

MISMATCH ESPACIAL NO MERCADO DE TRABALHO DA RMSP: A INSERÇÃO DA ACESSIBILIDADE AO EMPREGO PARA PARTICIPAÇÃO, DESEMPREGO, INFORMALIDADE E OVEREDUCATION

Leandro Batista Duarte

Professor da Universidade Estadual de Feira de Santana (UEFS) e Doutorando do PIMES/UFPE

Raul da Mota Silveira Neto

Professor da Universidade Federal de Pernambuco (UFPE)

Resumo

A localização das famílias na Região Metropolitana de São Paulo (RMSP) tanto pode refletir escolhas em função da renda, como dificuldades de acesso aos empregos (*mismatch* espacial). Com base nessa perspectiva, este estudo fornece uma análise da relação entre *mismatch* espacial e os resultados do mercado de trabalho em uma grande área metropolitana do Brasil, a RMSP. O desafio da pesquisa foi calcular o efeito da acessibilidade ao emprego na probabilidade de estar empregado, de estar desempregado, de pertencer ao setor informal e de chance de ser *overeducated*. O artigo utilizou dois bancos de dados: o Censo de População de 2010 e a Relação de Informações Anuais de Responsabilidade Social (RAIS). As equações estimadas foram com base nos modelos de probabilidade linear e *Probit*, com e sem variáveis instrumentais. Foram encontradas evidências importantes com disparidades espaciais em termos de acesso a oportunidades de emprego nas áreas de ponderação da RMSP, bem como efeitos de gênero heterogêneos na conectividade nos resultados trabalhistas. Os resultados mostraram que, em geral, o melhor acesso físico aos locais de emprego aumenta a participação no mercado, diminui o desemprego, a informalidade e a condição de *overeducated*.

Palavras-chave: *Mismatch* Espacial; Acessibilidade; RMSP.

Classificação JEL: R10; R30.

Área: Questões urbanas e metrópoles.

SPACE MISMATCH IN THE RMSP LABOR MARKET: THE INSERT OF EMPLOYMENT ACCESSIBILITY FOR PARTICIPATION, UNEMPLOYMENT, INFORMALITY AND OVEREDUCATION

Abstract

The location of families in the Metropolitan Region of São Paulo (RMSP) can reflect both choices based on income, as well as difficulties in accessing jobs (spatial mismatch). Based on this perspective, this study provides an analysis of the relationship between spatial mismatch and labor market results in a large metropolitan area in Brazil, the RMSP. The research challenge was to calculate the effect of accessibility to employment on the probability of being employed, being unemployed, belonging to the informal sector and the chance of being overeducated. The article used two databases: the 2010 Population Census and the Annual Social Responsibility Information List (RAIS). The estimated equations were based on the linear probability and Probit models, with and without instrumental variables. Important evidence was found with spatial disparities in terms of access to employment opportunities in the areas of consideration of the RMSP, as well as heterogeneous gender effects on connectivity in labor outcomes. The results showed that, in general, better physical access to places of employment increases market share, decreases unemployment, informality and being overeducated.

Keywords: Space Mismatch; Accessibility; RMSP.

JEL: R10; R30.

Area: Urban issues and metropolises.

1 Introdução

Com uma população de 20 milhões de habitantes em 2010, a Região Metropolitana de São Paulo (RMSP) é o centro urbano mais importante do Brasil contribuindo com aproximadamente 20% para o PIB nacional e concentrando 10% da população brasileira. Não muito diferente de algumas megacidades como Tóquio, Nova York, Seul e Cidade do México, a RMSP é conhecida por problemas urbanos crônicos associados a níveis muito altos de aglomeração, incluindo congestionamento de tráfego e muitas externalidades associadas (WHEELER, 2003). No Brasil, há a tendência da população de renda mais elevada morar nas regiões centrais, enquanto os grupos com renda mais baixa ocupam o entorno das aglomerações urbanas, na periferia. Este cenário não é diferente para a RMSP que exhibe uma divisão núcleo-periferia particularmente marcada (RAMOS, 2016). Moreno-Monroy e Ramos (2015) argumentam que a formação da RMSP se deu com um rápido crescimento populacional, tendo a população pobre e menos instruída ocupando um vasto cinturão periférico suburbano ao redor do centro formado por um processo de expansão centrífuga não planejada. Dessa forma, Nadalin e Iglioni (2010) argumentam que isso explica, pelo menos parcialmente, o padrão de localização das famílias de acordo com a renda, com as mais ricas, menos distantes dos centros de emprego.

A localização das famílias mais pobres na periferia da RMSP tanto pode refletir escolhas em função da renda, como dificuldades de acesso aos empregos (*mismatch* espacial). Segundo a hipótese do *mismatch*, trabalhadores localizados em regiões socioeconomicamente vulneráveis, enfrentam maiores dificuldades para encontrar empregos nos centros de negócios das cidades (KAIN, 1968). O *mismatch* espacial é a desconexão espacial entre trabalhadores e oportunidades de trabalho. Dessa forma, o *mismatch* espacial identifica vários mecanismos¹ através dos quais o acesso físico restrito aos empregos pode causar resultados insatisfatórios no mercado de trabalho.

A despeito do forte desenvolvimento da Economia Urbana e largo conjunto de trabalhos aplicados, é de interesse de diversos estudos buscarem entender as forças econômicas que são responsáveis por resultados no mercado de trabalho urbano. Para o Brasil, Motte *et al.*, (2016) testaram a influência de variáveis socioeconômicas e espaciais na distância e duração das viagens de deslocamento dos habitantes que trabalham nos setores formais e informais da região metropolitana do Rio de Janeiro (RMRJ). Especificamente para a região metropolitana de São Paulo (RMSP), Moreno-Monroy e Ramos (2015) estimam o impacto das expansões do sistema de transporte público sobre as taxas de informalidade local entre 2000 e 2010. Barufi e Haddad (2016) investigam a relação entre *mismatch* espacial, salários e desemprego para as grandes áreas urbanas do Brasil, que inclui a RMSP. Boisjoly *et al.*, (2017) exploram a relação entre a informalidade do trabalho e a acessibilidade ao emprego nos transportes públicos. Por fim, Haddad e Barufi (2017) exploram a relação entre tempo de deslocamento, acessibilidade para empregos e preços urbanos no contexto dos mercados urbanos.

Além da análise empírica acima, a hipótese do *mismatch* espacial permite estudar a relação com a literatura de *Overeducation*. Esta literatura também conhecida como *mismatch* educacional é caracterizada a partir da existência de indivíduos com escolaridade diferente daquela requerida pelo seu posto de trabalho, nesse sentido os trabalhadores podem ser classificados como *Overeducated* (sobre-educado) quando o trabalhador possuir o nível educacional superior ao que é requerido em sua ocupação. No Brasil, essa literatura tem-se desenvolvido mais recentemente e apresenta algumas contribuições importantes para o debate,

¹ Ver Kain (1968); Gobillon *et al.*, (2011); Gobillon & Selod (2014) para mais detalhes.

sobretudo em relação à discussão das evidências empíricas no tema (SANTOS, 2002; MACHADO, OLIVEIRA; DIAZ e MACHADO, 2008).

A aplicação empírica desta pesquisa está ligada a um grande número da literatura que analisa o efeito da acessibilidade ao emprego nos resultados do mercado de trabalho. Para o Brasil, são poucas evidências disponíveis sobre esta relação. Com base nesta perspectiva, este estudo fornece uma análise da relação entre acessibilidade e diferentes variáveis dependentes no mercado de trabalho em uma grande área metropolitana do Brasil, a RMSP. O desafio da pesquisa foi calcular o efeito da acessibilidade ao emprego na probabilidade de estar empregado, de estar desempregado, de pertencer ao setor informal e de chance de ser *overeducated*.

Apesar do maior interesse recente sobre o tema, a maior parte dos estudos empíricos ainda não chegou a um consenso sobre o papel dos fatores espaciais na explicação dos resultados individuais no mercado de trabalho. Como este estudo utiliza como estrutura conceitual um modelo espacial de equilíbrio geral de uso da terra com locais de trabalho endógenos em uma cidade circular (Fujita e Ogawa, 1982; Lucas e Rossi-Hansberg, 2002) e propriedades de eficiência da alocação de terras nessa mesma estrutura de cidade (Rossi-Hansberg, 2004) é necessário estar ciente do potencial viés de simultaneidade nas estimações empíricas para resultados no mercado de trabalho com base em medidas de acessibilidade. É possível lidar com a questão da simultaneidade incluindo instrumentos históricos ou geográficos que influenciaram a localização da infraestrutura de transporte dentro de uma cidade sem determinar diretamente a localização de trabalhadores e empresas. Esta vertente da literatura propõe abordar essas questões com o uso de variáveis instrumentais. Para o presente trabalho obteve-se por utilizar essa estratégia baseando-se na inclusão de uma variável geográfica (acesso à margem fluvial) como instrumento para acessibilidade a empregos.

O ensaio está estruturado da seguinte forma. Além desta introdução, a seção 2 descreve a área de estudo, banco de dados e a estratégia econométrica, enquanto na seção 3 serão analisados os resultados. Considerações finais seguem na seção 4.

2. Área de estudo e dados

A área de estudo é a Região Metropolitana de São Paulo (RMSP), a maior metrópole brasileira com cerca de 19,5 milhões de habitantes, onde cerca de, 30% da força de trabalho foi empregada informalmente em 2010. A unidade de análise é o indivíduo dentro da Área Espacial de Ponderação² (AEP), que é definida como uma área composta por um conjunto de zonas mutuamente exclusiva, projetado para dar a robustez estatística necessária à estratégia de amostragem do Censo de População. De acordo com Censo Demográfico de 2010, a RMSP é particionada em 633 AEPs, com uma população média de 31.096 habitantes e uma área média de 12,55 km².

