

Estimation des externalités d'agglomération à l'aide de variables historiques et géologiques.

P.-P. Combes, G. Duranton, L. Gobillon, S. Roux

GREQAM, University of Toronto, INED, CREST

Journées de Méthodologie Statistique

Objectif

Littérature importante sur l'estimation des effets d'agglomération, cf. Rosenthal et Strange, 2004 : le doublement de la densité des emplois augmente de 4 à 8% la productivité locale.

Certaines questions ne sont pas tranchées :

- Simultanéité et endogénéité
- Spécification : échelle spatiale ?
- Quelle variable dépendante à examiner ?

Deux biais potentiels importants

- endogénéité de la quantité de travail
 - Mécanisme : La densité urbaine serait la conséquence, et non la cause de la productivité plus élevée (biais d'hétérogénéité inobservée)
 - Solution : Utilisation d'instruments géologiques et historiques
- endogénéité de la qualité de travail
 - Mécanisme : Les plus grandes cités pourraient attirer les travailleurs qualifiés, l'effet de qualification étant non séparable de l'effet d'agglomération.
 - Solution : Utilisation d'effets fixes mesurant la "vraie" productivité des individus (estimés à partir d'équations de salaires)

Structure économique

- On considère la fonction de production

$$y_i = A_i k_i^\alpha l_i^{1-\alpha}$$

- La maximisation du profit amène \Rightarrow

$$w_i = (1 - \alpha) \left(\frac{\alpha}{r} \right)^{\alpha/(1-\alpha)} A_i^{1/(1-\alpha)}$$

- Salaires :

$$\ln w_i = \text{Constant} + \frac{1}{1 - \alpha} \log A_i$$

- Effets d'agglomération :

$$\log A_i = X_{a(i)}\varphi + \mu_i$$

- \Rightarrow Salaires :

$$\log w_i = \text{Constant} + X_{a(i)} \frac{\varphi}{1 - \alpha} + \mu'_i$$

- \Rightarrow PGF :

$$\ln y_i = \alpha \ln k_i + (1 - \alpha) \ln l_i + X_{a(i)}\varphi + \mu_i$$

Equations à estimer

- Idée : estimer des effets au niveau de la zone, W_a ou TFP_a .
- Les spécifications à estimer sont alors :

$$\ln W_a = Constant + X_a \varphi^W + \mu_a^W \quad (1)$$

et

$$\ln TFP_a = Constant + X_a \varphi^{TFP} + \mu_a^{TFP} \quad (2)$$

- Densité ($Dens$) est une variable explicative potentiellement endogène : il faut l'instrumenter !!
- Z est un bon instrument si
 - ① $Cov(Dens, Z) \neq 0$, vérifiable
 - ② $Cov(\mu_a^X, Z) = 0$, vérifiable si suffisamment d'instruments

Questions relatives à l'estimation

- Echelle spatiale (Zones d'emploi françaises et potentiel de marché)
- Quelles caractéristiques locales de contrôle ?
Urbanisation vs. effets de localisation
- Quantité de travail endogène (iv)
- Qualité de travail endogène / auto-sélection (effets fixes)
- Utilisation des salaires ou de la PGF :
 - PGF concept le plus direct de la productivité, mais au niveau de l'entreprise.
+ Problème d'endogénéité des inputs
 - Le salaire est une mesure indirecte de la productivité individuelle :
permet de contrôler l'hétérogénéité fixe inobservée

3 concepts de salaire

- Moyenne simple :

$$W_{at}^1 \equiv \ln \bar{w}_{at} \equiv \ln \left(\frac{1}{N_{at}} \sum_{j \in (a,t)} w_{jt} \right) \quad (3)$$

- Contrôle de l'hétérogénéité observable

$$\ln \bar{w}_{ast} = W_{at}^2 + \gamma_s + X_{ast} \varphi + \epsilon_{ast} \quad (4)$$

- Contrôle de l'hétérogénéité inobservable

$$\ln w_{it} = W_{a(it)t}^3 + \gamma_{s(it)} + X_{a(it)s(it)t}^1 \varphi_{s(it)}^1 + X_{it}^2 \varphi^2 + \theta_i + \epsilon_{it} \quad (5)$$

Concepts de PGF

- Equation de production

$$\ln va_{it} = \alpha \ln k_{it} + \beta \ln l_{it} + \sum_m \beta_m^S q_{imt} + \phi_t + \varepsilon_{it} \quad (6)$$

où k_{it} est le capital, l_{it} le nombre d'heures, q_{imt} la part du travail de qualification m , estimée par Olley-Pakes, d'où r_{it} , estimateur de ε_{it}

- Remarque : pas de contrôle de l'hétérogénéité individuelle
- On définit $r_{ast} \equiv \frac{1}{L_{ast}} \sum_{i \in (a,s,t)} l_{it} r_{it}$ où a est la zone, s le secteur.

$$\text{tfp}_{at}^1 \equiv \frac{1}{n_{at}} \sum_{s \in (a,t)} n_{ast} r_{ast},$$

où n_{at} est le nombre d'établissements dans la zone a à la date t .

