

Uma proposta de redesenho do mapa das regiões do Brasil e as projeções de crescimento econômico dos novos aglomerados para os próximos anos

Greisson Almeida Pereira
Doutorando em Economia - UCB

Tito Belchior Silva Moreira
Professor Doutor - UCB

RESUMO

Este trabalho tem como objetivo redesenhar as grandes regiões do Brasil com base em aglomerações de municípios com características socioeconômicas homogêneas. A partir dos dados dos dois últimos censos, 2000 e 2010, selecionou-se uma matriz de 41 variáveis socioeconômicas para cada um dos municípios brasileiros. Em seguida, reduziu-se a dimensão dessa matriz utilizando-se o método de análise de componentes principais (ACP). As componentes principais foram utilizadas para compor o algoritmo de aglomeração hierárquica que definiram os aglomerados municipais. Os resultados da aglomeração indicaram a existência de dez grandes regiões cujos municípios brasileiros apresentam características semelhantes. A partir dessa nova divisão do Brasil, projetou-se a distribuição do crescimento econômico nacional nesses aglomerados, utilizando as técnicas de painel de dados e regressão linear por mínimos quadrados ordinários. Os resultados indicaram maior crescimento econômico nos aglomerados Centro II, Norte e Sul II.

Palavras-chave: aglomeração municipal, PIB municipal, projeção

ABSTRACT

This work aims to redesign the big regions of Brazil based on agglomerations of municipalities with homogeneous socioeconomic characteristics. Based on data from the last two censuses, 2000 and 2010, a matrix of 41 socioeconomic variables was selected for each of the Brazilian municipalities. Then, the size of this matrix was reduced using the Principal Component Analysis (PCA) method. The main components were used to compose the algorithm of hierarchical agglomeration that defined the municipal agglomerates. The results of the agglomeration indicated the existence of ten large regions whose Brazilian municipalities present similar characteristics. From this new division of Brazil, the distribution of the national economic growth along these agglomerations was projected, using data panel techniques and linear regression by ordinary least squares. The results indicated higher economic growth in the Central II, North and South II agglomerations.

Key-Words: agglomeration, municipal GDP, forecasting

1 Introdução

O Brasil é um país de dimensões continentais. Considerando sua divisão político-administrativa, se divide em cinco grandes regiões, 27 unidades federativas e 5.570 municípios (IBGE, 2016a). Essa representação, baseada na junção legal dos territórios, pode não fornecer um nível de agregação ou desagregação que reflita a homogeneidade socioeconômica dos municípios que compõe cada uma dessas cinco regiões. Conforme salienta Lemos et al (2003), a junção atual dos municípios em regiões político-administrativas não corresponde à delimitação da aglomeração das economias urbanas e seus centros de polarização.

Ao longo de sua história, o País foi se desenvolvendo sócio e economicamente de forma dinâmica. Os atores econômicos foram se aglomerando em diversos grupos, que os identifica enquanto partes de um conjunto. Esses grupos refletem aglomerações de municípios que guardam semelhanças entre si, sejam elas perceptíveis ou não. A proposta de um novo desenho para as grandes regiões do Brasil agrega municípios vizinhos, homogêneos entre si, e que representam, com maior nitidez, a localidade em que estão inseridos. Portanto, o redesenho do Brasil não respeita, necessariamente, as divisões político-administrativas dos estados em que estejam inseridos. Doravante, chamaremos essas novas regiões de aglomerados.

De acordo com Marshall (1920), as firmas estão sempre dispostas a recorrer a qualquer lugar onde possam encontrar mão-de-obra especializada. Da mesma forma, os trabalhadores especializados vão procurar se manter em mercados onde há demanda para a oferta de trabalho. Assim, para o autor, a localização dos setores se dá em locais onde as forças sociais cooperam com as econômicas.

Além do custo do fator trabalho, há outros determinantes que vão além das fronteiras produtivas, como o custo de transporte de bens, de pessoas e de informações. Nessa perspectiva, de acordo com Claval (2005), surge o desenvolvimento da teoria das migrações humanas que contribui para a compreensão de situações onde se busca rendas mais altas ou a necessidade de se consumir em lugares diferentes. **O papel da geografia econômica no contexto atual da teoria econômica é bem maior que no passado. O espaço influencia o consumo, a produção e o fluxo de pessoas, de capitais e de bens.**

Para Porter (2000), a geografia econômica durante a denominada era global envolve um paradoxo. De um lado, temos o reconhecimento de que mudanças tecnológicas e de modelos de competição têm reduzido as regras tradicionais de localização. Do outro, novas influências da formação de *clusters* na competição têm ganhado importância na dinâmica econômica, sobretudo no papel da localização nas vantagens competitivas. Os *clusters*, segundo o autor, representam uma nova reflexão sobre o pensamento nacional, o estado e a economia local, além de exigirem novos papéis para empresas, governo e outras instituições, no sentido de melhorarem a competitividade. São concentrações geográficas interconectadas, compostas por empresas, fornecedores, provedores de serviços, universidades e associações comerciais, nas quais há competição e cooperação ao mesmo tempo. Esses agentes são ligados por sua homogeneidade e complementariedade. A competição, muitas vezes, não ocorre de forma direta e muitos participantes de *clusters* servem diferentes setores.

De acordo com Fujita, Krugman e Venables (2002 p. 15), aglomeração é “o agrupamento da atividade econômica, criada e sustentada por alguma lógica circular”. Ela pode ocorrer em muitos níveis, de regiões comerciais locais, que servem as áreas residenciais das proximidades das cidades, a regiões econômicas especializadas como o Vale do Silício que servem o mercado mundial como um todo. A geografia econômica compreende o estudo de onde a atividade econômica ocorre e por que deve ser levado em consideração nas análises econômicas. Entretanto, segundo os autores, no passado recente, esse assunto era negligenciado pelos economistas. A irregularidade espacial da economia real, como por exemplo, disparidades entre regiões industriais densamente populosas e regiões agrícolas com menos densidade demográfica, não é resultado das diferenças entre esses locais, mas de processos cumulativos, envolvendo algum tipo de retorno crescente, por meio do qual a concentração geográfica pode se auto reforçar, ou seja, cria um ambiente economicamente favorável que sustenta uma concentração ainda maior ou continuada.

Segundo Borowiecki (2013), tanto os indivíduos mais produtivos quanto as firmas mais produtivas se localizam em regiões de *clusters* geográficos, nos quais ele considera como aglomerações urbanas. O autor questiona se os *clusters* atraem esses indivíduos e firmas ou se as externalidades geradas pelos *clusters* os tornam mais produtivos. Se essa relação causal não for conhecida, as intervenções de políticas econômicas podem provocar distorções ao mercado. Ao analisar três aglomerações urbanas: Paris, Londres e Viena, composto por músicos clássicos nascidos entre 1750 e 1899, o autor encontrou evidências de que os *clusters* causaram aumento na produtividade dos músicos. Além disso, analisando a história de vida dos compositores, verificou-se que a aglomeração causou externalidades positivas persistentes ao longo do tempo. Para Glaeser e Mare (2001), apud Borowiecki (2013), nos locais com alta concentração de agentes, as evidências mostram que os trabalhadores têm maiores salários. No caso das firmas, segundo Nickell (1996), apud Borowiecki (2013), suas performances são melhores em regiões de *cluster*.

No estudo publicado pelo Ministério do Planejamento, Orçamento e Gestão (MPOG) (BRASIL, 2008), é ressaltado que o fluxo de capital, produção e comércio de uma região se estende além de sua fronteira político-administrativa, formando aglomerados que não respeitam essa rigidez espacial. Com isso, segundo aquele órgão, torna-se necessário discutir e propor uma nova regionalização para o Brasil, na qual poderá servir de base para uma atuação convergente dos vários entes federados e sua articulação com as iniciativas mais amplas da sociedade. O MPOG dividiu o país em seis grandes regiões, fundamentado em indicadores socioeconômicos, concentração da população no litoral, separação do bioma da floresta amazônica e da agroindústria do Nordeste. O foco do estudo do MPOG foi construir uma visão estratégica nacional para o horizonte de 2027 que orientasse uma proposta de organização territorial. Embora a pesquisa não delimitasse o agrupamento por municípios, colocando apenas uma linha imaginário, serviu de inspiração para a elaboração dos aglomerados neste trabalho.

Pesquisas sobre aglomeração municipal foram realizadas por alguns autores no Brasil. No entanto, o objetivo até então estudado era o de definir estratégias e dar subsídios para a implementação de políticas públicas, não se preocupando em redesenhar grandes regiões homogêneas. Chein, Lemos e Assunção (2007), dividiram o

país em 400 *clusters* homogêneos, não respeitando a limitação político-administrativa das regiões e estados, com o objetivo de buscar sinais de convergência ou agravamento de desigualdades regionais. Carvalho, Da Mata e Resende (2008), agruparam municípios com características socioeconômicas semelhantes, respeitando os limites territoriais de cada região, ou seja, mantendo o padrão das cinco grandes regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sul e Sudeste, com o objetivo de fornecer subsídios de política espacial e regional.

