

## Coaglomeração e externalidades marshallianas: Evidências para a indústria de transformação brasileira

Jullio Victor Pedrosa de Almeida<sup>1</sup>  | Roberta de Moraes Rocha<sup>2</sup>  | Wellington Ribeiro Justo<sup>3</sup> 

<sup>1</sup> PPGECON/UFPE. E-mail: juliovictor\_sport@hotmail.com

<sup>2</sup> PPGECON/UFPE, PPGGES/UFPE, PPGIT/UFPE. E-mail: roberta.mrocha@ufpe.br

<sup>3</sup> PPGECON/UFPE, PPGERU/URCA. E-mail: justowr@yahoo.com.br

### RESUMO

Este estudo visa quantificar a coaglomeração no setor da indústria de transformação entre distintos pares de indústrias, além de estimar a correlação entre a coaglomeração e os fatores marshallianos. A análise é conduzida em nível estadual e microrregional para as indústrias pertencentes à divisão CNAE de dois dígitos. O índice de coaglomeração proposto por Ellison et al. (2010) é empregado para mensurar a aglomeração entre esses pares de indústrias. A relação entre coaglomeração e externalidades marshallianas é estimada por meio de análise de regressão com dados em painel, controlando os efeitos fixos das indústrias. Esse método está fundamentado na literatura que associa vantagens produtivas, potencialmente geradoras de economias de aglomeração, à coaglomeração industrial. Os resultados indicam que a maioria das indústrias está coagulada com outras de mesma classificação tecnológica, sendo as indústrias de alta tecnologia mais propensas a esse fenômeno do que as de baixa tecnologia. Os resultados revelam que as forças marshallianas, incluindo *labor pooling* e *input linkage*, estão positivamente associadas ao nível de coaglomeração industrial. Isso sugere que uma maior dependência de compartilhamento de trabalhadores (*labor pooling*) e de insumos fornecidos por outras indústrias favorece a coaglomeração. Esses resultados contribuem para uma compreensão mais aprofundada das dinâmicas de coaglomeração na indústria de transformação.

### PALAVRAS-CHAVE

Coaglomeração Industrial, Indústria da Transformação, Economias de Aglomeração, Forças Marshallianas, Índice de coaglomeração EG

### Agglomeration and Marshallian externalities: Evidence for the Brazilian manufacturing industry

#### ABSTRACT

This study aims to quantify co-glomeration in the manufacturing industry sector between different pairs of industries, in addition to estimating the correlation between co-glomeration and Marshallian factors. The analysis is conducted at the state and micro-regional level for industries belonging to the two-digit CNAE division. The co-glomeration index proposed by Ellison, Glaeser and Kerr (2010) is used to measure the agglomeration between these pairs of industries. The relationship between co-glomeration and Marshallian externalities is estimated through regression analysis with panel data, controlling for industry fixed effects. This method is based on literature that associates productive advantages, potentially generating agglomeration economies, with industrial co-glomeration. The results indicate that most industries are co-clustered with others of the same technological classification, with high-technology industries being more prone to this phenomenon than low-technology industries. The results reveal that Marshallian forces, including *labor pooling* and *input linkage*, are positively associated with the level of industrial co-glomeration. This suggests that greater dependence on worker sharing (*labor pooling*) and inputs provided by other industries favors co-glomeration. These results contribute to a deeper understanding of co-glomeration dynamics in the manufacturing industry.

#### KEYWORDS

Industry Coagglomeration, Manufacturing Industry; Economies of Agglomeration, Marshallian Forces, EG Coagglomeration Index

### CLASSIFICAÇÃO JEL

R32, R23, L60

## 1. Introdução

A partir da virada do milênio, observa-se, na literatura da área de economia regional e urbana, um aumento significativo do interesse na identificação e mensuração do papel desempenhado por fatores locais como potenciais geradores de externalidades produtivas. Esse fenômeno tem como propósito explicar a tendência de concentração dos agentes produtivos, conforme abordado por Ellison e Glaeser (1999), bem como a aglomeração e coaglomeração de diversas indústrias em contextos geográficos específicos, como discutido por Jacobs et al. (2014), Diodato et al. (2018) e Steijn et al. (2022). Essa evolução na pesquisa reflete uma busca incessante por compreender os determinantes fundamentais que moldam a distribuição espacial das atividades econômicas, contribuindo, assim, para o avanço contínuo do campo da economia regional e urbana.

Embora a mensuração do nível de concentração geográfica da atividade produtiva tenha sido impulsionada na década de 60, o conceito de coaglomeração de Ellison et al. (2010) é relativamente recente e capta o nível de coaglomeração entre pares de indústrias (índice de EGK). O índice mede o quanto o padrão da distribuição geográfica de uma indústria se aproxima de outra, configurando um indicador de covariação da participação das indústrias do emprego regional. O índice de EGK é bastante consagrado pela literatura, principalmente por duas razões: primeiro, permite quantificar o nível de concentração de pares de indústrias; segundo, o índice tem um embasamento teórico na ideia de que indústrias distintas se concentram em determinada localidade para obter ganhos de produtividade e/ou escala, por exemplo quando as indústrias de um setor, por meio da relação de insumo-produto, integra a sua cadeia produtiva com outras indústrias.

As evidências empíricas apontam que trabalhadores e firmas, quando inseridos em *clusters* ou polos industriais, se beneficiam das economias de aglomerações, que estão diretamente relacionadas ao aumento: da produtividade total dos fatores de produção (Rosenthal e Strange, 2004; Tveteras e Battese, 2006; Andersson e Lööf, 2011) e do trabalho (Fingleton e Longhi, 2013; Andersson et al., 2014), da abertura comercial e das exportações (Becchetti e Rossi, 2000; Greenaway e Kneller, 2008; Antonietti e Cainelli, 2011), difusão tecnológica (Feldman, 1999; Rosenthal e Strange, 2008; Greenstone et al., 2010), e do crescimento e desenvolvimento econômico local (Puga e Venables, 1999; Martin e Ottaviano, 2001; Baldwin e Martin, 2004; Brülhart e Sbergami, 2009).

Marshall (1890), em seu trabalho seminal, apresentou três fatores associados à aglomeração de indústrias em determinada região, formando-se em *clusters* industriais devido à especialização produtiva. O primeiro se refere à redução de custos de transporte de insumos e produtos compartilhados pelas indústrias; o segundo relaciona-se ao compartilhamento da mão de obra local; e o terceiro, à difusão de conhecimento entre as indústrias. Esses fatores são conhecidos na literatura como

*externalidades marshallianas*, em respectivo, *input-output linkages*, *labor pooling* e *intellectual spillovers*. A partir dos conceitos apresentados por Marshall, algumas teorias surgiram, a exemplo das apresentadas por Jacobs (1969) e Porter (1990), que trazem relevantes contribuições ao defender que a diversificação de indústrias na mesma localidade levaria a ganhos de inovação e ao destacar que mercados mais competitivos, em detrimento a monopólios, seriam mais adequados para o crescimento das indústrias especializadas e geograficamente concentradas (Beaudry e Schiffauerova, 2009).

Nesta perspectiva de análise, tanto as evidências empíricas internacionais Kolko (2010); Gabe e Abel (2012); Mukim (2015); Howard et al. (2016) quanto as nacionais apresentam indicações de que as externalidades marshallianas são importantes para explicar o padrão de concentração entre diferentes indústrias ou coaglomeração industrial. Entre os primeiros estudos realizados no Brasil, destaca-se o de Resende (2012), que encontrou evidências de que o nível de coaglomeração industrial no estado do Rio de Janeiro estava associado ao fator de aglomeração labor pooling. Considerado como referência no país, Maciente (2013) investigou a coaglomeração entre todos os setores produtivos e mostrou que a relação insumo-produto, no Brasil, é importante para explicar a coaglomeração entre pares de indústrias. Além disso, o autor aponta que na Zona Franca de Manaus há uma significativa coaglomeração entre indústrias de alta tecnologia.

Buscando contribuir para essa área do conhecimento, este trabalho possui como objetivo geral: quantificar o nível de coaglomeração, entre pares de indústrias distintas, para todo o setor da indústria de transformação, relacionando o nível de coaglomeração industrial com os fatores de aglomeração marshallianos. A partir desses objetivos, e com as contribuições de Resende (2012) e Maciente (2013), é possível identificar, de maneira específica, quais indústrias apresentam uma maior tendência à coaglomeração e investigar os fatores locais que estão associados à coaglomeração no Brasil. Além disso, como as regiões brasileiras se diferenciam quanto à sua estrutura produtiva e intensidade dos fatores de produção, o artigo tem um olhar para a ciência regional, dado que o índice é construído a partir da colocação entre duas indústrias na mesma localidade, permitindo relacionar as evidências encontradas acerca dos padrões de coaglomeração com fatores produtivos dos estados e das microrregiões.

Para tanto, foi utilizada uma estratégia empírica que permite controlar os efeitos fixos de indústrias e de tempo. No campo metodológico, o trabalho tem duas limitações: o modelo foi estruturado para estimar efeitos estáticos das aglomerações e a variável *intellectual spillovers* carece de cuidado na análise, dada a dificuldade de mensuração para capturar um efeito de interação interpessoal. Além disso, a característica endógena dos fatores de aglomerações dificulta a análise até que se possam isolar os efeitos marshallianos, uma vez que as *externalidades marshallianas* interagem entre si influenciando a coaglomeração entre as indústrias.

