Journal of Urban Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

The dynamics of industry agglomeration: Evidence from 44 years of coagglomeration patterns[☆]

Mathieu P.A. Steijn[a,1,*], Hans R.A. Koster[a,2], Frank G. Van Oort[b,3]

[a] Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands
[b] Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA Rotterdam, Netherlands

A R T I C L E I N F O

JEL classification:
L14
L60
R32
F16
O33
R40

Keywords:
Coagglomeration
Labor market pooling
Input-output linkages
Knowledge spillovers

A B S T R A C T

Evidence abounds that agglomeration patterns have changed over time, but little is known about changes in the underlying determinants of agglomeration. We analyze 44 years of coagglomeration patterns of U.S. manufacturing industries and show that over time, input-output linkages and labor market pooling have become less important determinants of industry agglomeration, while knowledge spillovers have become more important. We show that trade and technology shocks are strongly associated with the decline in labor market pooling and the increase in knowledge spillovers. The downward trend in input-output linkages is associated with an increase in trade competition but not with a decrease in the transportation costs of goods.

1. Introduction

Economies of agglomeration are key in understanding the spatial distribution of economic activities (Ellison and Glaeser, 1999; Duranton and Overman, 2005). A well-established literature documents large changes in agglomeration patterns (see for example Glaeser, 2011; Moretti, 2012) but pays limited attention to the changing role of agglomeration determinants that may explain these patterns (Ellison et al., 2010; Moretti, 2012; Combes and Gobillon, 2015; Storper, 2018).[4]

A natural starting point from which to study changes in agglomeration determinants is the classification by Marshall (1890) into labor market pooling, input-output linkages, and knowledge spillovers. Ellison et al. (2010) were the first to empirically distinguish between the importance of each of these agglomeration determinants by regressing the pairwise coagglomeration intensity of U.S. manufacturing industries in 1987 on the extent to which they employ similar workers, sell or buy from each other, and use similar technologies. They find that input-output linkages matter most, followed by labor market pooling and knowledge spillovers.

Since Ellison et al. (2010) and the prior work by Dumais et al. (2002), there have been a number of studies that have examined coagglomeration; however, most of these studies neither allow for heterogeneity in the agglomeration determinants between industries nor measure

☆ We would like to thank in particular Bill Kerr, Ed Glaeser, Frank Neffke, Dario Diodato, and the editor Nathaniel Baum-Snow, but also David Autor, Pierre-Alexandre Balland, Ron Boschma, Jasmina Chauvin, Gilles Duranton, Koen Frenken, Henri de Groot, César Hidalgo, Jos van Ommeren, Anna Salomons, Michael Storper, Will Strange, two anonymous referees, and participants at seminars at NBER, Harvard University, MIT Media Lab, Utrecht University, Vrije Universiteit Amsterdam, UEA New York and Amsterdam, GEOINNO2018, and ERSA Groningen for useful comments on this paper. We also wish to thank Gilles Duranton, Sergio Petralia, and Bill Kerr for providing data. This work has benefited from grant 438-13-406 from JPI Urban Europe (Steijn and van Oort).

* Corresponding author.
E-mail addresses: m.p.a.steijn@vu.nl (M.P.A. Steijn), vanoort@ese.eur.nl (F.G. Van Oort).
1 Mathieu is also affiliated with the Department of Human Geography and Planning, Utrecht University, the Netherlands. Part of this work was written while Mathieu was a visiting research fellow at Harvard University.
2 Hans is also a research fellow at the Tinbergen Institute (the Netherlands) and affiliated with the Centre for Economic Policy Research (UK).
3 Frank is also affiliated with the Tinbergen Institute (the Netherlands).

4 Using historical data, but with different methodologies to ours, the studies by Kim (1995); Dumais et al. (2002); Klein and Crafts (2012, 2020) and Hanlon and Miscio (2017) suggest that agglomeration patterns, as well as their determinants, may change considerably over time.

https://doi.org/10.1016/j.jue.2022.103456
Received 4 December 2020; Received in revised form 1 April 2022

Please cite this article as: M.P.A. Steijn, H.R.A. Koster and F.G. Van Oort, The dynamics of industry agglomeration: Evidence from 44 years of coagglomeration patterns, Journal of Urban Economics, https://doi.org/10.1016/j.jue.2022.103456

changes over time (see *e.g.,* Jacobs et al., 2013; Behrens, 2016; Hanlon and Miscio, 2017; Aleksandrova et al., 2020). Notable exceptions are Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020), as these studies show that there is strong heterogeneity in the intensity of the agglomeration determinants between industries. Of particular interest here is the finding by Faggio et al. (2017) that technology and skill-intensive industries value knowledge spillovers more, while labor market pooling and input-output linkages are more relevant for low-skilled technology-extensive industries.[5]

Industries have become more technology and skill-intensive, which has likely changed their agglomeration determinants. Hence, we think that analyzing the heterogeneity between industries *over time* may be important in understanding the changing role of agglomeration determinants. Industry agglomeration changes through the growth, closure, opening and relocation of establishments. These establishment dynamics have likely been influenced by developments in trade competition, technological progress and transportation costs of goods (see Bloom et al., 2016; Brynjolfsson and Hitt, 2000 and Glaeser and Kohlhase, 2004); and these factors are also associated with an increase in technology and skill intensity.

In this paper, we study the dynamic nature of agglomeration economies. We assess the changes of Marshall's agglomeration determinants over time and document how these changes are related to changes in trade competition, technological change, and transportation costs. Our analysis consists of three steps. First, we exploit panel data to explain manufacturing agglomeration by proxies for labor market pooling and input-output linkages, as well as an improved proxy for knowledge spillovers. Second, we identify changes in sources of agglomeration economies over time by estimating year-by-year regressions. Third, we explore industry-year heterogeneity and test to what extent these three channels of economic change can be associated with changes in agglomeration determinants.

We invest considerable effort in digitizing hard-copy data in order to build a unique, balanced panel dataset with consistent geographical units and industries that covers relevant aspects of coagglomeration, occupations, input-output linkages and patented knowledge for the years for which data are available, *i.e.*, 1970, 1977, 1989, and then for every 5 years until 2014. Our main analyses focus on coagglomeration in Metropolitan Statistical Areas (MSA) of the U.S., as these areas approximately represent functional urban areas.[6]

Each of the three steps introduces innovative measures and produces novel insights. In the first step, we regress coagglomeration on proxies for Marshall's determinants of agglomeration. We build on Ellison et al. (2010) by defining coagglomeration and proxies for labor market pooling and input-output linkages. For knowledge spillovers, we improve on Ellison et al.'s (2010) measure by focusing on the cooccurrence of technologies employed in patented inventions rather than the patent citations between industries. We show that our so-called technological relatedness measure outperforms patent citations in explaining coagglomeration. Following Faggio et al. (2017), we control for the simultaneous dependencies of industry pairs on nonmanufacturing inputs that may be correlated with coagglomeration.

The preferred specification shows that labor market pooling is the most important determinant of agglomeration between 1970 to 2014. An increase of one standard deviation in the extent to which two industries can share workers is associated with an increase of 0.195 of a standard deviation in the extent to which these two industries are coagglomerated in the same MSA. The impacts of knowledge spillovers and input-output linkages are comparable in magnitude, as an increase of one standard deviation in the respective proxies leads to an increase in coagglomeration of 0.104 and 0.090, respectively.

In the second step, we investigate the dynamics in agglomeration determinants. We estimate year-specific regressions of coagglomeration on Marshall's sources of agglomeration economies. We find that knowledge spillovers have become more important, as since 1970 the coefficient on knowledge spillovers has almost doubled. This is strong support for the large stream of literature that suggests that the sharing of ideas is the reason that geographical proximity is still important despite the developments in transportation and communication technologies.[7] On the other hand, we find a clear downward trend in the importance of labor market pooling and input-output linkages, which decreased by approximately 45% and 90%, respectively. Hence, the large changes in agglomeration patterns that are documented in the literature may have been caused by drastic changes in agglomeration determinants.

In a third step, we explore why the determinants of industry agglomeration have changed over time and are different between industries. We estimate industry-year-specific coefficients for each of the agglomeration determinants and project these coefficients on proxies for three major economic trends that have considerably altered the composition of manufacturing industries. These trends are increased trade competition from low-wage countries, routine-biased technological change, and a large decrease in the transportation costs of goods (Glaeser and Kohlhase, 2004; Autor et al., 2013; 2015).[8] We then test whether trade, technology, and transportation shocks are associated with changes in the determinants of industry agglomeration. The results from the third step show that more intense trade competition is associated negatively with labor market pooling and positively with knowledge spillovers. Furthermore, the routine employment share of an industry is associated positively with labor market pooling, while it is associated negatively with knowledge spillovers. For example, a standard deviation increase in routine employment share and trade competition is associated with an increase of approximately 100% and 40% of the median coefficient on labor market pooling, respectively. Hence, the effects are sizable. A likely interpretation is that increased trade competition and technological progress have led to more knowledge-intensive and flexible industrial production processes, both through investments in surviving establishments and the closure of the least competitive establishments (see Brynjolfsson and Hitt, 2000; Bloom et al., 2016). As a result, the demand for standardized routine tasks, and therefore access to a 'common' labor pool, has decreased, while knowledge spillovers related to new (production) technologies have become more important. This interpretation is in line with the results on high-technology/high-education industries found by Faggio et al. (2017).

[5] They relate this outcome to the 'nursery city hypothesis' introduced by Duranton and Puga (2001). The nursery city hypothesis implies that firms first learn about their ideal production process by making prototypes. They then benefit from being located in diverse places. Once firms have found their ideal process, firms switch to mass production and relocate to specialized cities where production costs are lower. Faggio et al. (2017) hypothesize that industries in the early developmental phase of the industry life cycle coagglomerate because of knowledge spillovers; then, when industries become more mature and standardize their production process, they coagglomerate to take advantage of a common labor pool and input-output linkages.

[6] We show that similar results hold at the county level, which cover the entire U.S.

[7] This evidence is in line with suggestions made by, among others, Gaspar and Glaeser (1998); Storper and Venables (2004); McCann (2008); Glaeser (2011); Moretti (2012); Michaels et al. (2019) and Balland et al. (2020), who provide evidence that communication technologies complement face-to-face contact and that knowledge-intensive interactive activities increasingly concentrate in space.

[8] Regarding proxies for trade competition and technological progress, we closely follow Autor et al. (2013) and Autor et al. (2015). We measure trade through import competition from low-wage countries, and we measure technological progress by the share of workers with routine task-intensive jobs. For transportation costs, we follow Glaeser and Kohlhase (2004) by calculating iceberg-like transportation costs for goods.

Interestingly, we find that the transportation costs of goods are not strongly associated with input-output linkages. However, we do find evidence that more intense trade competition is negatively related to input-output linkages. A standard deviation increase in trade competition is associated with a decrease of approximately 50% of the median coefficient on input-output linkages. An extended analysis provides support for the idea that local input-linkages are replaced by input-linkages originating in low-wage countries. In contrast to Faggio et al. (2017), we do not find evidence that the decline in input-output linkages can be explained by industries becoming more technology and skill-intensive.

