

Econométrie des données de panel

Guillaume Horny*

*Banque de France

Master 2 MASERATI

Conclusion

Petite histoire des pratiques

- Historiquement, les modèles à effets fixes ont été les premiers à avoir été estimés. Les méthodes pour les modèles à effets aléatoires sont venues ensuite.
- Les modèles à effet aléatoire ont d'abord suscité un fort engouement, avant que les travaux empiriques ne reviennent au modèle à effet fixe.
- Bien que moins précis, les estimateurs du modèle à effet fixe restent convergents 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, dans la pratique de l'économétrie des panels, souvent en faveur des méthodes les plus simples.

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 :

- plus le phénomène étudié est maîtrisé (ex : résultat d'un protocole médical, d'un processus mécanique), plus on dispose d'un modèle détaillé, plus il y a de chance qu'un modèle à effet aléatoire soit convaincant.
- A l'inverse, plus les comportements sous-jacents nous échappent ou 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.

Organisation d'une démarche typique (1/6)

- 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

Organisation d'une démarche typique (2/6)

- **Connaissez vos données avant de faire quoi que ce soit !**
 - ▶ Comment ont-elles été collectées (données exhaustives ? enquête ? déclaration obligatoire ou volontaire ? ...) ?
 - ▶ combien de temps après l'événement qui vous intéresse ?
 - ▶ quelle est la période couverte ? la fréquence d'observation ?
 - ▶ si vous avez un panel, est-il cylindré ? les variables varient-elles dans le temps ou bien entre individus ?

Organisation d'une démarche typique (3/6)

- **Si vous avez des données de panel, nettoyez les scrupuleusement.**
 - ▶ Chercher les observations aberrantes, en utilisant plusieurs critères
 - ▶ étudiez la forme de l'attrition mais **évitez 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 (4/6)

- Quelle est la structure 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 ?
- Utilisez les estimateurs OLS, *within*, du modèle en différence, FGLS.
- Comparer les écarts entre les estimations et essayer de les expliquer. Quels résultats sont crédibles et lesquels ne le sont pas ? Comment s'expliquent les écarts ?
- Faites des tests d'Hausman, comparez les R^2 , regardez les profils des erreurs...

Organisation d'une démarche typique (5/6)

- 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 ?
- Les **tests d'effets individuels** conduisent-ils à les garder ?
- Soyez attentifs aux problèmes **d'hétéroscédasticité et d'autocorrélation** des erreurs

Organisation d'une démarche typique (6/6)

- 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 termes de coefficients estimés ou de t de Student.

Bonne chance !