A base de dados utilizada é a Relação Anual de Informações Sociais (RAIS), advinda

do Ministério da Economia. Utilizam-se também as matrizes insumo-produto calculadas por Guilhoto et al. (2010) e disponibilizadas pelo Núcleo de Economia Regional e Urbana da Universidade de São Paulo (NEREUS) e as classificações tecnológicas das indústrias apresentadas por Cavalvante (2014). As unidades geográficas de análise escolhidas foram os estados e as microrregiões, enquanto o período de análise compreende os anos de 2006 à 2015.

Este artigo está organizado em seis seções, incluindo esta introdução. A próxima seção apresenta o índice de coaglomeração e discute as evidências empíricas relevantes para o tema em questão. Nas seções três e quatro, são detalhados o modelo empírico e o conjunto de variáveis, juntamente com suas respectivas bases de dados, que serão empregados na investigação e decomposição dos fatores correlacionados à coaglomeração. A quinta seção aborda a análise da coaglomeração no contexto brasileiro e fornece uma discussão aprofundada dos resultados obtidos. Por fim, na última seção, são apresentadas as conclusões que emergem desta pesquisa, consolidando os principais insights e apontando para possíveis direções futuras de investigação.

## 2. Coaglomeração e Externalidades Marshallianas

Até meados da década de 1990, a maior parte da literatura quantificava a concentração produtiva no nível de aglomeração através de indicadores como o *Gini locacional*, o índice de concentração geográfica e o índice *Hirschman-Herfindahl*, os quais eram amplamente utilizados na época. No entanto, a partir do trabalho seminal de Ellison e Glaeser (1997), observou-se uma crescente sofisticação nos indicadores de concentração produtiva. Esses novos indicadores passaram a incorporar informações sobre a localização, o nível de especialização e a diversificação das atividades econômicas, levando em consideração fatores internos e externos às firmas, como o tamanho do mercado, a escala de produção e a população, conforme discutido por Nakamura e Paul (2019). Esse avanço metodológico contribuiu significativamente para uma análise mais abrangente e refinada da dinâmica de aglomeração e concentração produtiva.

Considerando os indicadores de concentração mais utilizados na literatura, é possível agrupá-los entre as medidas discretas (Ellison e Glaeser, 1997; Maurel e Sédillot, 1999; Ellison et al., 2010; Howard et al., 2016). e as medidas contínuas, para as quais é preciso obter informações da localização georreferenciada das firmas (Duranton e Overman, 2005).

Apesar de haver um consenso na literatura quanto à contribuição e importância dos indicadores discretos para quantificar o nível da concentração da atividade produtiva, também existem críticas destrutivas. As medidas discretas são criticadas pela limitação da sensibilidade dos resultados à escolha geográfica de análise, também chamado de *Modifiable Areal Unit Problem* (MAUP), enquanto as medidas contínuas apresentam questionamentos quanto à sua aplicabilidade e complexidade (Kominers,

2008).

Em relação às medidas discretas, Ellison et al. (2010) expandiram o índice de concentração de Ellison e Glaeser (1997) – que está associado a uma medida de dispersão e quantifica o nível de concentração de uma indústria –, para captar o nível de coaglomeração de pares de indústrias diferentes. Os autores argumentam que, devido à facilidade do cálculo e sua análise com apenas duas indústrias, as relações interseccionais existentes surgem de forma mais clara e precisa. O índice de coaglomeração é aplicado para as indústrias da manufatura dos Estados Unidos, e os autores comparam os resultados com os obtidos a partir da utilização do índice de Duranton e Overman (2005). A partir da comparação entre os resultados, os autores não evidenciaram diferenças significativas entre os resultados obtidos, validando, portanto, o índice EGK proposto para mensurar o nível da aglomeração das indústrias.

## 2.1 Índice de Coaglomeração

O índice proposto por Ellison et al. (2010) é a variável dependente do modelo, sendo denotada como  $EGK_{ij}$ , indica o nível de associação linear da participação do emprego entre pares de indústrias  $i$  e  $j$ , para determinada área geográfica, denotado por:

$$EGK_{ij} = \frac{\sum_{m=1}^M (S_{mi} - X_m)(S_{mj} - X_m)}{1 - \sum_{m=1}^M X_m^2} \quad (1)$$

Nesse índice,  $m$  indexa as localidades,  $S_{mi}$  e  $S_{mj}$  denotam a participação do emprego da indústria  $i$  e  $j$ , respectivamente, na localidade  $m$ , onde  $i \neq j$ . Por fim,  $X_m$  representa a participação do emprego de cada subárea  $m$  no total do emprego da área agregada  $M$ .

O índice EGK mede o grau de coaglomeração entre duas indústrias em relação a uma área agregada, para o qual quanto maior o grau de coaglomeração mais próximo é o padrão entre a distribuição geográfica das duas indústrias. Por exemplo, com as devidas proporções, se há dez localidades geográficas e duas indústrias se localizam apenas em uma localidade conjuntamente, esse valor tende a ser próximo de 1, enquanto que, se duas indústrias se localizam em cinco localidades cada uma, distintas entre si, o valor tende a ser menor que 0 (zero). Valores próximos a zero indicam que a coaglomeração acontece de forma aleatória. Isto é, pares de indústrias podem estar presentes conjuntamente em determinadas localidades e não conjuntamente em outras localidades.

Os valores dos índices também são afetados pela desagregação geográfica escolhida para a análise. Por exemplo, a desagregação em nível estadual é composta por 26 localidades (estados), enquanto a desagregação no nível microrregional possui 556 localidades (microrregiões), então é de se esperar que o índice de coaglomeração seja maior para o primeiro caso em comparação com as microrregiões. Porém, o que é mais

relevante em estudos como este é a variabilidade do índice e o ranking das indústrias conforme o seu nível de coaglomeração. Apesar disso, para adicionar robustez aos resultados, a análise foi realizada considerando as Unidades da Federação e as microrregiões.

Esse índice é estreitamente relacionado com a covariância do emprego industrial nas duas indústrias. O conceito é derivado da ideia de que indústrias distintas se concentram em determinada localidade para obter ganhos de produtividade e/ou escala, por exemplo, firmas em uma indústria que a relação de insumo-produto é importante tendem a se aglomerar próximas às indústrias que são seus consumidores ou produtores. E, dessa forma, é possível associar separadamente os pares de indústrias coaglomerados com as externalidades compartilhadas entre elas, sendo uma vantagem do índice em relação aos demais apresentados na seção anterior.

A fonte de dados predominantemente empregada neste estudo deriva da Relação Anual de Informações Sociais (RAIS), disponibilizada pelo Ministério da Economia (anteriormente conhecido como Ministério do Trabalho e Emprego (MTE)). A partir dessa fonte, foram adquiridos dados relativos ao emprego nas indústrias de transformação, classificadas por divisão de acordo com a Classificação Nacional de Atividades Econômicas 2.0 (CNAE 2.0), totalizando 24 setores industriais e 276 pares de industriais. A escolha do nível de desagregação por divisão foi motivada pela necessidade de se alinhar com as variáveis explicativas, as quais, devido a limitações de informações, também estão categorizadas por divisão.

Para a construção do índice de EGK, foram utilizados dados de emprego de cada indústria em cada localidade observada, permitindo verificar em qual localidade o emprego está mais concentrado e, conseqüentemente, quais pares de indústrias tendem a se localizar conjuntamente. Os índices foram calculados para cada ano, a partir de 2006 até 2015.

## 2.2 Evidências Empíricas

O índice de EGK tem sido aplicado para investigar a coaglomeração entre a indústria da transformação e o setor de serviços (Kolko, 2010), explorando os diferentes grupos de ocupações dos trabalhadores (Gabe e Abel, 2012), considerando os setores formal e informal (Mukim, 2015), e demais grupos de indústrias que atuam atraindo, em potencial, outras indústrias, como as intensivas em tecnologias.

Nesse contexto, a pesquisa empírica conduzida por Ellison et al. (2010) forneceu respaldo às externalidades marshallianas como fator determinante da coaglomeração entre pares de indústrias nos Estados Unidos, utilizando uma abordagem empírica com variáveis instrumentais. Os autores realizaram uma comparação entre os resultados relativos às forças marshallianas e o papel das vantagens naturais na explicação do nível de coaglomeração entre pares de indústrias. A conclusão alcançada foi a de que os fatores marshallianos demonstraram ter maior relevância como impulsionado-

res das aglomerações industriais nos EUA.

Com relação à contribuição da literatura empírica para a temática de pesquisa, destaca-se Kolko (2010), que foi um dos primeiros a aplicar o índice de EGK para outros setores, analisando a coaglomeração entre a indústria de serviços através das economias de urbanização. O autor demonstra que os estabelecimentos do setor de serviços são mais urbanizados e menos aglomerados que a indústria manufatureira, além de concluir que o labor pooling não é um potencial fator de explicação para as decisões de localização dos estabelecimentos do setor de serviços.

Por outro lado, diante da percepção da multifuncionalidade do índice, Gabe e Abel (2012) propuseram sua utilização para mensurar a coaglomeração entre ocupações nos Estados Unidos. A pesquisa revelou que a similaridade de conhecimento desempenha um papel determinante na coaglomeração de ocupações, especialmente em áreas metropolitanas em comparação com estados individuais. No artigo, os autores evidenciaram que empresas que empregam trabalhadores com perfis semelhantes tendem a se aglomerar em proximidade umas das outras.

