Section 04 : Les modèles panels (Séance 8)

GSF-6053: Économétrie Financière

Simon-Pierre Boucher[1]

[1]Département de finance, assurance et immobilier
Faculté des sciences de l'administration
Université Laval

15 Mars 2022

Références

Obligatoires:

- **Notes de cours:** Section 2 (Professeure: Marie-Hélène Gagnon)
- **Woolridge:** chapitres 13 et 14.

Complémentaires:

- **Gujarati et Porter:** chapitre 16

Plan de la séance

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Les modèles panels

- Il est possible de faire des estimations multivariées où Y est une matrice $T \times n$.
- Par exemple, nous pourrions vouloir estimer le modèle du CAPM sur plusieurs actifs à la fois.
- Au lieu de faire une série de régressions séparées sur plusieurs actifs financiers, nous formerons un système d'équations multivariées.
- Cette approche permet plusieurs innovations dans la modélisation qui peuvent être intéressantes :
  - Comparaison statistique de coefficients entre équations d'un même système.
  - Test conjoint des implications du modèle (telle que l'efficacité moyenne- variance) sur plusieurs portefeuilles ou actifs financiers.
  - Permets une plus grande flexibilité au niveau de la modélisation du terme d'erreur des équations du système.

Les modèles panels

Le CAPM univarié

$$R_t - R_{Ft} = \alpha + \beta(R_{Mt} - R_{Ft}) + U_t$$

Le CAPM multivarié

$$R_{\mathbf{j}t} - R_{Ft} = \alpha_{\mathbf{j}} + \beta_{\mathbf{j}}(R_{Mt} - R_{Ft}) + U_{\mathbf{j}t}$$

- $R_{\mathbf{j}t} - R_{Ft}$ représente le rendement excédentaire au temps t de l'actif j par rapport à un titre sans risque.
- On cherche a expliquer le rendement excédentaire du titre j avec la prime de risque du marché $R_{Mt} - R_{Ft}$.
- $\alpha_{\mathbf{j}}$ est une constante et $\beta_{\mathbf{j}}$ mesure la sensibilité du rendement excédentaire de l'actif j au rendement excédentaire du marché.

Les modèles panels

Le CAPM multivarié

- Selon le CAPM, les interceptes d'un modèle en rendements excédentaires devraient être nulles (hypothèse d'efficience moyenne variance).
  - $\alpha_\mathbf{j} = 0$
- $t = 1, \cdots, T$ représente l'indicateur temporelle. Pour chaque actif, nous aurons T rendement, allant de $t = 1$ jusqu'à $t = T$
- $j = 1, \cdots, N$ représente l'indicateur du titre ou portefeuille. Dans notre modèle univarié, nous aurons N titres ou portefeuilles.
- Nous n'avons pas encore spécifié la forme du terme d'erreur $u_\mathbf{jt}$

Les modèles panels

- Typiquement dans le cadre des données panel nous aurons deux éléments:
  - Plusieurs individus ou portefeuilles ou groupe (N)
  - Quelques observations temporelles (T)
- Un exemple de données panels serait des données sur
  - l'investissement (I)
  - la valeur de la firme (F)
  - la capital (C)
  - pour 20 années
  - Données sur les investissements de quatre entreprises **(Table 1.2 dans Gujarati et Porter.)**

