

# RESUMO EXECUTIVO

## Abertura de empresas por porte: Regressão espacial

### Sistema SEBRAE

Brasília - DF, 09 de Setembro de 2022





**Todos os direitos reservados.**

*A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação aos direitos autorais (Lei nº 9.610).*

**Serviço Brasileiro de Apoio às Micro e Pequenas Empresas – SEBRAE**

**Unidade de Gestão Estratégica e Inteligência**

SGAS 605 – Conjunto A – Asa Sul – Brasília/DF – CEP 70200-904

Tel.: 55 61 3348-7180

Site: <https://www.sebrae.com.br/>

**CONSELHO DELIBERATIVO NACIONAL**

**Presidente**

*José Roberto Tadros*

**DIRETORIA EXECUTIVA**

**Diretor-Presidente**

*Carlos do Carmo Andrade Melles*

**Diretor Técnico**

*Bruno Quick Lourenço de Lima*

**Diretor de Administração e Finanças**

*Eduardo Diogo*

**Gerente da Unidade de Gestão Estratégica e Inteligência**

*Adriane Ricieri Brito*

**Gerente Adjunto da Unidade de Gestão Estratégica e Inteligência**

*Fausto Ricardo Keske Casseiro*

**Coordenador do Núcleo de Pesquisa e Gestão do Conhecimento**

*Kennyston Costa Lago*

**Equipe Técnica**

*Tomaz Back Carrijo*

*Felipe Marcel Neves*

*Juliana Borges Vaz*

A regressão é um método estatístico usualmente utilizado para compreender ou prever um fenômeno numérico, tal como a abertura de empresas, possibilitando verificar quais fatores mais influenciaram sua ocorrência. Porém, métodos usuais de regressão não consideram o papel do espaço na influência das variáveis. De modo a compreender ainda mais os padrões espaciais da abertura de empresas de interesse do SEBRAE (Microempreendedor Individual - MEI, Microempresa - ME e Empresa de Pequeno Porte - EPP) entre os municípios brasileiros, foram analisados dados sobre a abertura de empresas através de métodos de regressão espacial em seu estado da arte.

Para tanto, dados sobre abertura de empresas por porte foram extraídos da base atualizada do Sebrae Nacional e Receita Federal. Os dados relativos às variáveis foram extraídos da base de dados do Índice Sebrae de Desenvolvimento Econômico Local - ISDEL (no estudo foram utilizadas 53 variáveis, oriundas do ano de 2019) e da Receita Federal (16 variáveis de faturamentos de CNAE, incluindo o valor do faturamento médio). As principais variáveis consideradas nos modelos estão na Tabela 1. Informações acessórias foram retiradas da base de dados pública do IBGE. As metodologias foram aplicadas utilizando o software python versão 3.9.12, e com os pacotes estatísticos apropriados.

Tabela 1 - Principais variáveis utilizadas no estudo.

Base de dados	Variável	Observação
ISDEL	Potencial de Consumo	-
	Densidade de telefonia móvel	-
	Densidade de banda larga	-
	Diversidade produtiva	-
	Emprego em economia criativa e turismo	-
	Serviços bancários	-
	Crédito <i>per capita</i>	-
	Emissão de gases <i>per capita</i>	-
	Complexidade econômica	-
	Depósitos de patentes	-
Receita Federal	Faturamento médio	Faturamento médio de todas as seções da CNAE
	Valor de faturamento - seção L	Valor de faturamento da seção Alojamento e alimentação
	Valor de faturamento - seção I	Valor de faturamento da seção Atividades imobiliárias

Fonte: ISDEL e Receita Federal.

Foram utilizados três modelos de regressão. O primeiro foi o modelo de regressão dos Mínimos Quadrados Ordinários - MQO, *Ordinary Least Squares* (OLS). Apesar de não considerar o papel do espaço, foi aplicado para determinar o número ótimo de combinação de variáveis (covariáveis). O segundo foi o de regressão espacial geograficamente ponderada (*Geographically Weighted Regression* - GWR), que consegue analisar fenômenos que sofrem variação conforme a área em que são estudados. O terceiro modelo é uma variação do GWR, chamada regressão multiescalar geograficamente ponderada (*Multiscale Geographically Weighted Regression* - MGWR). O MGWR é menos restritivo em suas suposições do

que o GWR, permitindo que uma covariável possa existir em diferentes escalas espaciais. Os modelos resultantes foram comparados por uma série de métricas para interpretar sua aplicabilidade.

