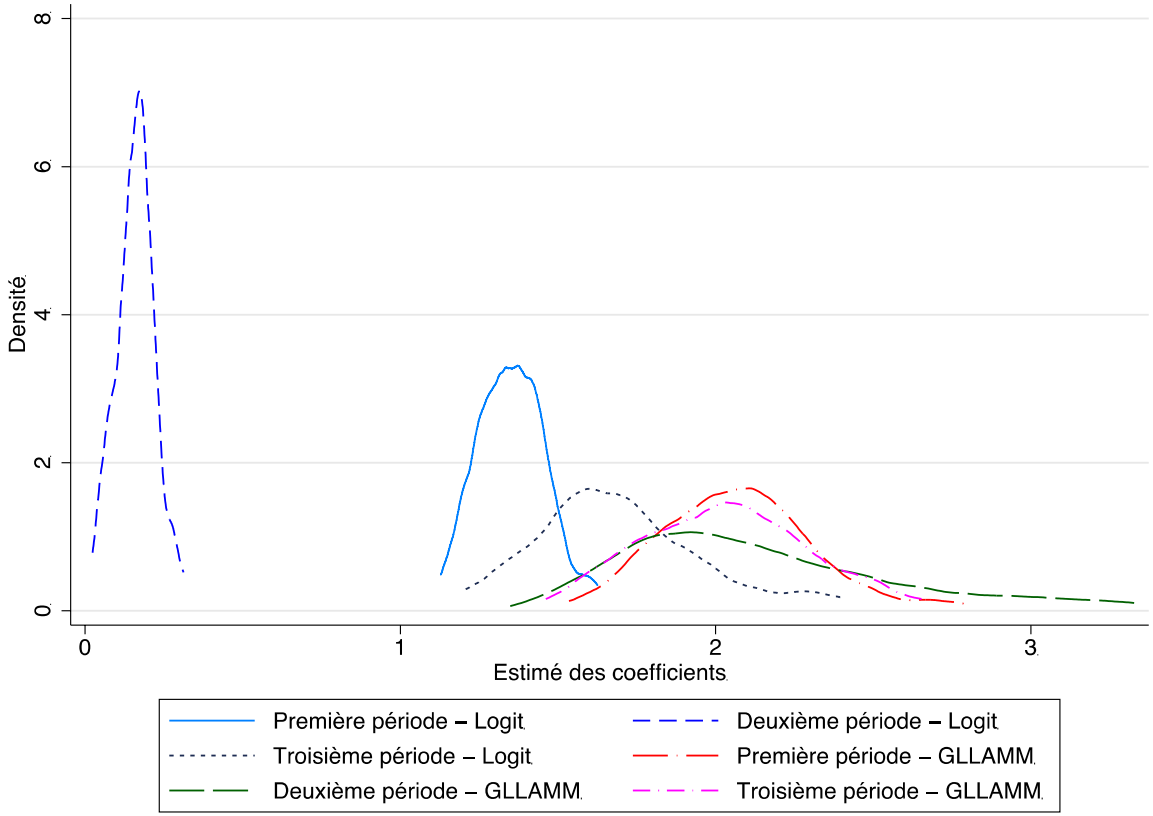
$$L = \sum_{c=1}^{C} \left\{ \prod_{t=1}^{T} F(x_i\beta + \alpha_t f_c)^{y_t}(1 - F(x_i\beta + \alpha_t f_c))^{1-y_t} \right\} g_c$$

Estimateur semi-paramétrique

- Un logit augmenté d'un terme aléatoire qui suit une distribution inconnue
  - L'appartenance à un groupe est fixe dans le temps (f_c)
  - L'effet de l'appartenance à un groupe varie en fonction de la transition (α_t)
  - La distribution peut être représentée par des points (les groupes)
  - On ne connaît pas le nombre de groupe qui représente le mieux la distribution(C) n'y le quelle est la distribution entre les groupe
- GLLAMM – commande Stata pour estimer le NPMLE

