The determinants of coagglomeration: Evidence from functional employment patterns

May 18, 2016

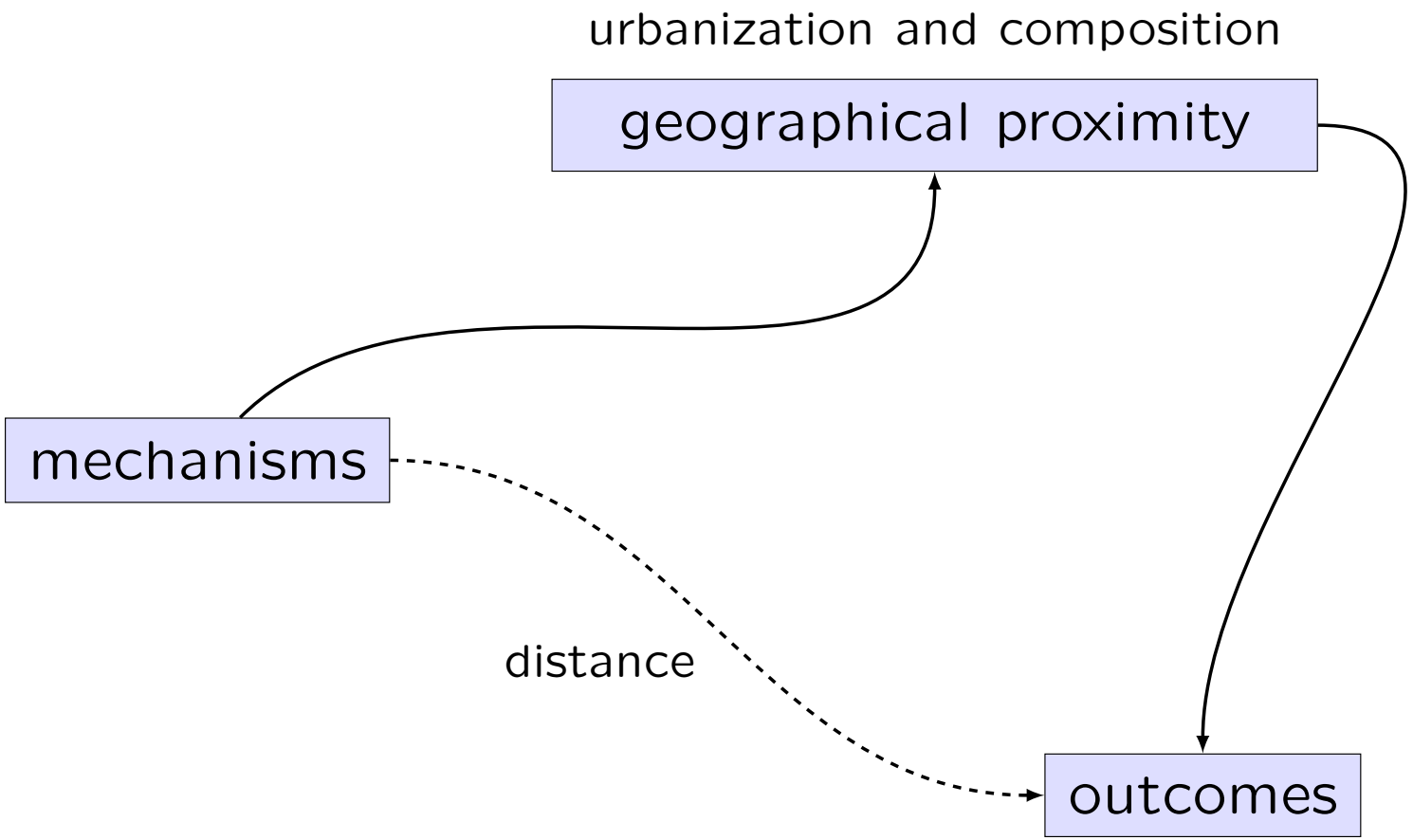
The World Bank & Cornell conference on 'Secondary Towns'