## 1 RESULTADOS GERAIS

Os modelos que consideram a configuração espacial dos municípios obtiveram um desempenho superior em todas as métricas, em particular, o modelo GWR para MEI e o modelo MGWR para ME e EPP. Os resultados mostram que existem padrões espaciais que impactam na influência de variáveis na abertura de empresas.

Deste modo, a apresentação dos resultados será por porte, sendo a análise do modelo OLS primariamente contemplada, seguida dos resultados dos modelos espaciais GWR e MGWR (no caso, o que obteve melhor desempenho entre os dois). O  $R^2$  local será apresentado em mapas, de modo a verificar como o ajuste do modelo varia ao longo do território nacional (possui valor de 0 a 1, quanto maior, melhor). Em seguida, os resultados dos mapas de coeficientes locais, mostram a distribuição espacial do impacto das variáveis explicativas examinadas sobre a abertura de empresas.

### 1.1 MEI

O modelo OLS estimou que 14 variáveis possuem alguma significância estatística na predição da abertura de MEIs. O número condicional final foi de 4,43 não indicando multicolinearidade, e o I de Moran dos resíduos foi significativo, demonstrando dependência espacial dos dados.

A distribuição da taxa de aberturas de MEIs no Brasil ocorreu com maior intensidade na região Sul, Sudeste e Centro-Oeste do País, e com menor intensidade no Norte e Nordeste. A taxa de abertura de empresas entre os municípios variou de 0,38 a 90 empresas abertas por mil habitantes. O  $R^2$  local do modelo GWR se mostrou alto no Norte, Nordeste, e em um trecho do Sul do país. Porém, no Centro-Oeste, Sudeste e interior do Sul o ajuste do modelo foi pior (Figura 1).

Podemos destacar entre os principais resultado obtidos pelo modelo (Figura 2):

- Potencial de consumo possui influência positiva significativa na abertura de MEIs em quase todo o território nacional, com exceção do Sul do país (82% dos municípios),
- Densidade de banda larga fixa e móvel possuem forte influência positiva no Sul, mas com uma maior influência de banda larga fixa no Nordeste e telefonia móvel no Centro-Oeste.
- Diversidade produtiva, possuiu, em geral, influência negativa na região Nordeste e Sul (55% dos municípios).
- Emprego em economia criativa e turismo teve influência positiva em grandes regiões do Brasil, excluindo, em geral, Norte e Nordeste (55% dos municípios).
- O valor de faturamento da seção I (Alojamento e alimentação), teve dois núcleos grandes de influência positiva no Nordeste, sendo que um deles se estende até o Sudeste (25% dos municípios).