Com outra perspectiva de análise, Mukim (2015) utilizou o índice de EGK para mensurar a coaglomeração entre emprego formal e informal dentro de cada indústria de manufatura da Índia. O autor concluiu que, dentro da mesma indústria, os vínculos comprador-vendedor conjuntamente com os *spillovers* tecnológicos explicam em boa parte a coaglomeração. Esses *spillovers* tecnológicos decorrem, principalmente, de onde as indústrias com maior tecnologia tendem a se coaglomerar.

Para o Vietnã, Howard et al. (2016) modificaram o índice de EGK para analisar a coaglomeração entre pares de indústrias utilizando as localizações das firmas e não mais o emprego. Os pesquisadores comparam os resultados do novo índice com os apresentados por Ellison et al. (2010), encontrando evidências divergentes. Todavia, o índice de EGK permite dar base metodológica à pesquisa, indicando a multifuncionalidade já destacada.

No caso dos estudos para o Brasil, Resende (2012) fez uma investigação dos determinantes de aglomeração na indústria de transformação do Rio de Janeiro no ano de 2010. Seguindo a linha do artigo de Ellison et al. (2010), ele obteve evidências de que o papel do *labor pooling* é mais forte do que as outras forças de aglomeração para as variáveis, aproximando-o da intensidade no uso de insumos. Por seu turno, Maciente (2013) em um capítulo da sua tese, investigou a coaglomeração no Brasil, de maneira geral, entre todas as indústrias da economia, mostrando que a relação insumo-produto neste país apresenta fortes evidências relacionadas à coaglomeração e que a Zona Franca de Manaus tem forte coaglomeração entre indústrias de alta tecnologia. Assim, este trabalho contribui com as evidências aumentando o escopo da análise para todo o Brasil, possibilitando melhor visualização das relações interindustriais. Através do uso de dados em painel, com o objetivo de controlar os efeitos ao longo do tempo e as características específicas das indústrias, são encontradas novas

evidências para o Brasil acerca da relação entre a coaglomeração e as externalidades marshallianas.

Os avanços mais recentes nessa literatura seguem no sentido de identificar o impacto, a partir de choques exógenos, dos fatores locais que estão associados à geração de economias de aglomeração no nível de concentração da atividade produtiva. Nesse sentido, citam-se as contribuições recentes de Steijn et al. (2022), os quais fazem uma análise da dinâmica e dos fatores produtivos locais relacionados à coaglomeração das indústrias dos Estados Unidos para um período de 44 anos (1970 a 2014). A principal inovação dos autores está na mensuração dos *spillovers* tecnológicos, captados através de dados de patentes. Os resultados dos autores são consistentes com o aumento da importância dos fatores de produção associados à geração de tecnologia e conhecimento para explicar a coaglomeração das industriais, quando comparados ao fator *labor pooling*. Ademais, os autores evidenciaram que choques comerciais e tecnológicos tenderam a reduzir o fator *labor pooling* e a aumentar os *spillovers* tecnológicos, o que deve repercutir no poder aglomerativo desses fatores.

### 3. Modelo Empírico

A estratégia empírica escolhida para associar *proxies* das externalidades marshallianas com o índice de coaglomeração industrial é similar à utilizada por Ellison et al. (2010), com o acréscimo de uma série de dados em painel com efeitos fixos de tempo, conforme especificado na Equação (2):

$$EGK_{ijt} = \beta_1 LAB_{ijt} + \beta_2 INPUT_{ijt} + \beta_3 OUT_{ijt} + \beta_4 INT_{ijt} + \alpha_{ijt} + \epsilon_{ijt} \quad (2)$$

Em que o termo  $EGK_{ijt}$  representa o índice de coaglomeração; o termo  $\alpha_{it}$  representa os efeitos fixos;  $\beta_1$  é o coeficiente que identifica o fator *labor pooling* representado por  $LAB_{ijt}$ ;  $\beta_2$  e  $\beta_3$  são os coeficientes que estão associados aos fatores *input* e *output linkages*, respectivamente dados por  $INPUT_{ijt}$  e  $OUT_{ijt}$ ;  $\beta_4$  é o coeficiente que está interligado as *proxies* referentes aos *intellectual spillovers* representados por  $INT_{ijt}$ ; os subscritos  $i$  e  $j$  denotam o par de indústrias e o subscrito  $t$  denota o ano do índice; por fim, o termo de erro está representado pela notação  $\epsilon_{it}$ .

Com base na Equação (2), um segundo modelo é estimado com a incorporação de variáveis *dummies* de intensidade tecnológica por Mínimos Quadrados Ordinário (MQO), similarmente a Ellison et al. (2010). Com a introdução das *dummies* pode-se avaliar a relação entre a similaridade tecnológica dos pares de indústrias e a coaglomeração entre si, comparando-a com a *dummy* omitida que representa a diferença tecnológica entre os pares de indústrias. A equação é dada agora por:

$$EGK_{ijt} = \beta_1 LAB_{ijt} + \beta_2 INPUT_{ijt} + \beta_3 OUT_{ijt} + \beta_4 INT_{ijt} + \beta_5 D_{AA} + \beta_6 D_{AB} + \beta_7 D_{BB} + \epsilon_{ijt} \quad (3)$$

Nela, o termo  $EGK_{ijt}$  representa o índice de coaglomeração;  $\beta_1$  é o coeficiente que identifica o fator *labor pooling* representado por  $LAB_{ijt}$ ;  $\beta_2$  e  $\beta_3$  são os coeficientes que estão associados aos fatores *input* e *output linkages*, respectivamente dados por  $INPUT_{ijt}$  e  $OUT_{ijt}$ ;  $\beta_4$  é o coeficiente que está interligado as *proxies* referentes aos *intellectual spillovers* representados por  $INT_{ijt}$ ;  $D_{AA}$  denota a *dummy* em que ambas as indústrias são consideradas de alta tecnologia;  $D_{BB}$  representa que ambas as indústrias são consideradas de baixa tecnologia; enquanto  $D_{AB}$  é a *dummy* em que o par de indústria apresenta classificação tecnológica distinta; os subscritos  $i$  e  $j$  denotam o par de indústrias e o subscrito  $t$  denota o ano do índice;  $\beta_5$ ,  $\beta_6$  e  $\beta_7$  são os coeficientes associados às *dummies*; por fim, o termo de erro está representado pela notação  $\epsilon_{it}$ .

O modelo da equação (2) e da Equação (3) é estimado para dois períodos distintos, visto que algumas variáveis apresentam diferentes séries de tempo. Foi utilizado o período de 2006 a 2015 para estimações dos fatores *Labor* e *Intellectual*, enquanto para as variáveis *Input* e *Output* foi utilizado o período de 2010 a 2013. Nas regressões multivariadas, o período de 2010 a 2013 foi empregado, representando um total de 1020 observações para 276 combinações de pares de indústrias.

## 4. Indicadores de Externalidades Marshallinas

Os dados utilizados para construção da variável *Labor Pooling* são segmentados por ocupações obtidas da Classificação Brasileira de Ocupações (CBO) ao nível dos principais subgrupos, contabilizando 48 ocupações, advindas das informações dos trabalhadores que compõem a RAIS.

Em relação à base de dados das variáveis *Input* e *Output linkages*, são utilizadas as informações das matrizes de insumo-produto advindas do Núcleo de Economia Regional e Urbana da USP (NEREUS). A matriz foi construída por Guilhoto et al. (2010) a partir de dados das contas nacionais para o período 1995-2013. No período 2006-2009, a agregação é maior que o período 2010-2013. Dessa forma, opta-se pelo período 2010-2013 para compatibilização com as demais variáveis.

As *proxies* para a variável *Intellectual Spillovers* (capital humano) também utilizam os dados da RAIS, visto que são disponibilizadas informações dos trabalhadores acerca do grau educacional em cada indústria, sendo possível a compatibilização para comparação das indústrias analisadas. O período observado para ambas as variáveis compreende os anos de 2006 a 2015.

### 4.1 Labor Pooling

O compartilhamento de mão de obra é um dos fatores de aglomeração abordado por Marshall (1890), que enfatiza a importância das economias de escala associadas ao mercado de trabalho. Diamond e Simon (1990) e Krugman (1991) argumentam que no decorrer do tempo, as empresas passam a ser mais ou menos produtivas e

a mobilidade dos trabalhadores entre os empregos proporciona ganhos de produtividade e diminuição do diferencial salarial. Helsley e Strange (1990) propuseram que as aglomerações facilitam o match entre emprego e trabalhador e suas aptidões. Combes e Duranton (2006) e Dahl e Klepper (2015) por sua vez, sugeriram que aglomerações com maior facilidade de mobilidade de trabalhadores entre as firmas incentivam a localização de novos empreendedores, decorrente da força de trabalho adequada, levando ao aumento da concentração do emprego.

Todas essas teorias partem do pressuposto de que as aglomerações ocorrem porque os trabalhadores são capazes de se mover entre as firmas e indústrias. Porém, isso pode acontecer apenas quando os trabalhadores apresentarem o mesmo perfil necessário para essas indústrias. Para essa variável foi escolhida uma *proxy* baseada na literatura existente, que fosse capaz de captar os efeitos na variação dos índices de coaglomeração. Dentre as opções disponíveis na literatura, optou-se por utilizar a mesma ideia dos artigos de Ellison et al. (2010) e Resende (2012). Desta forma, este trabalho se baseia no conceito de labor pooling, que capta a disponibilidade e a concentração de trabalhadores ocupados em atividades semelhantes.