Related literature. We contribute in several ways to the existing literature. First, most of the previous studies on coagglomeration are cross-sectional, while we use panel data (Ellison et al., 2010; Faggio et al., 2017; 2020). This approach enables us to make improvements regarding identification by including industry-by-year fixed effects, which matters for the results. Second, we improve on the proxy for knowledge spillovers by using a measure based on the co-occurrence of technologies mentioned in patents instead of patent citations. Third, compared to Diodato et al. (2018) who also analyze dynamics in coagglomeration patterns, we use more fine-grained data at the three-digit (SIC) industry level, as well as at the MSA level, and we include a proxy for knowledge spillovers. Fourth, in terms of investigating industrial heterogeneity in coagglomeration, we improve on Faggio et al. (2017) and Diodato et al. (2018) by explicitly explaining industry heterogeneity in a multivariate setting. This is similar to the approach used in Faggio et al. (2020), but we use more detailed data and also exploit temporal variation.

We think the paper relates to the broader literature regarding understanding changes in location patterns of (manufacturing) industries. Most notably, there is a large stream of literature that suggests that the increased demand for geographical proximity is due to an increasing importance of knowledge spillovers. Despite large improvements in communication and transportation technologies, knowledge spillovers still require face-to-face contact (Gaspar and Glaeser, 1998; Storper and Venables, 2004; Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011). The findings of the second step indeed confirm that localized knowledge spillovers have become more important in the last decades.

Contradictory predictions exist in the literature on labor market pooling. According to Moretti (2012), labor market pooling is expected to be on the rise due to the increase in skill levels of the workforce. By contrast, the results of Faggio et al. (2017) suggest that a shift towards more high-technology/high-education industries would lead to a decreasing importance of labor market pooling because the labor required is less standardized. Our results are in line with the latter study, as labor market pooling is becoming a less prominent determinant of industry agglomeration.

Contradictory predictions also exist regarding input-output linkages. On the one hand, Glaeser and Kohlhase (2004) suggest that because the transportation costs of goods have been greatly reduced, input-output linkages are currently likely to be less relevant. On the other hand, McCann and Fingleton (1996), Duranton and Storper (2008) and McCann (2008) argue that input-output linkages may have become more relevant as more competitive knowledge-intensive industries have required more frequent deliveries and more face-to-face interaction, which has led to higher coordination costs. Our results suggest that input-output linkages have become less important, although we cannot attribute this outcome to the reduction in transportation costs of goods.

The third step of is the analysis is related to the large stream of literature that aims to understand why and how the spatial organization of the economy changes. Recall that in this step, we project industry-year heterogeneity in regard to the agglomeration determinants on proxies for important economic trends related to trade, technology and transportation costs of goods. For example, the reduction of trade barriers allowed for more intense trade competition from low-wage countries and meant that low-skilled work has been offshored. Further, the computer revolution brought about fundamental changes in manufacturing. Entire production processes and value chains were reinvented to fully exploit the possibilities of the computer (Brynjolfsson and Hitt, 2000). In the process, computerized machinery took over much of the performance of routine tasks but raised the productivity of workers who perform abstract tasks, in particular those involving complex communication and coordination (Autor et al., 2003; Deming, 2017). As a result, both trade competition and technological progress have led to the downsizing and closure of establishments that were more low-skill labor intensive, low technology, and more likely to produce standardized products, while the surviving establishments have increased their investments in R&D, workers' skills, and capital (Brynjolfsson and Hitt, 2000; Bernard et al., 2006; Holmes and Stevens, 2014; Bloom et al., 2016; Pierce and Schott, 2016). This change in the composition of establishments within industries has altered colocation patterns and now particularly represents the location choices of knowledge and skill-intensive establishments. More specifically, we find that an increase in import competition and a decrease in the routine employment share are associated with stronger knowledge spillovers and weaker labor market pooling. Glaeser and Kohlhase (2004) demonstrate that over time, the transportation costs of goods have strongly decreased, which may have also incentivized establishments to change locations. However, we find little evidence that the decline in the transportation costs of goods can explain the decreasing trend in input-output linkages.[9]

The rest of the paper is organized as follows. In Section 2, we introduce the econometric framework, followed by a discussion of the various datasets used in the analyses in Section 3. We report and discuss the results in Section 4, and we conclude the paper in Section 5.

2. Empirical framework

This section outlines the econometric framework. We first focus on identifying the impact of Marshall's sources of agglomeration economies on coagglomeration patterns. Second, we aim to study changes over time in agglomeration determinants. Third, we project industry-year-level estimates of agglomeration determinants on proxies for trade, technology and transport costs of goods.

2.1. Step 1: Determinants of industry agglomeration

We aim to analyze the factors that impact coagglomeration of industries over time. Following Ellison and Glaeser (1997), the coagglomeration C_{ijt} of industries i and j in year t is as follows:

$$C_{ijt} = \frac{\sum_{m=1}^{M}(s_{imt} - x_{mt})(s_{jmt} - x_{mt})}{1 - \sum_{m=1}^{M} x_{mt}^2}, \tag{1}$$

where s_{imt} is the share of industry's i employment in location m in year t. More specifically, $s_{imt} = E_{imt}/\left(\sum_{m=1}^{M} E_{imt}\right)$, where E_{imt} captures the employment of industry i in location m. Further, x_{mt} is the size of location m in year t, which is measured by the employment share of the location in the total employment of the nation. Our main results are based on estimates from 1970 to 2014 at the MSA level, of which there are 363 in our data; however, we will also show results at the county level, which covers the entire U.S.[10]

[9] This does not mean that transportation costs do not matter, because the wages of skilled workers have strongly increased in the last decades (*e.g.,* due to technological progress Autor, 2019). This suggests that the transportation costs of *people* have likely increased as their value of time has increased (Glaeser and Kohlhase, 2004; Koster and Koster, 2015).

[10] Note that equation (1) implies that if both industries are not present in an area, this leads to positive coagglomeration values. We think this makes sense, as the industries then do not locate where the other is *not* present. As a robustness check, when calculating the coagglomeration index, we also removed areas in which both industries are not present, which led to very similar results. These results are available upon request.

Following the literature, we construct proxies for labor pooling, namely, $\mathcal{LP}_{ijt}$ and for input-output linkages, namely, $\mathcal{IO}_{ijt}$; we also construct an alternative proxy for knowledge spillovers, which we refer to as technological relatedness, *i.e.*, $\mathcal{TR}_{ijt}$. Regarding labor market pooling, firms can be located near firms that employ workers with similar skills and expertise. There are three benefits to this common labor pool. First, employing or laying off workers in the face of fluctuating demand becomes easier (Krugman, 1991); second, the matches between available job positions and workers improve (Helsley and Strange, 1990); third, workers will invest more in acquiring industry-specific skills (Rotemberg and Saloner, 2000). On the other hand, labor market pooling also increases the risk of losing workers to competitors, i.e., so-called labor poaching, which gives firms less incentive to coagglomerate (Matouschek and Robert-Nicoud, 2005; Combes and Duranton, 2006). With respect to input-output linkages, industries can coagglomerate with upstream industries or downstream industries to save on the costs of transporting inputs or outputs, respectively. Finally, the colocation of firms that employ similar technologies can lead to useful ideas 'spilling over' from one firm to another.

We address the potential correlation of unobservables with our variables of interest in multiple ways. First, following Faggio et al. (2017), we include dissimilarity indices, namely, $\mathcal{DS}_{ijt,k}$, to capture the shared dependence on various inputs, $k = 1, \ldots, 7$.[11] Second, we improve on the previous literature by controlling for the overall tendency of a three-digit industry to coagglomerate with other industries in a specific year. More specifically, we include industry i ×year and industry j ×year fixed effects. Our baseline specification is given by the following:

$$\mathcal{C}_{ijt} = \alpha\mathcal{LP}_{ijt} + \beta\mathcal{IO}_{ijt} + \gamma\mathcal{TR}_{ijt} + \sum_{k=1}^{7}\zeta_k\mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}, \quad (2)$$

where α, β, and γ are the main parameters of interest; ζ_k and $\forall k$ are additional parameters; λ_{it} and λ_{jt} are industry i-by-year and industry j-by-year fixed effects, respectively; and ϵ_{ijt} is an error term.[12]

Furthermore, we test for the presence of omitted variable bias by estimating bias-adjusted coefficients, following Oster (2019). She shows that the effects of the inclusion of observable control variables that lead to changes in R^2 and coefficient movements can be used to calculate the bias due to unobserved omitted variables. This process relies on two parameters, namely, $R^2_{\max}$, which is the R^2 from the hypothetical regressions of the dependent variables on all observables and unobservables, and δ, which depicts the relative degree of selection on observed and unobserved variables. Note that $R^2_{\max}$ is very unlikely to be 1 in most empirical applications due to measurement error in the dependent variable. Following Oster (2019), we set δ to 1 and $R^2_{\max}$ to 1.3 times the R^2 of the baseline regression with controls and industry-year fixed effects (see equation (2)).

Reverse causation may be another potential concern in equation (2). More specifically, while firms in industries with strong Marshallian links may choose to locate together, conversely, firms that are located close together may also forge Marshallian links.[13] To mitigate this issue we follow Ellison et al. (2010) by instrumenting the Marshallian agglomeration variables with proxies based on areas where one industry is present but the other is (virtually) not present, and *vice versa*. Even in industry pairs with high coagglomeration values, there will typically be some establishments that are not located near establishments in the other industry. By focusing on establishments in industry i that are not near establishments in industry j, their labor hiring decisions and knowledge spillovers are less likely to be driven by joint omitted factors or by the influence of proximity to the other industry; hence, coagglomeration does not affect labor market pooling or knowledge spillovers[14]

2.2. Step 2: Changes in agglomeration determinants

To identify changes in the determinants of industry agglomeration over time, we estimate a similar specification as that explained in (2), but we add year-specific coefficients for each of the agglomeration determinants as follows:

$$\mathcal{C}_{ijt} = \alpha_t\mathcal{LP}_{ijt} + \beta_t\mathcal{IO}_{ijt} + \gamma_t\mathcal{TR}_{ijt} + \sum_{k=1}^{7}\zeta_{t,k}\mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}, \quad (3)$$

where α_t, β_t, and γ_t are year-specific coefficients.

2.3. Step 3: Exploring industry-level and temporal heterogeneity

We expect that there is considerable heterogeneity in the agglomeration forces across industries, as shown by Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020). Hence, changes in the economy driven by developments in the *(i)* trade competition, *(ii)* technological progress, and *(iii)* transportation costs of goods likely influenced the importance of the determinants of industry agglomeration, which in turn altered coagglomeration patterns.

Trade competition is associated with the displacement of millions of workers within U.S. manufacturing industries in the last decades, even though the value added has kept increasing (Pierce and Schott, 2016). Thus, industries have become more technology and skill intensive, have focused more on niche products, and have invested more in R&D. These shifts happened because of *(i)* trade-induced technological changes within the surviving establishments, as well as *(ii)* a reduction in the employment and survival probability of more labor intensive low-technology establishments. Such changes in the composition of establishments within industries likely altered colocation patterns and now particularly represents the location choices of knowledge and skill-intensive establishments (Bernard et al., 2006; Holmes and Stevens, 2014; Pierce and Schott, 2016; Bloom et al., 2016). Following the literature, we construct a measure – import penetration – that is based on the exposure of an industry to imports from low-wage countries in year t.