Os microdados para as características individuais são obtidos no Censo de População de 2010 (Instituto Brasileiro de Geografia e Estatística - IBGE, 2010), juntamente com a Classificação Brasileira das Ocupações domiciliar e as descrições da Classificação Brasileira das Ocupações de 2002. Para o presente trabalho foi utilizado áreas de ponderação como definição de vizinhança, uma vez que, microdados ao nível individual também estão disponíveis para esta amostra. Também utilizou como fontes de recursos a Relação de Informações Anuais de Responsabilidade Social (RAIS) para o ano de 2010. Este é um dado administrativo mantido pelo Ministério do Trabalho do Brasil. A RAIS consiste em um painel

² Define-se área de ponderação como sendo uma unidade geográfica, formada por um agrupamento de setores censitários, para a aplicação dos procedimentos de calibração das estimativas com as informações conhecidas para a população como um todo (IBGE, 2010).

de alta qualidade com informações sobre as características dos contratos entre empresas e funcionários, para todos os indivíduos formalmente empregados, desagregados no nível do município. Com isso, para a pesquisa foi utilizado os microdados georreferenciados da RAIS para identificação dos centros de empregos (subcentros), que contêm informações sobre o endereço de cada estabelecimento registrado. O conjunto de dados georreferenciado foi obtido no Instituto de Pesquisa Econômica Aplicada (IPEA).

2.1 Especificações Econométricas

A estratégia empírica desenvolvida baseia-se na estimativa da relação entre diferentes medidas de acessibilidade e resultados do mercado de trabalho (probabilidade de participação no mercado, de estar desempregado, de estar na informalidade e na condição de *overeducated*). As estimativas são realizadas para indivíduos que residem em uma área metropolitana específica, a fim de capturar o efeito de cada variável em termos relativos dentro de uma estrutura urbana específica. Portanto, a equação a ser estimada é com base em um modelo de probabilidade, da seguinte maneira:

$$Prob[y_{ij} = 1] = F[\beta_1 + \beta_2 AC_j + \beta_3 X_{ij} + \beta_4 H_{ij} + \beta_5 K_j] \quad (1)$$

Nesta especificação, y_i refere-se ao *status* de emprego (é igual a 1 quando uma pessoa está na condição de interesse, PEA, desempregada, na informalidade ou na situação de *overeducated*) e F é uma função de probabilidade. Aqui, AC_j é o índice de acessibilidade por área de ponderação j , X_{ij} é um conjunto de características observadas para o indivíduo (sexo, idade, idade ao quadrado, cor ou raça, nível de escolaridade, se o indivíduo é casado, e se a pessoa é ou não a chefe da família), H_j é um vetor de características domiciliares; K_j é um vetor de características de vizinhança (composição social); e β são vetores de parâmetros que serão estimados.

Uma regressão linear simples que representa um Modelo de Probabilidade Linear estimado por OLS pode não ser apropriada, podendo ocorrer que as probabilidades condicionais calculadas não pertençam ao intervalo fechado 0 e 1. Por outro lado, a menor exigência com respeito à distribuição dos resíduos representa uma vantagem em relação a modelos mais tradicionais com variáveis dependentes binárias. Portanto, o estudo também considerou uma especificação projetada para controlar as exigências de modelos com variáveis dependentes binárias. Dessa forma, a probabilidade de se observar $y_i = 1$ é dada pelo modelo Probit (função de distribuição cumulativa da distribuição normal padrão), ou seja,

$$Prob[y_i = 1] = \int_{-\infty}^{\beta'x} \Phi(t) d_i = \Phi(\beta'x) \quad (2)$$

onde a função $\Phi(\cdot)$ é a notação usual para a distribuição normal padrão cumulativa e β é o vetor de parâmetros das variáveis explicativas consideradas. A equação geral do Modelo em função das variáveis explicativas adotadas é representada por:

$$\Phi(\beta'x) = \Phi(\beta_1 + \beta_2 AC_j + \beta_3 X_{ij} + \beta_4 H_{ij} + \beta_5 K_j) \quad (3)$$

Posteriormente, serão estimados os efeitos marginais, que serão considerados nas análises, a seguir. Os efeitos marginais representam, por exemplo, o quanto uma mudança na variável de acessibilidade pode afetar a probabilidade de um indivíduo participar no mercado de trabalho. Os efeitos marginais (a variação da probabilidade de o evento ocorrer quando

uma variável independente é modificada) são calculados por meio dos coeficientes estimados β 's.

$$\frac{\partial E[y|x]}{\partial x} = \Phi(x'\beta)\beta \quad (4)$$

É claro que modelos como a equação (3) tornam possível controlar algumas características individuais e domiciliares que podem influenciar tanto a escolha da vizinhança quanto os resultados individuais no mercado de trabalho. No entanto, é provável que algumas características individuais e familiares que não são observadas pelo pesquisador (e, portanto, não incluídas em X_i ou H_i), influenciem tanto o resultado do interesse quanto a escolha da vizinhança.

Assim, a acessibilidade aos empregos, que é a variável explicativa de interesse na análise, depende de alguns dos fatores que afetam as variáveis dependentes e também podem ser indiretamente afetados. Neste caso, o viés de simultaneidade surge à medida que os residentes e os locais das empresas são determinados ao mesmo tempo (HADDAD & BARUFI, 2017). Outras origens de endogeneidade podem surgir do comportamento dos trabalhadores. Por um lado, os trabalhadores são mais propensos a percorrer longas distâncias em troca de salários mais altos, o que influenciaria na maior participação da força de trabalho, menor a chance de estar desempregado, na informalidade ou na condição de *overeducated*. Por outro lado, indivíduos com rendimentos mais elevados podem se localizar mais longe de seus empregos para ter acesso a amenidades geralmente deslocadas nos arredores da cidade (MELO & GRAHAM, 2009).

Tanto a presença de fatores omitidos que afetam simultaneamente as variáveis dependentes e o indicador de acessibilidade, como a possibilidade de causalidade reversa fazem as estimativas de OLS provavelmente tendenciosas. Essas preocupações podem ser potencialmente eliminadas com uma estratégia de variáveis instrumentais (IV). Essa estratégia pode ser especificada como uma função linear de uma covariável exógena (instrumento) z e um componente aleatório μ_i .

$$AC_j = \beta_0 + \beta_1 z_j + \beta_2 X_{ij} + \beta_3 H_{ij} + \beta_4 K_j + \mu_j \quad (5)$$

Para alcançar a consistência, o instrumento z deve atender a duas condições: estar associada ao indicador de acessibilidade (ou seja, ter poder de explicação em uma regressão do primeiro estágio com variável dependente AC_j) e satisfazer a restrição de afetar o resultado exclusivamente através do indicador no segundo estágio. Na próxima seção será apresentada a medida de acessibilidade e, em seguida, o instrumento utilizado no primeiro estágio das regressões.

Uma questão fundamental nos estudos de *mismatch* espacial é como medir a acessibilidade residencial às oportunidades de emprego. Esta variável leva em conta a distribuição espacial dos empregos e a distância ou custo de acesso para alcançá-los. As medidas de acessibilidade são geradas ao nível de AEP, com base em uma medida de oportunidade cumulativa. Este tipo de medida conta o número de trabalhos que podem ser alcançados a partir de um AEP e é calculado da seguinte forma:

$$AC_j = \sum_{k=1}^n \frac{E_k}{d_{jk}^\alpha(t)} \quad (6)$$

Na equação (6), AC_j é a acessibilidade da região j (centróide da AEP), E_k corresponde à quantidade de empregos existentes em qualquer região k (centros de empregos), enquanto d_{jk} é o impedimento (por exemplo, distância) existente para viajar entre j e k .

Os valores de E_j foram obtidos da base de dados do Núcleo de Estudos Regionais e Urbanos (NERU) da UFPE, que utiliza informações da RAIS 2010 disponibilizadas pelo IPEA para o georeferenciamento das firmas e empregos. Logo, E_k corresponde aos empregos formais dos principais centros de emprego da RMSP, contabilizando 94 subcentros e 1 centro principal (CBD). Para identificar os subcentros presentes na RMSP, o trabalho faz uso de uma base de dados única que inclui o georeferenciamento de quase todos os empregos formais da região, algo apenas possível a partir de um minucioso e intensivo trabalho feito com base nas informações da RAIS (Figura 1).

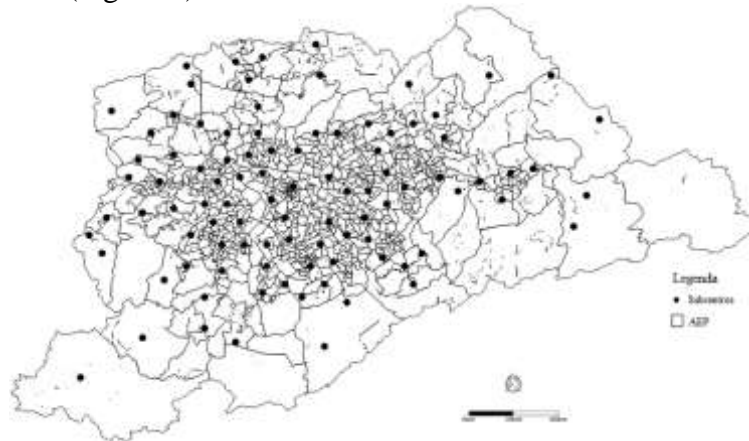


Figura 1 – Subcentros das áreas de ponderação da RMSP.
Fonte: Elaboração própria com base nos dados da RAIS 2010.

Em face da possibilidade de comparação com trabalhos recentes para cidades latino-americanas e da ausência ainda de consenso quanto à maior propriedade dos diferentes métodos, nesta pesquisa foi utilizado a metodologia³ desenvolvida por McMillen (2001) para identificar os centros e subcentros de emprego. Como argumentado pelo autor, esse procedimento permite estudar diferentes centros urbanos sem necessariamente ter um conhecimento prévio completo da região de estudo em análise.