Nesse sentido, o presente trabalho tem como objetivo redesenhar novas grandes regiões para o Brasil, no qual chamaremos de aglomerados, agrupando municípios contíguos e que sejam homogêneos nas suas características socioeconômicas. Adicionalmente, será estimada a distribuição do crescimento do PIB nacional nesses aglomerados para os próximos anos.

2 Referencial Teórico

De acordo com Porter (2000), as interconexões e as externalidades dentro de um cluster são mais importantes para o crescimento da produtividade do que o ganho de escala individual das firmas. Para ele, os governos, em suas várias instâncias, deveriam incentivar as vantagens competitivas de cada cluster e sua especialização. Os estados, as regiões metropolitanas e os municípios têm influências importantes nos negócios locais. Um cluster pode fazer com que empresas, governos e instituições locais ajam de maneira conjunta na construção de um diálogo, oferecendo um novo mecanismo para a colaboração entre mercado e governo, fazendo com que o todo seja melhor do que a soma das partes.

Para Monastério e Cavalcante (2011), apesar da importância do trabalho de Marshall (1920) para a literatura econômica regional, houve mais influência de Keynes e Schumpeter sobre o que eles convencionaram chamar de “teorias do desenvolvimento regional com ênfase nos fatores de aglomeração”. Para facilitar o entendimento dos principais teóricos regionais, o Quadro 1, baseado em Monastério e Cavalcante (2011), resume os principais autores e suas contribuições para a literatura regional e de aglomerações.

Para os estudiosos da Teoria Clássica da Localização, os fatores mais preponderantes na aglomeração eram os custos, tanto de transporte, como de produção e de pessoas. Esta pesquisa se baseia na Teoria do Desenvolvimento Regional com ênfase nos fatores de aglomeração, sobretudo no que se refere às ideias de Perroux (1995), na qual menciona que o crescimento não ocorre de forma homogênea no espaço e de Myrdall (1957), na qual afirma que quando uma região se desenvolve mais que a outra pode ter efeitos de retroação e de difusão. Perroux (1955) cita quatro formas de polarização por meio do qual as indústrias motrizes impulsionariam o desenvolvimento regional: i) técnica, referente aos efeitos de encadeamento entre a indústria motriz e outras firmas; ii) econômica, dada pela geração de emprego e renda decorrente da implantação da indústria motriz; iii) psicológica, movida pelo otimismo decorrido do sucesso da indústria motriz que impulsiona novos investimentos, e; iv) geográfica, relativa aos impactos nos sistemas urbanos do desenvolvimento da cidade onde se localiza a indústria.

Quadro 1 – Autores do século XIX e XX que discutiram a distribuição espacial da economia

Teoria	Autor	Contribuição
Teoria Clássica da Localização	Von Thünen (1826)	Elaborou o que ficou conhecido como anéis de Von Thünen, no qual o padrão de ocupação no espaço se dava em relação aos custos de transporte de cada mercadoria. Produtos com custo maior de transporte se localizam nas áreas próximas ao centro da localidade e aqueles com custo menor se localizam nas áreas periféricas.
	Alfred Weber (1909)	A contribuição de Weber à literatura ficou conhecida como triângulos de Weber. O autor argumenta que a decisão quanto à localização de atividades industriais decorreria da ponderação de três fatores: o custo de transporte, o custo da mão de obra e um "fator local" que seria decorrente das forças de aglomeração e desaglomeração.
	Christaller (1933)	Argumenta que o tratamento de uma região sobre a forma de círculos, apesar de reduzir a distância, deixa de fora algumas áreas. Se as firmas aproximam entre si para atender aos mercados, o que existe é uma forma linear e o autor apresenta uma formação de aglomeração que ficou conhecida como hexágonos de Christaller. O tamanho de cada área de mercado (hexágono) vai depender dos custos de transporte e elasticidades-preço do produto em questão.
	Lösch (1940)	Apresenta modelo com microfundamentos, semelhante ao demonstrado por Christaller. Tem-se um espaço de consumo, diferentemente de Christaller que apresentava um espaço de produção. Ficou conhecido como cones de Lösh.
	Isard (1956)	Primeira publicação em inglês das teorias da localização. Até Lösh eram em alemão. O autor propõe uma linha de pensamento que se convencionou chamar de " <i>regional science</i> ".
Teoria do Desenvolvimento Regional com ênfase nos fatores de aglomeração.	Marshall (1890)	Os benefícios resultantes da aglomeração seriam a possibilidade oferecida por um mercado local de viabilizar a existência de ganhos de escala, as vantagens decorrentes de uma oferta abundante de mão de obra e a troca de informações que ocorre quando empresas e trabalhadores situam-se próximos uns dos outros.
	Perroux (1955)	O crescimento não ocorre de forma homogênea no espaço, mas pode se manifestar em polos de crescimento, com variações de intensidade e que se expandem por diversos canais provocando efeitos variáveis em toda a economia.
	North (1955)	Dando ênfase à atividade exportadora das regiões, contesta a visão de que o desenvolvimento regional teria ocorrido em etapas sucessivas. Ele descreve o desenvolvimento regional a partir do surgimento de uma atividade de exportação baseada em fatores locais específicos.
	Myrdal (1957)	Causação circular e acumulativa. Quando uma região se desenvolve mais do que a outra, pode haver dois efeitos: i) retroação, uma região se torna mais atrativa e desenvolvida que outra, aumentando a desigualdade regional, e; ii) difusão, regiões centrífugas, levariam ao transbordamento do impulso de desenvolvimento para regiões atrasadas.
	Hirschman (1958)	Discute a questão regional usando os conceitos de efeitos para frente (oferta de insumos, tornando viáveis os setores que se posicionassem a jusante) e efeitos para trás (economias de escala necessárias à viabilização de empreendimentos).

Fonte: MONASTÉRIO e CAVALCANTI, 2011

Segundo Thisse (2011), há um consenso de que a economia espacial pode ser considerada como resultado de um *trade-off* entre diferentes tipos de economias de escala de produção e o custo da mobilidade de bens e de pessoas. Tal fato sugere que a localização das atividades econômicas é consequência de um complexo equilíbrio de forças que puxam e empurram consumidores e empresas em direções opostas. Nesse sentido, Fujita, Krugman e Venables (2002 p. 15) afirmam que as disparidades regionais são guiadas por uma "bola de neve", resultando em um esforço contínuo. Ele cita como exemplo a indústria manufatureira, que tende a se concentrar nos locais onde existe mercado para seus produtos, mas o mercado somente será grande o suficiente onde a indústria manufatureira for concentrada.

Para Perroux (1955), apud Jesus e Spínola (2015), o crescimento econômico não acontece em toda parte dentro de uma região. Ao contrário, pode manifestar-se em

pontos ou polos de crescimento, com intensidades variáveis, expandir-se por diversos canais e com efeitos finais sobre toda a economia. O processo de crescimento econômico não implica em equilíbrio como preconizava os economistas clássicos e neoclássicos. Na visão dos autores, o crescimento econômico pode ser um processo desequilibrado por natureza. O crescimento econômico não se faz de forma difusa no espaço, mas se manifesta em certos pontos chamados de polos de crescimento com intensidades variáveis, daí se propagando por diversos canais com efeitos terminais distintos. O polo de crescimento ocorre pelo surgimento de uma indústria motriz, sendo esta a indústria que realiza a separação dos fatores da produção, provoca a concentração de capitais e decompõe as tarefas.

De acordo com Lemos et al (2003 p. 666), *“uma regionalização para fins de planejamento e de políticas públicas deveria também incluir a análise da representação política, dada a divisão político-administrativa prévia do território brasileiro”*. Os autores propuseram uma nova regionalização para o Brasil, procurando delimitar polos econômicos e suas áreas de influência, utilizando modelo gravitacional. Identificaram a existência de onze áreas de polarização capitaneadas por Brasília, Goiânia, Manaus, Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre.

Desrochers e Sautet (2004), ao citar Porter (2000), menciona que os clusters surgem de forma independente, sem ação do governo, apesar de justificar uma política de intervenção. Entretanto, embora o governo deva estimular a criação de um ambiente propício para o surgimento dos clusters, ele deveria evitar a tentação de estimular um cluster somente porque pode ser a semente de um grande aglomerado. O governo deve estimular a formação de aglomerações através da redução de barreiras à inovação, investindo em capital humano, infraestrutura e dando suporte às firmas localizadas em aglomerações.