Technological changes, particularly those related to the computer revolution, have likely fueled changes in agglomeration determinants (see Glaeser, 2011; Moretti, 2012). Computers excel at performing so-called routine tasks and have thus replaced the often middle-skilled workers who perform these tasks. On the other hand, high-skilled workers are complemented by technological progress (Brynjolfsson and Hitt, 2000; Autor et al., 2003).[15] In particular, the demand for workers who

[11] Faggio et al. (2017) argue that these are the most obvious candidates for omitted variables, as other unobserved location characteristics need to have a very particular structure, which implies that industry agglomeration is correlated to both coagglomeration and the strength of the linkages between sectors measured by labor market pooling, input-output sharing and knowledge spillovers.

[12] We also considered exploiting temporal variation in coagglomeration by including industry fixed effects. However, there is too little meaningful variation to obtain reasonably precise coefficients.

[13] Note that Faggio et al. (2017) argue that coagglomeration that leads to productive links should be considered as agglomeration economies. For example, if two firms forge an input-output link *after* they have coagglomerated, then this is also a form of input-output sharing. Similar examples can be given for labor market pooling and knowledge spillovers. Hence, it is questionable whether reverse causation is really an issue that should be tackled. Fortunately, the OLS results are not fundamentally different from the IV results.

[14] Ellison et al. (2010) also mitigate reverse causality concerns by employing data from the United Kingdom to calculate Marshallian proxies in the U.K. to instrument for the corresponding U.S. variables. Apart from the limited historical data availability for the U.K., we consider the 'spatial' instruments to be more convincing, as similar reverse causation issues cannot be ruled out in the U.K. data.

[15] Note that technological progress is routine-biased rather than skill-biased; low-skilled workers, who generally perform manual tasks, fared better than middle-skilled workers during such progress, which is a phenomenon known as job polarization (Goos et al., 2009; Autor and Dorn, 2013).

perform interactive tasks increased (Autor et al., 2003; Deming, 2017; Balland et al., 2020). As these jobs rely on face-to-face contact, the demand for geographical proximity increases (Storper and Venables, 2004; McCann, 2008; Glaeser, 2011). Furthermore, technological progress allowed establishments to vertically disintegrate, relocate and outsource parts of their activities, thereby leading to fundamental changes in the distribution of the remaining establishments, which led to changes in coagglomeration patterns (Brynjolfsson and Hitt, 2000; Duranton and Puga, 2005; Glaeser and Ponzetto, 2007; Balland et al., 2020). As a proxy for exposure to technology, we therefore take the share of routine employment in a sector, following Autor et al. (2015).

Marshall (1890) already noted that geographical proximity implies reductions in transportation costs. Glaeser and Kohlhase (2004) show how the transportation costs of goods have decreased sharply in the last decades, which could be another reason for changes in agglomeration determinants. Following Glaeser and Kohlhase (2004), we proxy for the transportation costs of goods by the share of expenditure in each sector spent on transportation.

We apply a two-step estimation approach, similar to Faggio et al. (2020), to explain the observed variation in agglomeration determinants by trade competition, technological progress and transportation costs. First, we estimate a specification that is very similar to the baseline regression in (2) but contains *industry*-year-specific coefficients for each of the agglomeration determinants:

$$C_{ijt} = \alpha_{it}\mathcal{LP}_{ijt} + \beta_{it}\mathcal{IO}_{ijt} + \gamma_{it}\mathcal{TR}_{ijt} + \sum_{k=1}^{7}\zeta_{it,k}\mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}. \quad (4)$$

Equation (4) yields an estimated coefficient for each Marshallian force for each industry in each year.

Second, we regress $\hat{\alpha}_{it}$, $\hat{\beta}_{it}$, and $\hat{\gamma}_{it}$ of industry i in time period t on the three industry-level variables, thus capturing trade, technology, and transportation costs $\mathcal{IC}_{it}$ while controlling for time fixed effects μ_t:

$$\{\hat{\alpha}_{it}, \hat{\beta}_{it}, \hat{\gamma}_{it}\} = \sum_{\ell=1}^{3}\eta_{\ell}\mathcal{IC}_{it,\ell} + \mu_t + \xi_{it}, \quad (5)$$

where η_{ℓ} and $\forall\ell$ are parameters to be estimated, μ_t are year fixed effects and ξ_{it} is an error term.

We mitigate concerns related to omitted variable bias by estimating Oster-style bias-adjusted coefficients by adding variables on average establishment size and the capital-labor ratio. Given that these variables could be considered as proxy controls, the results should be interpreted as lower bounds (Angrist and Pischke, 2008).

One may be concerned that reverse causality is also an issue here. Agglomeration determinants are potentially correlated with local economic conditions that affect the demand for imports from low-wage countries. To instrument for import penetration, we follow Autor et al. (2015) by calculating import penetration in other high-wage countries.[16] The predicted part of import penetration due to trade shocks in all of these countries is likely due to a rising comparative advantage of low-wage countries and/or a decrease in trade costs rather than import demand changes due to local conditions.

Furthermore, the investment in technology may be dependent on the size the local labor pool, which makes it easier to employ or replace workers when necessary. Also, when firms are coagglomerated because of knowledge spillovers, the availability of new tacit knowledge is likely to encourage firms not to standardize but rather to continuously reinvent production processes (see Duranton and Puga, 2001; Faggio et al., 2017). For the routine employment share, we construct spatial instruments as described previously, *i.e.*, by focusing on areas where an industry does not coagglomerate to make use of a common labor pool, buyer-supplier relations, and knowledge spillovers.[17] In these MSAs, the share of routine employment is unlikely to be influenced by these agglomeration determinants.

Finally, the share of transportation expenditure may be influenced by the extent to which an industry coagglomerates with buyers or suppliers, because stronger coagglomeration leads to shorter distances and therefore less transport expenditures. Transportation expenditure as a share of total expenditure is dependent on not only transportation expenditure but also other expenditures. The latter are unlikely to be influenced by agglomeration determinants. We capture other expenditures by the mean value of a ton. Although using the mean value of a ton as an instrument for the share of transport expenditures addresses reverse causality, this approach may lead to omitted variable bias, as a higher value of a ton is likely correlated with more knowledge/skill intensive products, which may not be fully captured by our trade and technology measures. Therefore, we also add the value of a ton in 1970 as a control variable.

To obtain standard errors, and to address the issue that $\hat{\alpha}_{it}$, $\hat{\beta}_{it}$, and $\hat{\gamma}_{it}$ are estimated parameters, we bootstrap this two-step estimation procedure by randomly selecting industry $ij - ji$ pairs.

3. Data and descriptives

This section discusses the construction of the data. We discuss our proxies for agglomeration sources, in particular, with respect to our alternative proxy for knowledge spillovers, namely, technological relatedness. We further gather data related to industry-level proxies for trade competition, technology exposure, and transportation costs of goods. We close this section by reporting descriptive statistics in Section 3.3.

3.1. Determinants of industry agglomeration

Coagglomeration. We calculate industry-pair-specific coagglomeration measures using County Business Patterns (CBP) gathered by the U.S. Census Bureau; these data are available online from 1986 onwards. Raw data from before 1987 were kindly provided by Duranton et al. (2014). We construct a balanced panel dataset of consistent counties and industries using 3-digit SIC '87 classifications, which is discussed in more detail in Appendix A.1.

Labor market pooling. We use the National Industrial-Occupation Employment Matrix (NIOEM) published by the Bureau of Labor Statistics (BLS). These data are digitally available online from 1989 onwards for manufacturing employment. Data for 1970 and 1978 are obtained from hard-copy reports by the BLS (1981) using optical character recognition (OCR). We then manually checked the digitized data to avoid errors. We built a composite job classification, in which job occupations are consistent over time (see Appendix A.2.1 for further details). In line with the previous literature, we then calculate the correlation in the share of employees across occupations between each industry pair.

Input-output linkages. For the construction of input-output linkages, we employ the use tables of the U.S. Bureau of Economic Analysis (BEA), as well as their concordance tables. We define $Input_{i\leftarrow j}$ as the share of industry i's inputs that come from industry j, while $Output_{i\rightarrow j}$ is defined as the share of output sold to industry j by industry i.[18] We then define input-output linkages between industry i and j as $\mathcal{IO}_{ij} = \max(Input_{i\leftarrow j}, Output_{i\rightarrow j})$.

In contrast to previous work, we do not combine $\mathcal{IO}_{ij}$ and $\mathcal{IO}_{ji}$ into a single measure as these are directional variables and are therefore relevant for estimating industry-year-specific coefficients later on.

[16] Autor et al. (2015) use data from Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain and Switzerland. We use data from the U.N. Comtrade database for the same countries, but we also include France and the Netherlands, which is to be explained later in more detail.

[17] More specifically, we estimate the impact of agglomeration determinants *for each MSA m separately*. We then calculate the routine employment share of each industry in the 50 MSAs where $\hat{\alpha}_{imt}$, $\hat{\beta}_{imt}$, and $\hat{\gamma}_{imt}$ are the smallest.

[18] Similar to Ellison et al. (2010), we calculate these shares relative to all suppliers and customers, some of whom may be outside manufacturing; see Appendix for more details.

Knowledge spillovers. With respect to knowledge spillovers, we show results using the patent citation measure, which follows the existing literature, but we prefer the results based on a novel measure, namely, technological relatedness. Patent citations are used as a proxy for knowledge spillovers by linking technology classes to industries and then calculating the share of citations to each industry by patents of each industry. However, not all technologies, and therefore not all industrial knowledge, mentioned in cited patents are actually used in the citing patent. For example, in our sample, there are some patents that cite up to 1,500 other patents. Hence, it is likely that a large share of the technological knowledge of these cited industries is irrelevant for the described invention.

Furthermore, the share of patent citations of industry i compared to that of industry j only normalizes the results in terms of patents for the size of industry i and not for the size of industry j. This turns out to matter substantially. The correlation between the total number of patents of industry j and the share of patent citations received from industry i is 0.72. The industry with the most patents, *i.e.,* SIC356 (general industrial machinery and equipment), is one of the most important cited industries for approximately 65% of all the industries. As a result, the sheer size of industry SIC356 seems to suggest knowledge spillovers between industries that intuitively do not make much sense. We elaborate on this issue in Appendix A.2.3.