Amélioration - simulation

- Nous simulons des données où la vraie valeur de β est 2 et nous mettons deux groupes inobservés



Enquête auprès des jeunes en transition

- Enquête longitudinale, premier cycle en 1999, 6 cycles, terminée en 2008-2009
- Deux cohortes
  - Née en 1979-1981 (Cohorte B) et 1984 (Cohorte A)
  - 29 700 jeunes de la cohorte A et 23 000 de la cohorte B
- 80% de réponse par cycle – il restait 11 011 personnes au dernier cycle
- Un questionnaire à été passé aux parents de la cohorte A au premier cycle
- Les jeunes de la cohorte A ont répondu à l'enquête PISA

Déterminants socio-économiques du parcours scolaire – choix des variables

- Nous estimons l'effet des variables à chaque transition, ce qui multiplie les coefficients à estimer
- Donc nécessaire de choisir peu de variables
  - Niveau d'étude des parents, salaire, famille mono-parentale, sexe, statut de minorité visible, langue maternelle et lieu de résidence à 15 ans (rural ou urbain)
- Seul le test de lecture PISA est utilisé
- Toutes les variables sont fixes dans le temps

Résultats - 1

	Transitions					
	Obtention d'un diplôme d'études secondaires (N = 2084)		Inscription au niveau collégial (N = 1835)		Inscription au niveau universitaire (N = 1437)	
	Logit	NPMLE	Logit	NPMLE	Logit	NPMLE
Minorité visible	0,025	0,008	0,039	0,040	0,142**	0,101
	(0,048)	(0,068)	(0,059)	(0,059)	(0,063)	(0,063)
Revenu familial	0,00064***	0,00059**	0,00019	0,00017	0,00090***	0,00092***
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Femme	0,042**	0,045***	0,030	0,027	0,106***	0,109***
	(0,017)	(0,016)	(0,022)	(0,022)	(0,026)	(0,026)
Niveau d'études des parents[1]						
Pas de diplôme - secondaire	-0,072***	-0,066***	-0,071*	-0,063	-0,069	-0,092
	(0,022)	(0,022)	(0,038)	(0,040)	(0,055)	(0,060)
Diplôme - secondaire (réf.)	-	-	-	-	-	-
Diplôme - collégial	0,063***	0,068***	0,023	0,018	0,045	0,055*
	(0,021)	(0,019)	(0,027)	(0,029)	(0,033)	(0,033)
Diplôme - universitaire	0,081***	0,069***	0,059**	0,056*	0,223***	0,214***
	(0,025)	(0,026)	(0,028)	(0,028)	(0,033)	(0,032)
Score au test PISA	0,160***	0,164***	0,114***	0,102***	0,161***	0,177***
	(0,009)	(0,009)	(0,013)	(0,022)	(0,016)	(0,018)
Langues marternelles						
Autres langues	0,119***	0,106***	0,075**	0,070**	0,044	0,048
	(0,035)	(0,037)	(0,034)	(0,034)	(0,035)	(0,035)
Anglais	0,101**	0,116**	0,100*	0,091	0,061	0,094
	(0,043)	(0,057)	(0,058)	(0,060)	(0,064)	(0,063)
Français (réf.)	-	-	-	-	-	-
Rural	0,051**	0,036*	-0,001	-0,003	-0,031	-0,025
	(0,020)	(0,019)	(0,025)	(0,024)	(0,032)	(0,031)