2.2 Descrição das variáveis de indivíduos

O banco de dados construído permitiu contemplar diferentes aspectos sobre a RMSP. As variáveis que constituem esse banco de dados e suas descrições são apresentadas no Quadro 1, a seguir. Entre as variáveis contidas no quadro, a maioria foi criada em nível de indivíduos por área de ponderação com base no Censo Demográfico 2010. Primeiramente, têm-se as variáveis dependentes das estimativas (PEA, desemprego, informal e *overeducated*) referentes à condição do indivíduo no mercado de trabalho que, como se viu, por determinar um maior ou menor custo de associado ao deslocamento para o trabalho, pode afetar a localização das famílias. São também consideradas nas estimativas um conjunto de características pessoais (sexo, idade, raça e escolaridade), além do estado civil, que podem refletir diferenças nas preferências sobre a localização na cidade entre os indivíduos.

Mais especificamente, as variáveis individuais absorvem as diferentes preferências com relação ao espaço urbano de acordo com certas características. Por exemplo, quanto à *dummy* para gênero, regularmente, observa-se nos trabalhos empíricos que as mulheres

³ Para detalhes sobre o método ver McMillen (2001).

tendem a se localizar mais próximos aos locais de trabalho, o que é comumente atribuído à noção de maior responsabilidade para com as tarefas do lar atribuído às mulheres (CRAINE, 2007). De forma similar, a condição de casado pode atribuir um caráter mais permanente à localização dos domicílios e, assim, menor vinculação imediata ao local de trabalho. As variáveis quanto à idade e escolaridade traduzem possíveis influências da disposição à locomoção e preferências quanto à vizinhança por parte dos indivíduos. O tamanho da família captam a influência do poder aquisitivo da família e das preferências quanto à espaço da residência.

Quadro 1 – Descrição das variáveis dos modelos econométricos.

Amostra (pessoas da RMSP)	
VARIÁVEIS	DESCRIÇÃO
Variáveis dependentes	
PEA	Se a pessoa está na PEA (15 a 64 anos)
Desemprego	Se a pessoa está na PEA e desempregada
Informal	Se a pessoa está ocupada e na informalidade (sem carteira de trabalho)
Overeducated	Se a pessoa tem educação maior que a exigida no emprego
Variáveis Explicativas	
LnACi	Log do Índice de Acessibilidade (distância euclidiana)
LnACi_IV	Log do Índice de Acessibilidade (distância através dos rios)
Masculino	Variável <i>dummy</i> se for do sexo masculino
Idade	Em anos
Idade ²	<i>Proxy</i> para experiência
Educação	
i) Baixo	i) Sem ensino fundamental completo,
ii) Intermediário	ii) Com ensino fundamental completo, mas sem médio completo,
iii) Alto	iii) Com ensino médio completo, mas sem superior,
iv) Bem elevado	iv) Com superior completo ou mais
Casado	Variável <i>dummy</i> se for casado
Chefe	Variável <i>dummy</i> se for chefe da família
Raça	Branco Pardo Negro Amarelo
Tamanho da família	Numero de pessoas por domicilio
Tipo de Vizinhança	Muito Carente Carente Média Desenvolvida Muito Desenvolvida
Acesso à infraestrutura	Coleta de lixo Saneamento básico

Fonte: Censo/RAIS, 2010.

2.3 Definição das Características da vizinhança

Em relação às variáveis que caracterizam a composição social da vizinhança, foi utilizada a técnica de análise fatorial para produzir indicadores compostos. A vantagem de usar a análise fatorial para produzir indicadores compostos das características da vizinhança em vez de variáveis individuais é que esses procedimentos removem o impacto potencial da colinearidade ao usar esses indicadores como variáveis preditoras nas análises de regressão, já que muitos indicadores de composição de vizinhança são altamente inter-relacionados (O'REGAN & QUIGLEY, 1996). Para contornar esse problema, foi utilizados métodos de

análise fatorial padrão (Durjandin *et al.*, 2008) para resumir essas múltiplas características em diferentes tipos de vizinhança nas AEP da RMSP..

O método de análise fatorial foi usado para identificar áreas socialmente homogêneas dentro da RMSP, que serão subsequentemente usadas nas análises de regressão. Primeiro foi executado uma análise de componentes principais que define um número limitado de fatores não correlacionados resumindo um conjunto de informações da vizinhança como percentual de mulheres chefes, nível educacional, percentual de pessoas desocupadas, percentual de domicílios sem acesso à rede de saneamento e coleta de lixo (Tabela 1).

Tabela 1 -Variáveis usadas na análise de componentes principais e seus fatores.

	Fator 1	Fator 2
Autovalores	4,410	1,313
Porcentagem de variação explicada	55,14	16,42
<i>Variáveis</i>		
% de Mulheres chefes	0,362	-0,226
% de pessoas com baixo nível educacional	-0,456	0,062
% de pessoas com ensino intermediário	0,229	0,581
% de pessoas com alto nível educacional	0,428	-0,277
% de pessoas desocupadas	-0,339	0,274
% de domicílios sem acesso à rede de saneamento	0,291	0,389
% de pessoas sem acesso a coleta de lixo	0,136	0,550
% de pessoas com menos de 15 anos	-0,453	-0,001

Notas: Apenas fatores com valores próprios superiores ou iguais a 1 foram selecionados.

Especificamente, para as variáveis que representam o nível educacional, foi considerado “baixo” os indivíduos sem ensino fundamental completo e com ensino fundamental completo, mas sem médio completo; “intermediário” indivíduos com ensino médio completo, mas sem superior; por fim, “alto” indivíduos com superior completo ou mais. Em seguida, as áreas são agrupadas de acordo com suas coordenadas nos eixos fatoriais, usando uma classificação hierárquica ascendente (método Ward⁴, que minimiza a variância intragrupo). Como no trabalho de Durjandin *et al.* (2008), foram obtidos cinco tipos de vizinhança (Tabela 2). Este foi o número ideal de *clusters* de acordo com vários critérios, incluindo o critério de agrupamento cúbico.

O primeiro tipo de vizinhança compreende áreas muito carentes com percentual alto da população com baixo nível educacional e sem acesso a saneamento básico. Estas áreas também possuem os níveis de desemprego mais alto de toda a aglomeração. Eles estão socialmente próximos a um grupo de áreas carentes que apresentam características semelhantes, mas com uma proporção menor de famílias sem acesso à saneamento e uma situação menos grave em termos de educação e desocupação. O terceiro grupo têm um nível socioeconômico mais baixo, em particular uma super-representação de trabalhadores e indivíduos com menor escolaridade. No entanto, a taxa de desocupação está próxima da média da região. Os dois grupos restantes são caracterizados em média por níveis mais altos de educação e percentuais baixos de domicílios sem acesso a saneamento básico e com pessoas com menos de 15 anos.

⁴ Segundo Hair *et al.*, (2005), o método de Ward consiste em um procedimento de agrupamento hierárquico no qual a medida de similaridade usada para juntar agrupamentos é calculada como a soma de quadrados entre os dois agrupamentos feita sobre todas as variáveis. Esse método tende a resultar em agrupamentos de tamanhos aproximadamente iguais devido a sua minimização de variação interna. Em cada estágio, combinam-se os dois agrupamentos que apresentarem menor aumento na soma global de quadrados dentro dos agrupamentos.

Seja $x_{l,j,k}$: valor para a variável p na observação j pertencente ao cluster l . A soma total dos erros quadrados (agrupando os clusters l e i), será $SS_{l,i} = \sum_{k=1}^{n_l} \sum_{j=1}^p (x_{l,k,j} - \bar{x}_j)^2 + \sum_{k=1}^{n_i} \sum_{j=1}^p (x_{i,k,j} - \bar{x}_j)^2$.

Tabela 2 – Média das características dos tipos de vizinhança

	Muito Carente	Carente	Média	Desenvolvido	Muito Desenvolvido	Total
% de Mulheres chefes	11,7	13,3	10,9	15,1	20,1	12,9
% de pessoas com baixo nível educacional	72,8	60,3	79,1	46,5	29,4	64,6
% de pessoas com ensino intermediário	22,9	26,9	17,8	26,4	22,6	23,9
% de pessoas com alto nível educacional	4,2	12,6	3,1	27,1	47,9	11,4
% de pessoas desocupadas	5,1	3,9	4,8	2,9	2,3	4,4
% de domicílios sem acesso à rede de saneamento	17,6	6,1	6,2	1,9	0,1	15,1
% de pessoas sem acesso a coleta de lixo	0,3	0,0	2,7	0,0	0,0	0,3
% de pessoas com menos de 15 anos	25,5	20,1	27,8	15,4	11,2	22,3

Nota: Os 5 tipos de vizinhança foram definidos por uma classificação hierárquica ascendente.

2.5 A Endogeneidade da Localização Residencial e Variável Instrumental

O principal desafio da abordagem adotada nesta pesquisa é a forte preocupação teórica (Fujita e Ogawa, 1982; Lucas e Rossi-Hansberg, 2002) de que, os resultados no mercado de trabalho são determinados simultaneamente com a acessibilidade aos empregos, onde este será potencialmente correlacionado com o termo de erro, podendo levar ao viés e inconsistência na estimativa de OLS. A solução para este problema exige encontrar um instrumento adequado para a medida de acessibilidade. Desse modo, a estratégia de variáveis instrumentais baseia-se na inclusão de uma variável geográfica como o instrumento da acessibilidade à empregos, a saber, neste caso, a distância através do rio para as oportunidades de emprego (CBD e subcentros). Quanto à justificativa para a escolha desse instrumento, baseia-se na mesma adotada no estudo de Haddad e Barufi (2017). A geografia atuou como determinante da localização da infraestrutura de transporte na região, de modo que a rede rodoviária da cidade apresenta forte correlação espacial com "vias navegáveis" pré-urbanas. Atualmente, a grande maioria dos rios e riachos foi cobertas por asfalto e cimento e são praticamente desconhecidos pelos agentes econômicos (Figura 2).