Instead, we prefer to link *technologies* in patent documents to industries and then focus on the *co-occurrence of industries* in each patent. Note that a single patent can mention more than one technology class. We do so by estimating a network-based probability measure, known as relatedness, that normalizes for the size of both industry i and industry j (see Hidalgo et al., 2018).[19] Following Steijn (2021), we define *technological relatedness*, namely, $\mathcal{TR}_{ijt}$, between industry i and industry j in year t as follows:

$$\mathcal{TR}_{ijt} = \frac{\mathcal{O}_{ijt}}{\left(\frac{\mathcal{S}_{it}}{\sum_{n=1}^{N}\mathcal{S}_{nt}}\frac{\mathcal{S}_{jt}}{(\sum_{n=1}^{N}\mathcal{S}_{nt})-\mathcal{S}_{it}} + \frac{\mathcal{S}_{jt}}{\sum_{n=1}^{N}\mathcal{S}_{nt}}\frac{\mathcal{S}_{it}}{(\sum_{n=1}^{N}\mathcal{S}_{nt})-\mathcal{S}_{jt}}\right)\frac{\sum_{n=1}^{N}\mathcal{S}_{nt}}{2}}, \qquad i \neq j, \tag{6}$$

where $\mathcal{O}_{ijt}$ is the number of co-occurrences on patents between industry i and industry j in year t, *i.e.,* a count of the number of times that industries i and j are associated with the same patent. $\mathcal{S}_{it}$ and $\mathcal{S}_{jt}$ are the number of co-occurrences involving industry i and industry j in t, respectively, and $n = 1, \dots, N$ refer to industries. The rationale behind the measure is to divide the observed number of co-occurrences on patents between two industries by the expected number os co-occurrences if all the occurrences of industries would have been assigned to patents randomly. Hence, $\mathcal{TR}_{ijt} = 1$ indicates that exactly the same amount of co-occurrences has been observed as could be expected from a random distribution.

To calculate $\mathcal{TR}_{ijt}$, we then need to link technologies listed on patents to industries. We employ the concordance table by Kerr (2008) between the technology classification by the United States Patent and Trademark Office (USPTO) and the SIC classification. This concordance is based on a limited time period between 1990 and 1993 during which the Canadian Patent Office recorded both the technology classes involved in the invention on the patent and the industry class of the patenting firm (see Silverman, 2002; Kerr, 2008). Data on technology classes per patent is obtained from the USPTO (see Marco et al., 2015).[20]

Dissimilarity measures. Faggio et al. (2017) argue that two industries might be coagglomerated because both depend on an input that is unevenly distributed over space, such as transportation infrastructure (*e.g.,* ports) or natural resources. As a result, the two industries end up coagglomerating not because of Marshallian linkages but because of these location endowments. We follow their approach in addressing this issue by measuring how similar two industries are in their dependency on inputs from agriculture; mining; water; energy; transportation; finance, insurance, and real estate (FIRE); and other services. We refer to Appendix A.2.4 for details.

Spatial instruments. We develop spatial instruments for labor market pooling and knowledge spillovers following Ellison et al. (2010) to address the reverse causality concerns between coagglomeration and agglomeration determinants.[21] Using the CBP, we identify MSAs where industry i is present but industry j is (virtually) absent, and *vice versa*. Then, we calculate the correlation in labor share per occupation of industry i in the former MSAs with that of industry j in the latter MSAs by employing the IPUMS census samples by Ruggles et al. (2018). We use the combined 1970 1% *metro fm1* and 1970 1% *metro fm2* samples for 1970, the 1980 5% sample for 1977, the 1990 5% state sample for 1989 and 1994, the 2000 5% sample for 1999 and 2004, the 2009 *ACS 3yr* sample for 2009, and the 2014 *ACS 5yr* sample for 2014. We use the location of inventors from Petralia et al. (2016) for 1965 to 1975 and from Hall et al. (2001) for more recent years to construct a spatial instrument for technological relatedness.

For each industry i, Ellison et al. (2010) selected the 25 MSAs for which industry j is the least present but industry i is strongly present. We choose to increase the number of MSAs from 25 to 50, as the IPUMS sample size in 1970 is smaller than that of the 1990 data used in Ellison et al. (2010). Recall that there are a total of 363 MSAs in the data.

3.2. Trade, technology, and transportation costs

Thus far, we have discussed the data used to estimate the impact of agglomeration determinants on coagglomeration. In the third step of the analysis, we further aim to explore how these agglomeration determinants are related to proxies for trade, technology, and transportation costs.

Trade. The effects of trade on establishments, and therefore coagglomeration patterns, are thought to result from import competition from low-wage countries (Bernard et al., 2006; Bloom et al., 2016). Bernard et al. (2006) define these trade partners as countries that, across the entire time period, have a GDP per capita that is less than 5% of the GDP per capita in the U.S.; however, they also consider thresholds of 10% and 15%. We choose to use this last threshold as, due to our long sample period, China would fall out of the sample in 2014 if the other thresholds are used. Data on GDP per capita are obtained from the World Bank, and the full list of low-wage countries is shown in Appendix.

We deviate slightly from Autor et al. (2013, 2015) as our data are at the industry-level. We follow the value share approach of

[19] An argument in favor of using patent citations is that a citation captures an actual spillover. However, Jaffe et al. (2000) surveyed inventors and found that only 18% of the citations could be considered as an actual knowledge spillover, which is defined as stemming from direct contact between inventors or from attending a product demonstration.

[20] We caution that no patent-based measure can capture the full extent of knowledge spillovers and that part of the knowledge spillovers operate via input-output linkages, as suggested by Duranton and Storper (2008), and via labor market pooling, as shown by Serafinelli (2019). We also looked at the concordance table between technologies and industries of Goldschlag et al. (2020) used by Diodato et al. (2018). This crosswalk is based on keyword analysis in both technology and industry classifications. We found that for 1994, the correlation between technological relatedness following the concordance of Kerr (2008), and thus respectively of Goldschlag et al. (2020), is only 0.36. This low correlation coefficient does not give us enough confidence to use the Goldschlag et al. (2020) concordance, as 1994 is closest to the 1990–1993 reference years of the concordance by Kerr (2008), which means that Kerr (2008) should be the most accurate.

[21] Ellison et al. (2010) used material input trailers to construct spatial instruments for input linkages. These data are unfortunately unavailable to us. On the other hand, Ellison et al. (2010) did not have data with which to construct spatial instruments for knowledge spillovers.

Table 1
Descriptive statistics.

Statistic	Mean	St. Dev.	Min	Max
Coagglomeration	0.0001	0.007	−0.021	0.031
Labor market pooling	0.289	0.229	0.008	0.954
Input-output linkages	0.006	0.018	0.000	0.129
Technological relatedness	1.764	4.160	0.106	36.225
Patent citations	0.007	0.012	0.00001	0.066

Note: The number of observations is 155,680.

Table 2
Descriptive statistics - Second stage.

Statistic	Mean	St. Dev.	Min	Median	Max
Import penetration	0.181	0.231	0.00001	0.074	0.917
Routine employment share	0.212	0.134	0.066	0.165	0.673
Transportation costs	0.048	0.046	0.001	0.038	0.285

Notes: We report the independent variables here. The number of observations is 1120.

Bernard et al. (2006) and Bloom et al. (2016) by calculating the import penetration, namely, $\mathcal{IMP}_{it}$, for industry i in year t as the share of imports M_{it} from low-wage countries present in the total amount of imports in this industry; *i.e.*, $\mathcal{IMP}_{it} = M_{it}^{\text{Low-wage countries}} / M_{it}^{\text{World}}$. Trade data and concordance tables are obtained from the U.N. Comtrade database.

This measure simplifies import penetration as it ignores nationally produced and consumed goods. Another popular measure is the share of low-wage imports in 'apparent consumption', which is defined as the total of imports plus domestic production minus exports (Bloom et al., 2016). By employing the NBER-CES manufacturing database by Bartelsman and Gray (1996), we will show that similar results hold using this measure. However, we do not prefer this measure as it has limitations and incomplete industrial coverage (see Appendix A.3.1).

Technology. We follow Autor and Dorn (2013) and Autor et al. (2013, 2015) in defining technological progress as the decreasing share of workers performing routine tasks. To this end, we use the IPUMS census samples created by Ruggles et al. (2018).[22] The exact definition of routine task intensity is given in Appendix A.3.2.

Transport costs of goods. To obtain a proxy for the transport costs of goods, we employ the BEA's use tables to calculate the share spent on transportation sectors (SIC41-47) of the total use value of an industry i in year t. Note that the total use (*i.e.,* demand) value of an industry is equal to the total make (*i.e.,* supply) value. We correct for the underestimation of transport expenditure due to stronger use of private trucks in earlier time periods by employing the Commodity Flow Survey (CFS) (see Appendix A.3.3 for details).

3.3. Descriptive statistics

Determinants of industry agglomeration. Before we report descriptives, please note that the data contain some outliers, which are due to measurement errors and extreme values. The latter are a result of *e.g.,* strong dependencies in input-output linkages, or in the case of coagglomeration, they are due to extremely small industries. For instance, industry SIC237 (fur goods) had approximately 5,000 employees in 1970 but only 40 in 2014, which results in extreme coagglomeration values in the latter year.

In what follows, we cap outliers for all variables to limit the disproportionate impact of a few industries by setting values below the 1st percentile and above the 99th percentile to the 1st percentile and 99th percentile, respectively.

Table 1 reports the descriptive statistics of the main variables, while histograms of the variables, as well as the developments over time, are reported in Appendix A.4. By construction, the mean of the coagglomeration index is close to zero. The negative minimum value on labor market pooling reveals that some industries, such as SIC241 (logging) and SIC372 (aircraft and parts), have a negative correlation in employment shares per occupation, while industries SIC233 (women's, misses', and juniors' outerwear) and SIC231 (men's and boys' suits, coats, and overcoats) use virtually the same type of workers and therefore have a correlation value close to 1. The maximum value on technological relatedness indicates that technologies associated with SIC391 (jewelry, silverware, and plated ware) and SIC344 (secondary smelting and refining of nonferrous) are 36 times more likely to co-occur on a patent compared to a random assignment of technologies to patents. The maximum value on patent citations indicates that 6.6% of the citations by patents that are associated, for example, with SIC201 (meat products) also cite patents that are associated with SIC356 (general industrial machinery and equipment). The dissimilarity indices are measured as one half of the absolute difference in the share of inputs between industry i and industry j. For example, a maximum of 0.341 for the mining dissimilarity index indicates that there is an absolute difference of 68.2% in the share of inputs received from sectors related to mining between SIC201 (meat products) and SIC291 (petroleum refining).

Figure A2 in Appendix A.4 shows that both the mean of coagglomeration and each of the agglomeration determinants are relatively stable over time. Note that the mean of the coagglomeration index per time period in Fig. A2 is not indicative of changes in overall coagglomeration patterns as it is close to zero by construction. The correlation between the coagglomeration values in 1970 and 2014 is only 0.51, which strongly suggests that coagglomeration patterns have changed considerably. Furthermore, the variance has decreased by 60% since 1970. As Faggio et al. (2017) show that low-technology industries have more extreme coagglomeration values (*i.e.* a larger variance), this outcome is in line with industries becoming more technology-intensive over time. Note that the correlation between the values of 1970 and 2014 for labor market pooling, input-output linkages, and knowledge spillovers are 0.80, 0.62, and 0.98, respectively.

Trade, technology, and transportation costs. Table 2 gives the descriptive statistics of our proxies for trade, technology, and transportation costs. To limit the effect of outliers, we again cap the outliers. As we have 140 industries and 8 time periods, the number of observations is 1,120.

