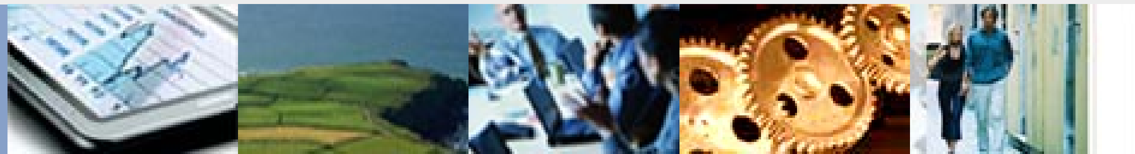
Pseudo maximum de vraisemblance et panels qualitatifs

Stéfan LOLLIVIER
Insee



JMS 17 mars 2009

Introduction

- De plus en plus de panels :
  - Effets dynamiques, causalités temporelles
  - Hétérogénéité inobservé
- Des variables le plus souvent qualitatives
  - Expliquées et explicatives
  - Avec endogénéité
- Les méthodes les plus simples sont parfois les plus robustes…
- … et les plus souvent généralisables à des systèmes multivariés ….
- … au prix d'une analyse théorique plus complexe.

Les m-estimateurs : maximisation d'un critère

- La vraisemblance
  - Le plus naturel
  - Mais fragile
- Les moments
  - Ou, comment trouver des instruments
- Le pseudo-maximum de vraisemblance
  - Comment remplacer une vraisemblance compliquée mais exacte par une vraisemblance simple mais approchée
  - On pratique parfois le PMV sans le savoir

Faire simple dans l’estimation

- Moindres carrés ordinaires plutôt que moindres carrés généralisés
  - Méthode plus robuste aux erreurs de spécification
  - En fait, une application du PMV
- Avec un panel qualitatif, probit simple empilé plutôt que simulation d’une vraisemblance complexe
  - Estimateur robuste à la spécification de la matrice de variance du terme d’erreur intertemporel (quitte à perdre en efficacité)
  - En fait utiliser un PMV
- Utiliser la méthode d’Heckman en deux étapes plutôt que l’estimation du modèle bivarié par le MV
  - Evite la spécification de la loi du terme d’erreur de l’équation quantitative

Recourir au bootstrap pour calculer la variance asymptotique

- Profiter de la propriété IID des observations pour les individus
- Procéder à un tirage avec remise de même taille que celui de l'échantillon
- Procéder à l'estimation
- La moyenne et la variance empirique des estimateurs bootstrap converge vers la vraie valeur
- Avec un panel, l'élément IID est l'individu (avec toutes ses dates) et pas l'observation individu*date

Modèles bivariés sur panel
Estimation en plusieurs étapes

- Equation instrumentale (ou de sélection selon le cas) : l’estimateur probit empilé est convergent
- On peut en tirer des résidus simulés des termes d’erreur conditionnels à l’observation (tirages dans une loi normale tronquée)
- Equation d’intérêt, sous la forme :

$$y_{it1}^{*} = x_{it1}\beta_1 + \alpha y_{it2} + \rho \varepsilon_{it2} + \sqrt{1-\rho^2}\, w_{it}$$

- Conditionnellement au terme d’erreur de l’équation 1 :

$$E(y_{it1} \mid x_i, y_{i2}; \varepsilon_{it2}) = \Phi\left(\frac{x_{it1}\beta_1 + \alpha y_{it2} + \rho \varepsilon_{it2}}{\sqrt{1-\rho^2}} \right)$$

Modèles bivariés sur panel
Estimation en plusieurs étapes

- non conditionnellement au terme d'erreur de l'équation 1 :

$$E(y_{it1} | x_i, y_{i2}) = \underset{\varepsilon_{it2}}{E}\left[\Phi\left(\frac{x_{it1}\beta_1 + \alpha y_{it2} + \rho\varepsilon_{it2}}{\sqrt{1-\rho^2}}\right)\right]$$

- Expression que l'on peut approximer par un simulateur

$$E(y_{it1} | x_i, y_{i2}) \approx \frac{1}{H}\sum_{h=1}^{H}\left[\Phi\left(\frac{x_{it1}\beta_1 + \alpha y_{it2} + \rho\varepsilon_{it2h}}{\sqrt{1-\rho^2}}\right)\right]$$

- Afin d'utiliser la méthode du pseudo-maximum de vraisemblance simulé, en mettant en regard l'observé et son espérance

Modèles bivariés sur panel
Estimation en plusieurs étapes

- Maximisation de la PV, analogue à celle d'un modèle binomial

$$L=\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it1}\log[E(y_{it1}|x_i,y_{it2})]+(1-y_{it1})\log[1-E(y_{it1}|x_i,y_{it2})]$$

- En recourant au moment d'ordre 1 simulé
- Estimateurs convergents pour un nombre de simulations suffisamment grand.
- Robustesse : pas d'hypothèse sur la matrice de la variance de la loi jointe des termes d'erreur (seulement la loi marginale en t)
- Calcul de la matrice de variance en recourant à la méthode du bootstrap par paires

Extensions de la méthode

- Equation de sélection polytomique ordonnée
  - On reste dans le champ des modèles bivariés
- Equation d'intérêt polytomique ordonnée
  - On reste encore dans le champ bivarié
  - La PV renvoie à un modèle multinomial
- Equation de sélection polytomique non ordonné
  - Difficulté à simuler les termes d'erreur du modèle de sélection car loi marginale multivariée
  - Utilisation possible de l'échantillonnage de Gibbs

Conclusion

- La méthode du PMV est triviale à utiliser dans le cas univarié
  - Elle est en outre robuste aux variantes de spécification
- La méthode du PMV simulé permet de traiter les modèles avec sélection ou endogénéité alors que le résultat est pratiquement hors de portée avec le MV
- On s'est limité ici aux modèles statiques, mais la méthode devrait aussi présenter un intérêt dans le cas des modèle dynamiques (avec T petit et endogénéité de la situation initiale).