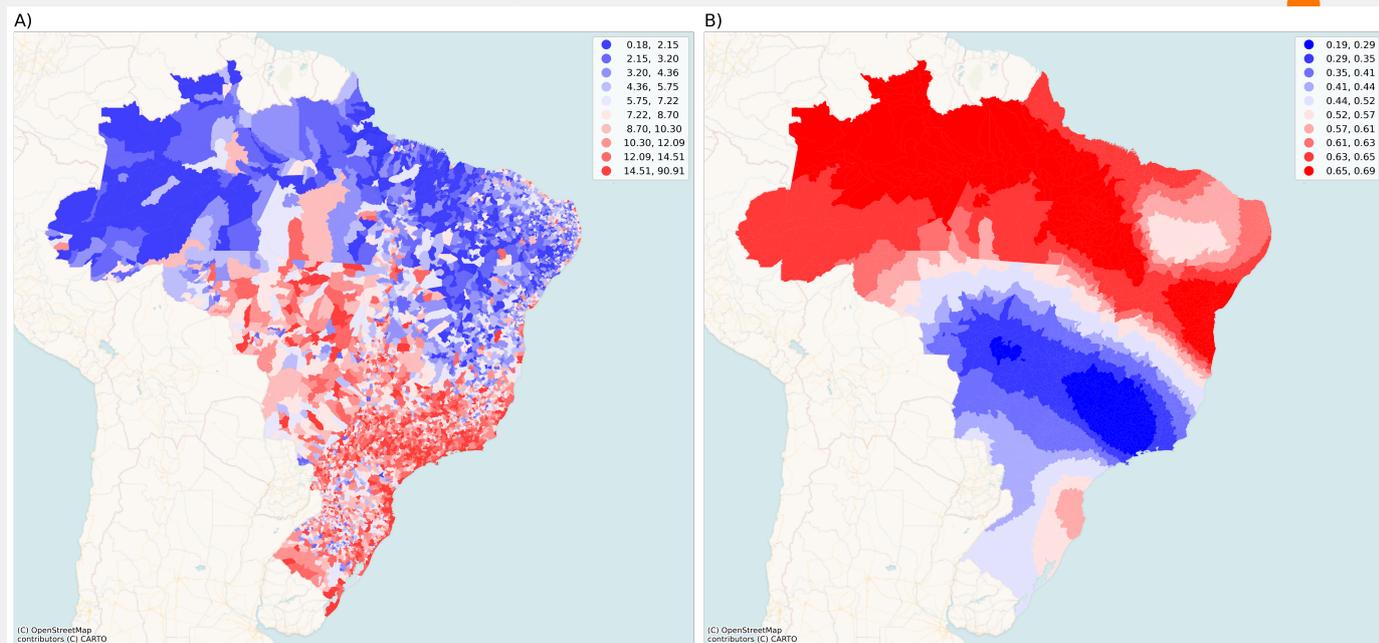
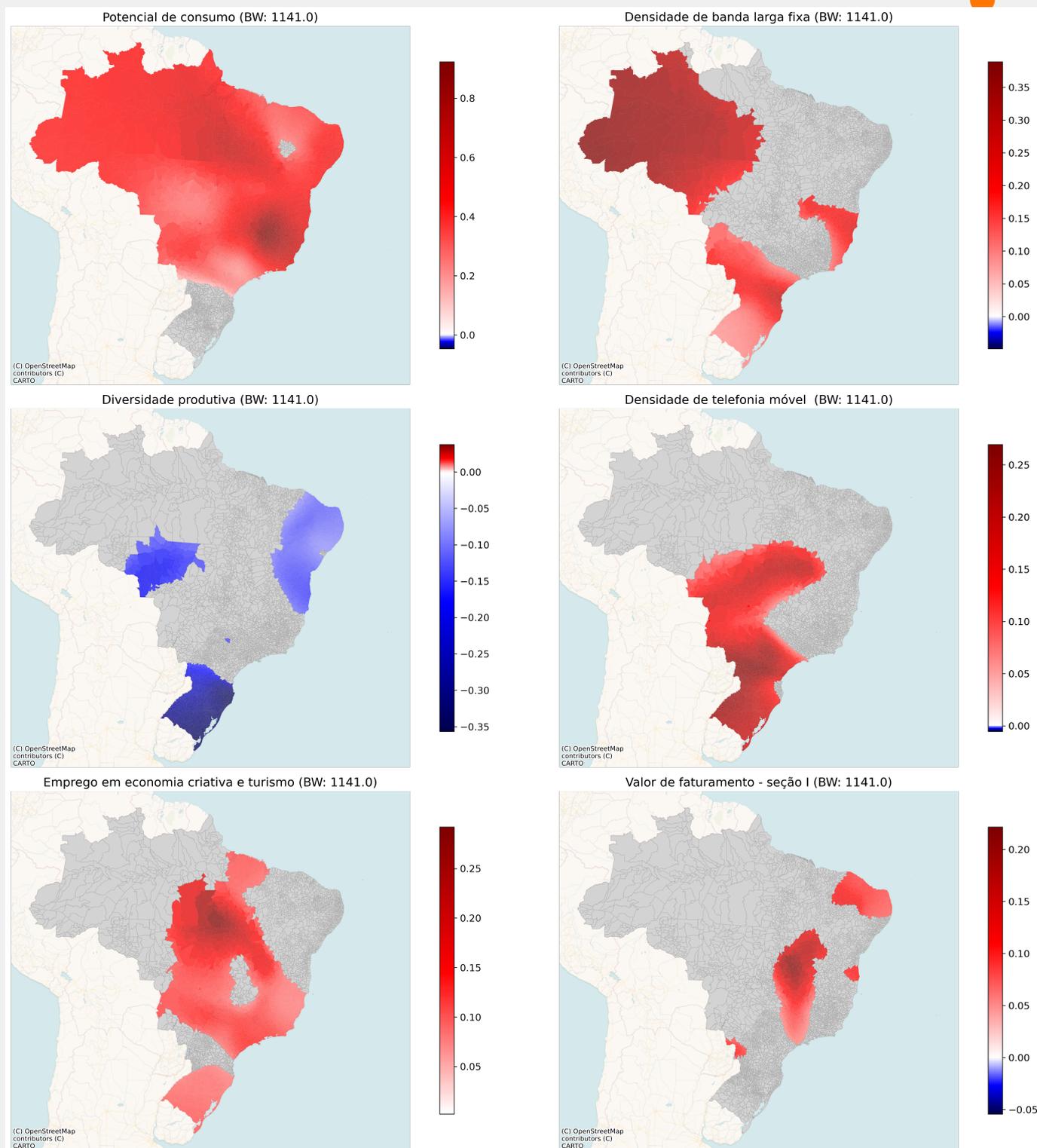