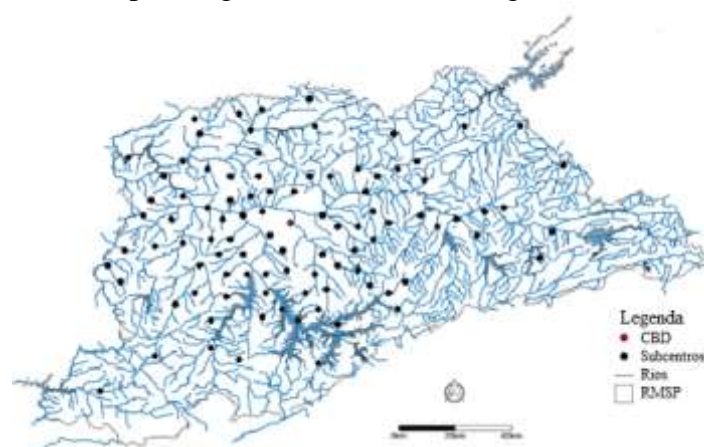


Figura 2 – Percurso dos rios e centros de empregos na RMSP.

Fonte: Elaboração própria com base nos dados do Censo 2010.

Haddad e Barufi (2017) com base numa perspectiva histórica, argumentam que em meados do século XIX, a cidade de São Paulo iniciou uma ocupação mais sistemática das várzeas dos principais cursos de água da cidade, tendo o motivo associado à implementação de infraestrutura regional e urbana (ferrovias e estradas). Em 1929, o Plano de Avenidas elaborado por Prestes Maia reforçou ainda mais a ocupação das planícies de inundação dos rios, concentrando-se no desenvolvimento de grandes avenidas nos rios da região (HADDAD e BARUFI, 2017). A *priori* o caso para a exogeneidade do instrumento parece, pois, plausível.

O instrumento foi construído a partir de uma base de dados GIS das rotas dos rios dentro da RMSP. A tarefa principal foi calcular a distância dos centróides das AEP's para os rios e depois para os centros de empregos para todas as unidades de observação no banco de dados. Assim, foi criada uma "Matriz Origem-Destino" pelo *software* QGIS, sendo que o parâmetro utilizado para construir a matriz foi o comprimento da rota (em quilômetros), assumindo que os indivíduos podem caminhar em ambas as direções ao longo das rotas.

3. Resultados

A análise estatística das variáveis selecionadas do banco de dados permite a obtenção de resultados preliminares para a RMSP, conforme a Tabela 3, a seguir. Para a média das variáveis dependentes, percebe-se uma elevada taxa percentual para a População Economicamente Ativa (PEA), tendo as variáveis para o desemprego, informalidade e *overeducated* valores menores, porém significativos para uma região metropolitana.

Tabela 3- Estatísticas resumidas das variáveis incluídas no modelo de regressão.

Variáveis	Média	Desvio Padrão	Mínimo	Máximo
PEA	0,71	0,45	0	1
Desemprego	0,08	0,28	0	1
Informalidade	0,15	0,36	0	1
<i>Overeducated</i>	0,14	0,34	0	1
LnACi_CBD	10,70	0,60	9,17	14,13
LnACi_sub	12,72	0,47	11,17	15,62
LnACi_CBD_IV	10,73	0,58	9,16	13,38
LnACi_sub_IV	12,76	0,45	11,15	13,95
Masculino	0,48	0,49	0	1
Idade	37,06	13,01	10	138
Idade ²	1543,40	1076,84	100	19044
Baixo	0,49	0,49	0	1
Intermediário	0,16	0,37	0	1
Alto	0,23	0,42	0	1
Bem Elevado	0,10	0,30	0	1
Casado	0,31	0,46	0	1
Chefe	0,30	0,45	0	1
Branco	0,57	0,49	0	1
Negro	0,06	0,24	0	1
Amarelo	0,01	0,12	0	1
Pardo	0,34	0,47	0	1
Muito Carente	0,55	0,49	0	1
Carente	0,21	0,41	0	1
Médio	0,07	0,25	0	1
Desenvolvida	0,12	0,33	0	1
Muito Desenvolvido	0,03	0,17	0	1
T_Família	4,08	1,93	1	42
Sanemamento	0,84	0,36	0	1
Coleta	0,94	0,23	0	1

Fonte: Elaboração própria.

Quanto às variáveis de acessibilidade, os indivíduos tem uma menor acessibilidade média aos empregos para o CBD do que para os subcentros. Em relação às variáveis educacionais, observa-se que o nível mais baixo é o que apresenta maior percentual na região. A variável de composição étnica aponta que a concentração de brancos é maior, seguidas por pardos. Quanto às variáveis de vizinhança, percebe-se que a variável “Muito Carente” prevalece na região metropolitana de São Paulo.

Iniciando as análises dos resultados, tem-se que os modelos IV visam lidar com a questão da endogeneidade levantada anteriormente. A acessibilidade aos empregos é instrumentalizada pela distância ao CBD e aos subcentros, seguindo um caminho formado por rios. Primeiramente, será analisada essa relação na participação do mercado, seguida pela condição de estar desempregado, na informalidade e, por fim, na chance de ser *overeducated*. A Tabela 4 mostra a regressão do 1º Estágio, bem como as estatísticas de Durbin e Wu-Hausman⁵, sugerindo que o estimador IV é preferível ao estimador⁶ OLS. De fato, o conjunto de estatísticas apresentadas para avaliar o instrumento fornece forte confiança para usá-lo. Primeiro, observe que a estatística F para o primeiro estágio é estatisticamente significativa e assumem valores elevados, segundo, os coeficientes de determinação R² apresentam valores altos.

Tabela 4 – Resultado da estimação para a probabilidade de participação no mercado.

Variáveis	CBD		Subcentros	
	MPL_IV	D.P.	MPL_IV	D.P.
1º Estágio				
LnACi_IV	0,9122***	0,0004	0,8154***	0,0006
Controles	Sim	Sim	Sim	Sim
R ² / Pseudo R ²	0,9250		0,8136	
Estat. F	542387.43***		191964.84***	
Durbin (score, <i>Chi</i>)	11,17***		9,23***	
Wu-Hausman (F)	11,17***		9,23***	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%. D.P: desvio padrão.

A Tabelas 5 exibe os resultados com base na equações 1 e 2 usando índices de acessibilidade. As colunas (1)-(4) exibem os resultados para o CBD, enquanto as colunas (5)-(8) são estimativas para subcentros. A tabela considera os efeitos marginais das variáveis para todos os indivíduos em idade de trabalhar no ponto médio da amostra. As estimativas dos regressores são obtidas a partir de modelos de Probabilidade Linear e Probit (efeitos marginais) com (colunas (3)-(4) e (7)-(8)) e sem (colunas (1)-(2) e (5)-(6)) variáveis instrumentais. Sobre os efeitos marginais, é importante destacar que a sua interpretação em modelos de variável dependente binária, é a seguinte: i) para variáveis contínuas, representa o impacto médio em termos de pontos percentuais na probabilidade de um individuo participar do mercado de trabalho para aumento unitário da variável explicativa contínua considerada; e ii) para as variáveis qualitativas, determina a mudança média na probabilidade de um indivíduo estar na PEA, em pontos percentuais, devido à presença da característica indicada pela *dummy* considerada.

De acordo com as estimativas, a principal variável de interesse neste estudo, acessibilidade ao trabalho, tem um impacto significativo na probabilidade de emprego. Tanto para o CBD quanto para os subcentros, um maior acesso à empregos aumenta a probabilidade de participação no mercado para todos os indivíduos da amostra. Este resultado é significativo comparado com a média de chance de participação que é de 0,71 com desvio padrão de 0,45.

⁵ O teste de Durbin-Wu-Hausman indica que os resíduos de uma regressão do índice de acessibilidade sobre as demais variáveis são estatisticamente significantes quando colocados como regressores de uma regressão da probabilidade de estar na PEA sobre todas as variáveis ditas explicativas.

⁶ Ambos os estimadores são tendenciosos nessa situação, mas apenas IV pode gerar estimativas consistentes.

Ou seja, tomando os efeitos marginais da coluna 7, tem-se que o aumento de 10% no índice de acessibilidade aumenta em aproximadamente 0,2% a probabilidade de se participar no mercado.

Tabela 5 – Efeitos Marginais para a probabilidade de participação no mercado.