De acordo com Hayek (1948), apud Desrochers e Sautet (2004), a principal razão para a persistência da aglomeração geográfica tem muito a ver com a geração e transmissão do conhecimento tácito, que está relacionado às circunstâncias particulares do espaço e do tempo. No meio industrial, o principal conhecimento não é aquele que pode ser formalizado, mas aquele que está inserido no capital humano e não é explícito, é adquirido pela experiência e passado àqueles ao redor. Os clusters são constituídos através de independências complexas que trabalham dentro, fora e na borda do cluster.

Khanna (2016) considera que as técnicas de desenvolvimento de mapas evoluíram bastante. No entanto, o autor adverte que ainda há espaço de conhecimento para preencher esses mapas. O autor afirma que *“...nós nunca podemos conhecer o mundo sem um mapa, nem o representar em apenas um.”* (Khanna, 2016 p. xxi). Acrescenta ainda que estamos nos movendo para uma era em que as cidades e suas aglomerações serão mais importantes e que o poder da oferta de suprimentos será mais do que o poder militar de um estado, sendo mais importante proteger uma rede de suprimentos do que simplesmente uma fronteira. Para ele, as fronteiras existem para delimitar cidades, estados e países. Entretanto, o que se vê, em muitos casos, são aglomerações formadas em regiões de fronteira, envolvendo mais de um território. Um exemplo disso são as conurbações metropolitanas, quase imperceptível de se distinguir a fronteira, a não ser pelas placas indicativas. Khanna (2016) salienta que nesses casos,

é irônico que os mapas deem mais importância às fronteiras políticas em detrimento aos limites demográficos e econômicos. Redesenhar os mapas com base nessas características é uma forma de tangibilizar fronteiras naturais face ao que é determinado politicamente. Assim, nessa linha de pensamento é que se baseia este trabalho, aglomerar municípios semelhantes dentro do território brasileiro.

3 Metodologia

3.1 Base de dados para análise dos aglomerados

Para identificação dos aglomerados compostos por municípios com características socioeconômicas semelhantes, selecionou-se 41 variáveis referentes aos 5.507 municípios existentes em 2000 e aos 5.565 municípios existentes em 2010. Conforme demonstrado na Tabela 1, as variáveis foram reunidas em oito grupos: demográfico e social, mercado de trabalho, econômico, produção agropecuária e extrativa vegetal, setor externo, setor bancário, serviços básicos e desempenho educacional. Como o menor nível de agregação de variáveis se dá em âmbito municipal, optou-se por utilizar os dados dos dois últimos censos (2000 e 2010), nos quais grande parte das variáveis que possam caracterizar estruturalmente os municípios estão disponíveis. Todas as variáveis foram trabalhadas em seu nível per capita.

3.2 Análise de componentes principais

A análise de componentes principais (ACP) é uma técnica estatística que permite reduzir o número de variáveis observadas, transformando-as em novas variáveis consideradas hipotéticas. De acordo com Pestana e Gageiro (1998), a análise fatorial pressupõe a redução de um número menor de variáveis observadas subjacentes aos dados, que expressam o que existe de comum nas variáveis originais. A diferença principal entre análise fatorial e ACP é que na primeira é considerada a variância comum entre as variáveis e na segunda é considerada a análise da variância total.

A análise fatorial é realizada com base nas seguintes etapas: i) padronizar as variáveis; ii) calcular os autovalores dos fatores como critério para extrair o número de fatores ou componentes, selecionando componentes e fatores que apresentam raízes latentes maiores que 1; iii) obter a matriz de fatores que contém as cargas fatoriais; iv) verificar a rotação da matriz de fatores que tem o objetivo de alcançar o critério da estrutura simples, na qual cada linha da matriz fatorial deve ter, pelo menos, um zero (ou carga fatorial próxima de zero).

Na ACP, a rotação de fatores tem como objetivo principal a transformação dos coeficientes desses componentes em uma estrutura simplificada. De acordo com Corrar, Paulo e Dias Filho (2007), apud Fávero et. Al (2009), a rotação dos fatores é possível dado que as cargas fatoriais podem ser representadas como pontos entre os eixos que, nesse caso, são os próprios fatores. Esses eixos podem ser girados sem alterar a distância entre os pontos. Todavia, as coordenadas do ponto em relação aos eixos são alteradas, ou seja, as cargas fatoriais são alteradas na rotação.

Os métodos de rotação podem ser os ortogonais ou os oblíquos. Nos ortogonais, possuem fatores que não estão correlacionados entre si, chamados de fatores ortogonais, sendo interpretados a partir de suas cargas. Na rotação oblíqua, por outro lado, os fatores estão correlacionados e, para interpretação da solução, torna-se necessária a consideração simultânea das correlações e das cargas. Entre os métodos rotacionais ortogonais tradicionais, merecem destaque o Varimax, o Quartimax e o Equamax. Nesse trabalho, utilizou-se o método mais usual, o Varimax, que busca maximizar o número de variáveis que tem altas cargas em um fator, simplificando a interpretação dos fatores.

Tabela 1 – Variáveis utilizadas na clusterização divididas em grupos

Grupo	Descrição das variáveis	Fonte
Demográfico e Social	Percentual da população do município residente na zona urbana; percentual da população do município residente na zona rural; razão de dependência; índice de mortalidade infantil; probabilidade de sobrevivência até 60 anos; percentual de pobres; desigualdade de renda (Índice de Gini); expectativa de vida ao nascer (longevidade) e taxa de fecundidade.	IBGE (Censos 2000 e 2010)
Mercado de Trabalho	Percentual da PEA; taxa de ocupação da população acima de 10 anos; percentual de pessoas ocupadas no setor agropecuário; percentual de pessoas ocupadas no comércio; percentual de pessoas ocupadas na construção civil; percentual de pessoas ocupadas no setor de serviços; percentual de pessoas ocupadas na indústria da transformação; taxa de emprego (número de ocupados / população total); percentual da renda proveniente dos rendimentos do trabalho.	IBGE (Censos 2000 e 2010)
Econômico	Renda per capita das famílias (considera os rendimentos oriundos do trabalho e transferências governamentais); PIB per capita do setor agropecuário; PIB per capita da indústria; PIB per capita dos serviços e da administração pública e impostos indiretos líquidos de subsídios per capita.	IBGE (Contas regionais 2000 e 2010)
Produção agropecuária e extrativa vegetal	Valor da produção pecuária per capita; valor da produção agrícola per capita; valor da produção extrativa vegetal per capita e valor da produção da silvicultura per capita.	IBGE (PAM, PPM e PEVS 2000 e 2010)
Setor Externo	Valor das exportações em US\$ milhões per capita e valor das importações em US\$ milhões per capita	MDIC 2000 e 2010
Bancário	Valor das captações totais per capita e valor do crédito total per capita.	Bacen (Estban 2000 e 2010)
Serviços básicos	Percentual da população residente em domicílios com banheiro e água encanada; percentual da população residente em domicílios com coleta de lixo; percentual da população residente em domicílios com energia elétrica; percentual de pessoas residentes em domicílios com abastecimento de água e esgotamento sanitário inadequados.	IBGE (Censos 2000 e 2010)
Desempenho educacional	Percentual de crianças entre 5 a 6 anos na escola; percentual de crianças de 6 a 14 anos no ensino fundamental sem atraso; percentual de crianças de 15 a 17 anos no ensino médio sem atraso; taxa de frequência bruta; expectativa de anos de estudo e taxa de analfabetismo das pessoas com 25 anos ou mais.	IBGE (Censos 2000 e 2010)

Elaborado pelos autores.

Dentre os municípios existentes no Brasil em 2000, identificou-se 12 outliers¹ que foram retirados da amostra. Segundo Fávero et al (2009), a análise de conglomerados é muito sensível a observações consideradas atípicas. Assim, a amostra levou em consideração 5.495 municípios. Da mesma forma, foram identificados 13 outliers² para o ano de 2010, restando uma amostra com 5.552 municípios.

Após a retirada dos outliers, procedeu-se à análise de significância das variáveis quanto à matriz de correlação. Primeiramente, retirou-se a variável “percentual da população residente na zona rural”, por apresentar colinearidade perfeita com o “percentual de população residente na zona urbana”. Em seguida, foram retiradas as variáveis “valor da produção extrativa vegetal per capita” e “valor da produção da silvicultura per capita”, nos dois períodos, e a variável “percentual de ocupação na indústria extrativa mineral” para o ano de 2000, por não apresentar significância ao nível de 10%.

