4th User Conference of the RDC

Agglomeration economies with consistent productivity estimates

Philipp Ehrl

University of Passau

8. April 2011

Outline

motivation

- Economic activity is spatially concentrated.
- Firms bear higher factor prices for land and labor in agglomerations.
- What are the advantages? How can we detect them?
- Marshall (1890): proximity to suppliers, large labor markets and knowledge spillovers.

→ Productivity is the most direct measure, in order to capture agglomeration economies (Puga, 2010).

- Glaeser and Gottlieb (2009): "the field has still not reached a consensus on the relative importance of different sources of agglomeration economies".

motivation

- Economic activity is spatially concentrated.
- Firms bear higher factor prices for land and labor in agglomerations.
- What are the advantages? How can we detect them?
- Marshall (1890): proximity to suppliers, large labor markets and knowledge spillovers.

→ Productivity is the most direct measure, in order to capture agglomeration economies (Puga, 2010).

- Glaeser and Gottlieb (2009): "the field has still not reached a consensus on the relative importance of different sources of agglomeration economies".

related literature

Productivity as depedent variable

- Combes, Duranton, Gobillon, Roux (2010):
  TFP increases in the **density** of regions.
- Greenstone, Hornbeck, Moretti (2010):
  evidence that **each** of the Marshallian forces is associated with higher TFP.

Discriminating between agglomeration mechanisms

- Rigby and Essletzbichler (2002):
  all three Marshallian forces are associated with higher **labor productivity**.
- Ellison, Glaeser, Kerr (2010):
  all three Marshallian forces foster **coagglomeration** of industries.

approach of the paper

- Several estimation strategies for total factor productivity (TFP) from firm level production functions are contrasted.
- Agglomeration variables for labor pooling, input relations and knowledge spillovers are constructed, based on theoretical models.
- Conducting univariate and multivariate regressions.
- Localization, urbanization and specialization economies are tested as well.

TFP estimation (OLS)

Cobb-Douglas production function

$$y_{jt} = \beta_0 + \omega_{jt} + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + \zeta z_{jt} + u_{jt}^p$$

problems:

1. unobserved productivity is positively correlated with input choices. (endogeneity bias)

$\Rightarrow$ upward biased coefficients

TFP estimation (OLS)

Cobb-Douglas production function

$$r_{jt} = y_{jt} + p_{jt} = \beta_0 + \omega_{jt} + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + \zeta z_{jt} + u_{jt}^p + p_{jt}$$

problems:

1. unobserved productivity is positively correlated with input choices. (endogeneity bias)
$\Rightarrow$ upward biased coefficients.
2. unobserved output price is negatively correlated with input choices. (price bias)
$\Rightarrow$ downward biased coefficients.

(KG) Klette and Griliches (1996)

Introduction of a demand system to replace unobserved output prices

$$Q_{jt} = \left(\frac{P_{jt}}{P_{It}}\right)^{-\sigma} \cdot \frac{R_{It}}{P_{It}}$$

and deflating revenues by an industry price index yields

$$\widetilde{r}_{jt} := y_{jt} + p_{jt} - p^s_{It} =$$

$$= \underbrace{\widetilde{\beta}_0 + \widetilde{\omega}_{jt} + \widetilde{\alpha}_k k_{jt} + \widetilde{\alpha}_l l_{jt} + \widetilde{\alpha}_m m_{jt} + \widetilde{\zeta} z_{jt}}_{\left(\frac{\sigma-1}{\sigma}\right)\left(\beta_0 + \omega_{jt} + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + \zeta z_{jt}\right)} + \frac{1}{\sigma}(r^s_{It} - p^s_{It}) + u_{jt}$$

(OP) Olley and Pakes (1996)

Intuition:

- An equation for investment demand $i_{jt} = i_t(k_{jt}, \omega_{jt}(G_t^k))$ is inverted to $\omega_{jt} = h_t(k_{jt}, i_{jt}, G_t^k)$ in order to proxy for unobserved productivity.
- ω_{jt} is influenced by regional deteminants (G_t^k).

$$r_{jt} = \beta_0 + h_t(k_{jt}, i_{jt}, G_t^k) + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + u_{jt}$$

- $h_t(k_{jt}, i_{jt}, G_t^k)$ is approximated by a second order polynomial.

(OP) Olley and Pakes (1996)

Intuition:

- An equation for investment demand $i_{jt} = i_t(k_{jt}, \omega_{jt}(G_t^k))$ is inverted to $\omega_{jt} = h_t(k_{jt}, i_{jt}, G_t^k)$ in order to proxy for unobserved productivity.
- ω_{jt} is influenced by regional deteminants (G_t^k).

$$r_{jt} = \beta_0 + h_t(k_{jt}, i_{jt}, G_t^k) + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + u_{jt}$$

- $h_t(k_{jt}, i_{jt}, G_t^k)$ is approximated by a second order polynomial.