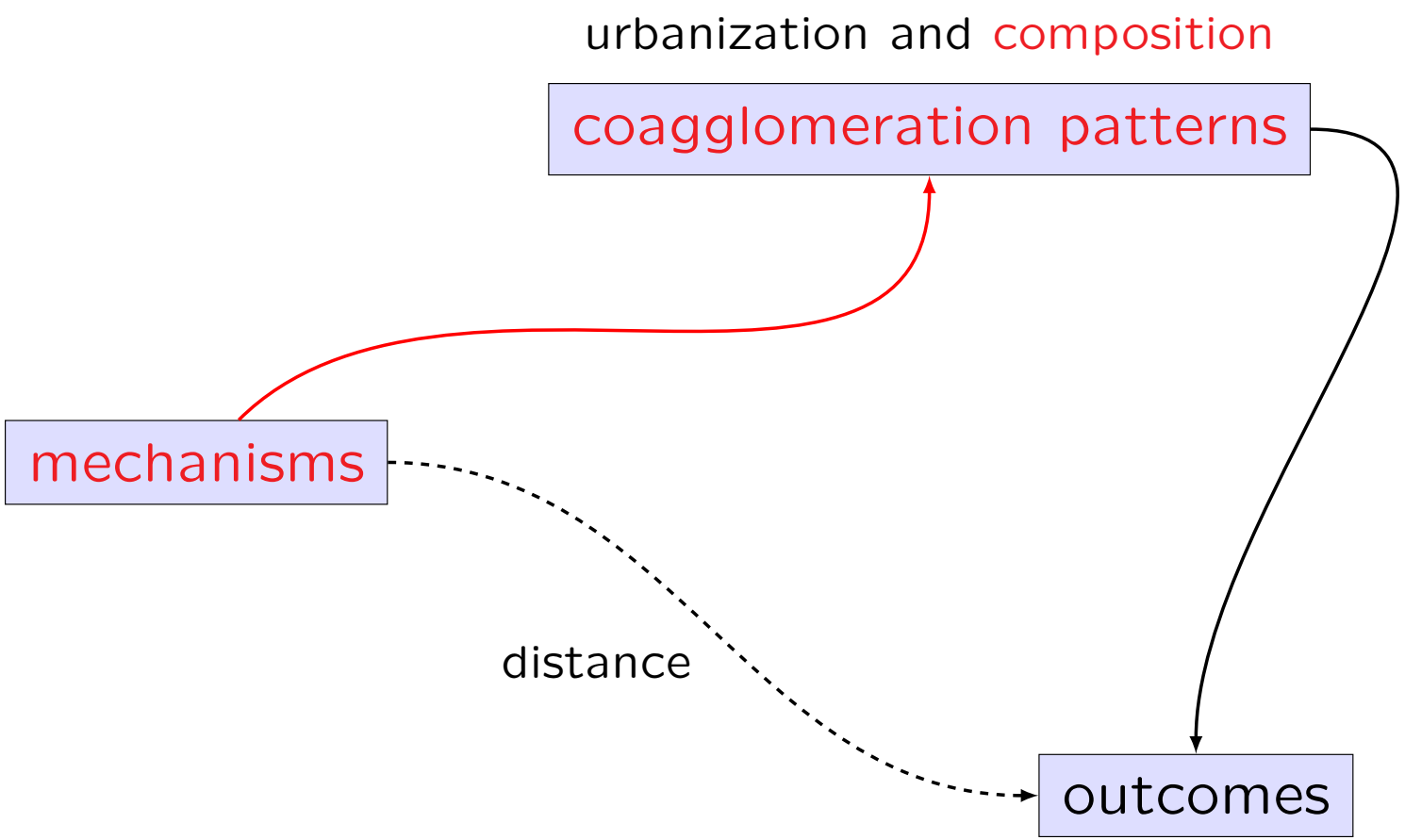
Kristian Behrens
UQÀM; NRU HSE; and CEPR

Rachel Guillain
Université de Bourgogne

"[. . .] the urbanization discourse tends to take place at an aggregative level, with the overall national rate of urbanization taking center stage [. . .] as an outcome to be explained [. . .]"

"Our central tenet is that the composition of urbanization is at least as important as its aggregate rate [. . .]"


urbanization and composition
geographical proximity
mechanisms
distance
outcomes


urbanization and composition
coagglomeration patterns
mechanisms
distance
outcomes

Understanding the mechanisms is important:

◦ to frame clearly the questions related to the links between agglomeration and outcomes;

◦ for the design local policies to act on outcomes (e.g., infrastructure investments vs knowledge-based clusters).

Few empirical studies on agglomeration and firm location in developing countries (but a lot of policy interest).

In the context of urbanizing countries, maybe large potential gains from clustering.