- On décompose $r_{ast} = \gamma_s + \iota_{ast}$. On définit alors :

$$\text{tfp}_{at}^2 \equiv \frac{1}{n_{at}} \sum_{s \in (a,t)} n_{ast} \hat{\iota}_{ast}.$$

tfp_{at}^2 indice de productivité net des secteurs

- On peut rajouter des variables explicatives X_{ast} . On a alors

$$r_{ast} = \text{tfp}_{at}^3 + \gamma_s + X_{ast}\varphi + \epsilon_{ast}$$

Construit de la même manière que W_{at}^2

Données

- Pour les salaires : dads, 1976-1994
- Pour la PGF :
 - Données d'entreprises brn et rsi, 1994-2002
 - Localisation de l'établissement : siren, 1994-2002
 - Plus dads, 1994-2002
- Agrégées pour créer les mesures par zone de productivité
- Autres contrôles introduits
- Utilisation d'instruments

Endogénéité de la quantité de travail

- IV1 : Variables endogènes retardées (Ciccone et Hall, AER 1996, Combes et al 2006)
- IV2 : Géologie (Rosenthal et Strange 2006)
- IV3 : Autres – localisation relative (Ciccone et Hall, AER 1996)

L'utilisation d'instruments très différents renforce la crédibilité des tests de suridentification et permet la comparaison des coefficients

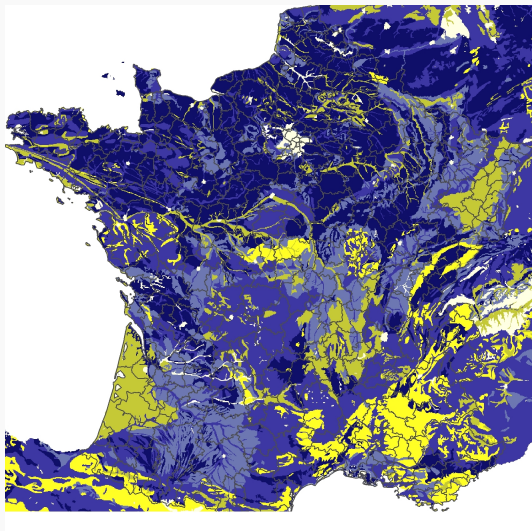
Variables historiques : instruments ?

- Pertinence :
 - Les variables de population passée sont de bons prédicteurs de la population actuelle (Eaton Eckstein, RSUE, 1997)
- Exogénéité :
 - Le niveau de population passée et ses déterminants n'ont pas de raison de causer directement la productivité actuelle
 - Des variables très retardées sont indiquées

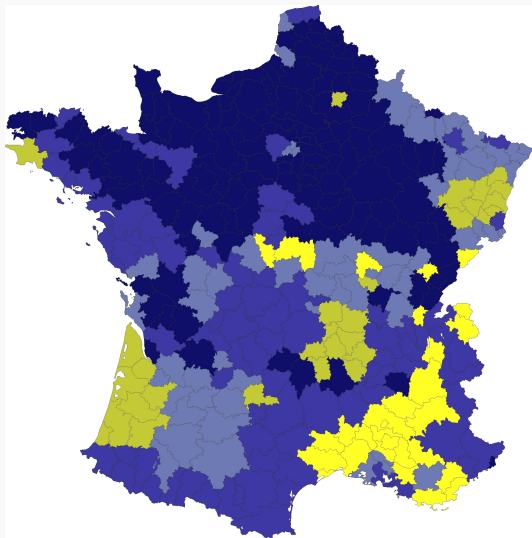
Variables géologiques : instruments

- Pertinence :
 - La nature des sols est un déterminant de long terme de l'installation des population et potentiellement de leur comportement démographique
- Exogénéité :
 - L'agriculture n'est plus une contrainte sur la localisation des populations (et n'est pas utilisé dans les données)