OP/KG combined

OP and KG procedures combined:

$$\tilde{r}_{jt} = \tilde{\beta}_0 + \tilde{h}_t(k_{jt}, i_{jt}, G_t^k) + \tilde{\alpha}_k k_{jt} + \tilde{\alpha}_l l_{jt} + \tilde{\alpha}_m m_{jt} + \frac{1}{\sigma}(r_{It}^s - p_{It}^s) + u_{jt}$$

gives consistent a TFP estimate

$$\hat{\omega}_{jt} = \left[\tilde{r}_{jt} - \hat{\tilde{\alpha}}_l l_{jt} - \hat{\tilde{\alpha}}_m m_{jt} - \hat{\tilde{\alpha}}_k k_{jt} - \left(\widehat{\frac{1}{\sigma}}\right)(r_{It}^s - p_{It}^s) - \hat{\zeta} z_{jt}\right]\left(\frac{\hat{\sigma}}{\hat{\sigma} - 1}\right)$$

which is then used as the dependent variable to detect agglomeration economies.

data

For the estimation of the production functions the IAB Establishment Panel (IABB) is used from 2000 to 2005.

- Main advantage: the location of the plant at NUTS 3-digit level (counties) and the industry classification are also available.
- The capital variable is constructed from plant investment behaviour employing the modified perpetual inventory method (Müller, 2008).
- Industry specific aggregate revenues and price indices ($r^s_{It} - p^s_{It}$) are taken from the Federal Statistical Office.

$\Rightarrow$ The study comprises 22 of 41 industries, between the 1-digit and 2-digit level. Not only manufacturing.

production function results

	OLS	OP	KG		OP/KG	
	$\widetilde{\alpha}$	$\widetilde{\alpha}$	$\widetilde{\alpha}$	α	$\widetilde{\alpha}$	α
materials	0.6522	0.6484	0.6512	0.8230	0.6474	0.8049
	(0.0063)	(0.0043)	(0.0063)		(0.0043)	
labor	0.3300	0.3287	0.3312	0.4186	0.3297	0.4100
	(0.0086)	(0.0057)	(0.0087)		(0.0057)	
capital	0.0472	0.0465	0.0472	0.0596	0.0458	0.0569
	(0.0037)	(0.0006)	(0.0035)		(0.0006)	
demand ela.	-	-	5.58		5.84	
			(1.1797)		(1.6137)	
west	0.1212	0.1012	0.1204	0.1522	0.1007	0.1272
	(0.0090)	(0.0090)	(0.0090)		(0.0061)	
high-skilled	0.1508	0.1465	0.1481	0.1872	0.1434	0.1783
share	(0.0178)	(0.0159)	(0.0177)		(0.0161)	
N	18569	18569	18569		18569	
R^2	0.9711	-	0.9711		-	

Notes: std. errors in parenthesis. Industry FE included.

labor market pooling

- Matching models like Coles and Smith (1998) or Berliant, Reed, Wang (2006): larger markets provide more opportunities to find suitable matches.
  $\Rightarrow$ expected productivity is higher.

$\rightarrow$ The closer the industry's profile is to the composition of the local labor market the better.

1. Empirical implementation: correlation between occupational structure in the industry under scrutiny and the occupational structure in all remaining industries in a county.

2. Coles and Smith (1998) also predict:
   high # of job vacancies $\Rightarrow$ higher productivity.

knowledge spillovers

- Job changes: worker bring their knowledge, experience or new ideas to their new employer (Fosfuri and Rønde, 2004).

1. From the BA Employment Panel: average job changes in each county. (Only skilled workers.)

- High innovativeness of a region might spill over to neighboring firms.

2. Funding to companies, universities and institutions for innovative R&D projects from the Federal Ministry of Education and Research (BMBF).
3. Patent applicants per worker in a county.

input linkages

- Models with an intermediate goods sector, e.g. Ethier (1982): high # of intermediate goods producers ⇒ high productivity in the final goods sector.
- supplier relations in industry *i*: $strength_{ij}$
  share of goods that industry *i* purchases from industry *j*
- is related to the industrial structure in each region

$$input\text{-}linkage_i^k = \sum_j \left(strength_{ij} \cdot \frac{E_j^k}{E^k} \right)$$