Figura 1 – Taxa de abertura de MEIs por município (A) e  $R^2$  local proveniente do modelo de regressão espacial GWR (B). A intensidade das cores varia dos valores mais baixos (coloração azul) até os mais altos (coloração vermelha).

Fonte: Elaborado pelos autores.



**Figura 2 – Mapas de coeficientes locais das variáveis mais importantes do modelo GWR para abertura de MEIs.** Coeficientes com valores positivos indicam influência positiva (cor vermelha), e negativos indicam influência negativa (cor azul). A cor cinza representa valores não significativos.  
 Fonte: Elaborado pelos autores.

## 1.2 ME

O modelo OLS estimou que 25 variáveis possuem alguma significância estatística na predição da abertura de MEs. Considerando os diagnósticos da regressão, o número condicional foi de 6,54 não indicando multicolinearidade, e o I de Moran dos resíduos foi significativo, também demonstrando dependência espacial dos dados.

A distribuição da taxa de aberturas de MEs no Brasil ocorreu similarmente a abertura de MEIs, com maior intensidade no Sul, Sudeste, e no caso dos MEs, com uma intensidade grande também no Centro-Oeste do País, e menor no Norte e Nordeste. A taxa de abertura de empresas entre os municípios variou de 0 a 21 empresas abertas por mil habitantes. O  $R^2$  local do modelo MGWR se mostrou alto no Norte, Nordeste (na maioria), e em um trecho do Sul do país, em especial no leste do estado do Paraná. Porém, apesar de visualmente se observar um melhor desempenho e homogeneização do ajuste comparado a abertura de MEIs, no Centro-Oeste, Sudeste e partes do Sul (Rio Grande do Sul e Santa Catarina) o ajuste do modelo aos dados obteve um desempenho pior (Figura 3).