Exemple — capacité hydraulique souterraine



Exemple — capacité hydraulique souterraine



Tab.: Estimation de première étape : Densité de l'emploi

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ln(1831 density)	0.906 (0.046) ^a								
ln(1881 density)		0.924 (0.030) ^a							
Ruggedness			- 0.710 (0.224) ^a						
Subsoil mineralogy	N	N	N	Y	N	N	N	N	N
Dominant parent material (6 categories)	N	N	N	Y	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	N	Y	N	N	N
Soil carbon content	N	N	N	N	N	N	Y	N	N
Depth to rock	N	N	N	N	N	N	N	Y	N
Soil differentiation	N	N	N	N	N	N	N	N	Y
R-squared	0.58	0.78	0.07	0.07	0.17	0.06	0.10	0.15	0.11
F-test (H_0 – All instruments zero)	395.7	1018.8	10.0	5.5	9.1	1.7	6.5	12.6	12.3
Partial R-squared	0.57	0.77	0.03	0.04	0.13	0.02	0.06	0.11	0.08

Dependent variable : ln(employment density). 306 observations for each regression.

All regressions include a constant and three amenity variables (sea, lake, and mountain).

Standard errors in parentheses. a, b, c : corresponding coefficient significant at 1, 5, 10%.

Tab.: Estimation de première étape : Potentiel de marché

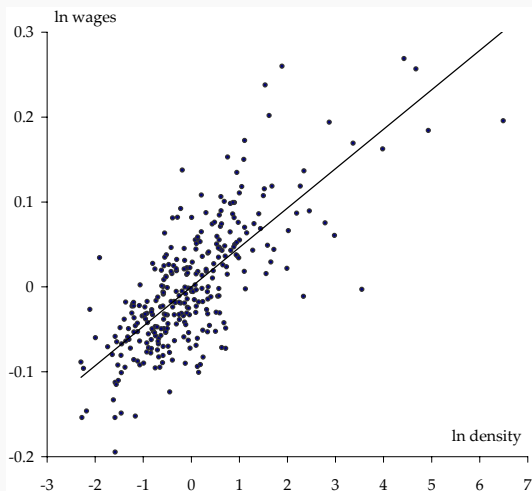
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ln(1831 market pot.)	1.026 (0.012) ^a								
ln(1881 market pot.)		0.970 (0.007) ^a							
Ruggedness			- 0.339 (0.111) ^a						
Subsoil mineralogy	N	N	N	Y	N	N	N	N	N
Dominant parent material (6 cat.)	N	N	N	N	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	N	Y	N	N	N
Soil carbon content	N	N	N	N	N	N	Y	N	N
Depth to rock	N	N	N	N	N	N	N	Y	N
Soil differentiation	N	N	N	N	N	N	N	N	Y
R-squared	0.97	0.99	0.23	0.24	0.43	0.41	0.28	0.44	0.31
F-test (H_0 – All instruments zero)	7106.47	21503.0	9.4	7.3	23.1	26.3	11.3	41.2	24.0
Partial R-squared	0.96	0.99	0.03	0.05	0.28	0.26	0.10	0.29	0.14

Dependent variable : ln(market potential). 306 observations for each regression.

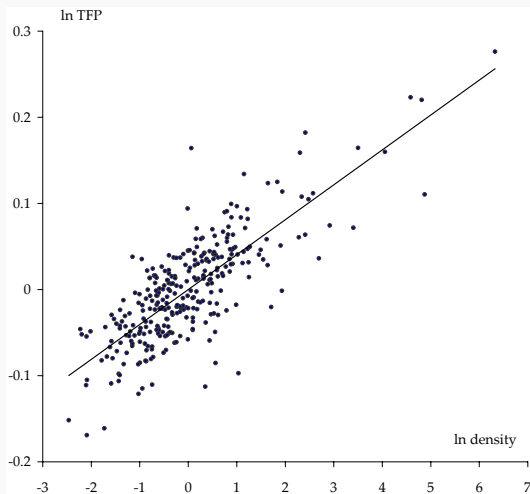
All regressions include a constant and three amenity variables (sea, lake, and mountain).

Standard errors in parentheses. a, b, c : corresponding coefficient significant at 1, 5, 10%.

Corrélations observées : salaires et densité de l'emploi



Corrélations observées : PGF et densité de l'emploi



Résultats Salaires instrumentés

Tab.: Local wages as a function of density : historical and geological instruments

Variable	[1] W^1 tsls	[2] W^2 tsls	[3] W^3 tsls	[4] W^3 gmm	[5] W^3 tsls	[6] W^3 tsls	[7] W^3 tsls	[8] W^3 tsls
ln(dens.)	0.040 (0.003) ^a	0.042 (0.002) ^a	0.027 (0.002) ^a	0.027 (0.002) ^a	0.027 (0.002) ^a	0.027 (0.002) ^a	0.027 (0.002) ^a	0.027 (0.002) ^a
Instruments :								
ln(1831 dens.)	Y	Y	Y	Y	Y	Y	Y	Y
Subsoil min.	Y	Y	Y	Y	N	Y	N	N
Ruggedness	N	N	N	N	Y	Y	Y	N
Hydro	N	N	N	N	N	N	Y	N
Topsoil water cap.	N	N	N	N	N	N	N	Y
1st stage stat.	138.7	138.7	138.7	116.2	208.7	108.1	69.8	76.2
Over-id p -value	0.98	0.83	0.15	0.13	0.31	0.21	0.53	0.02

306 observations for each regression.