	OLS	OP	KG	OP/KG
occ-corr	0.0154	0.0190	0.0235	0.0288
	(0.0010)	(0.0013)	(0.0000)	(0.0000)
R^2	0.0024	0.0053	0.0059	0.0086
vacancies	0.0078	0.0101	0.0096	0.0121
	(0.0283)	(0.0258)	(0.0074)	(0.0068)
R^2	0.0016	0.0046	0.0036	0.0065
job-change	0.0116	0.0151	0.0124	0.0158
	(0.0011)	(0.0008)	(0.0005)	(0.0004)
R^2	0.0022	0.0052	0.0041	0.0070
patents	**0.0069**	**0.0087**	0.0144	0.0178
	(0.1422)	(0.1437)	(0.0023)	(0.0024)
R^2	0.0015	0.0045	0.0045	0.0074
R&D	0.0151	0.0189	0.0182	0.0224
	(0.0010)	(0.0012)	(0.0001)	(0.0001)
R^2	0.0029	0.0058	0.0055	0.0083
input-linkage	0.0059	0.0080	**0.0044**	**0.0058**
	(0.0219)	(0.0156)	(0.0894)	(0.0718)
R^2	0.0013	0.0043	0.0030	0.0059

multivariate regressions

	OLS	OP	KG	OP/KG
occ-cor	**0.0098**	**0.0142**	0.0115	**0.0167**
	(0.0440)	(0.0037)	(0.0545)	(0.0047)
vacancies	0.0047	0.0051	0.0059	0.0062
	(0.1501)	(0.1207)	(0.1421)	(0.1163)
job-changes	**0.0092**	**0.0088**	**0.0116**	**0.0110**
	(0.0093)	(0.0132)	(0.0074)	(0.0106)
patents	-0.0004	0.0038	-0.0004	0.0048
	(0.9396)	(0.4203)	(0.9472)	(0.4009)
R&D	**0.0105**	**0.0108**	**0.0129**	**0.0131**
	(0.0237)	(0.0208)	(0.0238)	(0.0207)
input-linkage	0.0044	0.0037	0.0057	0.0046
	(0.0921)	(0.1547)	(0.0750)	(0.1460)
R^2	0.0041	0.0064	0.0066	0.0089
F	3.88	4.58	7.99	8.60
N	18569	18569	18569	18569

Notes: p-values in parethesis. Industry FE, year FE and constant included.

consequence of the price bias

Comparison between OP and OP/KG and accordingly between OLS and KG reveals the consequence of the price bias:
In the OLS and OP case the estimating equation is

$$\omega_{jt} + p_{jt} = \beta_0 + \gamma G_t^k + e_{jt}$$

- $\longrightarrow$ Lower coefficients are observed.
- $\Rightarrow$ p_{jt} and G_{jt}^k are negatively correlated.
  Moreover $corr(\omega_{jt}, G_{jt}^k) > 0$.
- $\Rightarrow$ High productivity firms quote lower prices
  Melitz, ECA (2003), Foster et al., AER (2008).

industrial environment proxies

urbanization

1. Size of a county in terms of employment (E^k).
2. Share of the largest three industries in a county k: $jacobs2^k$.
3. Diversity:

$$jacobs1^k = \sum_i \left(\frac{E_i^k}{E^k} - \frac{E_i}{E} \right)^2$$

localization

4. Own industry employment in a county (E_i^k).
5. Share of the own industry in a county (E_i^k / E^k).

industrial environment results

	OLS	OP	KG	OP/KG	OP/KG	OP/KG
E^k	**0.0124**	**0.0158**	**0.0149**	**0.0189**	**0.0175**	**0.0288**
	(0.0409)	(0.0092)	(0.0439)	(0.0098)	(0.0139)	(0.0000)
E_i^k	0.0029	0.0030	0.0042	0.0042	0.0044	0.0018
	(0.6371)	(0.6305)	(0.5826)	(0.5763)	(0.5566)	(0.8137)
E_i^k/E^k	-0.0049	0.0010	-0.0084	-0.0003	0.0002	**0.0177**
	(0.5749)	(0.9133)	(0.4297)	(0.9755)	(0.9867)	(0.0763)
$jacobs2$	**0.0213**	**0.0230**	**0.0262**	**0.0279**	**0.0293**	-
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
$jacobs1$	0.0027	0.0034	0.0035	0.0043	-	-
	(0.5732)	(0.4848)	(0.5566)	(0.4670)		
R^2	0.0064	0.0093	0.0089	0.0118	0.0117	0.0081
F	5.62	6.98	10.15	11.45	12.72	10.52
N	18569	18569	18569	18569	18569	18569

Notes: p-values in parethesis. Industry FE, year FE and constant included.

additional productivity measures

1. The Ackerberg, Caves, Frazer (2006) correction to Olley and Pakes (1996).
   Coefficients of L and M are identified differently than in OP.
2. Intermediate inputs as control function (Levinsohn and Petrin, 2003).
3. TFP estimated from value added production functions.
4. Labor productivity.