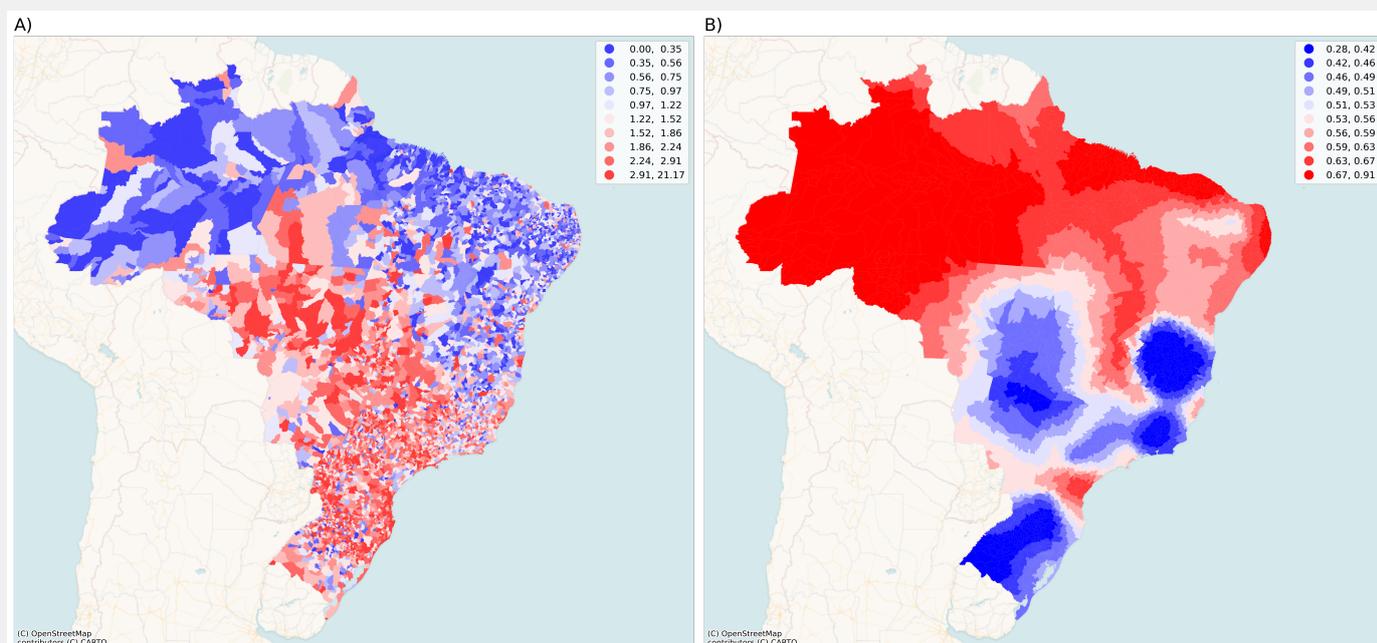


Figura 3 – Taxa de abertura de MEs por município (A) e  $R^2$  local proveniente do modelo de regressão espacial GWR (B). A intensidade das cores varia dos valores mais baixos (coloração azul) até os mais altos (coloração vermelha)

Fonte: Elaborado pelos autores.

Podemos destacar entre os principais resultado obtidos pelo modelo (Figura 4):

- ➔ Serviços bancários possuem influência negativa na abertura de MEs em vários municípios do território nacional (47% dos municípios brasileiros).

- ▶ Potencial de consumo possui influência positiva em todo território nacional. Crédito *per capita* teve expressiva influência no Norte, Centro-Oeste, Sul e Sudeste (26% dos municípios brasileiros).
- ▶ O valor de faturamento da seção L possui um padrão interessante, pois possui além de alguns municípios espalhados, pequenos agrupamentos concentrados (26% dos municípios brasileiros)
- ▶ Depósitos de patentes e emprego em economia criativa e turismo possuem, em geral, influência positiva, mas de maneira pulverizada (25% e 12% dos municípios brasileiros, respectivamente).
- ▶ O valor de faturamento da seção I (Alojamento e alimentação), teve dois núcleos grandes de influência positiva no Nordeste, sendo que um deles se estende até o Sudeste (25% dos municípios).