A variável será denotada como  $LAB_{ijt}$  e é definida como a proporção da atividade  $a$  nos perfis de trabalhadores das indústrias  $i$  e  $j$  de maneira análoga ao índice de coaglomeração. Mensura-se a similaridade dos trabalhadores nas indústrias por meio da correlação das participações das atividades  $a$  em ambas indústrias  $i$  e  $j$ , no tempo  $t$ . Assim:

$$LAB_{ijt} = \frac{\sum_{a=1}^A (L_{ai} - X_a)(L_{aj} - X_a)}{1 - \sum_{a=1}^A X_a^2} \quad (4)$$

Na equação,  $a$  indexa as atividades ocupacionais,  $L_{ai}$  e  $L_{aj}$  denotam a participação da atividade  $a$  nas indústrias  $i$  e  $j$ , respectivamente, e  $X_a$  representa a participação da atividade  $a$  no total das atividades agregadas  $A$ .

## 4.2 Input-Output Linkages

A proximidade entre fornecedores e clientes favorece a redução de custos de transportes, levando à verticalização entre as indústrias e incentivando a aglomeração, conforme demonstrado por Marshall (1890). Desta forma, a teoria aponta para uma aproximação entre as indústrias que tem forte relação insumo-produto. Fujita et al. (1999) apontaram a redução dos custos de transportes dos bens (insumo ou produto) como principal motivo a aglomerar. Para melhor identificação a variável foi dividida em duas composições: *Input Linkage* e *Output Linkage*. Deste modo, é possível captar os fatores referentes a correlação de utilização de insumos e produtos separadamente.

Define-se  $Input_{j \Rightarrow i}$  como a participação dos insumos da indústria  $i$  advindos da indústria  $j$ . E também se define  $Output_{i \Rightarrow j}$  como a participação dos produtos da indústria  $i$  que são vendidos para a indústria  $j$ , e vice versa. Essas participações são

calculadas entre todas as indústrias, desse modo o valor do índice será entre 0 e 1. Dado que há relações de ambas as direções, definem-se as variáveis  $INPUT_{ijt}$  e  $OUT_{ijt}$  como a média das participações dos pares de indústrias  $i \Rightarrow j$  e  $j \Rightarrow i$ . Denota-se:

$$INPUT_{ijt} = media\{Input_{i \Rightarrow j}; Input_{j \Rightarrow i}\} \quad (5)$$

$$OUT_{ijt} = media\{Output_{i \Rightarrow j}; Output_{j \Rightarrow i}\} \quad (6)$$

### 4.3 Intellectual Spillovers

A última variável proposta por Marshall (1890) diz respeito à flutuação de conhecimento. O autor enfatiza que trabalhadores podem aprender novas técnicas uns com os outros em clusters industriais, levando ao aumento da produtividade. Essa variável, todavia, apresenta um grande problema: o alto grau de dificuldade de mensuração, como pode ser detalhado em Henderson (2007). A literatura tenta resolver esse problema aplicando conceitos relacionados à difusão de conhecimento, sendo a % de trabalhadores com ensino superior Moretti (2004) e o nível de P&D das indústrias Autant-Bernard (2001) as principais *proxies* para capturar o efeito da variável. Trabalhos empíricos reforçam essa ideia mostrando que indústrias tecnológicas, com elevado nível de P&D e com alto grau de capital humano influenciam na aglomeração (Saxenian, 1996; Glaeser e Kahn, 2001; Kolko, 2010).

Além do problema acima mencionado, há também outra limitação importante dessa variável, dado que o modelo é endógeno e busca estimar a correlação entre a coaglomeração e as externalidades marshallianas, essa variável pode sofrer efeito das demais variáveis explicativas e também da variável dependente. Para minimizar essa possível influência, utilizam-se duas estratégias. A primeira é seguir a linha de capital humano dos trabalhos citados anteriormente, representado pelo grau escolar dos trabalhadores, e utilizando diferentes *proxies* para tentar ser mais assertivo possível sobre a relação estimada. A segunda estratégia é utilizar *dummies* representando a classificação tecnológica das indústrias, tendo em vista a dificuldade de base de dados de patentes e P&D no Brasil. Apesar dos esforços metodológicos e mesmo seguindo uma literatura consolidada, ressalta-se a importância de uma análise cuidadosa dessa variável.

Assim, a primeira *proxy* escolhida procura verificar se há correlação entre indústrias com maior proporção de trabalhadores qualificados e indústrias que apresentam menor proporção de trabalhadores qualificados, similar à Moretti (2004). Em seguida é gerada uma *proxy* para relacionar os efeitos das indústrias com perfis educacionais similares na coaglomeração, seguindo a linha de raciocínio apresentado por Kolko (2010). Por fim, a terceira *proxy* busca ponderar a proporção média de trabalhadores com ensino superior entre as indústrias.

A *proxy HighIntellect* é definida como a proporção máxima de trabalhadores com ensino superior ou acima entre pares de indústrias  $i$  e  $j$ . Sendo denotada por:

$$HighIntellect_{ij} = \max\{EducInd_i; EducInd_j\} \quad (7)$$

Em que  $\{EducInd_i; EducInd_j\}$  representam a proporção de trabalhadores com ensino superior ou acima para a indústria  $i$  e  $j$ , respectivamente.

A correlação entre o perfil educacional das indústrias  $i$  e  $j$  é definida pela *proxy SimIntellect*, denotada como:

$$SimIntellect_{ij} = \frac{\sum_{m=1}^M (E_{mi} - Ex_m)(E_{mj} - Ex_m)}{1 - \sum_{m=1}^M Ex_m^2} \quad (8)$$

Onde  $m$  indexa os níveis educacionais,  $E_{mi}$  e  $E_{mj}$  denotam a participação do nível educacional da indústria  $i$  e  $j$ , respectivamente, no nível educacional  $m$ , e  $Ex_m$  representa o tamanho agregado dos níveis educacionais  $m$  mensurados em termo da participação média do nível educacional entre o total das indústrias.

Por fim, define-se a variável *Intellect* como a correlação entre as proporções de trabalhadores com ensino superior nos pares de indústrias  $i$  e  $j$ , sendo denotado por:

$$Intellect_{ij} = media\{EducInd_i; EducInd_j\} \quad (9)$$

Onde,  $\{EducInd_i; EducInd_j\}$  representam a proporção de trabalhadores com ensino superior para a indústria  $i$  e  $j$ , respectivamente.

Para a análise referente à tecnologia das indústrias, foram criadas *dummies* levando em consideração a classificação proposta por Cavalcante (2014), em que o autor elenca as indústrias por grau tecnológico, sendo eles divididos entre baixa tecnologia intensiva, média-baixa, média-alta e alta, optando-se, neste trabalho, por agrupar e classificar em dois grupos: baixa tecnologia e alta tecnologia. Assim, foram feitas combinações entre os graus tecnológicos nos pares de indústrias acarretando na criação das *dummies*: baixa-baixa igual a um (zero para as demais combinações), alta-baixa igual a um (zero para as demais combinações), alta-alta igual a um (zero para as demais combinações). A lista das indústrias e a respectiva classificação tecnológica são dadas na Tabela 1.

$$\begin{cases} Baixa \iff Baixa = 1; & \text{Demais combinações} = 0 \\ Alta \iff Alta = 1; & \text{Demais combinações} = 0 \\ Baixa \iff Alta = 1; & \text{Demais combinações} = 0 \end{cases} \quad (10)$$

**Tabela 1.** Indústrias de transformação por divisão e intensidade tecnológica - OCDE 2014

Divisão	indústria (CNAE 2.0)	Class. OCDE
10	Fabricação de Produtos Alimentícios	Baixa
11	Fabricação de Bebidas	Baixa
12	Fabricação de Produtos do Fumo	Baixa
13	Fabricação de Produtos Têxteis	Baixa
14	Confecção de Artigos do Vestuário e Acessórios	Baixa
15	Preparação de Couros e Fabricação de Artefatos de Couro, Artigos para Viagem e Calçados	Baixa
16	Fabricação de Produtos de Madeira	Baixa
17	Fabricação de Celulose, Papel e Produtos de Papel	Baixa
18	Impressão e Reprodução de Gravações	Baixa
19	Fabricação de Coque, de Produtos Derivados do Petróleo e de Biocombustíveis	Média-baixa
20	Fabricação de Produtos Químicos	Média-Alta
21	Fabricação de Produtos Farmoquímicos e Farmacêuticos	Alta
22	Fabricação de Produtos de Borracha e de Material Plástico	Média-baixa
23	Fabricação de Produtos de Minerais Não Metálicos	Média-baixa
24	Metalurgia	Média-baixa
25	Fabricação de Produtos de Metal, Exceto Máquinas e Equipamentos	Média-baixa
26	Fabricação de Equipamentos de Informática, Produtos Eletrônicos e Ópticos	Alta
27	Fabricação de Máquinas, Aparelhos e Materiais Elétricos	Média-alta
28	Fabricação de Máquinas e Equipamentos	Alta
29	Fabricação de Veículos Automotores, Reboques e Carrocerias	Média-alta
30	Fabricação de Outros Equipamentos de Transporte, Exceto Veículos Automotores	Média-Alta
31	Fabricação de Móveis	Baixa
32	Fabricação de Produtos Diversos	Baixa
33	Manutenção, Reparação e Instalação de Máquinas e Equipamentos	Baixa

Fonte: Elaboração dos autores, com base em Cavalcante (2014).