agglomeration results for additional TFP measures

	OP/KG	ACF/OP/KG	LP/KG	VA/KG	L-prod
occ-cor	**0.0167**	**0.0304**	**0.0209**	**0.0588**	**0.0883**
	(0.0047)	(0.0000)	(0.0004)	(0.0000)	(0.0000)
vacancies	0.0062	0.0051	0.0060	0.0115	**0.0141**
	(0.1163)	(0.2151)	(0.1291)	(0.1113)	(0.0275)
job-changes	**0.0110**	**0.0096**	**0.0106**	**0.0216**	0.0117
	(0.0106)	(0.0341)	(0.0136)	(0.0089)	(0.1319)
patents	0.0048	**0.0176**	0.0084	0.0179	**0.0422**
	(0.4009)	(0.0033)	(0.1446)	(0.1012)	(0.0000)
R&D	**0.0131**	**0.0154**	**0.0139**	**0.0227**	**0.0225**
	(0.0207)	(0.0105)	(0.0151)	(0.0416)	(0.0432)
input-linkage	0.0046	**-0.0093**	0.0006	0.0000	**-0.0231**
	(0.1460)	(0.0112)	(0.8518)	(0.9953)	(0.0052)
R^2	0.0089	0.0416	0.0148	0.0116	0.2315
F	8.60	10.85	8.72	6.32	15.37
N	18569	18569	18569	18569	18569

Notes: p-values in parethesis. Industry FE, year FE and constant included.

industrial environment results for additional TFP measures

	OP/KG	ACF/OP/KG	LP/KG	VA/OP	L-prod
E^k	**0.0189**	**0.0255**	**0.0209**	**0.0573**	**0.0507**
	(0.0098)	(0.0009)	(0.0043)	(0.0001)	(0.0004)
E_i^k	0.0042	0.0070	0.0053	0.0084	**0.0382**
	(0.5763)	(0.3969)	(0.4957)	(0.5984)	(0.0202)
(E_i^k/E^k)	-0.0003	**0.0609**	0.0173	**0.0479**	**0.1431**
	(0.9755)	(0.0000)	(0.1038)	(0.0250)	(0.0000)
*jacobs*2	**0.0279**	**0.0251**	**0.0271**	**0.0482**	**0.0320**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0029)
*jacobs*1	0.0043	0.0079	0.0052	0.0050	0.0122
	(0.4670)	(0.1985)	(0.3743)	(0.5421)	(0.1757)
R^2	0.0117	0.0474	0.0182	0.0145	0.2357
F	12.72	16.45	12.18	8.87	18.41
N	18569	18569	18569	18569	18569

Notes: p-values in parethesis. Industry FE, year FE and constant included.

insights from these extensions

1. The results from OP/KG TFP measure are reinforced.
2. Some significant indication for all agglomeration proxies

→ does not appear stable.

3. Industrial diversity of a region is not associated with higher TFP in *any* of the regressions.
4. LP estimation method is very close to the OP procedure.
5. Labor productivity is imprecise

→ overestimates the agglomeration effects.

6. Agglomeration results from value added and revenue production functions are not comparable.

insights from these extensions

1. The results from OP/KG TFP measure are reinforced.
2. Some significant indication for all agglomeration proxies
   → does not appear stable.
3. Industrial diversity of a region is not associated with higher TFP in *any* of the regressions.
4. LP estimation method is very close to the OP procedure.
5. Labor productivity is imprecise
   → overestimates the agglomeration effects.
6. Agglomeration results from value added and revenue production functions are not comparable.

conclusion

1. It makes a difference how productivity is estimated.

→ Accounting for omitted output prices and endogeneity of inputs is important.

2. Examined seperately, almost all agglomeration variables show some significant indication.
3. Multivariate regressions: labor market pooling (occupational correlation), public R&D funding to innovative projects and job changes of qualified workers raise local TFP.
4. Specialized county strucure is beneficial to firms, whereas no evidence for Jacobs economies.
5. A doubling of employment in a county entails a 2-3% higher firm TFP.
6. Results are robust to variations in (1) the construction of agglomeration variables, (2) scope of industries and (3) estimation of TFP.

conclusion

1. It makes a difference how productivity is estimated.

→ Accounting for omitted output prices and endogeneity of inputs is important.

2. Examined seperately, almost all agglomeration variables show some significant indication.
3. Multivariate regressions: labor market pooling (occupational correlation), public R&D funding to innovative projects and job changes of qualified workers raise local TFP.
4. Specialized county strucure is beneficial to firms, whereas no evidence for Jacobs economies.
5. A doubling of employment in a county entails a 2-3% higher firm TFP.
6. Results are robust to variations in (1) the construction of agglomeration variables, (2) scope of industries and (3) estimation of TFP.