All regressions include a constant and three amenity variables (sea, lake, and mountain).

Standard errors in parentheses. *a*, *b*, *c* : corresponding coefficient significant at 1, 5, 10%.

Résultats avec potentiel de marché

Tab.: Local wages as a function of density and (endogenous) market potential : historical and geological instruments

Variable	[1] W^1 tsls	[2] W^2 tsls	[3] W^3 tsls	[4] W^3 tsls	[5] W^3 tsls	[6] W^3 tsls	[7] W^3 tsls	[8] W^3 tsls
ln(density)	0.033 (0.003) ^a	0.040 (0.003) ^a	0.020 (0.002) ^a	0.018 (0.002) ^a	0.019 (0.002) ^a	0.020 (0.002) ^a	0.020 (0.002) ^a	0.020 (0.003) ^a
ln(market pot.)	0.034 (0.006) ^a	0.020 (0.005) ^a	0.034 (0.003) ^a	0.048 (0.007) ^a	0.039 (0.005) ^a	0.036 (0.005) ^a	0.036 (0.006) ^a	0.033 (0.010) ^a
Instruments used :								
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N
Soil carbon content	N	N	N	Y	Y	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N
Depth to rock	N	N	N	N	N	Y	Y	N
Ruggedness	N	N	N	N	N	N	Y	Y
Soil differentiation	N	N	N	N	N	N	N	Y
First stage statistics	298.0	298.0	298.0	8.3	17.0	23.0	19.8	8.3
Over-id test p -value	0.57	0.36	0.67	0.62	0.19	0.36	0.54	0.11

306 observations for each regression.

All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c : corresponding coefficient significant at 1, 5, 10%.

Résultats PGF

Tab.: Local tfp (Olley-Pakes) as a function of density and (endogenous) market potential : historical and geological instruments

Variable	[1] tfp ¹ tsls	[2] tfp ² tsls	[3] tfp ³ tsls	[4] tfp ³ tsls	[5] tfp ³ tsls	[6] tfp ³ tsls	[7] tfp ³ tsls	[8] tfp ³ tsls
ln(density)	0.028 (0.003) ^a	0.030 (0.002) ^a	0.035 (0.002) ^a	0.034 (0.003) ^a	0.034 (0.003) ^a	0.034 (0.003) ^a	0.034 (0.003) ^a	0.034 (0.004) ^a
ln(market pot.)	0.025 (0.005) ^a	0.027 (0.004) ^a	0.026 (0.004) ^a	0.021 (0.009) ^b	0.023 (0.007) ^a	0.021 (0.006) ^a	0.022 (0.008) ^a	0.024 (0.013) ^c
Instruments used :								
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N
Soil carbon content	N	N	N	Y	Y	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N
Depth to rock	N	N	N	N	N	Y	Y	N
Ruggedness	N	N	N	N	N	N	Y	Y
Soil differentiation	N	N	N	N	N	N	N	Y
First stage statistics	230.6	230.6	230.6	8.2	16.4	22.8	20.2	8.0
Over-id test <i>p</i> -value	0.16	0.43	0.17	0.68	0.90	0.60	0.29	0.04

306 observations for each regression.

All regressions include a constant and three amenity variables (sea, lake, and mountain).

Standard errors in parentheses. *a*, *b*, *c* : corresponding coefficient significant at 1, 5, 10%.

Conclusion

- Quantité endogène de travail :
 - Les variables endogènes retardées sur longues période (historiques) sont des instruments valides
 - Les variables géologiques sont plus compliquées à manipuler
 - Les deux types de variables conduisent à la même conclusion : le biais d'endogénéité de la quantité de travail est relativement faible
- Qualité endogène de travail :
 - Les meilleurs salariés vont dans les meilleures zones
 - En contrôlant de l'hétérogénéité fixe des individus, l'élasticité des salaires à la densité de l'emploi est divisée par 2. On ne peut appliquer cette approche à la PGF.
- Ces résultats sont propres à la France. Ils peuvent différer pour des pays où la mobilité des individus est plus forte (Etats-Unis).