Variáveis	CBD				Subcentros			
	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
LnACi	0,0128*** (0,0010)	0,0133*** (0,0011)	0,0143*** (0,0011)	0,0058*** (0,0012)	0,0112*** (0,0013)	0,0109*** (0,0014)	0,0141*** (0,0016)	0,0054*** (0,0017)
Masculino	0,1611*** (0,0009)	0,1752*** (0,0009)	0,1611*** (0,0009)	0,1752*** (0,0009)	0,1611*** (0,0009)	0,1751*** (0,0009)	0,1611*** (0,0009)	0,1751*** (0,0009)
Idade	0,0505*** (0,0002)	0,0489*** (0,0002)	0,0505*** (0,0002)	0,0489*** (0,0002)	0,0505*** (0,0002)	0,0489*** (0,0002)	0,0505*** (0,0002)	0,0489*** (0,0002)
Idade ²	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)	-0,0006*** (0,0000)
Baixa	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Intermediária	0,0567*** (0,0013)	0,0547*** (0,0012)	0,0567*** (0,0013)	0,0547*** (0,0012)	0,0566*** (0,0013)	0,0546*** (0,0012)	0,0566*** (0,0013)	0,0546*** (0,0012)
Alta	0,1663*** (0,0012)	0,1601*** (0,0011)	0,1663*** (0,0012)	0,1601*** (0,0011)	0,1662*** (0,0012)	0,1600*** (0,0011)	0,1662*** (0,0012)	0,1600*** (0,0011)
Bem elevada	0,2410*** (0,0016)	0,2161*** (0,0010)	0,2411*** (0,0016)	0,2161*** (0,0010)	0,2407*** (0,0016)	0,2159*** (0,0010)	0,2407*** (0,0016)	0,2159*** (0,0010)
Casado	-0,0311*** (0,0010)	-0,0280*** (0,0011)	-0,0310*** (0,0010)	-0,0280*** (0,0011)	-0,0313*** (0,0010)	-0,0282*** (0,0011)	-0,0313*** (0,0010)	-0,0282*** (0,0011)
Chefe	0,0650*** (0,0010)	0,0780*** (0,0011)	0,0650*** (0,0010)	0,0780*** (0,0011)	0,0651*** (0,0010)	0,0781*** (0,0011)	0,0651*** (0,0010)	0,0781*** (0,0011)
Branco	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Negro	0,0429*** (0,0018)	0,0449*** (0,0018)	0,0428*** (0,0018)	0,0448*** (0,0018)	0,0430*** (0,0018)	0,0450*** (0,0018)	0,0429*** (0,0018)	0,0449*** (0,0018)
Amarelo	-0,0269 (0,0035)	-0,0332*** (0,0041)	-0,0269*** (0,0035)	-0,0335*** (0,0041)	-0,0265*** (0,0035)	-0,0327*** (0,0041)	-0,0265*** (0,0035)	-0,0329*** (0,0041)
Pardo	0,0228*** (0,0010)	0,0239*** (0,0011)	0,0228*** (0,0010)	0,0239*** (0,0011)	0,0228*** (0,0010)	0,0239*** (0,0011)	0,0228*** (0,0010)	0,0239*** (0,0011)
Muito Carente	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Carente	-0,0088*** (0,0012)	-0,0084*** (0,0014)	0,0037* (0,0022)	-0,0089*** (0,0014)	-0,0068*** (0,0022)	-0,0060*** (0,0014)	-0,0077*** (0,0012)	-0,0067*** (0,0014)
Média	-0,0136*** (0,0019)	-0,0149*** (0,0021)	-0,0132*** (0,0019)	-0,0147*** (0,0021)	-0,0135*** (0,0019)	-0,0151*** (0,0021)	-0,0126*** (0,0019)	-0,0142*** (0,0021)
Desenvolvida	-0,0216*** (0,0017)	-0,0229*** (0,0020)	-0,0229*** (0,0018)	-0,0233*** (0,0021)	-0,0172*** (0,0017)	-0,0178*** (0,0019)	-0,0191*** (0,0018)	-0,0182*** (0,0020)
Muito Desenvolvida	-0,0531*** (0,0031)	-0,0667*** (0,0040)	-0,0556*** (0,0032)	-0,0641*** (0,0041)	-0,0435*** (0,0029)	-0,0548*** (0,0037)	-0,0464*** (0,0031)	-0,0541*** (0,0038)
Tamanho da família	-0,0030*** (0,0002)	-0,0035*** (0,0002)	-0,0030*** (0,0002)	-0,0035*** (0,0002)	-0,0030*** (0,0002)	-0,0035*** (0,0002)	-0,0030*** (0,0002)	-0,0035*** (0,0002)
Saneamento	-0,0003 (0,0013)	0,0007 (0,0015)	-0,0004 (0,0013)	0,0005 (0,0015)	-0,0002 (0,0013)	0,0008 (0,0015)	-0,0006 (0,0014)	0,0006 (0,0015)
Coleta	0,0158*** (0,0020)	0,0577*** (0,0068)	0,0160*** (0,0020)	0,0190*** (0,0022)	0,0156*** (0,0020)	0,0185*** (0,0022)	0,0159*** (0,0020)	0,0187*** (0,0022)
Constante	0,535*** (0,0120)	-	-0,4740*** (0,0125)	-	-0,4658*** (0,0165)	-	-0,5006*** (0,0201)	-
Estat. F/Wald	8466,47	141745,08	8466,88	123761,12	8461,79	141674,39	8461,87	123711,32
R ² / Pseudo R ²	0,1614	0,1419	0,1614		0,1613	0,1418	0,1613	
Observações	835.804	835.804	835.804	835.804	835.804	835.804	835.804	835.804

Notas: *** Significativo ao nível de 1%, ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

A relação entre acessibilidades ao emprego e resultados do mercado de trabalho mostra padrões mais distintos quando permite a heterogeneidade entre os gêneros. Na Tabela 6 têm-se efeitos heterogêneos para homens e mulheres. Para os homens, observa-se a não significância quando considerado apenas o CBD, porém quando considerado os subcentros, incluindo o CBD, a relação entre acessibilidade à empregos e participação no mercado de

trabalho apresenta sinal oposto ao esperado. A insignificância do coeficiente estimado para acessibilidade pode está associado ao fato de que os homens, em geral, participam mais do mercado de trabalho (55% dos ocupados são do gênero masculino na amostra) que as mulheres já que são majoritariamente os chefes da família e podem sofrer maior pressão social para inserção no mercado de trabalho. Assim, seus resultados mostram que acesso mais difícil não é suficiente para inibir a participação destes. Para as mulheres, é interessante observar o efeito marginal que é quase o dobro daquele obtido para amostra em geral.

Tabela 6 – Efeitos marginais com e sem variáveis instrumentais para a probabilidade de participação no mercado (regressões para homens e mulheres)

Variáveis	CBD				Subcentros			
	HOMENS							
PEA	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
LnACi	0,0018 (0,0013)	0,0003 (0,0014)	0,0022 (0,0014)	-0,0013 (0,0015)	-0,0036** (0,0016)	-0,0056*** (0,0017)	-0,0044** (0,0020)	-0,0030 (0,0084)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	401.311	401.311	401.311	401.311	401.311	401.311	401.311	401.311
Estat. F/Wald	3993,23	59633,15	3993,26	56038,89	3993,41	59643,54	3993,42	53043,93
R ² / Pseudo R ²	0,1519	0,1494	0,1519		0,1519	0,1494	0,1519	
MULHERES								
LnACi	0,0235*** (0,0015)	0,0263*** (0,0048)	0,0259*** (0,0017)	0,0133*** (0,0019)	0,0256*** (0,0019)	0,0277*** (0,0021)	0,0325*** (0,0024)	0,0115*** (0,0024)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	434.493	434.493	434.493	434.493	434.493	434.493	434.493	434.493
Estat. F/Wald	3821,22	62253,10	3821,46	56232,27	3818,11	62198,34	3818,60	56213,74
R ² / Pseudo R ²	0,1367	0,1091	0,1367		0,1366	0,1090	0,1366	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

Na literatura, existem diversas explicações sobre por que as mulheres são mais sensíveis ao mercado de trabalho local do que homens, resumido em MacDonald (1999). Uma explicação é que as mulheres ganham menos do que os homens e os salários médios das mulheres variam menos no território, visto que as mulheres recebem uma compensação menos assalariada por viagens mais longas do que os homens. Outra explicação é que, na maioria dos casos, as mulheres não são chefes de famílias, portanto, buscam empregos parciais, sazonais ou temporários menos remunerados do que empregos em tempo integral.

Passando para a relação do indicador de acessibilidade na probabilidade de estar desemprego, tem-se primeiramente, a partir da Tabela 7, uma associação positiva e estatisticamente significativa entre o instrumento (acesso físico ao emprego por rios) e a variável dependente (probabilidade de estar desempregado), algo bem vindo por ter confiança em usar a variável como instrumento. Além disso, as estatísticas indicam que os resultados pelo modelo IV são mais consistentes.

Tabela 7 – Resultado da estimação para a probabilidade de desempregados.

Variáveis	CBD		Subcentros	
	MPL_IV	D.P.	MPL_IV	D.P.
1º Estágio				
LnACi_IV	0,9153***	0,0005	0,8121***	0,0007
Controles	Sim	Sim	Sim	Sim
R ² / Pseudo R ²	0,9239		0,8142	
Estat. F	394036,75***		142216,58***	
Durbin (score, <i>Chi</i>)	6,21**		25,53***	
Wu-Hausman (F)	6,21**		25,53***	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%. D.P: desvio padrão.

Analisando os resultados para os modelos da Tabela 8, tem-se que o melhor acesso à oportunidades de empregos diminui a probabilidade do indivíduo estar desempregado, sendo os efeitos do CBD semelhantes aos subcentros. Ainda nesta tabela, percebe-se que os resultados para os efeitos marginais são fortes quando comparado à média da chance de estar desempregado que é de 0,08 e o desvio padrão de 0,28 (Tabela 3). Tomando os maiores valores dos efeitos marginais que correspondem as colunas (3) e (7), tem-se que um aumento de 1% no índice de acessibilidade reduz a probabilidade de estar desemprego em 0,13% e 0,17% para o CBD e subcentros, respectivamente.

Tabela 8 – Efeitos marginais para a probabilidade de estar desempregado.