Les modèles panels

Données sur les investissements de quatre entreprises

TABLE 1.2 Investment Data for Four Companies, 1935–1954

Observation	I	F_{-1}	C_{-1}	Observation	I	F_{-1}	C_{-1}
GE				US			
1935	33.1	1170.6	97.8	1935	209.9	1362.4	53.8
1936	45.0	2015.8	104.4	1936	355.3	1807.1	50.5
1937	77.2	2803.3	118.0	1937	469.9	2673.3	118.1
1938	44.6	2039.7	156.2	1938	262.3	1801.9	260.2
1939	48.1	2256.2	172.6	1939	230.4	1957.3	312.7
1940	74.4	2132.2	186.6	1940	361.6	2202.9	254.2
1941	113.0	1834.1	220.9	1941	472.8	2380.5	261.4
1942	91.9	1588.0	287.8	1942	445.6	2168.6	298.7
1943	61.3	1749.4	319.9	1943	361.6	1985.1	301.8
1944	56.8	1687.2	321.3	1944	288.2	1813.9	279.1
1945	93.6	2007.7	319.6	1945	258.7	1850.2	213.8
1946	159.9	2208.3	346.0	1946	420.3	2067.7	232.6
1947	147.2	1656.7	456.4	1947	420.5	1796.7	264.8
1948	146.3	1604.4	543.4	1948	494.5	1625.8	306.9
1949	98.3	1431.8	618.3	1949	405.1	1667.0	351.1
1950	93.5	1610.5	647.4	1950	418.8	1677.4	357.8
1951	135.2	1819.4	671.3	1951	588.2	2289.5	341.1
1952	157.3	2079.7	726.1	1952	645.2	2159.4	444.2
1953	179.5	2371.6	800.3	1953	641.0	2031.3	623.6
1954	189.6	2759.9	888.9	1954	459.3	2115.5	669.7
GM				WEST			
1935	317.6	3078.5	2.8	1935	12.93	191.5	1.8
1936	391.8	4661.7	52.6	1936	25.90	516.0	0.8
1937	410.6	5387.1	156.9	1937	35.05	729.0	7.4
1938	257.7	2792.2	209.2	1938	22.89	560.4	18.1
1939	330.8	4313.2	203.4	1939	18.84	519.9	23.5
1940	461.2	4643.9	207.2	1940	28.57	628.5	26.5
1941	512.0	4551.2	255.2	1941	48.51	537.1	36.2
1942	448.0	3244.1	303.7	1942	43.34	561.2	60.8
1943	499.6	4053.7	264.1	1943	37.02	617.2	84.4
1944	547.5	4379.3	201.6	1944	37.81	626.7	91.2
1945	561.2	4840.9	265.0	1945	39.27	737.2	92.4
1946	688.1	4900.0	402.2	1946	53.46	760.5	86.0
1947	568.9	3526.5	761.5	1947	55.56	581.4	111.1
1948	529.2	3245.7	922.4	1948	49.56	662.3	130.6
1949	555.1	3700.2	1020.1	1949	32.04	583.8	141.8
1950	642.9	3755.6	1099.0	1950	32.24	635.2	136.7
1951	755.9	4833.0	1207.7	1951	54.38	732.8	129.7
1952	891.2	4924.9	1430.5	1952	71.78	864.1	145.5
1953	1304.4	6241.7	1777.3	1953	90.08	1193.5	174.8
1954	1486.7	5593.6	2226.3	1954	68.60	1188.9	213.5

Les avantages des données panel

- Permet de tenir compte de l'hétérogénéité entre les groupes (ou portefeuilles, etc.) en définissant des variables **sujet-spécifiques**.
- Les variables sont plus riches et informatives puisqu'elles contiennent plus de variabilité, possiblement moins de colinéarité et permettent plus de degrés de liberté dans les estimations ($T \times N$)

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Les avantages des données panel

- Certains sujets sont plutôt conçus pour les données panels :
  - Étude de l'emploi en économie ou des habitudes de consommation d'électricité par exemple.
  - En finance, plusieurs sujets en finance corporative et comptabilité (où les données sont souvent en fréquence annuelle) permettent d'évaluer la valeur de la firme en fonction de données comptables.
- Donne une dimension supplémentaire aux coupes transversales ou aux séries chronologiques.
- Permet d'agréger de l'information diffuse et de minimiser les biais dus à des agrégations trompeuses (en permettant l'hétérogénéité).

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Notation générale pour ces modèles panels

$$Y_{it} = X_{it}^{'}\beta + Z_i^{'}\alpha + u_{it}$$

$$Y_{it} = X_{it}^{'}\beta + c_i + u_{it}$$

- Où les régresseurs dans X n'incluent pas de constante.
- L'hétérogénéité vient de $Z_i^{'}\alpha$ où $Z_i^{'}$ contient un terme constant et une série de variables spécifiques au groupe (ou l'individu) i et constante dans le temps.
  - Celles-ci pouvant être soient observées ou non observées.
  - La formulation de $Z_i^{'}$ change le modèle adopté.

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Exemple de données panels

- On analyse des données sur les coûts de six compagnies aériennes pour la période 1970-1984, pour un total de 90 observations de données de panel.
  - I = identifiant de la compagnie aérienne (1 à 6 car 6 firmes)
  - T = l'année (15 années de données)
  - Q = Output (en mile-passagers)
  - C = cout
  - PF = prix de l'essence
  - LF = l'utilisation moyenne des avions, load factor
- Nous avons ici un panel balancé, car le nombre d'observations pour chaque firme est le même.
- Nous pourrions ajuster l'approche pour des panels non balancés.
- C'est aussi ce qu'on appelle un panel long, puisque T est plus grand que n.
- Un panel court serait un panel où n est plus grand que T.

Exemple de données panels

- Nous voulons trouver comment la fonction de coût des compagnies aériennes évolue en fonction de Q, PF, LF.
- La première approche serait de trouver par OLS 6 fonctions de coût une pour chaque compagnie.
  - Cela néglige de l'information, car celle-ci doit être semblable dans chaque compagnie.
  - Nous aurons peu d'observations pour chaque firme (15) ou alors très peu d'observation par coupe transversale (6) par année. Dans les deux cas, soit on
  - Ignore la dynamique dans le temps (coupe transversale) soit on ignore l'industrie (série chronologique).

Exemple de données panels

4 options se proposent à nous:

1-Modèle pooled

- Mettre toutes les observations ensemble et faire une seule estimation OLS sur $T \times n = 90$ observations sans s'occuper de la nature des données.