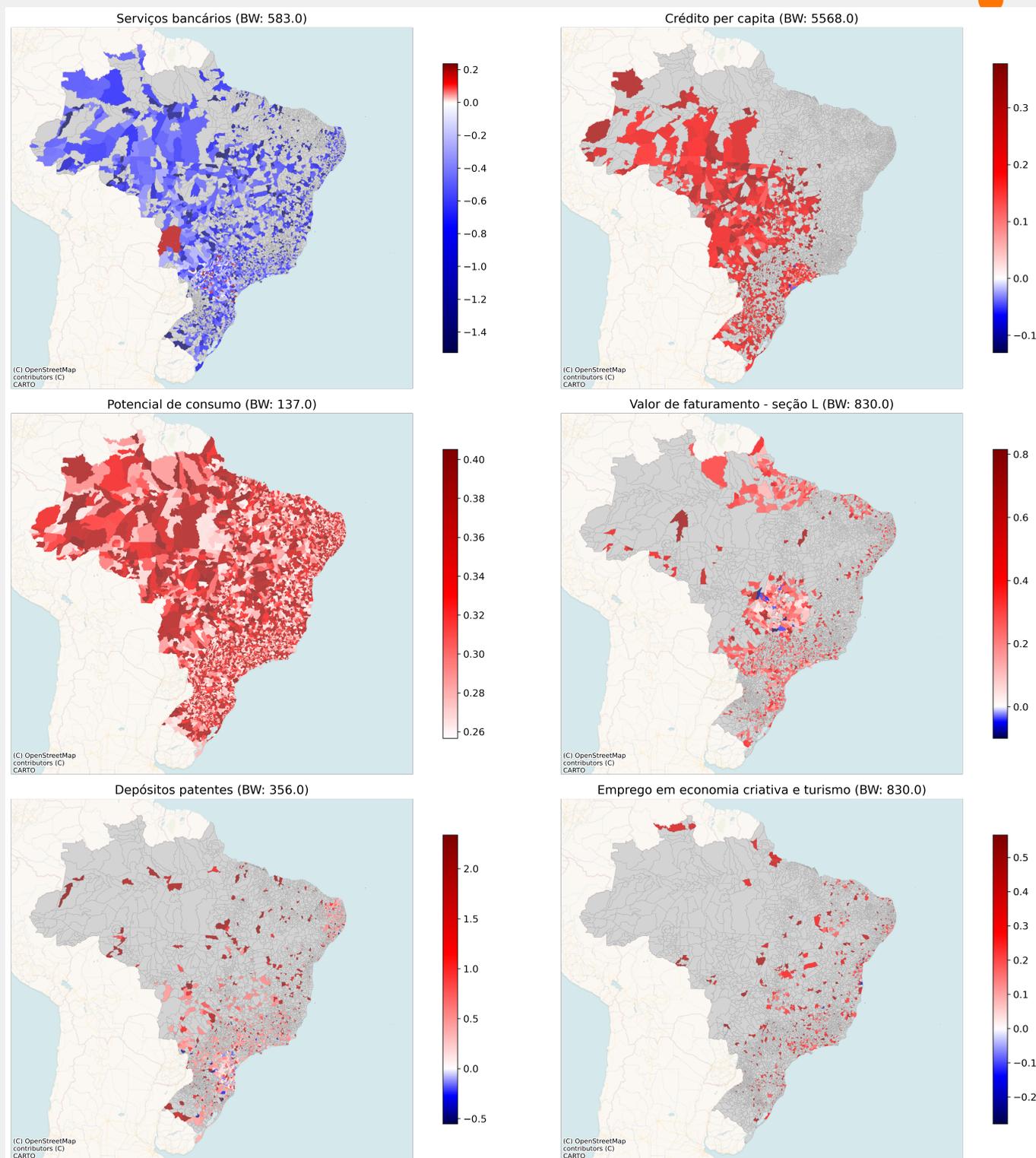


Figura 4 – Mapas de coeficientes locais das variáveis mais importantes do modelo MGWR para abertura de MEs. Coeficientes com valores positivos indicam influência positiva (cor vermelha), e negativos indicam influência negativa (cor azul). A cor cinza representa valores não significativos.

Fonte: Elaborado pelos autores.

### 1.3 EPP

O modelo OLS estimou que 14 variáveis possuem alguma significância estatística na predição da abertura de EPPs. O número condicional final foi de 4,80 não indicando multicolinearidade, e o I de Moran dos resíduos foi significativo, demonstrando dependência espacial dos dados.

A distribuição da taxa de aberturas de EPPs no Brasil ocorreu similarmente a abertura de MEs, com maior intensidade no Sul, Sudeste e Centro-Oeste do país, e com menor intensidade no Nordeste e em parte do Norte. A taxa de abertura de empresas entre os municípios variou de 0 a 2,64 empresas abertas por mil habitantes, o que, comparado aos outros portes, é muito pouco. O  $R^2$  local para EPP possui a pior desempenho, onde se mostrou alto no Norte, e em um trecho do Sul e Sudeste do país, em especial no leste. No Centro-Oeste, Sudeste, partes do Sul e Nordeste (em particular, no interior) o ajuste do modelo aos dados obteve um desempenho pior (Figura 5). No caso de EPPs, em especial, a frequência mais baixa de aberturas comparada aos outros portes, pode ser um fator que influencie na sua imprevisibilidade.