A ACP foi realizada para 37 variáveis no ano 2000 e 38 variáveis em 2010, com rotação ortogonal Varimax para uma amostra de 5.495 municípios em 2000 e 5.552 municípios em 2010. Conforme Tabela 3, o teste de normalidade Kolmogorov-Smirnov, indicado para grandes amostras, foi efetuado para todas as variáveis, que apresentaram distribuição normal. Apesar de a normalidade univariada não garantir a normalidade multivariada, de acordo com Hair, Anderson, Tatham e Black (2005), apud Fávero et al (2009 p. 247), “se todas as variáveis atendem essa condição, então quaisquer desvios da normalidade multivariada geralmente são inócuos”. Além disso, conforme Tabela 2, verifica-se conformidade de adequação da amostra de acordo com a medida de Kaiser-Meyer-Olkin e o teste de esfericidade de Bartlett, para os dois períodos.

Tabela 3 – Dados de adequação da amostra utilizada para ACP

Descrição	2000	2010
Número de municípios da amostra	5.507	5.565
Número de <i>outliers</i>	12	13
Número de municípios utilizados na ACP	5.495	5.552
Número total de variáveis	41	41
Número de variáveis não significativas na matriz correlação	4	3
Número de variáveis consideradas na análise ACP	37	38
Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)	0.876	0.861
Bartlett's Test of Sphericity		
Aproximação Chi-Square	258085.526	239774.697
Df	780	820
Significância	0.0000	0.0000
Critério Kaiser autovalor maior que um		
Número de componentes	6	9
Percentual de explicação da variância	72,5%	73,9%

¹ São Paulo (SP), São Sebastião (SP), São Francisco do Conde (BA), Anchieta (ES), Triunfo (RS), Barcarena (PA), Jaguariúna (SP), Osasco (SP), Fortaleza de Minas (MG), Alumínio (SP), Belo Oriente (MG) e Brasília (DF)

² Itabira (MG), Angra dos Reis (RJ), São Sebastião (SP), Sebastianópolis do Sul (SP), Triunfo (RS), Albertina (MG), Anchieta (ES), Porto Real (RJ), Gavião Peixoto (SP), Osasco (SP), Garruchos (RS), São Paulo (SP) e Alto Horizonte (GO).

Para o ano de 2000, a análise mostrou que seis componentes foram adequadas ao critério de Kaiser do autovalor (“eigenvalue”) maior que um e explicaram 72,5% da variância. O *screen plot* mostrou que os seis componentes estão posicionados antes da inflexão. Considerando o tamanho da amostra e a convergência entre o *screen plot* e o critério de Kaiser, este foi o número de componentes mantido na análise final. Para 2010, o mesmo critério de análise mostrou que nove componentes foram adequadas ao critério de Kaiser do autovalor (“eigenvalue”) maior que um e 73,9% da variância, com nove componentes posicionados antes da inflexão. Pelo mesmo critério de seleção para 2000, esse também foi o número de componentes mantido na análise final.

3.3 Análise de Aglomerados

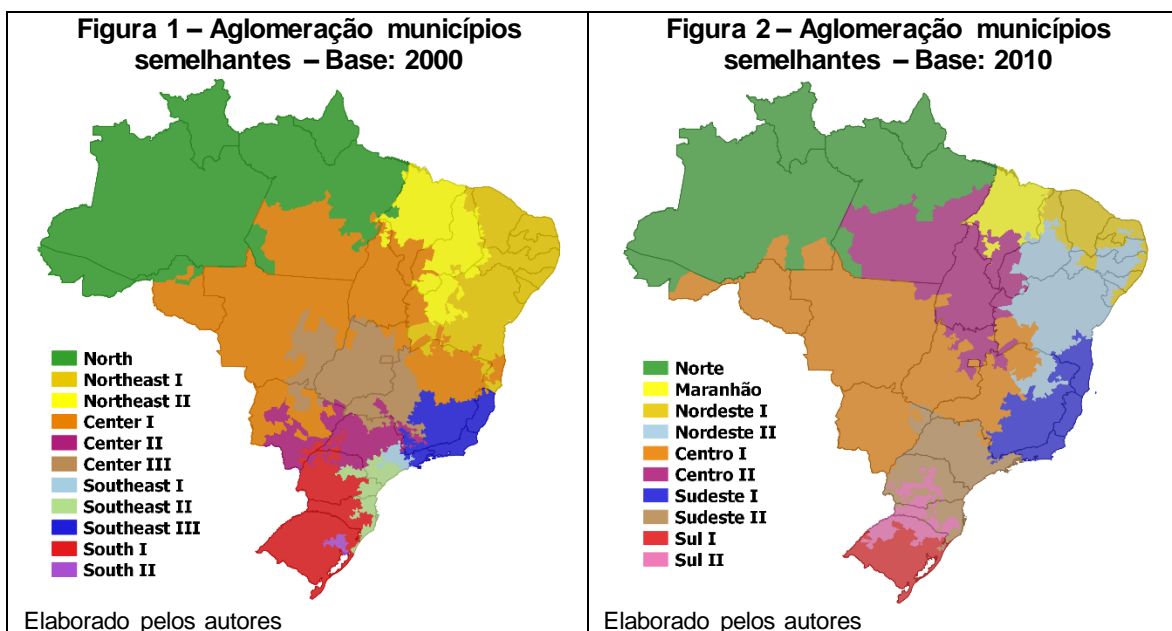
Para construção dos grupos de municípios que darão forma aos aglomerados, utilizou-se o software IpeaGEO, versão 2.1, desenvolvido pelos pesquisadores do Instituto de Pesquisa Econômica Aplicada (IPEA). O algoritmo, desenvolvido por Carvalho, Da Mata e Resende (2008), replicado neste estudo, pode ser classificado como um algoritmo combinatório e possui uma estrutura de formação do tipo hierárquica. Trata-se de uma modificação dos algoritmos de clusterização hierárquica comumente expostos na literatura, com restrições de contiguidade impostas. A aglomeração foi realizada pelas componentes principais, extraídas conforme item anterior.

De acordo com Carvalho, Da Mata e Resende (2008), nos algoritmos tradicionais, quando as unidades geográficas são agrupadas, não necessariamente os grupos homogêneos são formados por municípios ou setores censitários estritamente vizinhos. Com isso, municípios que são geograficamente separados podem pertencer a um mesmo grupo. Dependendo do tipo de pesquisa, isso não é complicador. No entanto, como no caso dessa pesquisa, em que o objetivo é selecionar aglomerados que sejam vizinhos, isso acaba sendo um problema e o uso do algoritmo modificado se faz necessário. As etapas que descrevem a modificação efetuada pelos autores e utilizadas neste trabalho são as seguintes: i) Parte-se de uma base C de N unidades geográficas que podem corresponder a agrupamentos iniciais de subunidades. Neste caso, 5.507 municípios para o ano de 2000 e 5.565 para o ano de 2010. De início, cada uma das N observações consiste em um cluster isoladamente e tem um conjunto de variáveis $(X_{i,1}, X_{i,2}, \dots, X_{i,n})$. Para cada uma dessas N unidades, encontra-se a sua lista de vizinhos. Nesta pesquisa, considerou-se como vizinhos municípios a matriz do tipo rook, quando apenas as fronteiras com extensão diferente de zero são consideradas. ii) Em seguida, calcula-se a distância entre todos os pares formados por municípios vizinhos na lista de N unidades. Utilizou-se neste trabalho a medida de Ward, no qual tende a formar cluster de tamanhos semelhantes. O número de pares testados leva em conta a contiguidade. iii) Sejam I e J dois municípios vizinhos apresentando a menor distância, ou dissimilaridade, entre eles. Agrupa-se I e J em um único cluster. O número de cluster passa a ser N-1. iv) Para definir um novo cluster, formados por I e J, são combinados, além das listas de variáveis, as listas dos vizinhos. Assim, será composta uma nova lista de municípios vizinhos a partir da união da lista de vizinhos dos municípios I com a lista de vizinhos dos municípios J. v) Para os N-1 clusters, depois da junção descrita nos itens iii e iv, calculam-se as distâncias

entre todos os pares de clusters vizinhos. Nesse caso, dois clusters, por exemplo, A e B, de municípios, são considerados vizinhos quando houver pelo menos um município em que A é vizinho de um município em B. Para o par de clusters com a menor distância, agrupam-se os elementos em um único novo cluster, de forma que o número de clusters existentes passe a ser $N-2$. A distância entre os clusters A e B refere-se unicamente à dissimilaridade entre as variáveis. A similaridade geográfica está explicitamente modelada quando são agrupados somente clusters vizinhos. vi) Repetem-se os passos ii a v até se obter um único cluster, que deverá conter todas as unidades geográficas originais. No final do processo, tem-se uma árvore caracterizando os agrupamentos decorridos em cada passo do algoritmo. (CARVALHO; DA MATA; RESENDE, 2008 p. 189-190)

4. Resultados da Aglomeração

As Figuras 1 e 2 mostram o resultado das aglomerações dos municípios com características socioeconômicas homogêneas com base nos censos de 2000 e 2010. Foram identificadas 11 aglomerações para o ano de 2000 e 10 para o ano de 2010. A mudança no desenho das aglomerações entre 2000 e 2010 é condizente com o que é afirmado por Harvey, Hawkins e Thomas (2012), de que os clusters podem se reconfigurar no espaço e no tempo, e que isso tem se tornado motivo de pesquisas sobre a difusão que ocorre em determinadas regiões, além do transbordamento das redes de interações. As linhas escuras nos dois mapas representam a linha divisória dos estados. É fácil de se notar que a aglomeração não respeita os limites estaduais. A força de concentração de municípios semelhantes vai além de suas fronteiras legais.