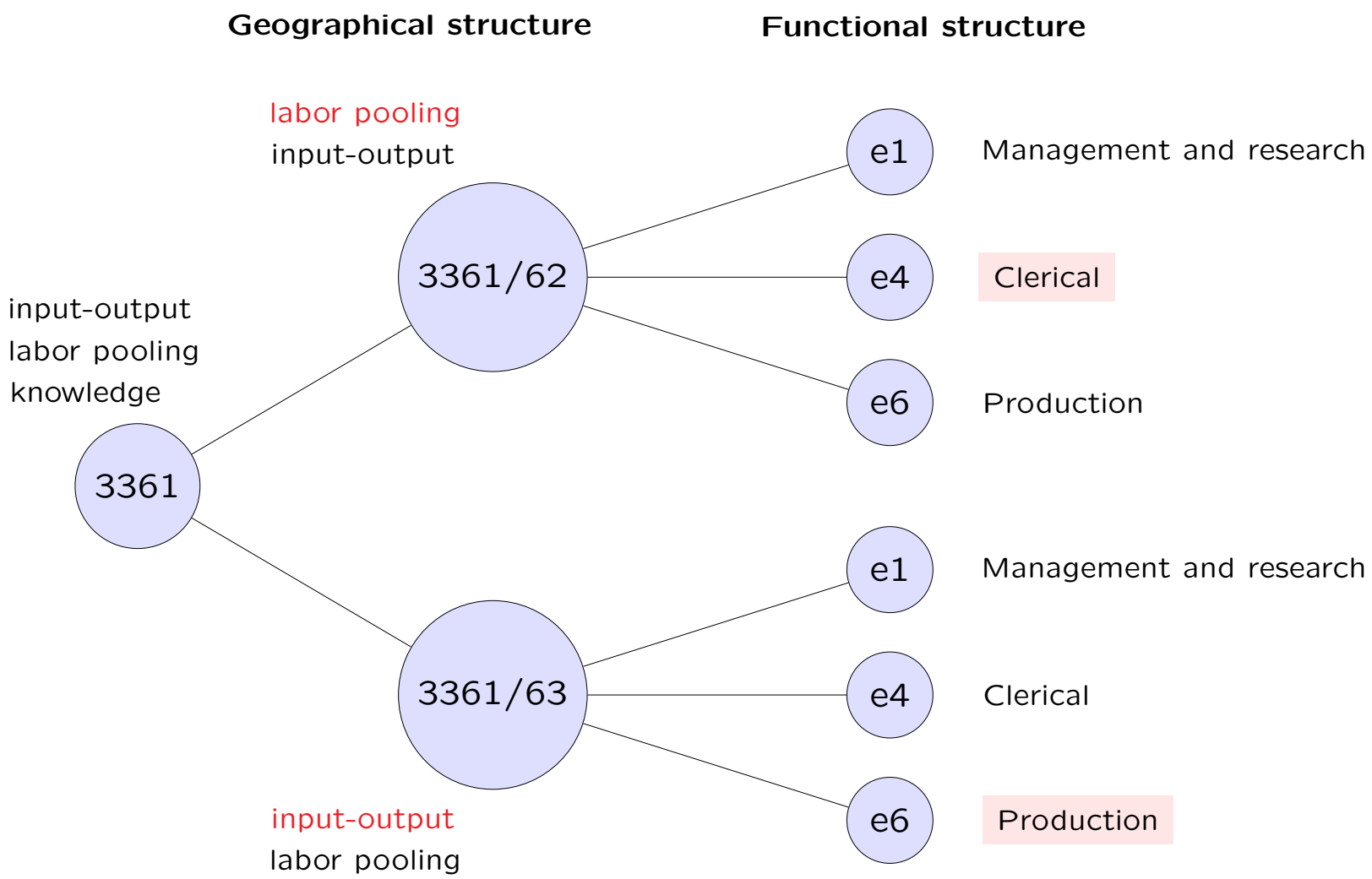
Our contribution:

◦ bring together coagglomeration (*locations differ in industrial composition*) and functional specialization (*locations differ in what they do within industries*);

◦ exploit between and within industry variation to shed more light on the determinants of coagglomeration;

◦ carefully address a number of neglected identification issues.

3361 = Motor vehicle 3362 = Motor vehicle body and trailer 3363 = Motor vehicle parts