Résultats - 2

Tableau 5.7 Résultats de la régression par NPMLE - Spécification 6

	Transitions									
	Obtention d'un diplôme d'études secondaires (N = 1656)		**Inscription au niveau collégial** (N = 1585)		**Obtention d'un diplôme de niveau collégial** (N = 1191)		**Inscription au niveau universitaire** (N = 883)		**Obtention d'un diplôme universitaire** (N = 663)	
Variables	Logit	NPMLE	Logit	NPMLE	Logit	NPMLE	Logit	NPMLE	Logit	NPMLE
Minorité visible	0,002	0,003	0,122*	0,113	0,098	0,066	0,098	0,098	-0,046	-0,058
	(0,046)	(0,063)	(0,072)	(0,084)	(0,075)	(0,099)	(0,081)	(0,083)	(0,078)	(0,119)
Revenu familial	-0,00002	0,00024	0,00031	0,00021	0,00001	0,00002	0,00086*	0,00084*	0,00025	0,00020
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Femme	0,022	0,028*	0,055**	0,043	0,121***	0,114***	0,069**	0,056	0,094**	0,107
	(0,017)	(0,016)	(0,024)	(0,040)	(0,029)	(0,034)	(0,033)	(0,045)	(0,041)	(0,134)
Niveau d'études des parents[1]										
Pas de diplôme - secondaire	-0,054***	-0,047	-0,086**	-0,061	-0,083	-0,093	-0,109	-0,100	-0,044	-0,045
	(0,021)	(0,033)	(0,041)	(0,051)	(0,060)	(0,062)	(0,068)	(0,075)	(0,113)	(0,113)
Diplôme - secondaire (réf.)	-	-	-	-	-	-	-	-	-	-
Diplôme - collégial	0,021	0,023	0,041	0,030	-0,033	-0,029	-0,043	-0,040	0,058	0,055
	(0,020)	(0,017)	(0,030)	(0,046)	(0,038)	(0,038)	(0,040)	(0,043)	(0,059)	(0,068)
Diplôme - universitaire	0,108***	0,081**	0,102***	0,078	0,084**	0,081**	0,156***	0,147***	0,095*	0,092
	(0,035)	(0,032)	(0,031)	(0,063)	(0,039)	(0,040)	(0,043)	(0,053)	(0,053)	(0,081)
Score au test PISA	0,063***	0,064***	0,146***	0,123**	0,098***	0,084*	0,148***	0,144***	0,110***	0,097**
	(0,009)	(0,010)	(0,013)	(0,051)	(0,019)	(0,045)	(0,020)	(0,029)	(0,026)	(0,041)
Langues marternelles										
Autres langues	0,045	0,048**	0,071*	0,055	-0,024	-0,025	0,013	0,019	-0,015	-0,028
	(0,032)	(0,022)	(0,040)	(0,056)	(0,042)	(0,043)	(0,048)	(0,051)	(0,059)	(0,112)
Anglais	0,010	0,023	0,026	0,016	0,035	0,037	0,092	0,085	0,036	0,039
	(0,040)	(0,067)	(0,069)	(0,090)	(0,079)	(0,093)	(0,086)	(0,087)	(0,090)	(0,146)
Français (réf.)	-	-	-	-	-	-	-	-	-	-
Rural	0,052**	0,032	0,005	-0,003	0,043	0,049	-0,091**	-0,098*	0,011	0,032
	(0,020)	(0,034)	(0,027)	(0,028)	(0,035)	(0,036)	(0,038)	(0,051)	(0,053)	(0,078)

Les difficultés du NPMLE

- Résultats très semblables au logit
  - Peu d'hétérogénéité inobservée?
  - Maximum local ou maximum global?
  - Heckman et Cameron trouvent des résultats différents
- Difficulté à estimer avec les outils actuels
  - Utilisation d'une commande fournie par les utilisateurs
  - La convergence s'effectue difficilement
  - Temps de calcul un peu long
  - Techniquement, on devrait calculer des écart-types à l'aide de bootstrap – beaucoup trop long
- Incertitude sur les résultats

Conclusion

- Les questions d'hétérogénéité sont très importantes
- Cette méthode présente des difficultés d'application
- Avenue intéressante pour une personne intéressée par l'économétrie et ayant un fort baguage en informatique
- Possibilité que les logiciels statistiques intègrent des commandes permettant l'estimation d'un tel estimateur plus facilement
- Possibilité d'utiliser d'autres méthodes de contrôle pour l'hétérogénéité