A alteração na forma do mapa entre os dois períodos indica que a forma como os municípios vão se aglomerando homogeneamente é dinâmica ao longo do tempo.

Algumas localidades podem se desenvolver de forma diferente das demais ao longo do tempo. Percebe-se, por exemplo, um avanço da região Centro-Oeste sobre a região Norte, a separação da região litorânea da região Nordeste e a delimitação da região Sul em dois blocos ao longo das duas décadas. O fato de haver mapas distintos para os dois períodos revela que o País tem que ser pensado de forma dinâmica. Possivelmente, os dados do próximo censo mostrarão um novo desenho.

Como a configuração dos aglomerados leva em consideração variáveis estruturais e conjunturais e, dado a importância da contemporaneidade das informações, doravante analisar-se-á os aglomerados em sua versão mais recente, 2010. Percebe-se, ao longo de uma década, que alguns municípios localizados ao sul da região Norte se mostraram mais semelhantes aos municípios da região Centro-Oeste. Na região Nordeste, observa-se a agregação de três sub-regiões. O aglomerado Nordeste I abrange a região litorânea que vai de Alagoas ao Ceará. O Nordeste II é formado pelos municípios localizados no semiárido. Por último, os municípios do estado do Maranhão, em 2010, não se assemelharam aos municípios da região Norte ou Nordeste e, quase a totalidade do estado, foi identificada como um aglomerado.

A região Sudeste, onde se concentrava mais da metade do PIB Nacional em 2010 (55,38%), no novo desenho, se transforma em dois aglomerados, Sudeste I e Sudeste II, sendo que a primeira concentra os municípios do estado do Rio de Janeiro, Espírito Santo, parte da Bahia e parte de Minas Gerais. Enquanto que a segunda representa grande parte dos municípios do estado de São Paulo e do Paraná, além de Santa Catarina, Mato Grosso do Sul e Rio Grande do Sul.

Se considerarmos os aglomerados Centro I e Centro II, como a região Centro-Oeste, Nordeste I, Nordeste II e Maranhão como a região Nordeste, Sul I e Sul II como região Sul, Sudeste I e Sudeste II como região Sudeste e Norte como região Norte, é possível ver que a região Centro-Oeste ganha força e avança sobre as demais regiões. Conforme Tabela 4, ao contrapormos os municípios dos aglomerados com os das regiões político-administrativas, pode-se dizer que cerca de 20% dos municípios não se assemelham mais com suas regiões.

Tabela 4 – Relação entre o número de municípios das Regiões Político-Administrativas e o número de municípios dos Aglomerados

	Regiões Político-Administrativas					Total
	Centro-Oeste	Norte	Nordeste	Sudeste	Sul	
Centro I	404	67	13	183	3	670
Centro II	55	176	40	4		275
Norte		201				201
Maranhão		5	215			220
Nordeste I			609			609
Nordeste II			832	79		911
Sudeste I			85	779		864
Sudeste II	7			623	453	1,083
Sul I					256	256
Sul II					476	476
Total	466	449	1,794	1,668	1,188	5,565

Elaborado pelos autores.

O novo mapa das grandes regiões proposto para o Brasil ressalta a importância de se rever políticas públicas em função das atuais delimitações territoriais. Por exemplo, a distribuição do FPM ou FPE, a destinação de recursos de Fundos Constitucionais e as áreas de atuação de bancos de desenvolvimento regionais. Dessa forma, é importante um novo olhar para o desenho das regiões, onde políticas públicas possam ser tomadas levando em consideração as diferenças regionais. Ademais, o fato de identificar aglomerados semelhantes favorece o setor privado na decisão de investimentos

regionalizados. Com essa proposta de novas regiões, o próximo passo é identificar como se dará a projeção da distribuição do crescimento econômico previsto para o país nos próximos anos nestas novas regiões.

5 Estimação da Distribuição do Crescimento do PIB nacional nos aglomerados para o período 2015-2020

A metodologia de elaboração do PIB no nível municipal vem sendo conduzida pelo IBGE desde 2001, com vistas à obtenção de estimativas coerentes e comparáveis entre si. Consiste em um processo descendente de repartição, pelos municípios, do valor adicionado das 15 atividades das unidades da federação. Primeiro, estima-se o valor estadual de cada agregado; em seguida, reparte-se esse valor pelos municípios, ou seja, uma vez estimado o valor adicionado por atividade de cada estado, procede-se à distribuição para as atividades municipais de acordo com indicadores de cada subsetor de atividade econômica (IBGE, 2004). Com isso, os dados das contas regionais dos municípios são divulgados com uma defasagem de dois anos. Portanto, a mais recente informação do PIB municipal é referente ao ano de 2014, publicado pelo IBGE em 2016 (IBGE, 2016b)

A projeção da distribuição do crescimento econômico regional se depara com alguns problemas conhecidos na literatura, como indisponibilidade de dados, poucas publicações frequentes e uma infinidade de revisões de séries de variáveis. De acordo com Lehmann e Wohlrabe (2014), essas barreiras podem impedir que possamos conduzir estudos mais intensivos em estimativas regionais. No entanto, segundo os autores, há uma quantidade crescente de estudos da literatura econômica lidando com esses problemas e desenvolvendo projeções econômicas regionais, estimando variáveis como PIB, taxa de desemprego regional, taxa de desempregos setorial e total.

Para projetar a distribuição do crescimento econômico do país, nos aglomerados, foram testados modelos de dados em painel de efeitos fixos, primeira diferença, *pooled* e regressão linear pelo método de mínimos quadrados ordinários.

5.1 Painel de Dados

O modelo de dados em painel é uma técnica altamente difundida para verificar o impacto das características individuais em exercícios econométricos. Uma das vantagens da estimação com dados em painel é a importância da heterogeneidade individual, sugerindo a existência de características dos indivíduos que os diferenciam dos demais. Tais características podem ou não ser constantes ao longo do período observado, de tal forma que estudos temporais ou seccionais que não considerem essa heterogeneidade poderão gerar resultados viesados. Além disso, a modelagem com dados em painel possibilita utilizar um conjunto de informações bem mais amplo. A inclusão da dimensão seccional em um estudo temporal permite mais variabilidade dos dados e maior quantidade de informação, aumentando, conseqüentemente, a qualidade da estimação.

Um modelo de regressão com dados em painel, com N observações em T períodos e j variáveis, pode ser representado da seguinte forma:

$$Y_{it} = \alpha_0 + \beta_{kj}X_{j,it} + \beta_{kj}Z_{j,t} + c_i + \varepsilon_{it}, \quad \text{com } t = 1, 2, \dots, T \text{ e } i = 1, 2, \dots, N \quad (1)$$

A hipótese a respeito da relação entre as variáveis explicativas ($X_{j,it}, Z_{j,t}$) e a heterogeneidade individual (c_i) é que ditará o modelo que deverá ser utilizado em uma estimação. Se essa relação for independente ($Cov(X_{j,it}, c_i) = 0$), as hipóteses clássicas de regressão são atendidas e o método de mínimos quadrados ordinários (MQO) poderia ser aplicado para a estimação. Entretanto, se ($Cov(X_{j,it}, c_i) \neq 0$), existe endogeneidade no modelo, fato que viola uma das hipóteses clássicas e inviabiliza a estimação por MQO.

De acordo com Wooldrige (2002), se as características individuais (c_i) forem correlacionadas com a variável explicativa ($X_{j,it}$), as estimativas obtidas via MQO serão não apenas viesadas, mas inconsistentes. Assim, quando a hipótese de que esses termos são independentes não pode ser considerada verdadeira, ou seja, ($Cov(X_{j,it}, c_i) \neq 0$), devem ser considerados estimadores alternativos ao MQO.