## 5. Resultados

Para análise exploratória dos índices de coaglomeração, analisaram-se 24 indústrias classificadas segundo divisão (2 dígitos) CNAE 2.0, resultando em 276 combinações, no período de 2006 a 2015. A análise nível estadual contempla 26 estados da Federação e o Distrito Federal. Já análise microrregional foi construída a partir da distribuição geográfica do ano base 2006, envolvendo, na época, 556 microrregiões.

### 5.1 Padrão de Coaglomeração Industrial

A magnitude dos índices de coaglomeração variou em torno de valores próximos de zero, assim como os obtidos pela literatura aqui citada, e 46,9% dos índices estimados apresentaram o sinal positivo. Baseando-se em Ellison, Glaeser e Kerr (2010), adotou-se um patamar acima de 0,01 para identificar os pares de indústrias mais intensamente coaglomeradas. Considerando esse valor, 11% dos pares de indústrias consideradas na análise são significativamente coaglomerados, sendo o índice calcu-

lado para as Unidades da Federação responsável por 92,37% desses pares de indústrias. Ressalta-se que esse resultado era previsto, uma vez que a escala do índice tende a aumentar com o nível de agregação geográfica. Porém, o que se deve considerar é o ranking dos índices para identificar quais as indústrias são potencialmente mais coaglomeradas.

Nesse sentido, ao se observa as Tabelas 2 e 3, pode-se notar a repetição das indústrias de Fabricação de Produtos do Fumo (12) e Fabricação de Produtos Químicos (20) entre os principais pares coaglomerados. Essas duas indústrias colaboram para reforçar a existência de padrões de coaglomeração em determinadas indústrias em maior intensidade e, conseqüentemente, a possível presença das forças de aglomeração Marshallianas.

**Tabela 2.** Maiores médias dos índices de coaglomeração - Nível estadual

Indústria i (CNAE Divisão)	Indústria j (CNAE Divisão)	$EGK_{i,j}$
Fabricação de Produtos do Fumo (12)	Prep. de Couros e Fab. de Artefatos de Couro, Artigos para Viagem e Calçados (15)	0.1370
Fabricação de Produtos Químicos (20)	Fabricação de Veículos Automotores, Reboques e Carrocerias (29)	0.0539
Fabricação de Produtos Químicos (20)	Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	0.0459
Fabricação de Produtos do Fumo (12)	Fabricação de Móveis (31)	0.0437
Fabricação de Máquinas e Equipamentos (28)	Fabricação de Veículos Automotores, Reboques e Carrocerias (29)	0.0430
Fab. de Equipamentos de Informática, Produtos Eletrônicos e Ópticos (26)	Fab. de Outros Equipamentos de Transporte, Exceto Veículos Automotores (30)	0.0410
Fabricação de Produtos do Fumo (12)	Fabricação de Produtos de Madeira (16)	0.0400
Fabricação de Produtos de Madeira (16)	Fabricação de Móveis (31)	0.0387
Fabricação de Produtos Químicos (20)	Fabricação de Veículos Automotores, Reboques e Carrocerias (29)	0.0386
Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	Fabricação de Produtos Diversos (32)	0.0375

Fonte: Elaboração dos autores.

**Tabela 3.** Maiores médias dos índices de coaglomeração - Nível microrregional

Indústria i (CNAE Divisão)	Indústria j (CNAE Divisão)	$EGK_{i,j}$
Fab. de Equipamentos de Informática, Produtos Eletrônicos e Ópticos (26)	Fab. de Outros Equipamentos de Transporte, Exceto Veículos Automotores (30)	0.0324
Impressão e Reprodução de Gravações (18)	Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	0.0182
Fabricação de Produtos Químicos (20)	Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	0.0160
Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	Fabricação de Veículos Automotores, Reboques e Carrocerias (29)	0.0159
Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	Fabricação de Produtos de Borracha e de Material Plástico (22)	0.0113
Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	Fabricação de Produtos Diversos (32)	0.0112
Impressão e Reprodução de Gravações (18)	Fabricação de Produtos Químicos (20)	0.0100

Fonte: Elaboração dos autores.

Ao comparar os níveis estadual e microrregional, observa-se que alguns dos índices microrregionais superiores a 0,01 mantêm a mesma lógica em relação às forças de aglomeração. A recorrência das indústrias nos níveis de Fabricação de Produtos Químicos (20) e Fabricação de Produtos Farmoquímicos e Farmacêuticos (21) indica que essas indústrias específicas têm uma marcante propensão à coaglomeração com indústrias diversas. Destaca-se a presença de três pares de indústrias com índices mais elevados em ambos os níveis, indicando uma forte inclinação para a aglomeração, independentemente da escala de desagregação. Esse resultado atenua a problemática do MAUP associada ao índice de coaglomeração, abordada anteriormente, reforçando que, independentemente da desagregação geográfica, alguns padrões permanecem consistentes.

Esses padrões são visualizados na tabela abaixo, que apresenta a percentagem de índices superiores a 0,01 para cada indústria, refletindo o grau de coaglomeração entre elas. Por exemplo, em 2006, a indústria de Fabricação de Outros Equipamentos

de Transporte, Exceto Veículos Automotores (30) estava coaglomerada com 52,17% das indústrias classificadas na CNAE 2.0, resultando em 12 das 23 combinações possíveis.

A análise da Tabela 4, a seguir, revela que algumas indústrias experimentaram significativa redução na coaglomeração. Destaca-se, especialmente, a indústria de Fabricação de Outros Equipamentos de Transporte, Exceto Veículos Automotores (30), que, em 2006, era a mais coaglomerada entre todas as analisadas, e a Fabricação de Produtos Têxteis (13), que sofreu uma queda abrupta, diminuindo de 34,78% para 8,70%.

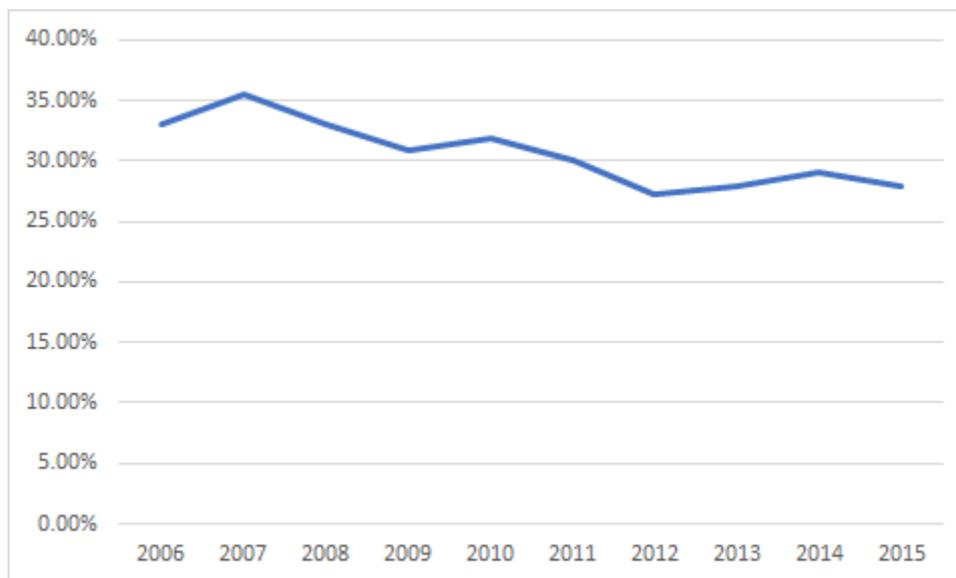
**Tabela 4.** Distribuição dos índices de EGK > 0.01 - Nível Estadual

Setor (CNAE Divisão)	% EGK > 0.01 - 2006	% EGK > 0.01 - 2015
Fabricação de Outros Equipamentos de Transporte, Exceto Veículos Automotores (30)	52.17%	26.09%
Fabricação de Produtos Químicos (20)	47.83%	43.48%
Fabricação de Produtos Farmoquímicos e Farmacêuticos (21)	47.83%	47.83%
Fabricação de Equipamentos de Informática, Produtos Eletrônicos e Ópticos (26)	47.83%	43.48%
Fabricação de Máquinas, Aparelhos e Materiais Elétricos (27)	47.83%	39.13%
Fabricação de Veículos Automotores, Reboques e Carrocerias (29)	47.83%	43.48%
Fabricação de Produtos Diversos (32)	47.83%	43.48%
Fabricação de Celulose, Papel e Produtos de Papel (17)	43.48%	34.78%
Fabricação de Produtos de Borracha e de Material Plástico (22)	43.48%	34.78%
Fabricação de Máquinas e Equipamentos (28)	43.48%	39.13%
Fabricação de Produtos de Metal, Exceto Máquinas e Equipamentos (25)	39.13%	26.09%
Fabricação de Produtos Têxteis (13)	34.78%	8.70%
Fabricação de Produtos de Madeira (16)	34.78%	26.09%
Fabricação de Bebidas (11)	30.43%	26.09%
Fabricação de Produtos Alimentícios (10)	26.09%	17.39%
Fabricação de Produtos do Fumo (12)	26.09%	26.09%
Preparação de Couros e Fabricação de Artefatos de Couro, Artigos para Viagem e Calçados (15)	26.09%	21.74%
Impressão e Reprodução de Gravações (18)	26.09%	21.74%
Fabricação de Móveis (31)	21.74%	17.39%
Manutenção, Reparação e Instalação de Máquinas e Equipamentos (33)	21.74%	17.39%
Confecção de Artigos do Vestuário e Acessórios (14)	17.39%	13.04%
Fabricação de Coque, de Produtos Derivados do Petróleo e de Biocombustíveis (19)	13.04%	17.39%
Fabricação de Produtos de Minerais Não-Metálicos (23)	13.04%	17.39%
Metalurgia (24)	8.70%	8.70%

Fonte: Elaboração dos autores.

