

Uma análise multidimensional do processo de convergência entre os estados brasileiros — 1970-2000*

André M. Marques**

Doutor em Economia pela Universidade Federal do Rio Grande do Sul (UFRGS), Professor associado do Departamento de Economia da Universidade Federal da Paraíba

Adelar Fochezatto***

Doutor em Economia pela UFRGS, Professor titular da Pontifícia Universidade Católica do Rio Grande do Sul

Resumo

Muitos estudos têm sido feitos sobre o processo de convergência regional no Brasil, nos últimos anos. A maioria desses estudos assume que a economia pode ser representada por uma função de produção agregada, com rendimentos decrescentes de escala. Em geral, esses estudos analisam a convergência apenas da renda *per capita*. Neste trabalho, o tema é analisado sob uma perspectiva diferente: o desenvolvimento é concebido como sendo multidimensional, e a análise de convergência é feita usando-se uma técnica estatística multivariada, a análise de agrupamentos. Esta análise é feita para quatro períodos, 1970, 1980, 1991 e 2000, e os resultados mostram que, quando múltiplos critérios são analisados, é constatada a formação de dois grupos bem distintos: um formado pelos estados do Norte e do Nordeste, outro, pelos estados do Sudeste, Sul e Centro-Oeste. Ao longo

* Artigo recebido em ago. 2015 e aceito para publicação em nov. 2017.

 Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-NãoComercial 4.0 Internacional.

Open Acces (Acesso Aberto)

Revisora de Língua Portuguesa: Tatiana Zismann

** E-mail: 00094751@ufrgs.br

*** E-mail: adelar@puccs.br

Os autores agradecem os comentários e sugestões dos dois pareceristas anônimos da Revista. Erros e omissões que persistam são da inteira responsabilidade dos autores.

das décadas, a distância entre os grupos aumentou e a dentro dos grupos tornou-se menor.

Palavras-chave

Convergência; divergência; desenvolvimento econômico

Abstract

Many studies have been done on the process of regional convergence in Brazil, in recent years. Most of these studies assume that economy can be represented by an aggregate production function with decreasing scale returns. In general, these studies tackle the income per capita only. In this paper, the issue is examined from a different perspective: development is conceived as being multidimensional, and the convergence analysis is performed by using a multivariate statistical technique, the cluster analysis. This analysis is carried out along four periods, 1970, 1980, 1991 and 2000, and the results show that there was a formation of two distinct groups, one comprising the northern and northeastern states and the other, the southeastern, southern and midwestern states. Over time, the distance between the groups increased and the one within the groups decreased.

Keywords

Convergence; divergence; economic development

Classificação JEL: O11, O18, O47

1 Introdução

A hipótese da convergência (condicionada ou incondicionada) entre países ou regiões vem recebendo especial atenção dos pesquisadores desde meados dos anos 90. A importância dessa suposição reside precisamente na possibilidade teórica de que países ou regiões podem vir a alcançar um

mesmo patamar de desenvolvimento em algum ponto do tempo, mesmo que iniciem seu processo de desenvolvimento em momentos diferentes.

A maioria dos trabalhos que investiga essa suposição atribui grande ênfase ao emprego de apenas um indicador de desenvolvimento, normalmente o Produto Interno Bruto (PIB) *per capita*. Isso porque, em princípio, essa seria uma medida indireta de bem-estar global da sociedade. Mais especificamente, essa seria uma medida da oferta disponível de bens e serviços por pessoa da comunidade. Não obstante, a utilização de um único indicador para mensurar o bem-estar tem sido contestada na literatura.