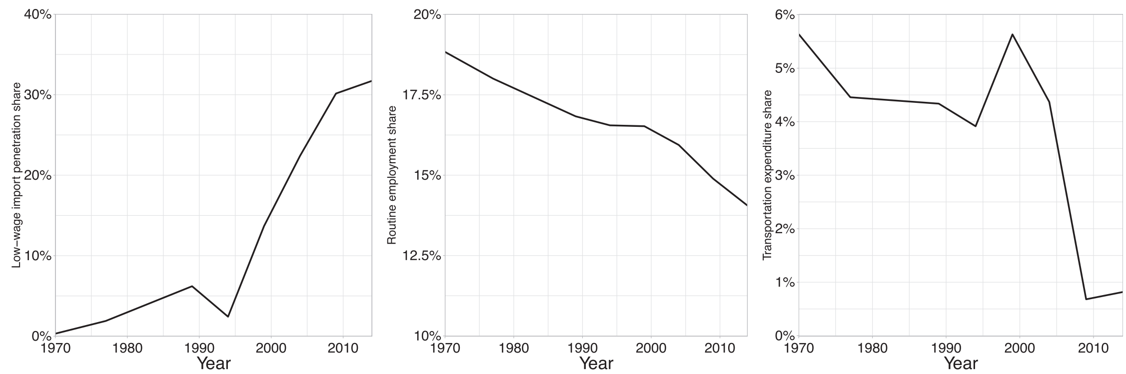
The mean of import penetration tells us that between 1970 and 2014, on average, 18.1% of the total imports in an industry originate from low-wage countries. SIC376 (guided missiles and space vehicles and parts) experiences the lowest (capped) import penetration at 0.1%, while the maximum of 91.7% is found in, *i.a.*, SIC302 (rubber and plastics footwear).

The mean routine employment share indicates that, on average, 21.2% of the employees have routine task-intensive jobs, which are susceptible to automation. The lowest shares are found in high-technology sectors such as SIC357 (computer and office equipment), while high values are found in sectors such as SIC231-239 (apparel and other textile products).

We show that, on average, 4.8% of all expenditures is spent on transportation and related services. The industries with the lowest expenditure share produce relatively expensive products (*e.g.,* SIC372 (aircraft and parts)). In contrast, SIC327 (concrete, gypsum, and plaster products) spends relatively the most on transportation.

Because of possible interdependence between trade, technology, and transportation costs, we check the correlation between our measures and find this value to be rather small. The Pearson correlation is 0.33 between trade and technology, −0.22 between trade and transportation costs, and −0.13 between technology and transportation costs.

[22] We use the same IPUMS census samples as those used for the spatial instruments.



(a) Trade (b) Technology (c) Transportation costs

Fig. 1. Median values of economic developments over time.

Fig. 1 plots the median value over time for each of the variables. As expected, the median import penetration rises over time, while the share of workers with routine tasks and the share of expenditure on transportation decreases over time.[23]

Our data provide several indications of structural changes in establishments within industries and their locations that drive changes in coagglomeration patterns. When comparing total employment per industry and MSA between 1970 and 2014, the correlation is only 0.58, which is in line with the low correlation found in the coagglomeration measure between 1970 and 2014. This suggests that large changes did indeed occur in the industry composition both within and between cities.

Furthermore, in the data used for robustness checks, we find that the mean average establishment size across industries in 1970 is approximately 186 workers, while that in 2014 is reduced to 77 workers; at the same time, the R&D expenditure per employee increased from approximately 3,282 to 13,659 (in 1987 dollars). This is in line with the literature; Brynjolfsson and Hitt (2000); Bernard et al. (2006); Holmes and Stevens (2014); Bloom et al. (2016) and Pierce and Schott (2016) show that larger establishments are more likely to engage in low-skilled standardized production processes and more likely to close down due to trade competition and technological change, while surviving establishments become more knowledge and skill intensive.

4. Results

4.1. Step 1: Determinants of industry agglomeration

The first step in the analysis is to replicate the results by Ellison et al. (2010), Faggio et al. (2017) and Diodato et al. (2018) for our data. All the variables have been standardized to have a mean of zero and a standard deviation of one. As $coagg_{ij}$ and $coagg_{ji}$ are identical, we cluster the standard errors at the industry pair $(ij - ji)$ by year level. We present the results in Table 3.

In Column (1), we estimate a naive specification where we only control for year fixed effects. We find that all proxies have a considerable positive effect on coagglomeration. For example, we find that, on average, a standard deviation increase in labor market pooling increases coagglomeration by 0.114 standard deviations. In contrast to the results of Ellison et al. (2010), which are based on 1987, we find a reverse order of importance for our pooled 1970–2014 data; *i.e.*, knowledge spillovers, as proxied by technological relatedness, is the most important determinant of industry agglomeration, followed by labor market pooling and input-output linkages.[24]

Column (2) adds dissimilarity measures, which is in line with Faggio et al. (2017). The results do not materially change, which suggests that either omitted variable bias is not a main issue or that having similar input requirements outside manufacturing is not a strong reason to coagglomerate.

In Column (3), we present our preferred specification, in which industry i×year and industry j×year fixed effects are included to control for the overall tendency of an industry to coagglomerate. The estimated coefficient on labor market pooling is significantly larger compared to those found in previous specifications. Here, a standard deviation increase in labor market pooling leads to an increase in coagglomeration of 0.195 of a standard deviation, which is almost twice as large as the one found in the previous specification. The coefficient on input-output linkages is essentially unaffected. By contrast, the coefficient on technological relatedness is approximately two-thirds the size of that found in the previous specifications. These results strongly suggest that it is important to control for unobservables at the industry-year level.

In Column (4), we use patent citations instead of technological relatedness as a proxy for knowledge spillovers. The coefficient on patent citations is less than half of that on technological relatedness shown in Column (3).[25] When including both proxies for knowledge spillovers in Column (5), the coefficient on patent citations is close to zero, while the coefficient on technological relatedness remains virtually unchanged. Hence, the technological relatedness measure strongly outperforms that of patent citations in explaining coagglomeration.

Column (6) presents the results when input linkages and output linkages are included separately. Output linkages appear to be a significantly stronger determinant of agglomeration than input linkages, while the coefficients on labor market pooling and technological relatedness are virtually unchanged.

In Column (7), we present the omitted variable bias-adjusted estimates, following Oster (2019). Reassuringly, the coefficients are not materially influenced when using this alternative estimation procedure. Unsurprisingly, the standard errors are higher, as this methodology is less efficient than OLS.

The instrumental variable regression results are reported in Column (8) of Table 3. The agglomeration determinants are instrumented by

[23] We double-checked the values in 1999 for transportation costs, as the values are somewhat higher than those in the preceding years, but we did not observe any peculiarities. Note that we have added year fixed effects to our regressions to mitigate average year-based measurement issues.

[24] We show results by year in Table in Appendix. In 1989, which is the closest to 1987, *i.e.*, the year used in Ellison et al. (2010), we find that labor market pooling is the most important agglomeration force, followed by input-output linkages and knowledge spillovers.

[25] Note that Ellison et al. (2010) do not use industry fixed effects and only find significant positive effects for patent citations using univariate regressions in their 1987 sample.

Table 3
Baseline results. *(Dependent variable: coagglomeration of industries i and j).*

	Naive specification (1)	*+ Dissimilarity measures* (2)	*+ Industry× year f.e.* (3)	*Patent citations* (4)	*Tech. rel & pat. cit.* (5)	*Separate input & output* (6)	*Bias-adjusted* (7)	*2SLS specification* (8)
Labor market pooling	0.114***	0.109***	0.195***	0.236***	0.195***	0.190***	0.148***	**0.294*****
	(0.008)	(0.008)	(0.013)	(0.013)	(0.013)	(0.013)	(0.026)	**(0.006)**
Input-output linkages	0.077***	0.076***	0.077***	0.090***	0.077***		0.073***	0.061***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)		(0.011)	(0.003)
Technological relatedness	0.161***	0.159***	0.104***		0.103***	0.099***	0.081***	**0.069*****
	(0.015)	(0.015)	(0.016)		(0.016)	(0.016)	(0.025)	**(0.003)**
Patent citations				0.051***	0.002			
				(0.014)	(0.013)			
Input linkages						0.048***		
						(0.006)		
Output linkages						0.063***		
						(0.007)		
Dissimilarity measures	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155,680	155,680	155,680	155,680	155,680	155,680	155,680	155,680
R^2	0.067	0.07	0.116	0.11	0.116	0.116		
R^2_{max}							0.151	
δ							1	
Kleibergen-Paap F-statistic								2632.19

Notes: Standard errors are clustered at the industry $ij - ji$ level and in parentheses. Instrumented variables are indicated in bold; *** $p < .01$, ** $p < .5$, * $p < .10$.

Table 4
First stage results (step 1).

Dependent variable:	*Labor market pooling* (1)	*Technological relatedness* (2)
Labor market pooling	0.718***	0.010***
(spatial instrument)	(0.010)	(0.002)
Input-output linkages	0.064***	- 0.001
	(0.005)	(0.001)
Technological relatedness	0.129***	0.983***
(spatial instrument)	(0.008)	(0.003)
Dissimilarity measures	Yes	Yes
Industry i × year fixed effects	Yes	Yes
Industry j × year fixed effects	Yes	Yes
Observations	155,680	155,680
R^2	0.758	0.982

Notes: Standard errors are clustered at the industry $ij - ji$ level and in parentheses; *** $p < .01$, ** $p < .5$, * $p < .10$.

values taken from areas where the latter industry is either not or hardly present. Recall that we do not have access to data with which to construct an instrument for input-output linkages; therefore, we only instrument for labor market pooling and knowledge spillovers. The first stage results are reported in Table 4. The coefficient on the spatial instrument of labor market pooling shown in Column (1) shows that a standard deviation increase in the instrument is associated with a 0.718 standard deviation increase in labor market pooling. Regarding the spatial instrument for technological relatedness shown in Column (2), the effect is almost equal to one. By looking at the Kleibergen-Paap F-statistic in Table 3, we can confirm that the instruments are strong.

Going back to the second-stage results in Column (8) of Table 3, we find a significantly higher coefficient for labor market pooling, which is in line with Ellison et al. (2010). By contrast, the coefficient on technological relatedness is somewhat lower.

4.2. Step 2: Changes in agglomeration determinants

Thus far, we have estimated the average of the coefficients between 1970 and 2014. However, we are particularly interested in how the determinants of agglomeration have changed over time. Therefore, we estimate year-specific coefficients for each agglomeration determinant. The estimated coefficients for each Marshallian determinant are plotted over time in Fig. 2, while the full regression results are shown in Appendix B.2.