No modelo de dados em painel de efeitos fixos (EF), cada unidade tem suas próprias características individuais que podem ou não influenciar os valores preditos. Assume-se que algo intrínseco às unidades de análise pode impactar o valor a ser estimado da variável de interesse. As características invariantes no tempo são únicas para o indivíduo e não devem ser correlacionadas com outras características individuais. O pressuposto básico dessa abordagem é eliminar o efeito não-observado (c_i), com base na suposição de que a $E(\varepsilon_{it}|x_i, c_i) = 0$, conhecida como condição de exogeneidade estrita. A transformação de EF, ou transformação *within*, é realizada da seguinte forma:

a) parte-se de um processo gerador de dados: $y_{it} = \beta x_{it} + c_i + \varepsilon_{it} \quad (2)$

b) tira-se a média da equação X no tempo: $\bar{y}_i = \beta \bar{x}_i + c_i + \bar{\varepsilon}_i \quad (3)$

c) subtrai-se (3) de (2) para cada t , obtendo-se a equação EF, removendo, assim, a heterogeneidade não-observada c_i : $y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (4)$

Supõe-se que para cada indivíduo i , há um intercepto diferente, respeitando a heterogeneidade entre os indivíduos, ao passo que os parâmetros de resposta são constantes para todos os indivíduos e em todos os períodos de tempo. De acordo com Hill et al (1999), as suposições do modelo são: $\beta_{0it} = \beta_{0i}$ e $\beta_{1it} = \beta_1 \dots \beta_{kit} = \beta_k$

Dessa forma, aplicando-se o estimador EF, sob as hipóteses de exogeneidade estrita e de que a heterogeneidade é constante para cada indivíduo, pode-se obter o modelo teórico especificado na equação (1). Nesse caso, as diferenças de comportamento entre os indivíduos que não foram controladas pelas variáveis explicativas seriam captadas pelo intercepto c_i , que pode ser interpretado como o efeito das variáveis individuais que foram omitidas do modelo.

O modelo de dados em painel primeira diferença (PD) é uma alternativa ao modelo de efeitos fixos quando a ($Cov(X_{j,it}, c_i) \neq 0$). No estimador PD, o pressuposto é eliminar o efeito não-observado, c_i , a partir da transformação em primeira diferença da equação (1): $y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \quad (5)$, com $\Delta y_{it} = \Delta \beta x_{it} + \Delta \varepsilon_{it} \quad (6)$

Sob as hipóteses de exogeneidade estrita e de que a heterogeneidade é constante para cada indivíduo, esse componente seria eliminado da equação ao tomar a primeira diferença em todos os termos. Dessa forma, o problema da endogeneidade gerado pela heterogeneidade individual é contornado, tornando possível a estimação da equação (6) por MQO. O modelo em PD pode ser especificado como:

$$\Delta Y_{it} = \Delta \beta_{kj} X_{j,it} + \Delta \beta_{kj} Z_{j,t} + \Delta \varepsilon_{it}, \quad \text{com } t = 1, 2, \dots, T \text{ e } i = 1, 2, \dots, N \quad (7)$$

Os termos considerados fixos para cada indivíduo (α_0 e c_i) são eliminados pela transformação em primeira diferença. Em muitos casos, os estimadores de efeitos fixos e primeira diferença são equivalentes. A diferença entre os modelos só se torna mais visível quando relaxamos a hipótese de exogeneidade estrita $E(\varepsilon_{it}|x_i, c_i) \neq 0$. Nesse caso, normalmente é mais fácil contornar o problema gerado pela violação aplicando-se o estimador em primeira diferença.

No modelo de efeitos aleatórios (EA), ao contrário do modelo de EF, a variação entre unidades é assumida como aleatória e não correlacionada com as variáveis a serem preditas ou independentes incluídas no modelo. Se houver convicção de que as diferenças entre entidades têm alguma influência sobre sua variável dependente, deve-se usar efeitos aleatórios. Nesse modelo, pode-se incluir variáveis invariantes no tempo. No modelo de efeitos fixos, ela seria absorvida pelo intercepto. O problema é que algumas variáveis não observadas são omitidas, causando vies nos estimadores. O modelo de efeitos aleatórios assume a forma: $y_{it} = \beta x_{it} + \alpha + u_{it} + \varepsilon_{it}$ (8).

O modelo de dados *Pooled* com erros-padrão robustos clusterizados consiste em uma estimação MQO com controle da correlação *within* do erro ao longo do tempo. Basicamente, trata-se de uma regressão linear com dados empilhados, dado pela equação: $y_{it} = \beta_0 + \beta x_{it} + u_{it}$ (9).

Adicionalmente aos modelos de dados em painel mencionados acima, efetuamos a estimativa utilizando uma regressão linear de MQO para cada município individualmente. De acordo com Greene (2012), no método de mínimos quadrados ordinários (MQO), os parâmetros não conhecidos da relação estocástica demonstrada abaixo são os objetos da estimação. $y_i = \alpha + x_i' \beta + \varepsilon_i$ (10)

Para estimar a distribuição do crescimento do PIB nacional pelos aglomerados serão utilizados os dados dos municípios foram agrupados em áreas mínimas comparáveis (AMC)³. O período da coleta de dados do PIB dos municípios vai de 1999 a 2014. Optou-se em utilizar AMC ao invés de municípios pois foram criados diversos municípios no período pesquisado, o que poderia comprometer a estimação, tanto dos novos municípios, que teriam poucas observações, quanto para os municípios que foram desmembrados em outros. Das 5.478 AMCs existentes em 1999, foram desconsideradas aquelas que apresentaram variação muito elevada do PIB em algum ano do período analisado, superiores a 50%. Ao todo, foram retiradas 203 AMCs, restando 5.275 (N=5.275). Os dados observados do PIB dos municípios, agrupados por AMC, estão disponíveis de 1999 a 2014 (T=16). Assim, tem-se 84.400 observações para os 10 aglomerados existentes em 2010.

Para as 203 AMCs que foram retiradas da estimação (*outliers*), o cálculo do PIB foi realizado de acordo com a participação da AMC no PIB nacional em 2014, aplicada para os períodos posteriores, conforme abaixo:

$$PIB_{outlier_i,t} = PartPIB_{BR_{2014}} \times PIB_{BR_t}, \quad \text{com } t = 2015, \dots, 2020 \quad (11)$$

Para projetar o PIB das AMCs para o período de 2015 a 2020, utilizou-se os dados do PIB observado para o Brasil em 2015 e 2016. Para o período de 2017 a 2020, utilizou-

³ As áreas mínimas comparáveis se referem à área agregada do menor número de municípios necessários para que as comparações intertemporais sejam geograficamente consistentes.

se os dados projetados pelo mercado, publicados pelo Banco Central do Brasil na pesquisa de expectativas de mercado⁴, realizada junto às instituições financeiras (BACEN, 2017).

Os dados da população estimada de cada município, calculada pelo IBGE, estão disponíveis até 2016 (IBGE, 2016c). Para as unidades da federação, o IBGE divulga a projeção da população estadual até 2030. Dessa forma, com base na participação de cada município na população do seu estado em 2016, calculamos a estimativa da população do município para o período 2017-2020, para que essa variável pudesse ser utilizada na projeção do PIB:

$$Pop_{it} = PartPopUF_{2016} \times PopUF_t, \quad com \ t = 2017, \dots, 2020 \quad (12)$$

O PIB do Brasil apresentou uma queda brusca ocorrida em 2009, um ano após a crise do *subprime*. Assim, para que o modelo pudesse capturar essa descontinuidade, incluiu-se na estimação uma *dummy* a partir deste período (se ano < 2009, *dummy*=0, se ano >=2009, *dummy*=1). Foram realizados testes para a *dummy* somente em 2009 (se ano=2009, *dummy*=1).

Em resumo, conforme Tabela 5, estimou-se o PIB das AMC que compõe cada aglomerado, utilizando os modelos de dados em painel mostrados anteriormente e, adicionalmente, o modelo MQO individual. Os dados foram tomados em sua forma logarítima, como exceção do painel em primeira diferença. O Breusch-Pagan rejeitou a hipótese nula de que há adequação do modelo *pooled* em relação ao modelo de efeitos aleatórios. O teste F de Chow rejeitou a hipótese nula de que há igualdade de interceptos e inclinações para todos as AMC. Por último, o teste de Hausman rejeita a hipótese nula de que o modelo de efeitos aleatórios oferece estimativas mais consistentes.