Entretanto, essa redução não parece ser um caso isolado, conforme evidenciado pelo Gráfico 1, a seguir, que ilustra a evolução dos índices com valores superiores a 0,01 entre os anos de 2006 e 2015. Notavelmente, observa-se uma tendência de decréscimo na coaglomeração durante o período analisado, sugerindo a possível influência de outros fatores, como a variação negativa do emprego nesse intervalo. É relevante ressaltar que, na maioria das indústrias, não ocorreram alterações significativas entre 2006 e 2015. Contudo, aquelas que apresentaram variações majoritariamente exibiram uma tendência à redução da coaglomeração, com exceção das indústrias de Fabricação de Coque, de Produtos Derivados do Petróleo e de Biocombustíveis (19) e Fabricação de Produtos de Minerais Não-Metálicos (23), que demonstraram uma leve variação ascendente.

Outro ponto relevante é que, de acordo com os dados da RAIS, das 556 microrregiões existentes no Brasil em 2006, as 15 microrregiões com o maior número de empregos concentraram 40,2% do emprego total na indústria de transformação. Essas microrregiões incluem as cidades de São Paulo, Rio de Janeiro, Campinas, Porto Alegre, Belo Horizonte, Curitiba, Blumenau, Caxias do Sul, Fortaleza, Joinville, Gua-

**Figura 1.** Gráfico - Evolução da proporção de índices > 0.01 - 2006 a 2015

Fonte: Elaboração dos autores.

ruhos, Sorocaba, Manaus, Osasco e São José dos Campos. Em 2011, essas mesmas microrregiões detinham 39,27% dos empregos na indústria de transformação e, em 2016, essa participação diminuiu para 35,77%, indicando uma tendência à desconcentração industrial. Esses dados preliminares destacam a importância da metodologia adotada para a estimação da relação entre as externalidades marshallianas e a coaglomeração.

Seguindo o raciocínio exposto, o índice oferece insights valiosos para compreender a dinâmica das indústrias, destacando a relevância das localidades no processo de coaglomeração. Isso ocorre devido ao cálculo considerar a proporção do emprego na indústria e na localidade em relação ao total agregado. Por exemplo, uma indústria presente em todos os estados, mas com uma grande participação do emprego, pode não apresentar padrões de coaglomeração, pois sua presença pode estar dispersa de maneira aleatória pelos estados ou as demais indústrias podem estar distribuídas de forma aleatória na localidade ou microrregião. Essa análise mais refinada ressalta a importância de considerar a distribuição geográfica para uma compreensão abrangente dos padrões de coaglomeração industrial.

Nesse cenário, pode-se ilustrar o caso da indústria de alimentos (10) em 2015, que detinha 21,9% de todo o emprego na indústria de transformação brasileira, mas exibiu um padrão de coaglomeração de apenas 17,4%. Esse padrão pode ser atribuído ao fato de que essa indústria possui uma significativa participação em todos os estados, alcançando uma expressiva fatia de 78,2% no setor de transformação em Alagoas. Além disso, ao analisar a distribuição do emprego da indústria alimentícia em relação ao total, constata-se que 63,6% estão concentrados nos estados de São Paulo, Minas Gerais, Paraná, Rio Grande do Sul e Santa Catarina.

A análise dos dados sugere que a indústria poderia apresentar um padrão de coaglomeração. No entanto, o índice considera também a participação nas localidades observadas, o que pode levar indústrias com alta presença em todos os estados a exibirem um padrão de coaglomeração que, na realidade, é aleatório. Essa perspectiva ressalta a importância de uma interpretação cuidadosa dos padrões de coaglomeração à luz da distribuição geográfica específica das indústrias.

O cenário oposto pode ser observado nas indústrias de produtos químicos (20) e de produtos farmoquímicos e farmacêuticos (21). Apesar de representarem apenas 5,2% do emprego total no setor de transformação, essas indústrias apresentaram, em 2006 e 2015, um padrão de coaglomeração acima de 43%, com uma leve variação. Ambas as indústrias estão concentradas em poucos estados, principalmente em São Paulo, com uma distribuição de trabalhadores de 48,3% e 55,2%, respectivamente, nesse estado. De fato, as indústrias localizadas predominantemente nas regiões sul-sudeste tendem a exibir padrões de coaglomeração mais elevados. Isso se deve principalmente ao fato de que, em 2015, 68% do emprego no setor de transformação estava concentrado nessas regiões. Esse contraste destaca a influência da distribuição geográfica específica das indústrias nos padrões de coaglomeração observados.

Ao examinar os dados brutos de emprego da RAIS, observa-se que a concentração significativa do emprego da indústria no estado de São Paulo aumenta a propensão a coaglomerar-se com as demais, como ilustrado no caso mencionado. Esse resultado inicial sugere indícios de que fatores marshallianos podem estar correlacionados com os padrões de coaglomeração. Contudo, outras indústrias situadas predominantemente nas regiões sul-sudeste podem também apresentar uma considerável participação em outros estados. Esse fator é evidenciado pelas indústrias de fabricação de equipamentos de informática, produtos eletrônicos e ópticos (26) e fabricação de outros equipamentos de transporte, exceto veículos automotores (30), que têm uma expressiva presença no estado do Amazonas. A primeira indústria (26) está distribuída com 77,4% do emprego na região sul-sudeste, 19,3% no Amazonas e 3,3% no restante do país. A segunda apresenta 75,4% do seu emprego na região sul-sudeste, 14,6% no estado do Amazonas e 10% no restante do país. Esses dados explicam por que essas duas indústrias apresentam padrões de coaglomeração elevados e são coaglomeradas tanto a nível estadual quanto a nível microrregional.

A relevância desses dados torna-se ainda mais clara quando se considera que essas duas indústrias, em conjunto, representam 40% do emprego no setor de transformação do estado do Amazonas. Esse impacto expressivo é principalmente atribuído à Zona Franca de Manaus, que incentiva a produção desses tipos de produtos. Nesse contexto, é possível inferir que não apenas as forças marshallianas podem estar presentes, mas também que a guerra fiscal provavelmente influencia na coaglomeração dessas duas indústrias. Essa interação complexa entre incentivos fiscais, forças marshallianas e padrões de coaglomeração destaca a necessidade de considerar múltiplos fatores na compreensão desses fenômenos industriais.

Alguns outros setores também estão concentrados na região sul-sudeste, são eles: fabricação de celulose, papel e produtos de papel (17) com 82%; impressão e reprodução de gravações (18) com 76,2%; fabricação de produtos de borracha e de material plástico (22) com 83,5%; metalurgia (24) com 86,2%; fabricação de produtos de metal, exceto máquinas e equipamentos (25) com 85,6%; fabricação de máquinas, aparelhos e materiais elétricos (27) com 85,9%; fabricação de máquinas e equipamentos (28) com 94,1%; fabricação de veículos automotores, reboques e carrocerias (29) com 91,3%; fabricação de móveis (31) com 82,1%; fabricação de produtos diversos (32) com 86,4%; manutenção, reparação e instalação de máquinas e equipamentos (33) com 78,7%; e fabricação de produtos de fumo (12) com 83,5%.

Esse último detém o maior valor absoluto do índice para o nível estadual, justamente por se localizar predominantemente no estado do Rio Grande do Sul, em conjunto com a indústria de preparação de couros e fabricação de artefatos de couro, artigos para viagem e calçados (15), que possui 40,9% e 32% do seu emprego total concentrado nesse estado. As indústrias relacionadas ao segmento têxtil e de vestuário também estão majoritariamente localizadas na região sul-sudeste, mas também têm presença no Nordeste, principalmente nos estados do Ceará e Pernambuco.

Diante das evidências apresentadas até aqui, pode-se afirmar que existem indústrias que demonstram padrões de coaglomeração, algumas mais intensamente do que outras. Intuitivamente, os fatores marshallianos parecem estar associados, de certa forma, com a coaglomeração entre alguns pares de indústrias, no Brasil. Em relação à literatura empírica, os resultados são consistentes, com os índices tendendo a zero e apenas alguns setores apresentando padrões de coaglomeração.

## 5.2 Fatores correlacionados

A partir da Tabela 5, que apresenta as estatísticas descritivas referentes às variáveis utilizadas no modelo, é possível observar que as variáveis explicativas entre pares de indústrias seguem a tendência esperada pela literatura, exibindo índices médios próximos de zero. A variável "Intellect", que mensura a média das participações de alto nível educacional, ilustra a existência de pares de indústrias com baixa participação de trabalhadores ocupados com Ensino Superior completo, assim como pares de indústrias em que há uma expressiva participação de trabalhadores com alto nível educacional.

Por outro lado, em relação à classificação tecnológica, nota-se uma maior quantidade de pares de indústrias relacionada à baixa tecnologia em comparação com as de alta tecnologia. Essas observações iniciais das variáveis explicativas fornecem insights importantes para a compreensão da dinâmica entre os pares de indústrias, alinhando-se com as expectativas da literatura e indicando potenciais correlações que serão exploradas no modelo empírico.