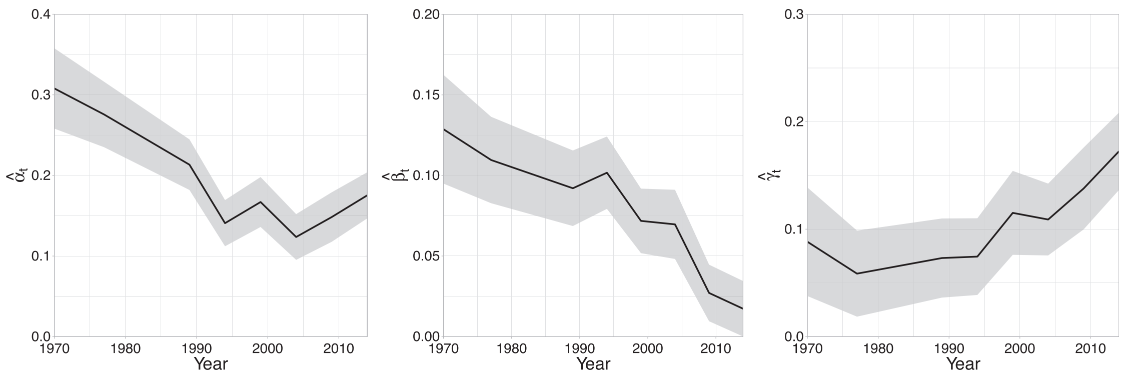
The graphs show a clear and more or less steady decline in labor market pooling and input-output linkages as determinants of agglomeration, whereas knowledge spillovers are shown to be relatively stable until 1994 and then significantly increase. This positive trend in knowledge spillovers is strong evidence that firms aim to increase their geographical proximity to share ideas, despite improvements in communication technologies (see Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011; Moretti, 2012; Balland et al., 2020).

The decrease in labor market pooling is surprising as it is not in line with Moretti (2012) and Diodato et al. (2018).[26] By contrast, our results are in line with Faggio et al. (2017), who suggest that a shift towards more high-technology/high-education industries will lead to not only a decrease in labor market pooling and input-output linkages but also an increase in knowledge spillovers because production is less standardized. However, Section 4.4 will show that the decrease in input-output linkages cannot be explained by the increase in technology and skill intensity.

4.3. Robustness of Step 1 and 2

Constant definition of agglomeration determinants. Our results are robust regarding several different specifications. A concern may be

[26] Note that the suggestion by Moretti (2012) that labor market pooling has become more important is based on the “thickness” of the labor market, while coagglomeration analyses look at the importance of each force through the relative shares of industries within cities. Hence, coagglomeration in a small MSA counts the same as that in a large MSA. Therefore, the results herein cannot be interpreted as implying that larger labor markets do not matter more than smaller ones; rather, they suggest that labor market pooling has become a less important determinant of coagglomeration. Furthermore, the decreasing importance of labor market pooling may not be contradictory to the results of Diodato et al. (2018), as a close inspection of Figure 6 in their work reveals the possibility of a decrease in labor market pooling since 1970; however, the presence of large standard errors prevents definite conclusions.



(a) Labor market pooling (b) Input-output linkages (c) Technological relatedness

Fig. 2. Estimated coefficient per agglomeration determinant over time Note: The shaded areas indicate 95% confidence bands.

that the trends in the Marshallian proxies found are driven by changes in the measurement quality of the proxies over time.

To test whether this alternative explanation is important, in Appendix B.1.1, we hold all variables, except coagglomeration, constant at their 1994 values and reproduce the main regressions. We choose 1994, as this year is in the middle of our time period and the original data from 1994 use SIC 87 classifications, which reduces the risk of concordance errors.

These tests show similar results over time, which indicates that the results are not driven by changes in variable definitions or the exact measurement of independent variables.[27]

Industry-pair fixed effects. We also consider the inclusion of industry-pair fixed effects. Including these effects implies that we solely rely on temporal variation in coagglomeration and agglomeration determinants to identify the effects of interest. However, industry-pair fixed effects also amplify measurement error in the agglomeration measures, as there is clearly more measurement error in the variables *within* industry pairs than in those *between* industry pairs. Hence, our estimates are expected to be biased towards zero.

We report the results with industry-pair fixed effects in Appendix B.1.2. The inclusion of these fixed effects captures most of the variation, as suggested by the R^2 of over 0.7. Nonetheless, labor market pooling is positive and significant in both columns, whereas the effects of the other agglomeration forces are positive but statistically insignificant. The coefficients on the time trends have the expected sign but are not statistically significant. Hence, despite the presence of large standard errors, the coefficients seem to confirm the findings with industry-by-year fixed effects.

Coagglomeration at the county level. In Appendix, we show that similar results hold at a more refined geographical level. More specifically, we calculate the coagglomeration index at the county level instead of at the MSA level, which leads to very comparable results.

Weighted regressions. The baseline results present the results for the average industry pair. However, industries vary greatly in size. In Appendix B.1.4, we reproduce the main results using weighted regressions in which we weight observations by the log of employees, number of establishments, and value added. Because the results would be entirely driven by a few very large industries, we take the log instead of weighting by the levels of industry size. We find that the weighted results are not statistically significantly different from those of the baseline regressions.

Two-way clustering of standard errors. In Appendix B.1.5, we reproduce the main results using two-way clustering by industry i and industry j instead of clustering at the industry pair (ij - ji) level. Unsurprisingly, these results show the same coefficients but standard errors that are considerably larger than those seen in the main results. Still, all the coefficients remain statistically significant.

Table 5
Descriptive statistics - Second stage.

Statistic	Mean	Median	St.Dev.	Min	Max
$\hat{\alpha}$	0.157	0.111	0.281	−0.476	1.199
$\hat{\beta}$	0.066	0.034	0.129	−0.180	0.575
$\hat{\gamma}$	0.255	0.095	0.570	−0.946	2.624

Notes: We report the estimated dependent variables here. $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are the coefficients obtained in the first stage on, respectively, labor market pooling, input-output linkages, and technological relatedness. The number of observations is 1120.

Dealing with outliers. In the baseline results, we have capped the values of the dependent and independent variables to the 1st percentile or the 99th percentile. This threshold is obviously somewhat arbitrary. In Appendix, we show that the results are largely robust when extreme observations are dropped and different thresholds are chosen.

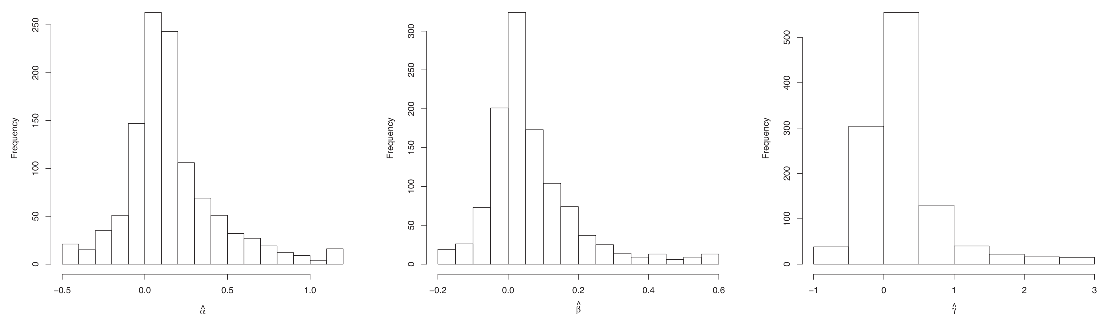
4.4. Step 3: Exploring industry-level and temporal heterogeneity

In this section, we explore industry-year-level heterogeneity and investigate whether our measures for trade, technology, and transportation costs are associated with the (trends in the) importance of agglomeration determinants. In the first step, we obtain industry-year specific-coefficients for each of the determinants of agglomeration (see Eq. (4)). Table 5 reports the descriptive statistics of the industry-year-specific coefficients, and Fig. 3 shows histograms. As described previously, we limit the effect of outliers by capping these to the 1st percentile or the 99th percentile. As we have 140 industries and 8 time periods, the number of observations is 1,120.

Recall that the variables in the first step are standardized to have a mean value of 0 and a standard deviation of 1. As such, the mean value of the coefficient on labor market pooling $\hat{\alpha}$ indicates that an increase of one standard deviation in the extent to which labor can be pooled is associated with an increase in coagglomeration of 0.157 of a standard deviation. The means of the other coefficients, *i.e.*, input-output linkages $\hat{\beta}$ and technological relatedness $\hat{\gamma}$, also show that on average, industries have a positive appreciation for the respective agglomeration determinants. Still, 23.6%, 27.1%, and 30.4% of the values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, respectively, are negative, although only very few of these estimates are statistically significant. We find some negative and significant labor market pooling effects for relatively skilled industries, such as SIC366 (communications equipment), where labor poaching may be a concern.

Each of the industry-year-specific coefficients is regressed on proxies for trade, technology, and transportation costs (see Eq. (5)). For each of

[27] We obtain similar results when using 1970 or 2014 instead of 1994.



(a) Labor market pooling (b) Input-output linkages (c) Technological relatedness

Fig. 3. Histograms of the estimated coefficients obtained in the first step.

Table 6
Trade, technology, and transportation costs results.

	Labor market pooling			*Input-output linkages*			*Technological relatedness*		
Dependent variable:	$\hat{\alpha}$			$\hat{\beta}$			$\hat{\gamma}$		
	OLS	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Import penetration	–0.042***	–0.057***	**–0.051*****	–0.019***	–0.008	**–0.017***	0.022*	0.049**	**0.023**
	(0.008)	(0.009)	**(0.014)**	(0.007)	(0.009)	**(0.010)**	(0.013)	(0.021)	**(0.021)**
Routine employment share	0.105***	0.142***	**0.128*****	–0.002	0.009	**–0.001**	–0.049***	–0.090	**–0.053****
	(0.014)	(0.020)	**(0.016)**	(0.009)	(0.008)	**(0.011)**	(0.016)	(0.118)	**(0.021)**
Transportation costs				–0.001	–0.001	**0.013**			
				(0.008)	(0.009)	**(0.014)**			
Value of a ton in 1970 *(log)*						0.004			
						(0.004)			
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extra controls	No	Yes	No	No	Yes	No	No	Yes	No
Observations	1120	1120	1120	1120	1120	1120	1120	1120	1120
R^2	0.133			0.045			0.028		
R^2_{max}		0.172			0.059			0.036	
δ		1			1			1	
Kleibergen-Paap *F*-statistic			112.73			16.15			118.24

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. Instrumented variables are indicated in bold. Extra controls consist of the natural logarithm of the average establishment size and the capital labor ratio. A dummy variable is added to indicate missing data in the CFS on the instrumental variable value of a ton for industries belonging to SIC27 but its coefficient is not reported in the table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

the coefficients on Marshall's agglomeration determinants, Table 6 reports the estimates from an OLS regression, bias-adjusted estimates, and estimates relying on instrumental variables. The independent variables are again standardized to have a mean of 0 and a standard deviation of 1, whereas the dependent variables are taken as they are since these are derived from regressions with standardized variables. Standard errors are obtained through bootstrapping.

Column (1) reports the OLS results for the coefficient on labor market pooling ($\hat{\alpha}$); an increase of a standard deviation in import penetration is associated with a decrease of 0.042 in the size of the coefficient on labor market pooling. As the median coefficient on labor market pooling is 0.111, a standard deviation increase in import penetration is associated with a decrease in labor market pooling of approximately 38% of the median. A standard deviation increase in routine employment share is associated with an increase of 0.105, which is almost equal to 100% of the median coefficient on labor market pooling. Hence, the effects are sizable and suggest that industries facing little trade competition and those with highly routinized job tasks benefit more from a common labor market pool. The reduction in the number of routine task-intensive jobs due to technological progress and the rise in import penetration complement each other and can explain the decreasing trend in labor market pooling.