Empirical approach

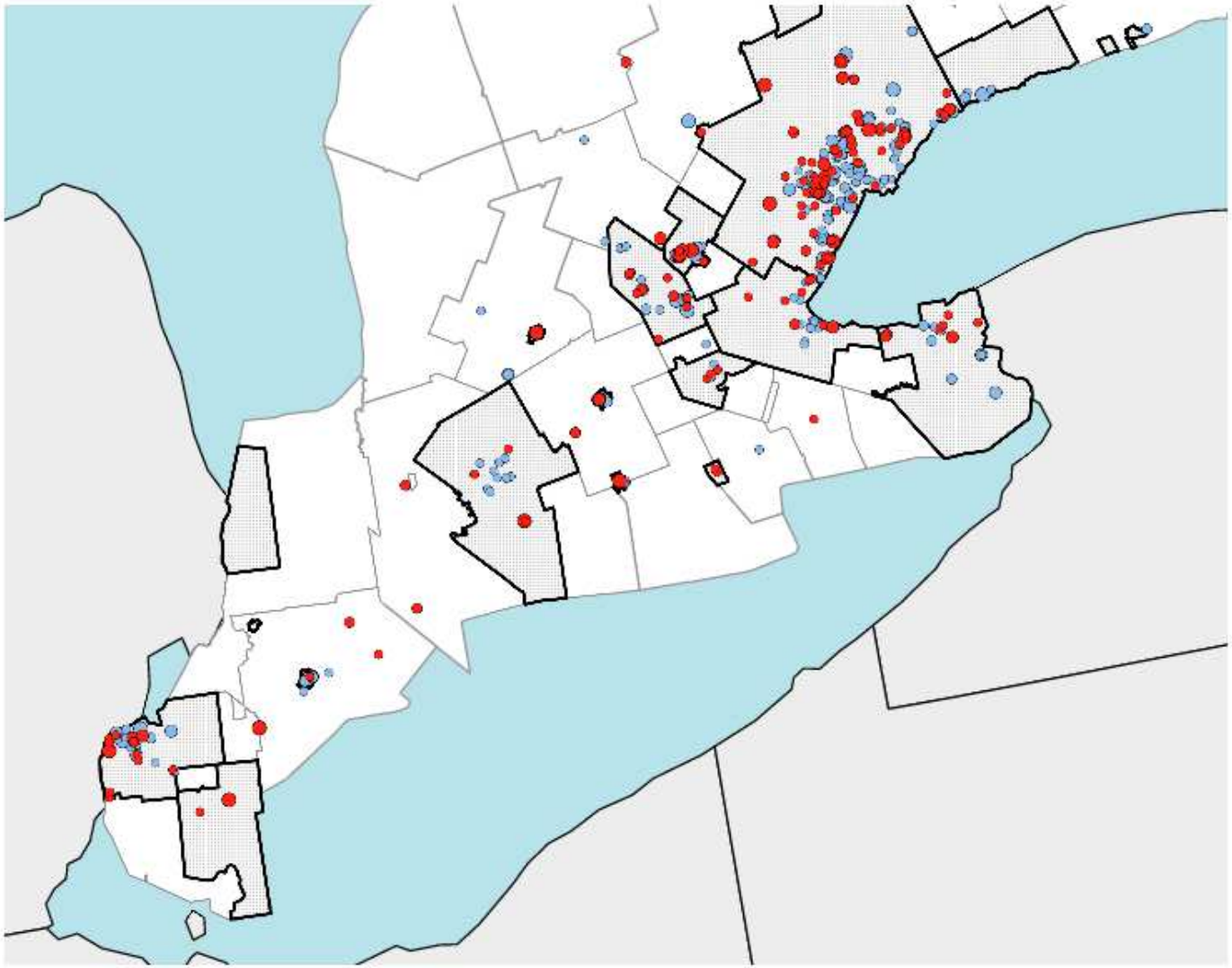
Two key datasets.

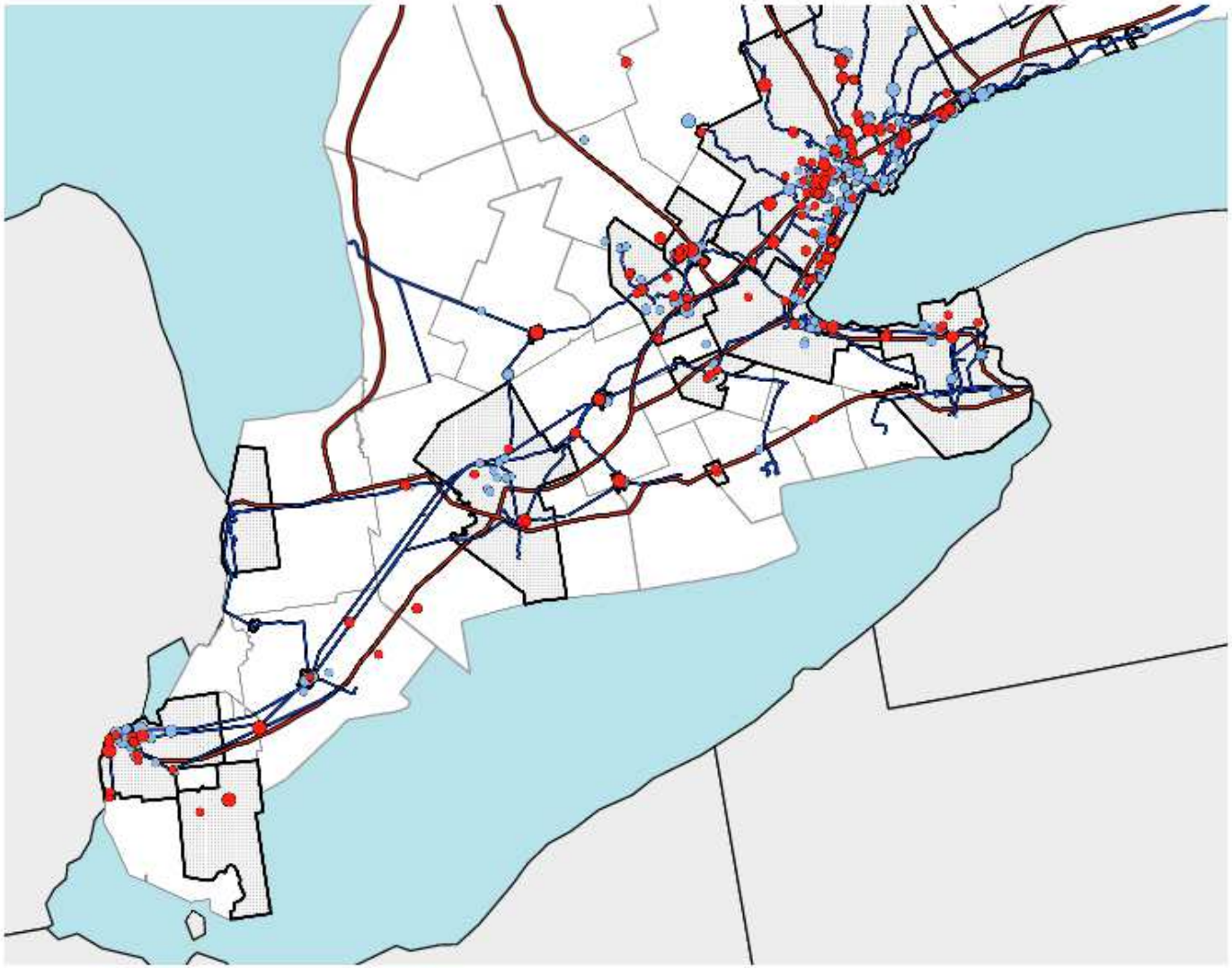
[1] Special census tabulations, split industry-level employment by census division, functional type, and rural-urban status:

- ‘Management and research’
- ‘Clerical’
- ‘Retail and services’
- ‘Production’

[2] Business Register plant-level data, with extensive coverage of the manufacturing sector (80-90% of plants):

- Geocoded to postal code centroids
- Primary and secondary 6-digit NAICS industry codes
- Employment





1. Split plant employment into types (details in the paper).

2. Compute coagglomeration K-density CDFs and 'excess localization measures' (Duranton and Overman, 2005, 2008).

3. Run pooled cross-section regressions of the form:

$$\text{coagglo}_{ijt}^{f} = \alpha_{io}\text{io}_{ijt} + \alpha_{oes}\text{oes}_{ijt}^{f} + \alpha_{know}\text{know}_{ijt} + \mathbf{X}_{ijt}\beta + \xi_i + \xi_j + \delta_t + \epsilon_{ijt},$$

(similar to Ellison, Glaeser, and Kerr, 2010; Faggio, Silva, and Strange, 2015).

Proxies for the agglomeration forces ('goods, people, and ideas'):

a. Input-output: Disaggregated I-O matrices, maximum of the input-output shares between i and j.

b. Labor market pooling: Either correlation of industries' occupation shares (554 occupations); or measure of labor movement across industry pairs (from CPS data).

c. Knowledge sharing: Following Kerr (2008), we use metrics derived from the NBER Patent Citation database.

Identification concerns

Concern 1. Locational advantage

Industries may spuriously colocate. Construct a baseline distribution of plants based on locational advantage (Ellison and Glaeser, 1999; Klier and McMillen, 2008):

◦ Industry regressions of chosen vs non-chosen sites using site-specific variables (distance to highway, rail, coast; population in 25km; distance to US industries...);

◦ Plants are assigned to sites in decreasing order of predicted probabilities, with older plants being assigned first.

◦ Compute the associated coagglomeration measures.

Concern 2. Natural advantage and industrial organization

◦ Control for industries' share of intermediates sourced from primary and from service industries.

◦ Control for industries' share of output sold to primary and to service industries.

◦ Control for the share of multiunit plants in each sector (industries in which multiunit firms are more present can more easily split functions across establishments/space).