Ao examinar a correlação entre as variáveis, a Tabela 6 aponta para uma cor-

**Tabela 5.** Estatísticas descritivas das variáveis utilizadas nas regressões

Variável	Observações	Média	Desvio-Padrão	Mínimo	Máximo
EGK	2,760	-0.0011	0.0228	-0.0801	0.1436
Labor	2,760	-0.0653	0.0643	-0.1986	0.3653
Input	1020	0.0112	0.0219	0.0000	0.2244
Output	1020	0.0130	0.0261	0.0000	0.2394
Intellect	2,760	0.0956	0.1823	0.0001	0.7956
HighIntellect	2,760	0.1357	0.0763	0.0146	0.3970
SimiIntellect	2,760	0.0013	0.0205	-0.0827	0.0857
Alta-Alta	256	0.1098	0.3132	0	1
Alta-Baixa	256	0.4745	0.5003	0	1
Baixa-Baixa	256	0.4156	0.4938	0	1

Fonte: Elaboração dos autores.

relação moderada entre o fator de aglomeração *labor pooling* e as demais variáveis explicativas do modelo. Além disso, observa-se uma correlação significativa entre as variáveis input e output, o que intuitivamente era esperado, dada a relação entre essas variáveis na composição da matriz de insumo-produto das indústrias. Essas correlações iniciais entre as variáveis destacam potenciais interações e padrões que serão explorados mais a fundo no contexto da análise empírica, fornecendo uma base sólida para a investigação das relações entre os fatores de coaglomeração e as características específicas das indústrias.

**Tabela 6.** Correlações entre as variáveis

	EGK	Labor	Intellect	HighIntellect	SimiIntellect	Input	Output
EGK	1.0000						
Labor	0.0774	1.0000					
Intellect	0.1109	0.1768	1.0000				
HighIntellect	-0.0040	0.0709	0.0573	1.0000			
SimiIntellect	0.0058	0.0595	-0.0187	0.0628	1.0000		
Input	0.0793	0.3832	-0.0993	0.0662	-0.0429	1.0000	
Output	0.0618	0.3590	-0.0508	0.0624	-0.0485	0.8245	1.000

Fonte: Elaboração dos autores.

Nas regressões subsequentes, utilizando a metodologia proposta, foram estimados modelos com diferentes formas, incluindo e excluindo variáveis para conferir maior robustez ao modelo. Observa-se, pelas estimações apresentadas na Tabela 7, que, mesmo com a dificuldade de isolar os efeitos de forma separada, os resultados indicam evidências significativas para a literatura nacional. O modelo com efeitos fixos reforça a importância de controlar as variações ao longo dos anos para efeitos não relacionados às externalidades marshallianas. É altamente provável que fatores como a guerra fiscal e vantagens naturais também apresentem correlação com a coaglomeração, e um avanço futuro na literatura seria explorar a correlação dessas variáveis com a coaglomeração no Brasil. No entanto, a ausência dessas variáveis é controlada pela metodologia adotada, permitindo a análise do objetivo proposto de forma consistente. Essa abordagem robusta contribui para a compreensão das complexas

interações subjacentes aos padrões de coaglomeração industrial no contexto brasileiro.

Ao observar a Tabela 7, fica evidente que as regressões com efeitos fixos apresentam um ajuste consideravelmente melhor do que a regressão OLS, destacando a importância da metodologia escolhida. Os principais resultados a serem destacados estão relacionados às variáveis de Input e Labor, que demonstram alta significância em todas as regressões. A variável Output, por outro lado, não se mostra significativa, reforçando a ideia de complementaridade entre as forças marshallianas, sendo relevantes no modelo conjuntamente. Percebe-se também que as *proxies* de capital humano apresentam consistência ao atuarem conjuntamente na aglomeração, embora seja necessário um detalhamento mais aprofundado sobre os resultados dessa variável. Separadamente, a análise mais importante está relacionada ao sinal negativo encontrado para a *proxy* "Intellect", que representa a participação de trabalhadores com Ensino Superior em ambas as indústrias. No entanto, é relevante notar que a *proxy* que mensura alta proporção de trabalhadores qualificados em pelo menos uma indústria não se mostrou significativa. Esses resultados ressaltam a complexidade das interações entre as variáveis e apontam para a necessidade de uma análise mais aprofundada para uma compreensão completa dos padrões de coaglomeração observados.

O sinal negativo pode ser interpretado de três formas possíveis. Primeiramente, é possível que a variável não esteja capturando os efeitos corretamente, uma vez que a literatura da área é consolidada quanto ao sinal positivo do capital humano em relação à coaglomeração. Nesse contexto, o resultado difere dos encontrados por Moretti (2004), Gabe e Abel (2012), em que o capital humano desempenha um papel fundamental como *proxy* para a difusão de conhecimento. Assim, é provável que o viés de seleção esteja afetando os resultados encontrados para essa *proxy* de capital humano.

Em contrapartida, a segunda explicação é basicamente que indústrias que apresentam conjuntamente alto capital humano tendem a não coaglomerar, o que é difícil de aceitar, considerando os resultados apontados na literatura apresentada. Todavia, dado que há outros fatores na decisão locacional da firma, é possível que o capital humano não seja um fator relevante na decisão das indústrias se aglomerarem em conjunto. Neste trabalho, no entanto, não é possível afirmar tal sentença, limitando-se apenas a apresentação de alternativas para o entendimento dos resultados encontrados.

A terceira explicação, com probabilidade menor de ser correta, refere-se ao fato de que a *proxy* pode estar capturando dois efeitos distintos, dado que a construção parte da média. Assim, pode haver uma indústria com alta proporção de trabalhadores qualificados e outra não, apresentando uma média alta de todo modo. Esse pensamento é reafirmado pela não significância da *proxy* "HighIntellect", que mostra que não basta apenas uma indústria apresentar alto nível de capital humano dos seus trabalhadores para estar coaglomerada com outra.

**Tabela 7.** Painel de regressões multivariadas com efeitos fixos

	OLS	Efeitos Fixos (1)	Efeitos Fixos (2)	Efeitos Fixos (3)	Efeitos Fixos (4)
Labor	0.011 (0.013)	0.1582*** (0.026)	0.1611*** (0.027)	0.1586*** (0.026)	0.1555*** (0.026)
Intellect	0.224*** (0.100)	-0.1155*** (0.030)			-0.1144*** (0.030)
HighIntellect	-0.005 (0.008)		0.0001 (0.001)		-0.0001 (0.001)
SimilIntellect	0.012 (0.034)			0.0152*** (0.004)	0.013*** (0.004)
Input	0.1036* (0.061)	0.1765** (0.074)	0.1753** (0.074)	0.1846** (0.074)	0.1848** (0.073)
Output	-0.0226 (0.042)	-0.003 (0.025)	0.000 (0.026)	-0.002 (0.025)	-0.0074 (0.025)
R <sup>2</sup>	0.0213	0.0704	0.0525	0.0661	0.0813
Observações	1020	1020	1020	1020	1020

Fonte: Elaboração dos autores.

Diante dessas limitações e com o intuito de proporcionar maior robustez ao fator "Intellectual Spillover" foi desenvolvida uma segunda abordagem em que o foco passa a ser verificar a importância da similaridade tecnológica entre os pares de indústrias. Nas próximas tabelas, são apresentados os resultados com a visão tecnológica como alternativa para o capital humano.

Assim, ao utilizar a Equação 10 e acrescentar *dummies* de intensidade tecnológica na regressão, busca-se conferir robustez ao resultado encontrado nas estimações anteriores, nas quais se estabeleceu a relação entre coaglomeração e similaridade tecnológica. Para tanto, foi necessário estimar regressões OLS no lugar de um painel, uma vez que as *dummies* não variam com o tempo. Isso, no entanto, não representa um problema, dado que o objetivo desta segunda abordagem é puramente analisar o comportamento da coaglomeração através da combinação pela classificação tecnológica entre os pares de indústrias analisados.

Analisando a Tabela 8, observa-se que as variáveis *dummies* de combinação tecnológica similar têm uma relação positiva com a coaglomeração, quando comparadas com as combinações de Alta-Baixa tecnologia. Em outras palavras, indústrias com o mesmo perfil de intensidade tecnológica tendem a se coaglomerar em maior proporção do que indústrias com diferentes perfis tecnológicos. Os pares de indústrias de alta tecnologia apresentam um coeficiente maior do que os pares de indústrias de baixa tecnologia, indicando que indústrias mais tecnológicas têm uma propensão maior a se coaglomerar. Vale ressaltar que mesmo após a incorporação das variáveis de controle no modelo, as variáveis apresentaram coeficientes estatisticamente significativos, com nível de confiança inferior a 5%. Esse resultado robusto reforça a importância da similaridade tecnológica como um fator determinante na coaglomeração das indústrias analisadas.

**Tabela 8.** Painel de regressões OLS

	Estimação OLS (1)	Estimação OLS (2)	Estimação OLS (3)	Estimação OLS (4)
Labor		0.066*** (0.008)	0.040*** (0.008)	0.062*** (0.018)
Intellect		-0.011*** (0.001)	-0.006*** (0.001)	-0.104*** (0.029)
HighIntellect			0.0000 (0.003)	-0.0001 (0.001)
SimiIntellect			0.028*** (0.017)	0.014*** (0.004)
Input				0.069 (0.051)
Output				-0.001 (0.024)
Alta-Alta	0.032*** (0.004)	0.029*** (0.004)	0.023*** (0.004)	0.025*** (0.004)
Baixa-Baixa	0.009*** (0.002)	0.008*** (0.002)	0.008*** (0.002)	0.005*** (0.002)
R <sup>2</sup>	0.0001	0.0332	0.0587	0.0677
Observações	2760	2760	2760	1020

Fonte: Elaboração dos autores.