O Banco Mundial, por exemplo, desde os anos 90, introduziu o Índice de Desenvolvimento Humano (IDH), uma medida resumida de três dimensões do desenvolvimento humano: educação, saúde e renda (acesso a bens e serviços), visando capturar a qualidade de vida das pessoas ou o progresso a longo prazo da sociedade. Existe ainda o índice de progresso social, que também considera múltiplas dimensões do desenvolvimento econômico, apresentando um escopo mais amplo do que o IDH, ao incorporar aspectos como água e saneamento, moradia, segurança pessoal, sustentabilidade dos ecossistemas e oportunidades.¹

Na mesma direção, Hobijn e Franses (2001) enfatizam o emprego de vários indicadores de qualidade de vida da população para analisar a hipótese de convergência como forma de mensurar o desempenho dos países. Os autores apresentam evidência de que convergência no PIB *per capita* (se houver) não implica convergência na qualidade de vida, medida pelo consumo diário de calorias, proteínas, mortalidade infantil, expectativa de vida ao nascer, etc. Após análise de um grande número de estudos, a principal objeção às análises que empregam apenas uma única variável para mensurar o desenvolvimento foi assim sintetizada por Brida *et al.* (2011, p. 321):

These studies yield evidence in favour of the view that a unique interpretative model is likely to be inadequate to describe growth and development experiences. Also, that standard GDP index is unable to capture the real inequalities among countries in terms of the different dimension of well-being of populations.

Por esses motivos, empiricamente, a limitação básica dos trabalhos que investigam a hipótese de convergência é o emprego do PIB *per capita* como medida única, haja vista que ela não informa sobre a utilização dos bens e serviços da economia (isto é, a qualidade de vida das pessoas), apenas informa a sua disponibilidade. Por exemplo, pode ser que o PIB *per capita* seja bastante elevado em certo país, estado ou município, entretanto, nesses mesmos locais, a renda e a propriedade podem ser concentradas ao

¹ Sobre esse ponto, ver: <<http://www.progressosocial.org.br/>>.

ponto de a maior parte da população se apropriar de uma parcela demasiadamente pequena da oferta total de bens e serviços. Esse é o caso, por exemplo, do Município de Triunfo, no Rio Grande do Sul, que possui uma alta renda *per capita* por causa da presença do Polo Petroquímico, mas a distribuição da renda é muito desigual.

Assim, apesar da alta disponibilidade de bens e serviços, uma parte significativa da população pode não ter acesso a esses bens. Isso pode se refletir nos indicadores educacionais (baixa escolaridade, analfabetismo, etc.), uma vez que o serviço de educação não é utilizado por todos igualmente, e, também, nos indicadores de saúde, habitação, saneamento, etc., afetando, assim, o desenvolvimento econômico de um país, estado ou município. Em suma, o PIB *per capita* analisado isoladamente é uma medida muito limitada para analisar se o desenvolvimento econômico alcança a todos os países, municípios ou estados de forma satisfatória.²

Desse modo, o presente trabalho adota o enfoque multidimensional para medir o desempenho dos estados brasileiros nas últimas quatro décadas, em termos de desenvolvimento econômico (BRIDA *et al.*, 2011; CRACOLICI; CUFFARO; NIJKAMP, 2010; HOBIJN; FRANSES, 2001). Espera-se, com essa abordagem, captar outros aspectos da qualidade de vida da população que têm sido ignorados na literatura empírica que analisa a hipótese de convergência e/ou divergência com o enfoque univariado.

O objetivo central do presente trabalho é verificar se está ou não havendo convergência de desenvolvimento entre os estados brasileiros, considerando múltiplos critérios. No total, 14 variáveis socioeconômicas são analisadas, constituindo seis dimensões do desenvolvimento econômico. A ideia básica é considerar outras dimensões do desenvolvimento — além da oferta de bens e serviços por pessoa (PIB *per capita*) —, que informem também a sua utilização pela sociedade.

Vários indicadores de desenvolvimento econômico são empregados na análise procurando capturar as múltiplas dimensões em que o desenvolvimento pode se manifestar, incluindo condições de moradia, desenvolvimento industrial, pobreza e desigualdade, renda, longevidade e educação, oferta de alimentos e nível de emprego. A análise é feita em quatro momentos do tempo ao longo do período que vai de 1970 a 2000.

A análise multivariada é importante porque nem todos os indicadores de desenvolvimento e bem-estar apresentam alta correlação com a renda

² O PIB *per capita* é normalmente considerado uma medida **indireta** do desenvolvimento econômico, enquanto o crescimento do PIB não contém informação relacionada ao grau de desenvolvimento dos países (GREMAUD; VASCONCELLOS; TONETO JÚNIOR, 2010; LUCAS, 1988). Uma medida **direta** do desenvolvimento são os indicadores sociais e econômicos, como aqueles empregados no presente trabalho (GREMAUD; VASCONCELLOS; TONETO JÚNIOR, 2010).

per capita, que é a variável usada na maioria dos estudos sobre o tema. Observando as matrizes de correlações em anexo, é possível verificar, por exemplo, que o índice de desigualdade de renda possui baixa correlação com o IDH (dimensão renda).