2-Modèle LSDV

- Mettre toutes les observations ensemble $T \times n = 90$ observations, mais on permet de l'hétérogénéité dans la fonction de cout en fonction de la compagnie aérienne.
- Nous permettrons à chaque compagnie aérienne d'avoir un intercepte propre à lui même.

Exemple de données panels

4 options se proposent à nous:

3-Modèle à effet fixe

- On exprime chaque variable en déviation par rapport à sa moyenne de **groupe** (ici la compagnie aérienne).
- On explique donc les variations dans la fonction de cout à l'intérieur d'un groupe **(within group effect)**

4-Modèle à effet aléatoire

- Où les effets de groupe ne sont pas corrélés avec les régresseurs.
- L'impact du groupe est donc à modéliser dans le terme d'erreur.

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Le Modèle pooled

- Si Z_i' contient seulement un terme constant (le même pour tous les groupes), alors les moindres carrés ordinaires nous donneront des estimateurs convergents de α et α.
- L'hypothèse de base de ce cas est qu'il n'y a pas d'hétérogénéité dans les groupes i et il est pertinent d'avoir un estimateur conjoint.

Le Modèle pooled

- Le modèle aura la forme suivante:

$$Y_{it} = \alpha + \beta' X_{it} + u_{it}$$

  - Les erreurs sont indépendantes, d'espérance nulle et de variance σ_u^2.
  - Les régresseurs u sont toujours strictement exogènes et non corrélés avec le terme d'erreur.
  - C'est le modèle qui représente la moyenne de la population, étant donné ces coefficients contraints.
  - Les estimateurs MCO sont convergents.

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Le modèle avec effets fixes

- Soit le modèle de base avec les caractéristiques inobservables de chaque groupe i:

$$Y_{it} = \beta' X_{it} + c_i + u_{it}$$

- On fait l'hypothèse que les c_i sont correlés avec les régresseurs.
- Dans ce cas, nous savons que les estimateurs OLS seront biaisés.

Le modèle avec effets fixes

- On règle cette question en ajoutant une constante propre à chaque groupe qui représente tous les effets observables du groupe i.

$$Y_{it} = \beta' X_{it} + \alpha_i + u_{it}$$

- Si la variance conditionnelle au régresseur des c_i est constante, ce modèle devient une régression linéaire classique.
- On appelle cette formulation le modèle avec effet fixe puisqu'il implique que la différence entre les groupes peut être capturée par la différence dans le terme constant.

Les modèles panels

Les avantages des données panel

Notation générale pour ces modèles panels

Exemple de données panels

Le Modèle pooled

Le modèle avec effets fixes

Least square dummy variable model

Least square dummy variable model

- Si le nombre de groupe n'est pas trop grand par rapport au nombre d'observations (si vous avez un panel très long).
- Le modèle peut être estimé par moindres carrés avec le LSDV, Least square dummy variable model où les α_i sont libres.
- L'idée est des générer des variables dichotomiques D_i qui prennent la valeur 1 quand nous sommes dans le groupe i et zéro sinon.

Least square dummy variable model

Modèle LSDV:

$$y = X\beta + D\alpha + u$$

- y est un vecteur $(NT \times 1)$
- X est une matrice $(NT \times k)$
- D est une matrice $(NT \times N)$
- On empile les Y dans un grand vecteur y.
- On empile également les régresseurs un par dessus les autres dans le même ordre et on fait de même avec les variables dichotomiques.

Least square dummy variable model

Variable dichotomiques:

$$D = \begin{bmatrix} d_1 & d_2 & \cdots & d_N \end{bmatrix}$$

- Sachant que d_1, d_2 et d_N sont tous des vecteurs $NT \times 1$
- Les y sont définis selon le modèle empilé et d_i est un vecteur composé de zéros partout sauf pour indiquer le groupe i.

Vecteur α: $(N \times 1)$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}$$

Least square dummy variable model

Matrice X**:** $(NT \times K)$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Vecteur u**:** $(NT \times 1)$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

Least square dummy variable model

- On estime ce modèle par OLS sur toutes les observations (NT).
- Pour trouver un estimateur de la variance, on obtient le vecteur de dimension ($NT \times 1$) des résidus qu'on notera $\hat{u}$.
- L'estimé de la variance pour le terme d'erreur (bruit blanc) est donc :

$$\hat{\sigma}_u^2 = \frac{\hat{u}'\,\hat{u}}{NT - N - K}$$

Least square dummy variable model

Dummy variable trap

- Vous ne pouvez pas inclure une dummy pour chaque groupe si vous voulez avoir une constante au modèle.
- Si vous voulez une constante dans le modèle, il faut laisser tomber une des dummy et l'utiliser comme point de référence.
- Supposons que le groupe 1 est votre groupe de référence et que vous incluez une constante.
  - La constante du modèle est l'effet du groupe 1
  - Les coefficients α_i pour les dummy sont la différence entre le groupe 1 et le groupe i.
  - Pour avoir l'effet du groupe n par exemple, il faudra additionner les deux:

$$\textbf{cts} + \alpha_n$$