Column (2) presents the results of the same specification with the inclusion of the omitted variable bias-adjusted approach proposed by Oster (2019).[28] The results confirm those shown in the previous column. Column (3) presents the 2SLS results. For import penetration, we use the import penetration of other high-wage countries as an instrument, while we use the routine employment share calculated for areas in which industries do not coagglomerate to make use of a common labor pool as an instrument for the routine employment share. The first-stage results show plausible signs and are reported in Appendix B.3.6; they are also very similar to those shown in Column (1).

Column (4) of Table 6 focuses on input-output linkages. It shows that import penetration is negatively and significantly associated with input-output linkages. A standard deviation increase in trade competition is associated with a considerable decrease of 56% of the median coefficient on input-output linkages. The coefficient on routine employment share is close to zero and highly statistically insignificant. Perhaps more surprisingly, this also holds for the coefficient on transportation

[28] We compare a regression without controls to a regression with additional controls, *i.e.*, the natural logarithm of the average establishment size and the capital-labor ratio, to observe movements in coefficient size and R^2, while the R^2_{max} is estimated to be 0.172. Recall that these extra controls can be seen as proxy controls since they are partly capturing omitted variables but also partly capturing the effect of trade and technological progress (see Angrist and Pischke, 2008). This procedure is therefore expected to provide a lower bound of the true effects of the variables of interest.

costs of goods.[29] In Column (5), we add extra controls and estimate bias-adjusted coefficients, following Oster (2019). In this specification, none of the coefficients are statistically significant; however, they are also not significantly different from the OLS results given in the previous column.

The 2SLS results are reported in Column (6); we use the same instruments as those used before for trade and technology. For transportation costs, we use the natural logarithm of the average value of a ton as an instrument. We control for the value of a ton in 1970, which mitigates the issue of omitted variable bias, as the value of a ton also captures the complexity of a product. The coefficient on transportation cost is larger and positive when instrumented but is not statistically significantly different from the one given in the OLS results.

Hence, the results unequivocally suggest that the 'pure' transportation costs of goods are not a relevant factor in coagglomerating with suppliers or customers, which is contradictory to the expectations of Glaeser and Kohlhase (2004) and Diodato et al. (2018). However, total trade costs consist of much more than only the 'physical' transportation costs of goods (Glaeser and Kohlhase, 2004). For instance, McCann and Fingleton (1996) and Duranton and Storper (2008) show that face-to-face contact and coordination are also important in sustaining input-output linkages.

By contrast, we show that import penetration reduces the demand for input-output linkages. We provide suggestive evidence in Appendix that localized input linkages are replaced by input linkages with low-wage countries, as trade competition for input negatively affects input linkages while it is positively associated with output linkages (although the coefficients are imprecise).

The decline in input-output linkages cannot be explained by industries becoming more technology and skill intensive (see Faggio et al., 2017), as other industry characteristics closely related to industry skill and technology levels are unrelated to input-output linkages (see Appendix B.3.1).

Column (7) in Table 6 explores whether the importance of knowledge spillovers can be linked to changes in trade and technology. Import penetration is positively associated with the intensity of knowledge spillovers. According to Column (7) the effect is approximately 23% of the median coefficient on technological relatedness (which is equal to 0.095). By contrast, the routine employment share shows a negative association of approximately 52% of the median coefficient. This suggests that the increases in import penetration and ongoing computerization that have led to more innovative skill technology-intensive manufacturing establishments have also raised the need to coagglomerate in proximity to establishments that use similar technologies. This is likely due to the increased relevance of new ideas in the production process and the need to meet face-to-face to exchange ideas (see Holmes and Stevens, 2014; Storper and Venables, 2004). Column (8), which displays bias-adjusted results, shows stronger but not significantly different coefficients. The 2SLS results in Column (9) show very similar results as those found in the bias-adjusted estimates.

All in all, the results presented in this section indicate a complementary impact of increasing import competition and decreasing routinization of labor tasks on labor market pooling and knowledge spillovers. While input-output linkages seem unaffected by the decrease in the pure transportation costs of goods, the increase in import competition does seem to have a negative effect.

4.5. Robustness of Step 3

In Appendix B.3.1, we further investigate the robustness of the results by showing *(i)* the effects of the transportation costs of goods on labor market pooling and knowledge spillovers, *(ii)* the effects of the control variables (*i.e.*, average establishment size and the capital-labor ratio) used in the bias-adjusted estimation procedure, and *(iii)* the effects of increased R&D expenditure and skill intensity (following Bloom et al., 2016 and Pierce and Schott, 2016), which are closely associated to trade competition and technological progress. The results show that the main results are robust to these specifications. We further find evidence that the decreasing importance of labor market pooling and the increasing importance of knowledge spillovers are both likely related to the rise of the high technology/high education firms, as R&D expenditures are strongly associated with labor market pooling and knowledge spillovers. By contrast, none of the additional industry characteristics capturing technology and skill levels are statistically significantly associated with input-output linkages.

Further, in the main results in Table 6, we attach equal weight to each industry-by-year observation. By contrast, Faggio et al. (2020) weight each observation by the inverse of the standard deviation of the coefficient obtained in the first step. The results reported in Appendix B.3.2 are largely similar.

In Appendix B.3.3, we employ an alternative measure of import penetration, namely, the imports from low-wage countries divided over the so-called apparent consumption in the U.S. The apparent consumption is equal to domestic production minus exports plus imports. The results are not significantly different from our main results shown in Table 6. This demonstrates that there is no issue in using the value share measure, for which more data and instruments are available.

Appendix B.3.4 shows similar effects if we estimate everything in one step instead of using our proposed two-stage approach (recall the Eqs. (4) and (5)); the signs on the coefficients all point in the same direction, with only minor differences in significance levels and effect sizes.

Finally, in Appendix B.3.5, we further explore the results on input-output linkages when *(i)* using the value of a ton as a proxy for transportation costs instead of as an instrument, *(ii)* using the import penetration within sectors from which inputs are obtained instead of the import penetration within each sector itself, and *(iii)* calculating separate coefficients for input linkages and output linkages using these variables as separate dependent variables in the second step. The results are similar to the main results and provide further evidence that the decrease in transportation costs of goods is not a relevant factor in explaining input-output linkages. By contrast, import competition within both producing industries and supplying industries influences input linkages.

5. Conclusion

In the last 50 years, the economy has undergone large and fundamental changes due to more intense trade competition, technological progress, and reductions in the transportation costs of goods. There is abundant evidence that these developments have resulted in large changes in agglomeration *patterns*. In this paper, we assess changes in agglomeration *determinants* over time and explore whether industry-year-level heterogeneity can be explained by changes in trade competition, technological progress and reductions in transport costs.

Using an alternative proxy for knowledge spillovers, we find that between 1970 and 2014, knowledge spillovers became more important. This is strong evidence that geographical proximity is becoming more relevant for exchanging ideas, despite the presence of strong improvements in communication technologies. On the other hand, we find that labor market pooling and input-output linkages have become less important agglomeration determinants.

Furthermore, we show that trade competition and technological progress are strongly related to labor market pooling and knowledge spillovers. These results suggest that the computer revolution and trade competition, which led to less standardized, less vertically integrated and more knowledge-intensive establishments, altered the composition of industries and therefore the relevance of labor market pooling and knowledge spillovers in explaining agglomeration. Maybe surprisingly,

[29] One may argue that by controlling for the dissimilarity variable capturing transport inputs in the first stage, the coefficient of transport costs may be reduced. Thus, we have estimated regressions in which we excluded the transport dissimilarity measure in the first stage, which led to nearly identical results.

we do not find that the transportation costs of goods are associated with input-output linkages. On the other hand, we do find a negative effect of increasing trade competition. We present suggestive evidence that this outcome is likely due to the import substitution of local inputs in input-output linkages.

Our study opens up avenues for further research. First, future studies could look more closely into the heterogeneity of the agglomeration benefits of establishments *within* industries, related to *e.g.*, the skill and capital-intensity of the establishments. Second, we note that the current framework overlooks the effects of agglomeration size; this is because the coagglomeration index is a relative measure.[30] Third, the two-step methodology introduced to explain changes in the determinants of agglomeration could be expanded to include various other industrial or regional characteristics. Fourth, a more obvious step forward would be to include measures of knowledge spillovers in the services industry. Finally, we note that Duranton and Puga (2004) distinguish between sharing, matching, and learning, rather than use Marshall's categorization. While Duranton and Puga's categorization may be conceptually more intuitive, to date, it has not been possible to develop meaningful empirical metrics for this categorization approach. Thus, future research could aim to find meaningful proxies for sharing, matching, and learning.

Supplementary material

The National Industry-Occupation Employment Matrix gives the share of workers per job occupation in each industry. These matrices are essential in calculating a proxy for the extent to which industries can share workers in coagglomeration studies focussed on the U.S.A.. such as 'The Dynamics of Industry Agglomeration: Evidence from 44 years of Coagglomeration Patterns'. However, the Bureau of Labor Statistics (BLS) only has the data since around 1989 online. But the original 1981 report of the BLS with the data for 1970, 1978 and a projection for 1990 was available in printed form and was scanned and put online by FRASER, see here. By using Optical Character Recognition (OCR) and datascraping techniques in R, among others the R package 'Tabulizer', I managed to gather the data in this document. Note that I manually corrected wrongly recognized characters, which were relatively easy to find as totals should add up to 100%. This dataset was used for the labor market pooling measure in the years 1970/1971 and 1977/1978. Please cite the paper when using the data.

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2022.103456.

CRediT authorship contribution statement

Mathieu P.A. Steijn: Conceptualization, Methodology, Software, Supervision, Formal analysis, Investigation, Visualization, Writing – original draft. **Hans R.A. Koster:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Supervision, Writing – original draft. **Frank G. Van Oort:** Conceptualization, Writing – original draft, Supervision, Funding acquisition.