Variáveis	CBD				Subcentros			
	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
Desemprego								
LnACi	-0,0101*** (0,0007)	-0,0098*** (0,0007)	-0,0109*** (0,0008)	-0,0059*** (0,0008)	-0,0098*** (0,0015)	-0,0089*** (0,0009)	-0,0135*** (0,0012)	-0,0027*** (0,0011)
Masculino	-0,0410*** (0,0007)	-0,0384*** (0,0006)	-0,0410*** (0,0007)	-0,0384*** (0,0006)	-0,0410*** (0,0007)	-0,0384*** (0,0006)	-0,0410*** (0,0007)	-0,0384*** (0,0006)
Idade	-0,0132*** (0,0001)	-0,0084*** (0,0007)	-0,0132*** (0,0001)	-0,0084*** (0,0001)	-0,0132*** (0,0001)	-0,0084*** (0,0001)	-0,0132*** (0,0001)	-0,0084*** (0,0001)
Idade ²	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0005*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)
Baixa	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Intermediária	0,0090*** (0,0010)	0,0029*** (0,0009)	0,0090*** (0,0067)	0,0029*** (0,0009)	0,0090*** (0,0010)	0,0030*** (0,0009)	0,0090*** (0,0010)	0,0030*** (0,0009)
Alta	-0,0211*** (0,0009)	-0,0171*** (0,0007)	-0,0212*** (0,0009)	-0,0172*** (0,0007)	-0,0210*** (0,0009)	-0,0170*** (0,0007)	-0,0211*** (0,0009)	-0,0170*** (0,0007)
Bem elevada	-0,0438*** (0,0012)	-0,0452*** (0,0008)	-0,0438*** (0,0012)	-0,0452*** (0,0008)	-0,0435*** (0,0012)	-0,0450*** (0,0008)	-0,0435*** (0,0012)	-0,0450*** (0,0008)
Casado	-0,0136*** (0,0007)	-0,0167*** (0,0007)	-0,0137*** (0,0007)	-0,0167*** (0,0007)	-0,0134*** (0,0007)	-0,0166*** (0,0007)	-0,0135*** (0,0007)	-0,0166*** (0,0007)
Chefe	-0,0093*** (0,0007)	-0,0140*** (0,0007)	-0,0093*** (0,0007)	-0,0141*** (0,0007)	-0,0094*** (0,0008)	-0,0141*** (0,0007)	-0,0094*** (0,0008)	-0,0141*** (0,0007)
Branco	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Negro	0,0066*** (0,0013)	0,0067*** (0,0013)	0,0066*** (0,0013)	0,0067*** (0,0013)	0,0065*** (0,0013)	0,0066*** (0,0013)	0,0066*** (0,0013)	0,0067*** (0,0013)
Amarelo	-0,0025 (0,0026)	-0,0051* (0,0027)	-0,0025 (0,0026)	-0,0049* (0,0027)	-0,0028 (0,0026)	-0,0053** (0,0027)	-0,0028 (0,0026)	-0,0051* (0,0027)
Pardo	0,0072*** (0,0007)	0,0064*** (0,0007)	0,0072*** (0,0007)	0,0064*** (0,0007)	0,0072*** (0,0007)	0,0064*** (0,0007)	0,0072*** (0,0007)	0,0065*** (0,0007)
Muito Carente	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Carente	-0,0091*** (0,0009)	-0,0076*** (0,0008)	-0,0088*** (0,0009)	-0,0074*** (0,0008)	-0,0104*** (0,0009)	-0,0090*** (0,0008)	-0,0092*** (0,0009)	-0,0083*** (0,0008)
Média	-0,0105*** (0,0015)	-0,0083*** (0,0012)	-0,0107*** (0,0015)	-0,0084*** (0,0012)	-0,0109*** (0,0015)	-0,0084*** (0,0012)	-0,0120*** (0,0015)	-0,0092*** (0,0012)
Desenvolvida	-0,0097*** (0,0013)	-0,0119** (0,0012)	-0,0089*** (0,0013)	-0,0118*** (0,0012)	-0,0126*** (0,0013)	-0,0147*** (0,0011)	-0,0102*** (0,0013)	-0,0144** (0,0011)
Muito Desenvolvida	-0,0013*** (0,0023)	-0,0067*** (0,0023)	0,0000 (0,0024)	-0,0079** (0,0024)	-0,0080*** (0,0022)	-0,0132** (0,0020)	-0,0042*** (0,0023)	-0,0140*** (0,0021)
Tamanho da família	0,0057*** (0,0002)	0,0035*** (0,0001)	0,0057*** (0,0002)	0,0035*** (0,0001)	0,0057*** (0,0002)	0,0035*** (0,0001)	0,0057*** (0,0002)	0,0035*** (0,0001)
Saneamento	-0,0061*** (0,0010)	-0,0043*** (0,0009)	-0,0060*** (0,0010)	-0,0042*** (0,0009)	-0,0060*** (0,0010)	-0,0042*** (0,0009)	-0,0055*** (0,0010)	-0,0041*** (0,0009)
Coleta	-0,0091*** (0,0015)	-0,0076*** (0,0014)	-0,0092*** (0,0015)	-0,0077*** (0,0014)	-0,0090*** (0,0015)	-0,0074*** (0,0014)	-0,0093*** (0,0015)	-0,0077*** (0,0014)
Constante	0,5108*** (0,0088)	-	0,5195*** (0,0095)	-	0,5282*** (0,0127)	-	0,5734*** (0,0155)	-
Estat. F/Wald	1854,09	31742,53	1854,03	28853,39	1850,58	31674,94	1851,73	23828,52
R ² / Pseudo R ²	0,0540	0,0875	0,0540		0,0539	0,0873	0,0539	
Observações	616.754	616.754	616.754	616.754	616.754	616.754	616.754	616.754

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

Os resultados da Tabela 9, a seguir, referem-se à probabilidade de estar desempregado para homens e mulheres. Observa-se claramente que o efeito do coeficiente é mais forte para as mulheres do que para os homens quando analisado o índice de acessibilidade.

Tabela 9 – Efeitos marginais com e sem variáveis instrumentais para a probabilidade de estar desempregado (regressões para homens e mulheres)

Variáveis	CBD				Subcentros			
	HOMENS							
Desemprego	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
LnACi	-0,0056*** (0,0009)	-0,0054*** (0,0010)	-0,0056*** (0,0010)	-0,0056*** (0,0010)	-0,0046*** (0,0012)	-0,0043*** (0,0012)	-0,0058*** (0,0014)	-0,0021*** (0,0014)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	333.347	333.347	333.347	333.347	333.347	333.347	333.347	333.347
Estat. F/Wald	926,50	14311,00	926,22	13297,58	925,37	14294,93	925,43	13289,09
R ² / Pseudo R ²	0,0476	0,0873	0,0476	-	0,0476	0,0869	0,0476	-
MULHERES								
LnACi	-0,0159*** (0,0013)	-0,0169*** (0,0013)	-0,0179*** (0,0014)	-0,0073*** (0,0014)	-0,0170*** (0,0016)	-0,0160*** (0,0015)	-0,0236*** (0,0020)	-0,0042** (0,0020)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	283.407	283.407	283.407	283.407	283.407	283.407	283.407	283.407
Estat. F/Wald	915,98	15258,06	916,42	13852,05	913,44	15198,89	914,92	13837,23
R ² / Pseudo R ²	0,0550	0,0782	0,0550	-	0,0548	0,0779	0,0548	-

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

Analisando agora os resultados para a informalidade, tem-se que, como explicado anteriormente, a validade do instrumento proposto baseia-se em sua relevância e exogeneidade em relação à variável de resultado. Para avaliar sua relevância, estimou-se também as regressões apresentadas na Tabela 10. Como pode ser visto nas duas primeiras linhas, o instrumento é estatisticamente significativo e positivo. O instrumento parece ser relevante pelos testes apresentados.

Tabela 10 – Resultado da estimação para a probabilidade de se estar na informalidade.

Variáveis	CBD		Subcentros	
1º Estágio	MPL_IV	D.P.	MPL_IV	D.P.
LnACi_IV	0,9163***	0,0005	0,8112***	0,0007
Controles	Sim	Sim	Sim	Sim
R ² / Pseudo R ²	0,9231		0,8146	
Estat. F	355840.82***		130290.52***	
Durbin (score, <i>Chi</i>)	0,15		27,88***	
Wu-Hausman (F)	0,15		27,89***	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%. D.P: desvio padrão.

A Tabela 11 aponta para o acesso desigual aos empregos em termos espaciais como determinante da informalidade, e sugere que a baixa acessibilidade aos empregos formais, devido às características da estrutura espacial, determina quem vai acabar trabalhando no setor informal. Observando os efeitos marginais das variáveis de acesso aos empregos e tomando o valor da coluna 7 da tabela, temos: um aumento de 10% do índice de acessibilidade está associado a uma redução de quase 1% na chance de se estar na informalidade. Os sinais encontrados dos parâmetros corroboram com a literatura do *mismatch* espacial que sugerem que o melhor acesso aos centros de emprego contribuem na redução da informalidade indo na mesma linha de diversas evidências empíricas sobre o tema, tais como os trabalhos de Miltra (2005), Suárez et al. (2016), Chen et al. (2017), Moreno-Monroy e Ramos (2015), Motte et al., (2016) e Boisjoly et al., (2017), apesar dessas pesquisas apresentarem níveis de análise distintos e outras metodologias.

Tabela 11 – Efeitos marginais para a probabilidade de trabalhadores informais.