## 6. Conclusões

O estudo teve como propósito examinar os padrões de coaglomeração nas indústrias de transformação no Brasil, identificando os fatores produtivos associados a esses padrões. Por meio da aplicação de modelos de regressão com dados em painel, buscou-se investigar a associação dos fatores marshallianos com o índice de coaglomeração entre os pares de indústrias.

Os resultados da análise de regressão em painel corroboram os achados anteriores de Ellison et al. (2010) e Resende (2012), fornecendo uma visão mais detalhada da situação brasileira. Além disso, evidenciaram que indústrias que compartilham a mesma tecnologia apresentam padrões mais pronunciados de coaglomeração em comparação com aquelas que não compartilham. Notavelmente, os pares de indústrias de alta tecnologia demonstraram uma maior propensão à coaglomeração, reforçando a relação entre aglomeração e produtividade, conforme proposto por Henderson (2003); Rosenthal e Strange (2004); Lööf e Andersson (2010); Combes et al. (2012).

Quanto às *proxies* para a variável de *intellectual spillovers*, os resultados demandam análise cautelosa. Indicam que o alto capital humano não é um fator determinante para a coaglomeração entre as indústrias, enquanto a média de trabalhadores qualificados entre duas indústrias apresenta uma relação negativa com a coaglomeração, agindo de forma contrária. Em contrapartida, a *proxy* de similaridade entre os níveis educacionais compartilhados pelas indústrias se mostrou relevante na coaglomeração, apoiando descobertas anteriores de Kolko (2010) e Howard et al. (2016) indicando que indústrias de mesma intensidade tecnológica são mais propensas à coaglomeração.

Ressalta-se, porém, que o artigo apresenta limitações, como a possível causalidade reversa e a não inclusão de vantagens naturais ou incentivos fiscais como variáveis explicativas. No entanto, essas limitações podem ser exploradas em trabalhos futuros. Em última análise, o estudo contribui para a compreensão dos fatores associados à coaglomeração no Brasil, oferecendo insights para formuladores de políticas públicas que buscam identificar indústrias propensas à coaglomeração e facilitar a disseminação de externalidades marshallianas, além de promover a análise de clusters industriais e atrair indústrias tecnologicamente similares para determinadas regiões, com base nos resultados obtidos.

## Referências

- Andersson, M., Klaesson, J., e Larsson, J. P. (2014). The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science*, 93(4):727–747.
- Andersson, M. e Löf, H. (2011). Agglomeration and productivity: evidence from firm-level data. *The annals of regional science*, 46(3):601–620.
- Antonietti, R. e Cainelli, G. (2011). The role of spatial agglomeration in a structural model of innovation, productivity and export: a firm-level analysis. *The Annals of Regional Science*, 46(3):577–600.
- Autant-Bernard, C. (2001). The geography of knowledge spillovers and technological proximity. *Economics of Innovation and New Technology*, 10(4):237–254.
- Baldwin, R. E. e Martin, P. (2004). Agglomeration and regional growth. In: *Handbook of regional and urban economics*, volume 4, Página 2671–2711. Elsevier.
- Beaudry, C. e Schiffauerova, A. (2009). Who's right, marshall or jacobs? the localization versus urbanization debate. *Research policy*, 38(2):318–337.
- Becchetti, L. e Rossi, S. P. (2000). The positive effect of industrial district on the export performance of italian firms. *Review of industrial organization*, 16(1):53–68.
- Brülhart, M. e Sbergami, F. (2009). Agglomeration and growth: Cross-country evidence. *Journal of Urban Economics*, 65(1):48–63.
- Cavalvante, L. R. (2014). Classificações tecnológicas: uma sistematização.
- Combes, P.-P. e Duranton, G. (2006). Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1):1–28.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., e Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

- Dahl, M. S. e Klepper, S. (2015). Whom do new firms hire? *Industrial and Corporate Change*, 24(4):819–836.
- Diamond, C. A. e Simon, C. J. (1990). Industrial specialization and the returns to labor. *Journal of Labor Economics*, 8(2):175–201.
- Diodato, D., Neffke, F., e O'Clery, N. (2018). Why do industries coagglomerate? how marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, 106:1–26.
- Duranton, G. e Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106.
- Ellison, G. e Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of political economy*, 105(5):889–927.
- Ellison, G. e Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review*, 89(2):311–316.
- Ellison, G., Glaeser, E. L., e Kerr, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195–1213.
- Feldman, M. P. (1999). The new economics of innovation, spillovers and agglomeration: A review of empirical studies. *Economics of innovation and new technology*, 8(1-2):5–25.
- Fingleton, B. e Longhi, S. (2013). The effects of agglomeration on wages: Evidence from the micro-level. *Journal of Regional Science*, 53(3):443–463.
- Fujita, M., Krugman, P. R., e Venables, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.
- Gabe, T. M. e Abel, J. R. (2012). Specialized knowledge and the geographic concentration of occupations. *Journal of Economic Geography*, 12(2):435–453.
- Glaeser, E. L. e Kahn, M. E. (2001). Decentralized employment and the transformation of the american city. Texto para discussão, National Bureau of Economic Research.
- Greenaway, D. e Kneller, R. (2008). Exporting, productivity and agglomeration. *European economic review*, 52(5):919–939.
- Greenstone, M., Hornbeck, R., e Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.
- Guilhoto, J. et al. (2010). Estimação da matriz insumo-produto utilizando dados preliminares das contas nacionais: Aplicação e análise de indicadores econômicos para

- o brasil em 2005 (using data from the system of national accounts to estimate input-output matrices: An application using brazilian data for 2005). *Available at SSRN 1836495*.
- Helsley, R. W. e Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*, 20(2):189–212.
- Henderson, J. V. (2007). Understanding knowledge spillovers. *Regional Science and Urban Economics*, 37(4):497–508.
- Henderson, V. (2003). The urbanization process and economic growth: The so-what question. *Journal of Economic growth*, 8(1):47–71.
- Howard, E., Newman, C., e Tarp, F. (2016). Measuring industry coagglomeration and identifying the driving forces. *Journal of Economic Geography*, 16(5):1055–1078.
- Jacobs, J. (1969). *The economy of cities*. Texto para discussão, Vintage Books: A Division of Random House.
- Jacobs, W., Koster, H. R., e van Oort, F. (2014). Co-agglomeration of knowledge-intensive business services and multinational enterprises. *Journal of Economic Geography*, 14(2):443–475.
- Kolko, J. (2010). Urbanization, agglomeration, and coagglomeration of service industries. In: *Agglomeration economics*, Página 151–180. University of Chicago Press.
- Kominers, S. D. (2008). Measuring agglomeration. In: *Harvard Urban and Social Economics Seminar (Economics 2800b)* [http://www.scottkom.com/articles/measure\\_agglomeration.pdf](http://www.scottkom.com/articles/measure_agglomeration.pdf). Citeseer.
- Krugman, P. R. (1991). *Geography and trade*. MIT press.
- Lööf, H. e Andersson, M. (2010). Imports, productivity and origin markets: The role of knowledge-intensive economies. *World Economy*, 33(3):458–481.
- Maciente, A. (2013). *The determinants of agglomeration in Brazil: input-output, labor and knowledge externalities*. Tese de Doutorado, University of Illinois at Urbana-Champaign.
- Marshall, A. (1890). *Principles of economics* macmillan. London (8th ed. Published in 1920).
- Martin, P. e Ottaviano, G. I. (2001). Growth and agglomeration. *International economic review*, 42(4):947–968.
- Maurel, F. e Sédillot, B. (1999). A measure of the geographic concentration in french manufacturing industries. *Regional Science and Urban Economics*, 29(5):575–604.
- Moretti, E. (2004). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

- Mukim, M. (2015). Coagglomeration of formal and informal industry: evidence from india. *Journal of Economic Geography*, 15(2):329–351.
- Nakamura, R. e Paul, C. J. M. (2019). Measuring agglomeration. In: *Handbook of regional growth and development theories*. Edward Elgar Publishing.
- Porter, M. E. (1990). The competitive advantage of nations. *Harvard business review*, 68(2):73–93.
- Puga, D. e Venables, A. J. (1999). Agglomeration and economic development: Import substitution vs. trade liberalisation. *The Economic Journal*, 109(455):292–311.
- Resende, M. (2012). Co-aglomeração industrial no estado do rio de janeiro: Um estudo empírico. Texto para discussão, Séries working paper BNDES/ANPEC.
- Rosenthal, S. S. e Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In: *Handbook of regional and urban economics*, volume 4, Página 2119–2171. Elsevier.
- Rosenthal, S. S. e Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.
- Saxenian, A. (1996). *Regional advantage*. Harvard University Press.
- Steijn, M. P., Koster, H. R., e Van Oort, F. G. (2022). The dynamics of industry agglomeration: Evidence from 44 years of coagglomeration patterns. *Journal of Urban Economics*, 130:103456.
- Tveteras, R. e Battese, G. E. (2006). Agglomeration externalities, productivity, and technical inefficiency. *Journal of regional science*, 46(4):605–625.

 Este artigo está licenciado com uma CC BY 4.0 license.