References

Aleksandrova, E., Behrens, K., Kuznetsova, M., 2020. Manufacturing (co)agglomeration in a transition country: evidence from russia. J Reg Sci 60 (1), 88–128. doi:10.1111/jors.12436.
Angrist, J.D., Pischke, J.S., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton.
Autor, D.H., 2019. Work of the past, work of the future. AEA Papers and Proceedings 109, 1–32. doi:10.1257/pandp.20191110.
Autor, D.H., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US labor market. American Economic Review 103 (5), 1553–1597. doi:10.1257/aer.103.5.1553.
Autor, D.H., Dorn, D., Hanson, G.H., 2013. The geography of trade and technology shocks in the united states. American Economic Review 103 (3), 220–225. doi:10.1257/aer.103.3.220.
Autor, D.H., Dorn, D., Hanson, G.H., 2015. Untangling trade and technology: evidence from local labour markets. Economic Journal 125 (584), 621–646. doi:10.1111/ecoj.12245. https://academic.oup.com/ej/article/125/584/621-646/5077881
Autor, D.H., Levy, F., Murnane, R.J., 2003. The skill content of recent technological change: an empirical exploration. Q J Econ 118 (4), 1279–1333. doi:10.1162/003355303322552801.
Balland, P.-A., Jara-Figueroa, C., Petralia, S.G., Steijn, M.P.A., Rigby, D.L., Hidalgo, C.A., 2020. Complex economic activities concentrate in large cities. Nat. Hum. Behav. 4 (3), 248–254. doi:10.1038/s41562-019-0803-3.
Bartelsman, E.J., Gray, W., 1996. The NBER manufacturing productivity database. NBER working paper series, 205.
Behrens, K., 2016. Agglomeration and clusters: tools and insights from coagglomeration patterns. Canadian Journal of Economics/Revue canadienne d'économique 49 (4), 1293–1339. doi:10.1111/caje.12235.
Bernard, A.B., Jensen, J.B., Schott, P.K., 2006. Survival of the best fit: exposure to low-wage countries and the (uneven) growth of u.s. manufacturing plants. J Int Econ 68 (1), 219–237. doi:10.1016/j.jinteco.2005.06.002.
Bloom, N., Draca, M., Van Reenen, J., 2016. Trade induced technical change? the impact of Chinese imports on innovation, IT and productivity. Rev Econ Stud 83 (1), 87–117. doi:10.1093/restud/rdv039.
BLS, 1981. The National Industry-Occupation Employment Matrix, 1970, 1978, and Projected 1990, Vols. I and II, Bulletin 2086. Technical Report.
Brynjolfsson, E., Hitt, L.M., 2000. Beyond computation: information technology, organizational transformation and business performance. Journal of Economic Perspectives 14 (4), 23–48. doi:10.1257/jep.14.4.23.
Combes, P.-P., Duranton, G., 2006. Labour pooling, labour poaching, and spatial clustering. Reg Sci Urban Econ 36 (1), 1–28. doi:10.1016/J.REGSCIURBECO.2005.06.003.
Combes, P.-P.P., Gobillon, L., 2015. The Empirics of Agglomeration Economies. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics Vol.5, Vol. 5. Elsevier, Amsterdam, pp. 247–348. doi:10.1016/B978-0-444-59517-1.00005-2.
Deming, D.J., 2017. The growing importance of social skills in the labor market. Q J Econ 132 (4), 1593–1640. doi:10.1093/qje/qjx022.
Diodato, D., Neffke, F., O'Clery, N., O'Clery, N., 2018. Why do industries coagglomerate? how Marshallian externalities differ by industry and have evolved over time. J Urban Econ 106, 1–26. doi:10.1016/J.JUE.2018.05.002.
Dumais, G., Ellison, G., Glaeser, E.L., 2002. Geographic concentration as a dynamic process. Review of Economics and Statistics doi:10.1162/003465302317411479.
Duranton, G., Morrow, P.M., Turner, M.A., 2014. Roads and trade: evidence from the US. Rev Econ Stud 81 (2), 681–724.
Duranton, G., Overman, H.G., 2005. Testing for localization using micro-Geographic data. Rev Econ Stud 72 (4), 1077–1106. doi:10.1111/0034-6527.00362.
Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. American Economic Review 91 (5), 1454–1477.
Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. Economic Policy 4 (04), 2063–2117. doi:10.1016/S1574-0080(04)80005-1.
Duranton, G., Puga, D., 2005. From sectoral to functional urban specialisation. J Urban Econ 57 (2), 343–370. doi:10.1016/j.jue.2004.12.002.
Duranton, G., Storper, M., 2008. Rising trade costs? agglomeration and trade with endogenous transaction costs. Canadian Journal of Economics/Revue canadienne d'économique 41 (1), 292–319. doi:10.1111/j.1365-2966.2008.00464.x.
Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: A Dartboard approach. Journal of Political Economy 105 (5), 889. doi:10.1086/262098. arXiv:1011.1669v3
Ellison, G., Glaeser, E.L., 1999. The geographic concentration of industry: does natural advantage explain agglomeration? American Economic Review: Papers and Proceedings 89 (2), 311–327.
Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? evidence from coagglomeration patterns. American Economic Review 100 (6), 1195–1213.
Faggio, G., Silva, O., Strange, W.C., 2017. Heterogeneous agglomeration. Review of Economics and Statistics 99 (1), 80–94. doi:10.1162/REST_a_00604.
Faggio, G., Silva, O., Strange, W.C., 2020. Tales of the city: what do agglomeration cases tell us about agglomeration in general? Journal of Economic Geography 20 (5), 1117–1143. doi:10.1093/jeg/lbaa007. https://academic.oup.com/joeg/article/20/5/1117/5825466
Gaspar, J., Glaeser, E.E.L., 1998. Information technology and the future of cities. J Urban Econ 43 (1), 136–156. doi:10.1006/juec.1996.2031. http://ac.els-cdn.com/S0094119096920318/1-s2.0-S0094119096920318-main.pdf?_tid=2e8eaac4-16a8-11e6-b43b-00000aab0f26&acdnat=1462882407_3cb6f4fa7fe4c6a8fec876b79e4aa8f4
Glaeser, E., Ponzetto, G.A.M., 2007. Did the Death of Distance Hurt Detroit and Help New York? Technical Report. National Bureau of Economic Research, Cambridge, MA doi:10.3386/w13710.
Glaeser, E.L., 2011. Triumph of the City: How our Greatest Invention makes US Richer, Smarter, Greener, Healthier and Happier. Penguin Press, New York.
Glaeser, E.L., Kohlhase, J.E., 2004. Cities, regions and the decline of transport costs. Papers in Regional Science 83 (1), 197–228.
Goldschlag, N., Lybbert, T.J., Zolas, N.J., 2020. Tracking the technological composition of industries with algorithmic patent concordances. Economics of Innovation and New Technology 29 (6), 582–602. doi:10.1080/10438599.2019.1648014.
Goos, M., Manning, A., Salomons, A., 2009. Job polarization in Europe. American Economic Review 99 (2), 58–63. doi:10.1257/aer.99.2.58.

[30] The coagglomeration index considers the relative joint presence of two industries in a city. Hence, coagglomeration in a small city is just as important as coagglomeration in a large city.

Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NBER patent citation data file: lessons, insights and methodological tools. NBER working paper series, 8498.

Hanlon, W.W., Miscio, A., 2017. Agglomeration: a long-run panel data approach. J Urban Econ 99, 1–14. doi:10.1016/j.jue.2017.01.001.

Helsley, R., Strange, W.C., 1990. Matching and agglomeration economies in a system of cities. Reg Sci Urban Econ 20 (2), 189–212.

Hidalgo, C.A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D.F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., Zhu, S., 2018. The Principle of Relatedness. In: Unifying Themes in Complex Systems IX. Springer, Cham, pp. 451–457. doi:10.1007/978-3-319-96661-8_46.

Holmes, T.J., Stevens, J.J., 2014. An alternative theory of the plant size distribution, with geography and intra- and international trade. Journal of Political Economy 122 (2), 369–421. doi:10.1086/674633.

Jacobs, W., Koster, H.R.A., Van Oort, F.G., 2013. Co-agglomeration of knowledge-intensive business services and multinational enterprises. Journal of Economic... 14 (2), 443–475. http://joeg.oxfordjournals.org/content/early/2013/01/07/jeg.lbs055.short.

Jaffe, A.B., Trajtenberg, M., Fogarty, M.S., 2000. Knowledge spillovers and patent citations: evidence from a survey of inventors. American Economic Review 90 (2), 215–218.

Kerr, W.R., 2008. Ethnic scientific communities and international technology diffusion. Review of Economics and Statistics 90 (3), 518–537.

Kim, S., 1995. Expansion of markets and the geographic distribution of economic activities: the trends in u. s. regional manufacturing structure, 1860–1987. Q J Econ 110 (4), 881–908. doi:10.2307/2946643.

Klein, A., Crafts, N., 2012. Making sense of the manufacturing belt: determinants of u.s. industrial location, 1880–1920. Journal of Economic Geography 12 (4), 775–807. doi:10.1093/jeg/lbr023. https://academic.oup.com/joeg/article/12/4/775/946998

Klein, A., Crafts, N., 2020. Agglomeration externalities and productivity growth: US cities, 1880–1930. Econ Hist Rev 73 (1), 209–232. doi:10.1111/ehr.12786.

Koster, P.R., Koster, H.R.A., 2015. Commuters' Preferences for fast and reliable travel: a semi-parametric approach. Transportation Research Part B: Methodological 81, 289–301.

Krugman, P., 1991. Geography and Trade. MIT Press, Cambridge, MA.

Marco, A.C., Carley, M., Jackson, S., Myers, A.F., 2015. The USPTO historical patent data files two centuries of innovation. SSRN Electronic Journal doi:10.2139/ssrn.2616724.

Marshall, A., 1890. Principles of Economics. MacMillan and Co., London.

Matouschek, N., Robert-Nicoud, F., 2005. The role of human capital investments in the location decision of firms. Reg Sci Urban Econ 35 (5), 570–583. doi:10.1016/J.REGSCIURBECO.2004.09.001.

McCann, P., 2008. Globalization and economic geography: the world is curved, not flat. Cambridge Journal of Regions, Economy and Society 1 (3), 351–370. doi:10.1093/cjres/rsn002.

McCann, P., Fingleton, B., 1996. The regional agglomeration impact of just-in-time input linkages: evidence from the Scottish Electronics Industry. Scott J Polit Econ 43 (5), 493.

Michaels, G., Redding, S. J., Rauch, F., 2019. Task specialization in U.S. Cities from 1880 to 2000. https://academic.oup.com/jeea/article/17/3/754/4922084.

Moretti, E., 2012. The new geography of jobs. Houghton Mifflin Harcourt Publishing Company, New York.

Oster, E., 2019. Unobservable selection and coefficient stability: theory and evidence. Journal of Business & Economic Statistics 37 (2), 187–204. doi:10.1080/07350015.2016.1227711.

Petralia, S.G., Balland, P.-A.A., Rigby, D.L., 2016. Unveiling the geography of historical patents in the United States from 1836 to 1975. Sci Data 3, 1–14. doi:10.1038/sdata.2016.74.

Pierce, J.R., Schott, P.K., 2016. The surprisingly swift decline of US manufacturing employment. American Economic Review 106 (7), 1632–1662.

Rodríguez-Pose, A., Crescenzi, R., 2008. Mountains in a flat world: why proximity still matters for the location of economic activity. Cambridge Journal of Regions, Economy and Society 1 (3), 371–388. doi:10.1093/cjres/rsn011.

Rotemberg, J.J., Saloner, G., 2000. Competition and human capital accumulation: a theory of interregional specialization and trade. Reg Sci Urban Econ 30 (4), 373–404. doi:10.1016/S0166-0462(99)00044-7.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., Sobek, M., 2018. IPUMS USA: Version 8.0 [dataset]. IPUMS.

Serafinelli, M., 2019. æGoodg firms, worker flows, and local productivity. J Labor Econ 37 (3), 747–792. doi:10.1086/702628.

Silverman, B.S., 2002. Technological resources and the logic of corporate diversification. Routledge, London.

Steijn, M.P.A., 2021. Improvement on the association strength: implementing a probabilistic measure inspired by combinations without repetition. Quantitative Science Studies 2 (2), 778–794. doi:10.1162/qss_a_00122.

Storper, M., 2018. Separate worlds? explaining the current wave of regional economic polarization. Journal of Economic Geography 18 (2), 247–270. doi:10.1093/jeg/lby011.

Storper, M., Venables, A.J., 2004. Buzz: face-to-face contact and the urban economy. Journal of Economic Geography 4 (4), 351–370. doi:10.1093/jnlecg/lbh027.