Variáveis	CBD				Subcentros			
	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
Informalidade								
LnACi	-0,0048*** (0,0010)	-0,0042*** (0,0010)	-0,0048*** (0,0011)	-0,0034*** (0,0011)	-0,0080*** (0,0013)	-0,0078*** (0,0013)	-0,0133*** (0,0016)	0,0021 (0,0016)
Masculino	-0,0741*** (0,0009)	-0,0724*** (0,0009)	-0,0741*** (0,0009)	-0,0724*** (0,0009)	-0,0741*** (0,0009)	-0,0724*** (0,0009)	-0,0741*** (0,0009)	-0,0724*** (0,0009)
Idade	-0,0093*** (0,0001)	-0,0079*** (0,0001)	-0,0093*** (0,0001)	-0,0079*** (0,0001)	-0,0093*** (0,0001)	-0,0079*** (0,0001)	-0,0093*** (0,0001)	-0,0079*** (0,0001)
Idade ²	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)
Baixa	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Intermediária	-0,0443*** (0,0014)	-0,0353*** (0,0011)	-0,0443*** (0,0014)	-0,0353*** (0,0011)	-0,0443*** (0,0014)	-0,0353*** (0,0011)	-0,0443*** (0,0014)	-0,0353*** (0,0011)
Alta	-0,1114*** (0,0012)	-0,0944*** (0,0010)	-0,1114*** (0,0012)	-0,0944*** (0,0010)	-0,1114*** (0,0012)	-0,0944*** (0,0010)	-0,1114*** (0,0012)	-0,0944*** (0,0010)
Bem elevada	-0,1521*** (0,0016)	-0,1241*** (0,0010)	-0,1521*** (0,0016)	-0,1241*** (0,0010)	-0,1519*** (0,0016)	-0,1240*** (0,0001)	-0,1519*** (0,0016)	-0,1240*** (0,0010)
Casado	-0,0309*** (0,0010)	-0,0326*** (0,0010)	-0,0309*** (0,0010)	-0,0326*** (0,0010)	-0,0309*** (0,0010)	-0,0326*** (0,0046)	-0,0310*** (0,0010)	-0,0327*** (0,0010)
Chefe	0,0018*** (0,0010)	-0,0020** (0,0010)	0,0018* (0,0010)	-0,0020** (0,0010)	0,0018* (0,0010)	-0,0020** (0,0010)	0,0018* (0,0010)	-0,0021** (0,0010)
Branco	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Negro	0,0147*** (0,0018)	0,0140*** (0,0019)	0,0147*** (0,0018)	0,0140*** (0,0019)	0,0147*** (0,0018)	0,0141*** (0,0019)	0,0148*** (0,0018)	0,0142*** (0,0019)
Amarelo	-0,0015 (0,0035)	-0,0032 (0,0037)	-0,0015 (0,0035)	-0,0031 (0,0037)	-0,0017 (0,0035)	-0,0033 (0,0037)	-0,0017 (0,0035)	-0,0030 (0,0037)
Pardo	0,0101*** (0,0010)	0,0095*** (0,0010)	0,0101*** (0,0010)	0,0095*** (0,0010)	0,0102*** (0,0010)	0,0096*** (0,0010)	0,0102*** (0,0010)	0,0096*** (0,0010)
Muito Carente	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Carente	0,0021* (0,0013)	0,0011 (0,0013)	0,0021* (0,0013)	0,0012 (0,0013)	0,0026** (0,0013)	0,0018 (0,0013)	0,0043*** (0,0013)	0,0031*** (0,0013)
Média	0,0068*** (0,0020)	0,0063*** (0,0020)	0,0068*** (0,0020)	0,0063*** (0,0020)	0,0056*** (0,0021)	0,0050** (0,0020)	0,0039* (0,0021)	0,0034* (0,0020)
Desenvolvida	0,0065*** (0,0018)	0,0042** (0,0018)	0,0065*** (0,0018)	0,0043** (0,0019)	0,0073*** (0,0017)	0,0054*** (0,0018)	0,0107** (0,0018)	0,0061*** (0,0019)
Muito Desenvolvida	0,0099*** (0,0031)	0,0082** (0,0035)	0,0100*** (0,0032)	0,0079** (0,0035)	0,0102*** (0,0029)	0,0091*** (0,0033)	0,0155*** (0,0031)	0,0076*** (0,0034)
Tamanho da família	0,0038*** (0,0002)	0,0031*** (0,0002)	0,0038*** (0,0002)	0,0031*** (0,0002)	0,0038*** (0,0002)	0,0031*** (0,0002)	0,0038*** (0,0002)	0,0031*** (0,0002)
Saneamento	-0,0190*** (0,0014)	-0,0168*** (0,0014)	-0,0190*** (0,0014)	-0,0167*** (0,0014)	-0,0185*** (0,0014)	-0,0162*** (0,0062)	-0,0177*** (0,0015)	-0,0159*** (0,0015)
Coleta	-0,0131*** (0,0021)	-0,0110*** (0,0021)	-0,0131*** (0,0091)	-0,0110*** (0,0021)	-0,0133*** (0,0021)	-0,0113*** (0,0021)	-0,0137*** (0,0021)	-0,0117*** (0,0021)
Constante	0,535*** (0,0120)	-	0,5364*** (0,0129)	-	0,5854*** (0,0173)	-	0,6499*** (0,0212)	-
Estat. F/Wald	1411,85	26042,45	1411,70	24430,34	1412,68	26060,38	1414,08	24475,41
R ² / Pseudo R ²	0,0454	0,0533	0,0454	-	0,0455	0,0533	0,0454	-
Observações	563.463	563.463	563.463	563.463	563.463	563.463	563.463	563.463

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

A Tabela 12 apresenta os resultados para homens e mulheres. Observa-se que a relação entre o melhor acesso físico ao CBD para homens não foi significativo, enquanto que para as mulheres se obteve uma relação inversa e significativa. Quando analisado os subcentros como oportunidades de empregos, a relação foi significativa para ambos os gêneros, sendo bem mais forte para o sexo feminino. A diferença sugere que existe uma segmentação de gênero nos mercados de trabalho, com predominância da informalidade para as mulheres.

Tabela 12 – Efeitos marginais com e sem variáveis instrumentais para a probabilidade de estar na informalidade (regressões para homens e mulheres)

Variáveis	CBD				Subcentros			
	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
HOMENS								
Informalidade								
LnACi	-0,0014 (0,0013)	-0,0007 (0,0013)	-0,0016 (0,0014)	-0,0046 (0,0073)	-0,0044*** (0,0016)	-0,0037** (0,0016)	-0,0083*** (0,0020)	-0,0378*** (0,0101)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	310.912	310.912	310.912	310.912	310.912	310.912	310.912	310.912
Estat. F/Wald	596,16	10012,90	596,17	9646,69	596,49	10017,63	597,02	9660,16
R ² / Pseudo R ²	0,0334	0,0422	0,0334		0,0334	0,0422	0,0334	
MULHERES								
LnACi	-0,0102*** (0,0017)	-0,0100*** (0,0017)	-0,0104*** (0,0018)	-0,0086*** (0,0019)	-0,0148*** (0,0002)	-0,0148*** (0,0022)	-0,0224*** (0,0027)	-0,0001*** (0,0027)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	252.551	252.551	252.551	252.551	252.551	252.551	252.551	252.551
Estat. F/Wald	823,34	14364,35	823,08	13467,59	823,92	14376,81	825,13	13506,34
R ² / Pseudo R ²	0,0554	0,0581	0,0554		0,0555	0,0581	0,0554	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

Por fim, tem-se a análise da relação entre a acessibilidade e a condição de *overeducated*. Os resultados para o 1º estágio, bem como as estatísticas sugerem para o uso das estimativas IV.

Tabela 13 – Resultado da estimação para a probabilidade de *overeducated*.

Variáveis	CBD		Subcentros	
	MPL_IV	D.P.	MPL_IV	D.P.
1º Estágio				
LnACi_IV	0,9163***	0,0005	0,8112***	0,0103
Controles	Sim	Sim	Sim	Sim
R ² / Pseudo R ²	0,9231		0,8146	
Estat. F	375610,09***		137529,13***	
Durbin (score, <i>Chi</i>)	21.07***		8,36***	
Wu-Hausman (F)	21.07***		8,36***	

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%. D.P: desvio padrão.

Na Tabela 14, os coeficientes do índice de acessibilidade para o CBD são positivos e significativos nos modelos sem variáveis instrumentais. Esses resultados não são esperados de acordo com a teoria do *mismatch* espacial e educacional. Ao analisar os subcentros de emprego, percebe-se que os coeficientes são negativos e significativos para todos os modelos estimados. Observando os efeitos marginais das variáveis de acesso aos empregos e tomando o valor da coluna 7 da tabela, temos: um aumento de 10% do índice de acessibilidade está associado a uma redução de 0,56% na chance de se estar na informalidade.

Tabela 14 – Efeitos marginais para a probabilidade de *overeducated*.

Variáveis	CBD				Subcentros			
	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
Overeducated								
LnACi	0,0035*** (0,0012)	0,0047*** (0,0013)	0,0010 (0,0013)	0,0058 (0,0045)	-0,0045*** (0,0016)	-0,0036** (0,0017)	-0,0079*** (0,0020)	-0,0026** (0,0019)
Masculino	-0,0096*** (0,0011)	-0,0087*** (0,0012)	-0,0096*** (0,0011)	-0,0087*** (0,0012)	-0,0045*** (0,0016)	-0,0087*** (0,0012)	-0,0097*** (0,0011)	-0,0087*** (0,0012)
Idade	-0,0057*** (0,0002)	-0,0059*** (0,0002)	-0,0057*** (0,0002)	-0,0059*** (0,0002)	-0,0057*** (0,0002)	-0,0059*** (0,0002)	-0,0057*** (0,0002)	-0,0059*** (0,0002)
Idade ²	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0001*** (0,0000)	0,0000*** (0,0000)	0,0001*** (0,0000)
Baixa	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Intermediária	0,3774***	0,3995***	0,3773***	0,3995***	0,3772***	0,3993***	0,3771***	0,3993***