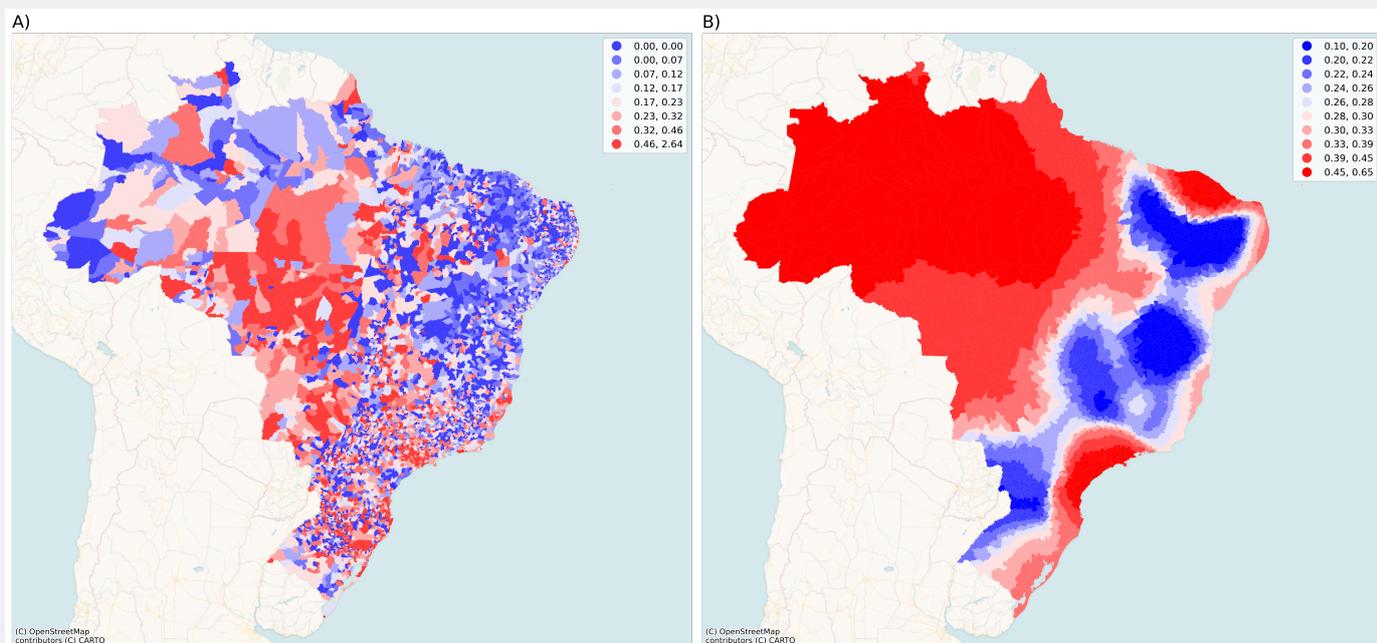


Figura 5 – Taxa de abertura de EPPs por município (A) e  $R^2$  local proveniente do modelo de regressão espacial GWR (B). A intensidade das cores varia dos valores mais baixos (coloração azul) até os mais altos (coloração vermelha).

Fonte: Elaborado pelos autores.

Podemos destacar entre os principais resultado obtidos pelo modelo (Figura 6):

- ➔ Complexidade econômica e faturamento médio possuem influência positiva em todo o território nacional.

- Valor de faturamento da seção L possui influência positiva e distribuição não aglomerada em todo o país, assim como potencial de consumo (55% e 38%, respectivamente).
- Emissão de gases *per capita* possui influencia positiva no Norte e Centro-Oeste, crédito *per capita* também, em menor território, mas em maior número de municípios (8% e 10%, respectivamente).