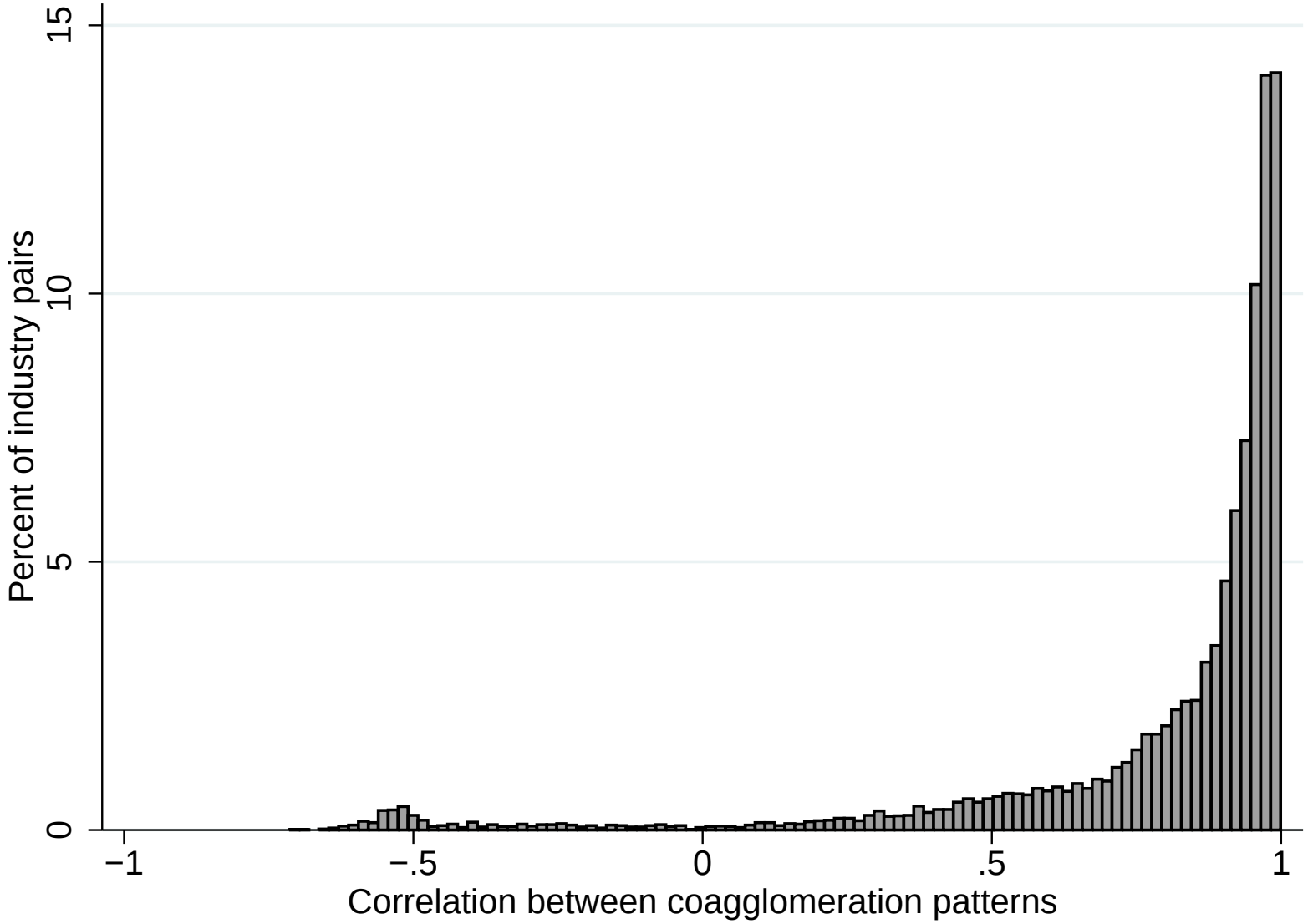
◦ Control for 'within firm agglomeration' (internalization), based on establishments that report operating in multiple industries.

Concern 3. Third industry effects

◦ If plant pairs of A and B are close, and if plant pairs of C and B are close, then pairs in A and C are close because of B.

Raises problem of how to think about the coagglomeration of A and C (is it spurious because of B?)

We include the correlation coefficient of the coagglomeration measures of industries i and j with all other industries. Note that this 'control' may be overly strong.


Percent of industry pairs
0
5
10
15
−1
−.5
0
.5
1
Correlation between coagglomeration patterns

Concern 4. Omitted variables and reverse causality

◦ We include industry fixed effects in most regressions to deal with omitted variables.

◦ I-O links may reflect coagglomeration patterns (input substitution). We instrument with US benchmark I-O tables.

◦ Labor pooling uses US data, and labor markets are local. So unlikely that there is a big issue here.

◦ Patent data is very difficult to instrument (see Ellison, Glaeser, and Kerr, 2010; Howard et al., 2015). We also use US data.

Results: Total employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(IV)
input-output	0.086^a	0.037^a	0.058^a			0.037^a	0.043^a	0.059^a
	(0.013)	(0.007)	(0.009)			(0.007)	(0.007)	(0.016)
occupation	0.136^a	0.066^a	0.074^a	0.077^a	0.064^a	0.066^a		0.055^a
	(0.011)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)		(0.010)
knowledge	-0.007	0.009	0.010^c	0.010^c	0.009^c		0.010^c	0.007
	(0.011)	(0.005)	(0.006)	(0.005)	(0.005)		(0.005)	(0.005)
input				0.016^b				
				(0.007)				
output					0.046^a			
					(0.007)			
knowledge (make)						0.008		
						(0.006)		
labor movement							0.036^a	
							(0.006)	
Industry FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,292	10,292	9,729	10,292	10,292	10,292	10,292	10,292
R-squared	0.033	0.846	0.847	0.846	0.847	0.846	0.845	0.846
First-stage IV:								
max_io_share_US								0.535^a
								(0.045)
oes_corr								0.306^a
								(0.021)
max_flow_uctp								0.050^b
								(0.011)
First-stage R^2								0.520
First-stage F statistic								144.35