Tabela 5 – Equações estimadas em cada modelo

Modelo	Equação a ser estimada
Efeitos Fixos	$\ln PIB_{it} = \ln \beta_1 PIB_{BR,t} + \ln \beta_2 POP_{i,t} + D_{2009_t} + c_i + u_{it} + \varepsilon_{it}$
Efeitos Aleatórios	$\ln PIB_{it} = \ln \beta_1 PIB_{BR,t} + \ln \beta_2 POP_{i,t} + D_{2009_t} + \alpha + u_{it} + \varepsilon_{it}$
Primeira Diferença	$\Delta \ln PIB_{it} = \Delta \ln \beta_1 PIB_{BR,t} + \Delta \ln \beta_2 POP_{i,t} + \Delta \varepsilon_{it}$
<i>Pooled</i>	$\ln PIB_{i,t} = \alpha_i + \ln PIB_{BR,t} + \ln Pop_{i,t} + D_{2009_t} + \varepsilon_{i,t}$
MQO individual	$\ln PIB_t = \alpha + \ln PIB_{BR,t} + \ln Pop_t + D_{2009_t} + \varepsilon_t$

Para painel de dados com T pequeno e N grande, Harris e Tsavallis (1999) derivaram um teste de raiz unitária que assume a dimensão T como fixo (HT). Eles assumiram que ε_{it} é independente e identicamente distribuído (i.i.d), com distribuição normal e variância constante. Eles derivaram a média e o desvio-padrão sob a hipótese nula: $H_0: \rho = 1$ quando nem a média e nem a tendência temporal são incluídas (requer ausência de constante), quando somente a média foi incluída e quando ambos, média e tendência temporal foram incluídas. A distribuição assintótica da estatística do teste é justificada como $N \rightarrow \infty$. Então, deve-se ter um número relativamente grande de N quando se usa este teste. Além disso, o teste exige que seja balanceado. Neste trabalho

⁴ A expectativa de mercado é realizada pelo Banco Central com base no levantamento diário das previsões de cerca de 120 bancos, gestores de recursos e demais instituições para a economia brasileira. (BACEN, 2017)

utilizou-se o teste HT, com tendência e média. Os resultados não apontaram raiz unitária para a variável dependente.

A Tabela 6 apresenta o resumo do resultado dos coeficientes estimados para o modelo pooled, que mais se aproximou com os resultados do Brasil. Em resumo, a maioria dos coeficientes foram significativos, incluindo a dummy em 2009.

Tabela 6 – Resultado estimação – Pooled

	Sul I	Sul II	Sudeste I	Sudeste II	Centro I	Centro II	Nordeste I	Nordeste II	Maranhão	Norte	
Número Observações	3536	6896	13248	16720	9952	4064	9376	14096	3408	3104	
Sem dummy	Log PIB Brasil	0.9712 ***	1.3263 ***	1.1517 ***	1.0194 ***	1.3432 ***	1.5973 ***	1.3052 ***	1.4085 ***	1.4073 ***	1.4911 ***
	Log População	0.9008 ***	0.8928 ***	0.968 ***	1.0129 ***	0.6872 ***	0.8102 ***	0.9349 ***	0.8518 ***	0.8154 ***	0.5951 ***
	Constante	-24.4909 ***	-34.9881 ***	-31.016 ***	-26.9884 ***	-33.653 ***	-42.7943 ***	-35.6737 ***	-38.1061 ***	-37.8771 ***	-37.6529 ***
Dummy 2009=1	Log PIB Brasil	0.9487 ***	1.3401 ***	1.1170 ***	0.9864 ***	1.2312 ***	1.4961 ***	1.2436 ***	1.3525 ***	1.2903 ***	1.1663 ***
	Log População	1.0583 ***	1.0564 ***	1.2060 ***	1.1288 ***	1.0674 ***	1.0525 ***	1.1358 ***	1.1373 ***	1.0852 ***	1.0675 ***
	Dummy 2009=1	-0.0584 ***	-0.0630 ***	-0.0444 ***	0.0003	-0.0144 **	0.0360 ***	-0.0017	0.0084 **	0.0377 **	-0.4199 ***
	Constante	-25.3249 ***	-36.8015 ***	-32.2555 ***	-27.1298 ***	-33.9122 ***	-42.0357 ***	-35.7971 ***	-39.1601 ***	-37.0569 ***	-32.8687 ***
Dummy 2009/2020=1	Log PIB Brasil	0.5806 ***	1.2699 ***	0.8434 ***	0.8022 ***	1.0010 ***	1.3093 ***	1.0913 ***	1.1500 ***	1.0842 ***	0.8527 ***
	Log População	1.0586 ***	1.0563 ***	1.2061 ***	1.1288 ***	1.0645 ***	1.0526 ***	1.1359 ***	1.1375 ***	1.0853 ***	1.0678 ***
	Dummy 2009/2020=1	0.1461 ***	0.0242 ***	0.1085 ***	0.0751 ***	0.9285 ***	0.0786 ***	0.0620 ***	0.0832 ***	0.0866 ***	0.1249 ***
	Constante	-14.5997 ***	-34.7586 ***	-24.2842 ***	-21.7620 ***	-27.2062 ***	-36.5920 ***	-31.3576 ***	-33.2587 ***	-31.0501 ***	-23.7323 ***

Como a estimação nesta pesquisa se refere à distribuição do PIB do Brasil entre os aglomerados, compostos pelas AMCs, é de se esperar que o melhor modelo seja aquele em que a soma do PIB projetado para as AMCs seja o mais próximo possível do que é projetado para o Brasil, de acordo com os dados observados e com as expectativas de mercado. Dessa forma, a Tabela 7, resume o resultado da variação do crescimento do PIB do Brasil dado pelo somatório do valor estimado para cada AMCs, calculado pelos modelos MQO, Pooled, PD e EF, em três formas: com dummy somente em 2009, com dummy no período 2009 a 2020 e sem dummy. Os resultados mostraram que os resultados que mais se aproximaram do projetado para o Brasil foram os modelos com inclusão de dummy. Para os anos 2015 e 2016, o modelo que mais se aproximou do PIB nacional foi o Pooled com inclusão de dummy em 2009. Para o ano de 2017, o modelo que mais se aproximou foi o MQO, também com dummy em 2009. Para os anos de 2018 a 2020, o modelo que mais se aproximou foi o Pooled com inclusão de dummy para o período 2009/2020.











Tabela 7 – Comparativo variação do PIB nacional observada para 2015 e 2016 e projetada pelas expectativas de mercado para o período de 2017 a 2020 e a variação da soma das AMCs

Ano	Crescimento PIB Brasil (%)	Crescimento do PIB do Brasil pelo somatório do PIB das AMCs (%)											
		Modelos com inclusão Dummy (2009/2020=1)				Modelos com inclusão Dummy (2009=1)				Modelos sem inclusão Dummy			
		MQO	Pooled	Primeira Diferença	Painel Efeitos Fixos	MQO	Pooled	Primeira Diferença	Painel Efeitos Fixos	MQO	Pooled	Primeira Diferença	Painel Efeitos Fixos
2015	-3.77	-4.88	-2.69	-0.99	-3.83	-5.56	-3.44	-1.03	-4.43	-5.73	-3.43	-1.03	-4.42
2016	-3.60	-4.50	-2.55	-0.78	-3.65	-5.13	-3.27	-0.80	-4.22	-5.30	-3.25	-0.80	-4.21
2017	0.50*	0.73	1.69	1.01	1.14	0.54	1.79	1.05	1.20	0.56	1.79	1.05	1.20
2018	2.50*	3.30	3.11	1.44	3.16	3.47	3.62	1.50	3.55	3.58	3.61	1.50	3.55
2019	2.50*	3.34	3.08	1.42	3.15	3.50	3.59	1.48	3.54	3.61	3.58	1.48	3.54
2020	2.50*	3.38	3.05	1.41	3.13	3.53	3.56	1.46	3.53	3.65	3.55	1.46	3.52

* Projeção Mercado - Banco Central

A Tabela 8 apresenta os resultados médios da participação no PIB nacional, da variação do PIB e do PIB per capita dos aglomerados para os períodos: 2000-2009, 2010-2014 e 2015-2020. Optou-se por essa divisão de períodos para delimitar a queda do PIB nacional ocorrida em 2009 e a rápida recuperação em 2010, com variação positiva até 2014. O período seguinte 2015-2020, compreende as quedas no PIB ocorridas em 2015, 2016 e a expectativa de recuperação prevista para os próximos anos. Para composição dos valores médios, levou-se em consideração os modelos mais aderentes em cada ano da projeção, conforme Tabela 8. Os resultados mostraram que os aglomerados Centro II, Norte e Sul II são os que vão apresentar melhor crescimento nos próximos anos. Por outro lado, os resultados indicaram que os aglomerados Sul I e Sudeste I vão ter crescimento menor do que os seus pares, nessa mesma base de comparação. No tocante ao PIB per capita, estima-se queda em quase todos os aglomerados, exceção apenas do Maranhão, que se manterá praticamente estável, quando se compara a média esperada para o período 2015-2020 em relação à média 2010-2014. As maiores quedas per capita do produto econômico são projetadas para os aglomerados do Centro e Sudeste.

Tabela 8 – Participação no PIB Nacional, variação PIB e PIB per capita dos aglomerados com dados observados até 2014 e projeção Efeitos Fixos 2015-2020.