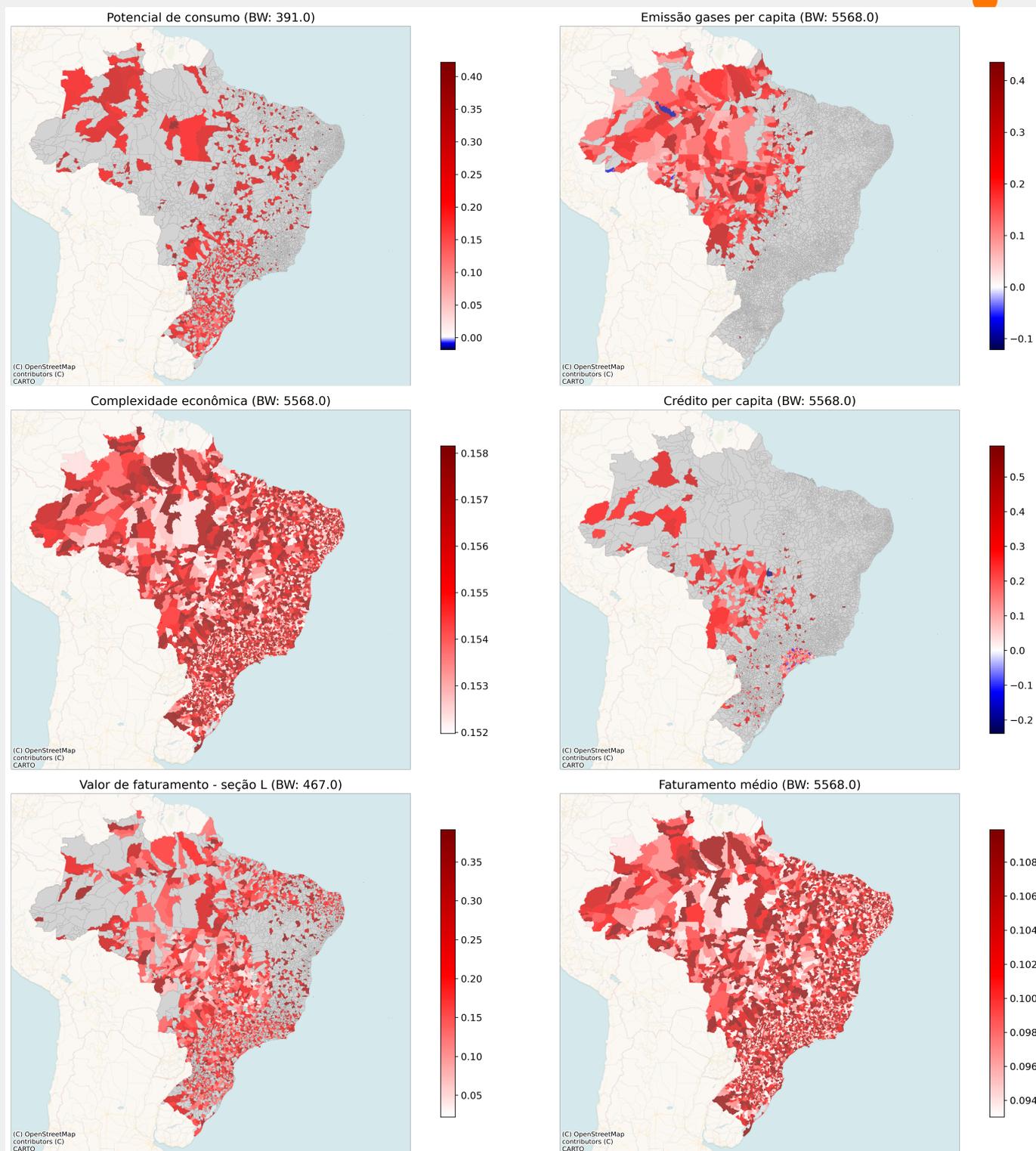


Figura 6 – Mapas de coeficientes locais das variáveis mais importantes do modelo MGWR para abertura de EPPs. Coeficientes com valores positivos indicam influência positiva (cor vermelha), e negativos indicam influência negativa (cor azul). A cor cinza representa valores não significativos.

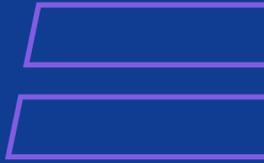
Fonte: Elaborado pelos autores.

## 2 CONSIDERAÇÕES FINAIS

Neste estudo foi possível observar que algumas variáveis influenciaram de maneira relativamente moderada a intensidade de abertura de empresas nos municípios brasileiros. Em particular, padrões em partes da região Sul, Sudeste e Centro-Oeste foram difíceis de serem explicados pelas variáveis presentes no modelo, como verificado pelo  $R^2$  local. Isto indica que estas regiões podem sofrer de influências externas desconhecidas, ou que possuam complexidades que acabam por justificar sua falta de previsibilidade.

Em relação aos portes, os modelos para as taxas de aberturas em EPPs tiveram pior desempenho, o que pode ser explicado pelo fato de EPPs possuírem uma taxa de abertura bem menor que os outros portes, que diminui sua previsibilidade. Dentre as variáveis mais importantes examinadas, a que teve influência positiva em todos os portes, e na maioria do território nacional, foi potencial de consumo. No entanto, de forma geral, a importância das variáveis dentro de cada porte foi divergente. Por exemplo, a variável de valor de faturamento da seção L (Atividades imobiliárias) foi importante somente para MEs e EPPs, o que pode ser explicado pelo fato de que, em geral, empreendimentos deste CNAE não se enquadram em MEI.

As variáveis tiveram distribuições espaciais dos coeficientes bem diversificadas, podendo representar uma ligação interessante em como a abertura de empresas está associada com o desenvolvimento de certos municípios e suas características socioeconômicas regionais.



**SEBRAE**

**50+50**

