

Econométrie des données de panel

Guillaume Horny*

*Banque de France

Master 2 MASERATI

Conclusion

Plan

Plan

Petite histoire des pratiques

Historiquement, les modèles à effets fixes ont été les premiers à voir été estimés. Les méthodes pour les modèles à effets aléatoires sont venues ensuite. Elles ont tout d'abord suscités un fort engouement, avant que les travaux empiriques ne reviennent aux estimateurs *within* et en différence première. Bien que moins précis, ceux-ci restent convergent en présence de corrélation entre les caractéristiques inobservables et les variables explicatives. De plus, les gains de précision associés aux estimateurs FGLS sont souvent considérés comme minimes. L'arbitrage coût-bénéfice penche donc ici souvent en faveur des méthodes les plus simples.

La pratique de l'économétrie des panels aujourd'hui

La littérature empirique contemporaine accorde beaucoup d'importance à l'évaluation des effets "causaux" d'une variable X_k sur la variable dépendante Y . Un grand nombre de projets de recherche utilisent des estimateurs *within*, et portent un soin tout particulier aux question d'endogénéité de X_k (utilisation de variables instrumentales), ainsi qu'à la mesure des écarts-types des coefficients (estimations successives avec multiples corrections pour différentes structures d'hétéroscédasticité et d'autocorrélation des erreurs).

Quelle utilité pour les FGLS ?

Dans le cadre d'une utilisation non-académique, l'arbitrage est moins clair. *In fine*, tout va dépendre du domaine d'application, de la question posée, ainsi que de la richesse des données disponibles.

Très grossièrement, on peut dire que plus le phénomène étudié est maîtrisé (ex : résultat d'un protocole médical, d'un processus mécanique) et plus on en dispose d'une modélisation précise, plus il y a de chance qu'un modèle à effet aléatoire soit convaincant. A l'inverse, plus les comportements sous-jacents nous échappent et plus les variables explicatives potentielles sont rares, plus les inobservables vont jouer un rôle important et plus il est probable qu'elles soient corrélées avec les (quelques) variables explicatives.

Plan

Organisation d'une démarche typique (1/)

- Si vous avez la possibilité de choisir les questions sur lesquelles vous allez devoir travailler, **réfléchissez bien au type de données qu'il vous faudrait pour pouvoir répondre à votre question.**
On a besoin de suivre des individus pour étudier l'évolution de leurs comportements, d'observer les offreurs et les demandeurs pour étudier l'équilibre d'un marché, le temps que passent les individus dans un état pour étudier la probabilité qu'ils en sortent...
- **Connaissez vos données avant de faire quoi que ce soit !**
Comment ont-elles été collectées ? qui les a déclarées ? combien de temps après l'événement qui vous intéresse ? à quel point sont-elles fiables ? quelle est la période couverte ? la fréquence d'observation ? si vous avez un panel, la variance est-elle plutôt au niveau intra- ou inter- individuel ?

Organisation d'une démarche typique (2/)

- **Si vous avez des données de panel, nettoyez les scrupuleusement.** Chercher les observations aberrantes, la forme de l'attrition... **Eviter de cylindrer !** Vous perdrez des observations et restreignez l'échantillon aux individus qui répondent tout le temps (les plus vieux, qui restent tout le temps dans l'état étudié...), alors qu'ils sont généralement loin d'être représentatifs.
- Avant d'écrire votre modèle, revenez à la question de départ. Quel paramètre du modèle permet d'y répondre ? Quelle prévision ? Quelle distribution ? Qui sont les individus concernés ?

Organisation d'une démarche typique (3/)

- Quelle est forme de **l'hétérogénéité inobservée** ? Entre individus ? entre groupes d'individus ? Est-ce qu'elle affecte le niveau de Y ? sa variance ? les deux ? Sa mesure a-t'elle une importance ?
- Dans le cas général, restez prudents et partez du principe que les **inobservables sont corrélés avec les observables**. Utilisez les estimateurs OLS, *within*, du modèle en différence, *between*, FGLS.
- Les résultats sont-ils crédibles ? Comparer les écarts entre les estimations et essayer de les expliquer. D'où vient le problème expliquant les écarts ? Faites des tests d'Hausman, comparez les R^2 , regardez les profils des erreurs...

Organisation d'une démarche typique (4/)

- Les **tests d'effets individuels** conduisent-ils à les garder ?
- Soyez attentifs aux problèmes **d'hétéroscédasticité et d'autocorrélation** des erreurs
- les résultats sont-ils **robustes** à l'introduction de nouvelles variables ?
Au changement du mode de calcul de la variable dépendante ou d'une des explicatives ?

Organisation d'une démarche typique (5/)

- C'est à ce moment seulement que vous pouvez privilégier un estimateur plutôt qu'un autre ! Effectuer votre classement sur la base de **tous** les critères précédents. Quand vos résultats sont obtenus, n'oubliez pas que vous traitez un problème économique qui doit avoir une réponse en termes **économiques**, et pas seulement en terme de coefficients estimés ou de t de Student.
- Si votre problème est de type **causal**, réfléchissez attentivement aux hypothèses d'exogénéité et à leur bien-fondé. Si elles ne tiennent pas pour une variable, comment traiter ce problème **d'endogénéité** ? Est-ce qu'on disposerait d'instruments ? Seraient-ils valides ? seraient-ils faibles ? Auraient-ils un sens économique ? Si oui, tester-les.

Bonne chance !