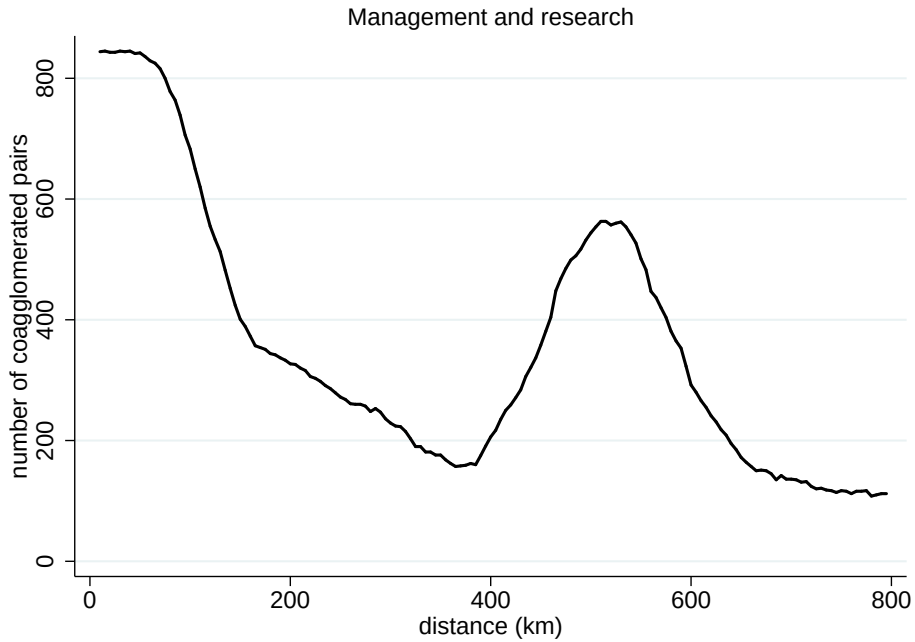
Notes: All regressions include years 2001, 2003, and 2005 with year fixed effects.

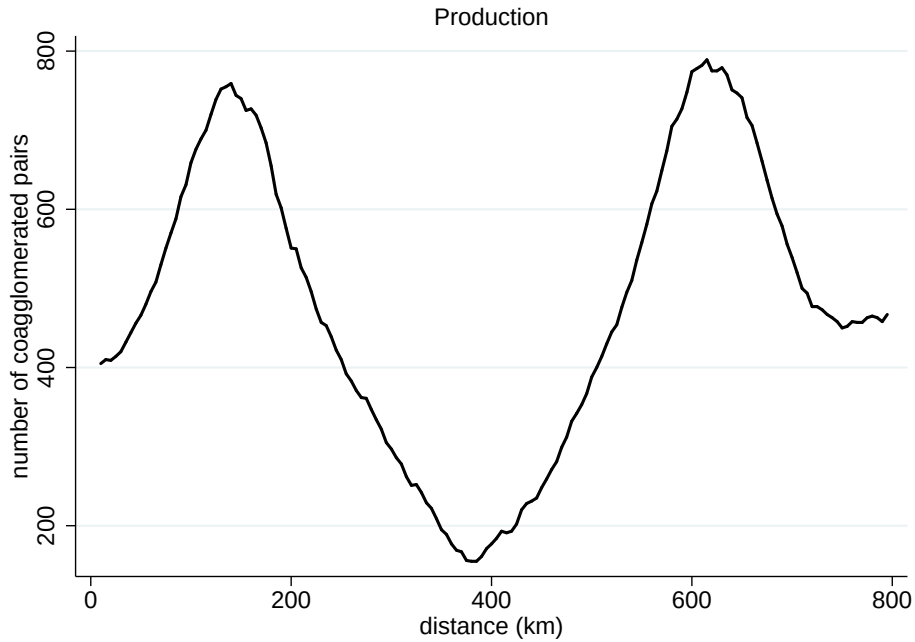
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
input-output	0.044^{a}	0.022^{b}	0.028^{a}	0.026^{a}	0.031^{a}	0.025^{a}	0.007
	(0.011)	(0.010)	(0.010)	(0.006)	(0.007)	(0.006)	(0.004)
occupations	0.127^{a}	0.038^{a}	0.046^{a}	0.031^{a}	0.048^{a}	0.024^{a}	0.009
	(0.011)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.006)
knowledge	-0.015	0.004	0.001	0.003	0.008	0.003	0.005
	(0.010)	(0.008)	(0.008)	(0.004)	(0.005)	(0.004)	(0.004)
locational advantage	0.217^{a}	0.166^{a}	0.161^{a}	0.072^{a}		0.072^{a}	0.059^{a}
	(0.011)	(0.010)	(0.010)	(0.005)		(0.005)	(0.004)
primary input share		-0.475^{a}	-0.418^{a}	-0.221^{a}		-0.210^{a}	-0.185^{a}
		(0.008)	(0.008)	(0.015)		(0.015)	(0.014)
primary output share		-0.182^{a}	-0.183^{a}	-0.071^{b}		-0.068^{b}	-0.060^{c}
		(0.005)	(0.005)	(0.033)		(0.033)	(0.033)
service input share			0.044^{a}	-0.010		-0.010	-0.014
			(0.008)	(0.011)		(0.011)	(0.009)
service output share			0.168^{a}	-0.053^{a}		-0.052^{a}	-0.038^{a}
			(0.009)	(0.014)		(0.014)	(0.012)
multiunit share					-0.047^{a}	-0.031^{a}	-0.033^{a}
					(0.008)	(0.007)	(0.006)
within agglo. share					0.023^{a}	0.008^{c}	0.003
					(0.006)	(0.005)	(0.004)
rho_3rd_ind							0.516^{a}
							(0.017)
Observations	10,292	10,292	10,292	10,292	10,292	10,292	10,292
R-squared	0.078	0.339	0.368	0.856	0.847	0.856	0.895