Aglomerados	Participação no PIB Nacional			Crescimento PIB (%)			PIB per capita (R\$ mil)		
	Média 2000-2009	Média 2010-2014	Média 2015-2020	Média 2000-2009	Média 2010-2014	Média 2015-2020	Média 2000-2009	Média 2010-2014	Média 2015-2020
 Norte	3.19	3.25	3.22	4.71	3.74	1.61	16.68	19.51	17.67
 Nordeste I	6.13	6.57	6.68	4.12	4.62	1.37	14.23	18.45	17.83
 Nordeste II	4.70	4.70	4.69	3.93	2.68	1.36	11.67	14.46	13.77
 Maranhão	1.23	1.34	1.41	5.29	5.26	1.30	9.56	12.46	12.54
 Centro I	11.06	11.52	11.72	3.62	3.79	1.49	30.69	37.58	35.20
 Centro II	1.71	2.33	2.28	7.96	6.83	1.97	15.99	24.51	21.96
 Sudeste I	21.09	21.32	20.94	3.62	3.59	1.15	29.69	37.20	34.98
 Sudeste II	43.34	41.80	41.87	3.01	2.80	1.20	40.00	47.18	44.36
 Sul I	5.44	5.06	5.09	2.04	3.19	0.61	31.13	37.11	35.96
 Sul II	2.11	2.10	2.11	3.24	4.74	1.60	25.57	32.92	32.00

6 Considerações Finais

Os resultados da aglomeração de municípios socioeconomicamente semelhantes em onze regiões em 2000 e dez no ano de 2010, mostraram que a atual divisão político-

administrativa não reflete a junção de municípios homogêneos. Ao longo do tempo, a dinâmica econômica e social vai moldando novos territórios. Portanto, delimitar uma região somente pela sua fronteira pode não fazer mais sentido. O conhecimento dessa nova forma de pensar a divisão regional proporciona uma nova forma de tratar a realidade brasileira, seja para implementação de políticas públicas ou nas decisões de investimentos privados e públicos. A metodologia utilizada neste trabalho pode ser replicada para vários fins, podendo ser gerados novos desenhos (aglomerados), dependendo apenas do objetivo do pesquisador e das variáveis a serem utilizadas.

Em um olhar mais aguçado sobre o aglomerado mais contemporâneo (2010), vemos que a distribuição econômica no Brasil, mesmo com o novo desenho, reflete a desigualdade existente no País. Por exemplo, o aglomerado Sudeste II tem mais de 40% de participação na riqueza nacional, distribuída entre os seus 1083 municípios e concentrada em 30% da população nacional. Junto com o Sudeste I, essas duas regiões concentram mais de 60% da riqueza nacional e quase 50% da população.

Nos próximos anos, em que os prognósticos do mercado mostram que a economia voltará a crescer, ainda que em ritmo bem inferior ao passado recente. A distribuição desse crescimento ao longo do território brasileiro beneficiará mais as regiões Centro II e do Norte. Quanto ao PIB per capita, nos próximos anos, os resultados indicaram que nenhuma região terá recuperado a média do período 2010-2014, com quedas acima da média nacional para várias regiões.

Bibliografia

- BANCO CENTRAL DO BRASIL. **Sistema de Expectativas de Mercado**. Brasília, 2017. Disponível. Acesso em 21 mar 2017.
- _____. **ESTBAN - Estatística Bancária Mensal por município**. Brasília, 2016
- BOROWIECKI, Karol Jan. Geographic clustering and productivity: an instrumental variable approach for classical composers. **Journal of Urban Economics**, n. 73, 2013.
- BRASIL. Ministério do Planejamento, Orçamento. **Estudo da Dimensão Territorial para o Planejamento, volume III**. Brasília, 2008
- CARVALHO, A. X. Y.; DA MATA, D.; RESENDE, G. M. Clusterização dos Municípios Brasileiros. In: CARVALHO, A. X. Y. et al (Org.). **Dinâmica dos Municípios**. Brasília: IPEA, 2008. p. 181-207
- CHEIN, F.; LEMOS, M.; ASSUNÇÃO, J. J.. Desenvolvimento desigual: evidências para o Brasil. **RBE**, Rio de Janeiro, v. 61, n. 3, p. 301–330, Jul-Set 2007
- CLAVAL, P. Geografia Econômica e Economia. **GeoTextos**, Salvador, n. 1, v. 1, 2005.
- DESROCHERS, P.; SAUTET, F.. Cluster-Based Economic Strategy, Facilitation Policy and the Market Process. *The Review of Austrian Economics*, n. 17, p. 233-245, 2004.
- FÁVERO, L. P. et. al. **Análise de Dados. Modelagem Multivariada para Tomada de Decisões**. Rio de Janeiro: Elsevier, 2009.
- FÁVERO, L. P. et. al. **Métodos Quantitativos com Stata**. Rio de Janeiro: Elsevier, 2014.
- FUJITA, M; KRUGMAN, P.; VENABLES, A.J. **Economia Espacial: urbanização, prosperidade econômica e desenvolvimento humano no mundo**. São Paulo: Futura, 2002

- Instituto Brasileiro de Geografia e Estatística (IBGE). **Produto Interno Bruto dos Municípios**. Série Relatórios Metodológicos, v. 29, 2004. Rio de Janeiro (RJ)
- _____. **Perfil dos Municípios Brasileiros 2015**. Rio de Janeiro, 2016a
- _____. **Produto Interno Bruto dos Municípios 2010-2014**. Contas Nacionais, n. 54, Rio de Janeiro, 2016b.
- _____. **Projeção da População das Unidades da Federação por sexo e idade: 2000-2030**. População, Projeção da População, Rio de Janeiro, 2014. Atualização: 26/03/2014.
- _____. **Número de municípios nos Censos Demográficos, segundo as Grandes Regiões e as Unidades da Federação – 1960/2010**. Disponível em: <<http://www.censo2010.ibge.gov.br>>. Acesso em 10 out 2016.
- JESUS, Josias Alves; SPINOLA, Noelio Dantasié. Seis décadas da teoria dos polos de crescimento: revisitando Perroux. **Revista de Desenvolvimento Econômico**, n. 2, p. 935-952, dez. 2015.
- KHANNA, Parag. **Connectography: mapping the future of global civilization**. New York: Random House, 2016
- LEHMANN, Robert; WOHLRABE, Klaus. Regional Forecasting: State-of-the-Art Methodology and Future Challenges. **CESifo Working Paper**, n. 5145, Munich, dez. 2014.
- LEMOS, Mauro Borges; DINIZ, Clelio Campolina, GUERRA, Leonardo Pontes. A Nova Configuração Regional Brasileira e sua Geografia Econômica. **Estudos Econômicos**, n. 4, v. 33, p. 665-700, out-dez/2003
- MALHOTRA, N. **Pesquisa de Marketing**. 6 ed. Porto Alegre: Bookman, 2012
- MARSHALL, Alfred. **Principles of Economics: An introductory volume**. 8. ed. Londres: The Macmillan Press Ltda, 1920
- MENDES, Constantino Cronemberger; MATTEO, Miguel. Formação e Evolução do Planejamento Regional no Brasil. In: CRUZ et al (Org.). **Economia Regional e Urbana: teorias e métodos com ênfase no Brasil**. Brasília: IPEA, 2011
- MONASTÉRIO, Leonardo; CAVALCANTI, Luiz Ricardo. Fundamentos do Pensamento Econômico Regional. In: CRUZ et al (Org.). **Economia Regional e Urbana: teorias e métodos com ênfase no Brasil**. Brasília: IPEA, 2011
- MYRDAL, G. **Teoria econômica e regiões subdesenvolvidas**. Belo Horizonte: Biblioteca Universitária – UFMG, 1960. Edição original de 1957.
- PERROUX, F. O conceito de pólo de desenvolvimento. In: SCHWARTZMAN, J. (Org.). **Economia Regional: textos escolhidos**. Belo Horizonte: Cedeplar, 1977. Ed de 1955.
- PESTANA, Maria Helena; GAGEIRO, João Nunes. **Análise de Dados para Ciências Sociais: a Complementariedade do SPSS**. 6. Ed. Lisboa: Edições Sílado
- Porter, Michael E. Location, Competition, and Economic Development: Local Clusters in a Global Economy. **Economic Development Quarterly**, n. 1, vol. 14, 15-34, fev. 2000.
- THISSE, Jacques-François. Geografia Econômica. In: CRUZ et al (Org.). **Economia Regional e Urbana: teorias e métodos com ênfase no Brasil**. Brasília: IPEA, 2011
- WOOLDRIDGE, Jeffrey M. **Introdução à Econometria: uma abordagem moderna**. São Paulo: Cengage Learning, 2016.