	(0,0013)	(0,0014)	(0,0013)	(0,0014)	(0,0013)	(0,0014)	(0,0013)	(0,0014)
Alta	0,2590***	0,3185***	0,2589***	0,3185***	0,2589***	0,3183***	0,2589***	0,3183***
	(0,0017)	(0,0021)	(0,0017)	(0,0021)	(0,0017)	(0,0021)	(0,0017)	(0,0021)
Casado	-0,0047***	-0,0033**	-0,0048***	-0,0033**	-0,0049***	-0,0036**	-0,0050***	-0,0036**
	(0,0012)	(0,0013)	(0,0012)	(0,0013)	(0,0012)	(0,0013)	(0,0012)	(0,0013)
Chefe	-0,0001	-0,0010	-0,0001	-0,0010	-0,0001	-0,0010	-0,0001	-0,0010
	(0,0012)	(0,0014)	(0,0012)	(0,0014)	(0,0012)	(0,0014)	(0,0012)	(0,0014)
Branco	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Negro	0,0063***	0,0031	0,0064***	0,0031	0,0064***	0,0032	0,0065***	0,0032
	(0,0022)	(0,0025)	(0,0022)	(0,0025)	(0,0022)	(0,0025)	(0,0022)	(0,0025)
Amarelo	-0,0062	-0,0031	-0,0062	-0,0031	-0,0062	-0,0030	-0,0062	-0,0030
	(0,0042)	(0,0045)	(0,0042)	(0,0045)	(0,0042)	(0,0045)	(0,0042)	(0,0045)
Pardo	0,0173***	0,0157***	0,0174***	0,0157***	0,0174***	0,0157***	0,0174***	0,0157***
	(0,0012)	(0,0014)	(0,0012)	(0,0014)	(0,0012)	(0,0014)	(0,0012)	(0,0014)
Muito Carente	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Carente	-0,0242***	-0,0231***	-0,0230***	-0,0231***	-0,0211***	-0,0198***	-0,0199***	-0,0198***
	(0,0015)	(0,0016)	(0,0015)	(0,0016)	(0,0015)	(0,0016)	(0,0015)	(0,0016)
Média	-0,0057**	-0,0067**	-0,0064**	-0,0067**	-0,0080***	-0,0090***	-0,0091***	-0,0090***
	(0,0024)	(0,0028)	(0,0024)	(0,0028)	(0,0024)	(0,0028)	(0,0025)	(0,0028)
Desenvolvida	-0,0731***	-0,0673***	-0,0707***	-0,0673***	-0,0669***	-0,0613***	-0,0647***	-0,0613***
	(0,0021)	(0,0021)	(0,0021)	(0,0021)	(0,0020)	(0,0020)	(0,0020)	(0,0020)
Muito Desenvolvida	-0,1483***	-0,1268***	-0,1441***	-0,1268***	-0,1379***	-0,1185***	-0,1344***	-0,1185***
	(0,0037)	(0,0029)	(0,0038)	(0,0029)	(0,0035)	(0,0028)	(0,0035)	(0,0028)
Tamanho da família	0,0021***	0,0021***	0,0021***	0,0021***	0,0021***	0,0021***	0,0021***	0,0021***
	(0,0003)	(0,0003)	(0,0003)	(0,0003)	(0,0003)	(0,0003)	(0,0003)	(0,0003)
Saneamento	0,0089***	0,0120***	0,0092***	0,0120***	0,0100***	0,0131***	0,0104***	0,0131***
	(0,0017)	(0,0019)	(0,0017)	(0,0019)	(0,0017)	(0,0019)	(0,0017)	(0,0019)
Coleta	0,0061**	0,0092**	0,0059**	0,0092**	0,0055**	0,0086***	0,0053**	0,0086***
	(0,0026)	(0,0029)	(0,0026)	(0,0029)	(0,0026)	(0,0029)	(0,0026)	(0,0029)
Constante	0,2282***	-0,8621***	0,2550***	-0,8621***	0,3221***	-0,5763***	0,3652***	-0,5763***
	(0,0144)	(0,0477)	(0,0156)	(0,0477)	(0,0210)	(0,0696)	(0,0257)	(0,0696)
Estat. F/Wald	5658,61	90187,15	101850,59	82875,69	5658,63	90179,85	101865,85	82887,97
R ² /Pseudo R ²	0,1531	0,1311	0,1531	-	0,1531	0,1311	0,1531	-
Observações	563.463	563.463	563.463	563.463	563.463	563.463	563.463	563.463

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

A Tabela 15 mostra os resultados das regressões separadas por gêneros. Os coeficientes para os subcentros novamente apresentam sinais esperados, com efeito, um pouco mais forte para o gênero feminino quando significativos. Note que estes novos resultados estão em linha com resultados anteriores: acesso é muito mais importante para as mulheres.

Tabela 15 – Efeitos marginais com e sem variáveis instrumentais para a probabilidade de overeducated (regressões para homens e mulheres)

Variáveis	CBD				Subcentros			
	HOMENS							
Overeducated	MPL (1)	Probit (2)	MPL_IV (3)	Probit_IV (4)	MPL (5)	Probit (6)	MPL_IV (7)	Probit_IV (8)
LnACi	0,0049*** (0,0017)	0,0064*** (0,0018)	0,0023 (0,0018)	0,0116* (0,0006)	-0,0044** (0,0021)	-0,0030 (0,0023)	-0,0067** (0,0026)	-0,0158* (0,0089)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	310.912	310.912	310.912	310.912	310.912	310.912	310.912	310.912
Estat. F/Wald	3145,34	47148,84	53466,69	43812,38	3145,05	47138,17	53470,93	43814,64
R ² /Pseudo R ²	0,1468	0,1267	0,1467	-	0,1467	0,1267	0,1467	-
MULHERES								
LnACi	0,0013 (0,0019)	0,0020 (0,0021)	-0,0012 (0,0021)	-0,0025 (0,0068)	-0,0053** (0,0024)	-0,0050** (0,0026)	-0,0105*** (0,0030)	-0,0048 (0,0030)
Controles	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Observações	252.551	252.551	252.551	252.551	252.551	252.551	252.551	252.551
Estat. F/Wald	2842,27	42971,19	48321,58	38928,59	2842,57	42973,73	48333,37	38939,33
R ² /Pseudo R ²	0,1606	0,1364	0,1606	-	0,1606	0,1364	0,1606	-

Notas: *** Significativo ao nível de 1%; ** 5% e * 10%.

Fonte: Elaboração própria com base na média dos efeitos marginais.

4. Considerações Finais

Este artigo mede a acessibilidade ao emprego e sua relação com os resultados trabalhistas dos residentes da região metropolitana de São Paulo. Uma das principais contribuições é fornecer estimativas consistentes da acessibilidade aos empregos usando um instrumento exógeno para analisar o impacto em diferentes variáveis dependentes, a saber, participação no mercado, condição de estar desempregado, na informalidade e de ser *overeducated*. Além disto, o estudo buscou analisar esta relação separada por gêneros.

Os resultados deste trabalho enfatizam a necessidade de analisar resultados do mercado de trabalho na Economia Urbana. Foram encontradas evidências importantes com disparidades espaciais em termos de acesso a oportunidades de emprego nas áreas de ponderação da RMSP, bem como efeitos de gênero heterogêneos na conectividade nos resultados trabalhistas. Os resultados mostraram que, em geral, o melhor acesso físico aos locais de emprego aumenta a participação no mercado, diminui o desemprego, a informalidade e a condição de *overeducated*.

A acessibilidade às oportunidades de emprego é um dos muitos desafios. Famílias mais pobres localizados em áreas periféricas e de difíceis acessos, geralmente têm dificuldade de se proteger e se engajar em atividades econômicas significativas. Ao nível mais agregado da cidade, a má conectividade evita o crescimento de *clusters* e a produtividade urbana. O aumento da acessibilidade não só é importante para os habitantes da cidade, mas também é essencial para estimular economias de aglomeração e crescimento econômico em longo prazo.

Outros estudos poderiam sugerir uma melhor solução para o problema de endogeneidade de acessibilidade, através da disponibilidade de bons instrumentos e da estimativa de equações estruturais. Embora na literatura ainda não haja consenso sobre a forma adequada de modelar essas equações estruturais, bem como as decisões de localização residencial e local de trabalho.

Referências

BARUFI, A. M. B.; HADDAD, E. A. Spatial Mismatch, Wages and Unemployment in Brazilian Metropolitan Areas. 2016.

BOISJOLY, G., MORENO-MONROY, A. I., & EL-GENEIDY, A. (2017). Informality and accessibility to jobs by public transit: Evidence from the São Paulo Metropolitan Region. *Journal of Transport Geography*, 64, 89-96.

CRANE, R. (2007) "Is There a Quiet Revolution in Women's Travel? Revisiting the Gender Gap in Commuting", *Journal of the American Planning Association*, Summer 2007, 73: 298-316.

DIAZ, M. D. M.; MACHADO, L. Overeducation e Undereducation no Brasil: Incidência e Retornos. *Estudos Econômicos*. São Paulo, v.38, n.3, p. 431-460, jul.-set. 2008.

DUJARDIN, C., SELOD, H. AND THOMAS, I. (2008) Residential segregation and unemployment: the case of Brussels, *Urban Studies*, 45, pp. 89–113.

FUJITA, M., & OGAWA, H. (1982). Multiple equilibria and structural transition of nonmonocentric urban configurations. *Regional Science and Urban Economics*, 12(2), 161-196.

- HADDAD, E. A; BARUFI, A. M. B. (2017). From rivers to roads: Spatial mismatch and inequality of opportunity in urban labor markets of a megacity. *Habitat International*, 68, 3-14.
- IBGE. Base de informações do Censo Demográfico 2010: Resultados do Universo por área de ponderação. Rio de Janeiro: IBGE, 2011.
- KAIN, J. (1968). Housing segregation, negro employment and metropolitan decentralization. *Quarterly Journal of Economics* 82: 175–197.
- LUCAS, R. E; ROSSI-HANSBERG, E. (2002). On the internal structure of cities. *Econometrica*, 70 (4), 1445-1476.
- MACHADO, A. F.; OLIVEIRA, A. M. H. C.; CARVALHO, N. F. Tipologia de qualificação da força de trabalho: uma proposta com base na noção de incompatibilidade entre ocupação e escolaridade. *Nova economia*, Belo Horizonte, v. 14, n. 2, p. 11-33, 2004.
- MCMILLEN, D. P. (2001). Nonparametric employment subcenter identification. *Journal of Urban economics*, 50(3), 448-473.
- MELO, P. C., & GRAHAM, D. J. (October 2009). Agglomeration economies and labour productivity: Evidence from longitudinal worker data for GB's travel-to-work areas. SERC Discussion Paper, No. 31.
- MORENOMONROY, A., RAMOS, F., 2015. The impact of public transport expansions on informality: The case of the São Paulo Metropolitan Region.
- MOTTE, B., AGUILERA, A., BONIN, O., NASSI, C.D., 2016. Commuting patterns in the metropolitan region of Rio de Janeiro. What differences between formal and informal jobs? *Journal of Transport Geography* 51, 59-69.
- NADALIN, Vanessa Gapriotti; IGLIORI, Danilo Camargo. *Evolução urbana e espraiamento na Região Metropolitana de São Paulo*. 2010.
- O'REGAN, K. M. QUIGLEY, J. M. (1996) Spatial effects upon employment outcomes: the case of New Jersey teenagers, *New England Economic Review*, May/June, pp. 41–58.
- RAMOS, Frederico Roman. Identificando subcentros de emprego na região metropolitana de São Paulo: potenciais locacionais e consequências para a estrutura urbana. *Anais*, 2016, 1-21.
- ROSSI-HANSBERG, E. (2004). Optimal urban land use and zoning. *Review of Economic Dynamics*, 7, 69-106.
- SANTOS, A. Overeducation no Mercado de trabalho brasileiro. *Revista Brasileira de Economia de Empresas*, v.2, n.2, p. 1-22. 2002.
- WHEELER, C. H. (2003). Evidence on agglomeration economies, diseconomies, and growth. *Journal of Applied Econometrics* 18: 79–104.