Notes: All regressions include years 2001, 2003, and 2005 with year fixed effects.

Results: By functional types

Different functions have different coagglomeration profiles





	Results for 'management and research'								
	(No FE)	(FE)	(Excl3)	(I)	(O)	(Make)	(Move)	(Labor)	(IV)
input-output	0.095^a	0.038^a	0.048^a			0.038^a	0.030^a	0.025^a	0.054^a
	(0.012)	(0.006)	(0.009)			(0.006)	(0.006)	(0.006)	(0.011)
occupations (type)	-0.038^a	0.016^c	0.038^a	0.025^a	0.015^c	0.017^c			0.007
	(0.009)	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)			(0.010)
knowledge	0.020	0.013^a	0.014^b	0.015^a	0.013^a		0.012^b	0.011^b	0.012^b
	(0.012)	(0.005)	(0.006)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)
input				0.021^a					
				(0.006)					
output					0.046^a				
					(0.006)				
knowledge (make)						0.012^b			
						(0.005)			
labor movement							0.025^a		
							(0.005)		
occupations (all)								0.048^a	
								(0.006)	
Observations	10,292	10,292	9,729	10,292	10,292	10,292	10,292	10,292	10,292
R-squared	0.011	0.844	0.845	0.843	0.845	0.844	0.845	0.845	0.844

Notes: All regressions include year and industry fixed effects.

	Results for 'production'								
	(No FE)	(FE)	(Excl3)	(I)	(O)	(Make)	(Move)	(Labor)	(IV)
input-output	0.071^a	0.039^a	0.060^a			0.039^a	0.042^a	0.035^a	0.063^a
	(0.013)	(0.007)	(0.009)			(0.007)	(0.007)	(0.007)	(0.017)
occupations (type)	0.174^a	0.064^a	0.065^a	0.073^a	0.063^a	0.064^a			0.052^a
	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)			(0.010)
knowledge	-0.012	0.007	0.010^c	0.008	0.007		0.008	0.007	0.005
	(0.011)	(0.005)	(0.006)	(0.005)	(0.005)		(0.005)	(0.005)	(0.006)
input				0.020^a					
				(0.007)					
output					0.046^a				
					(0.007)				
knowledge (make)						0.006			
						(0.006)			
labor movement							0.043^a		
							(0.007)		
occupations (all)								0.075^a	
								(0.008)	
Observations	10,292	10,292	9,729	10,292	10,292	10,292	10,292	10,292	10,292
R-squared	0.042	0.818	0.820	0.817	0.818	0.818	0.817	0.818	0.817

Notes: All regressions include year and industry fixed effects.

Differences in the patterns of significant coefficients

	Management & research		**Production**	
	Without controls	With controls	Without controls	With controls
Input-output	83%	100%	100%	100%
Occupations	83%	92%	100%	100%
Knowledge	92%	50%	8%	0%

Where do we stand?

◦ Coagglomeration patterns of different functional employment types display different spatial profiles.

◦ Input-output links and labor market pooling are on par.

◦ Input-output links are through the bench the most robust mechanism.

◦ Knowledge sharing is important for 'management and research', but not for 'production'. Labor market pooling is more important for production.

Functional splits allow to more cleanly identify the Marshallian forces (not 'average effects' across heterogeneous functions).

Some thorny identification issues remain to be addressed more fully (third-industry effects; within-firm agglomerations).

We cannot exclude that a small number of 'focal industries' may drive most of the coagglomeration patterns. Generally limited explanatory power of the Marshallian determinants.

Caution. Our analysis may not straightforwardly extend to the context of developing countries:

◦ the source of agglomeration economies may be the plant or 'entrepreneur' (Howard et al., 2015)

◦ agglomeration forces may need to be measured differently, especially for skilled labor and knowledge