Em relação aos trabalhos que utilizam análise de regressão com a renda *per capita* como variável dependente, este estudo acrescenta informação, pois identifica quais variáveis contribuem para a formação de clubes de convergência ou divergência, informação ausente em estudos de convergência que empregam apenas uma variável ou poucas explicativas (GAZONATO; GOMES; REIS, 2014; GOMES; ESPIRIDIÃO, 2016).

Ademais, ao se detectarem os estados mais parecidos e diferentes entre si, nenhuma suposição *a priori* é feita, como a imposição de *dummies* regionais ou estaduais exógenas, a exemplo do estudo de Gazonato, Gomes e Reis (2014). Ao contrário, são os próprios indicadores de desenvolvimento que geram os clubes de convergência e divergência, sem imposições *a priori* do pesquisador.

Outra característica importante deste estudo é a comparação entre as várias décadas, indicando que a formação de clubes de desenvolvimento econômico não é fenômeno recente na economia brasileira. Isso representa outro acréscimo de informação em relação aos estudos que se limitam a apenas uma década (a exemplo de Guimarães e Almeida (2017) e Gazonato, Gomes e Reis (2014), Gomes e Espiridião (2016)).

A metodologia empregada para detectar o padrão de convergência ou divergência multidimensional entre os estados baseia-se na análise de agrupamentos (*cluster analysis*). Para testar a significância das diferenças entre os grupos formados, utiliza-se a estatística Λ de Wilks e, para testar a igualdade das distâncias entre os estados ao longo do tempo, emprega-se o teste de Wilcoxon.

Em contraste com a análise de regressão, a análise de agrupamentos não está baseada na existência de dependência entre as variáveis, não utiliza modelos probabilísticos. Ela é um método exploratório empregado para gerar hipóteses a serem testadas posteriormente ao identificar os casos mais semelhantes e dessemelhantes a partir de múltiplos critérios. Posteriormente, para que essas hipóteses sejam testadas, pode-se empregar a estatística Λ de Wilks, conhecida como Análise de Variância Multivariada (Manova) (MINGOTI, 2007; REIS, 2001).

Especificamente, o objetivo é procurar capturar, de forma sintética, uma medida de separação ou aproximação ao longo dos anos, considerando todas as 14 variáveis. Por isso, a escolha de uma técnica da estatística multivariada, a análise de agrupamentos. Essa técnica permite sintetizar uma grande quantidade de variáveis correlacionadas, medidas nos casos

(estados), em uma única matriz de distâncias, que, por fim, resulta em uma árvore de agrupamentos hierárquicos, dos mais semelhantes para os mais dessemelhantes.

Uma abordagem similar foi empregada por Saint-Arnaud e Bernard (2003), para verificar se estaria ocorrendo convergência entre os regimes de bem-estar de diferentes países. Porém, na literatura sobre convergência entre regiões brasileiras não se tem conhecimento, até o momento, da utilização dessa metodologia, que é inteiramente compatível com a perspectiva multidimensional para medir o desempenho dos países (CRACOLICI; CUF-FARO; NIJKAMP, 2010; WORLD BANK, 2001). Em certo sentido, o presente trabalho complementa o estudo de Brida *et al.* (2011, p. 345), pois os autores reconhecem explicitamente a limitação de seu trabalho, que utiliza apenas duas variáveis (nível e taxa de crescimento do PIB *per capita*), e sugerem a inclusão de variáveis institucionais — indicadores sociais econômicos adicionais — para testar a hipótese de convergência.

O artigo possui cinco seções e está organizado do seguinte modo: após esta **Introdução**, a seção dois traz uma breve revisão da literatura. A seção três descreve a metodologia e a base de dados. A seção quatro analisa os resultados encontrados, e, por fim, a seção cinco apresenta os comentários finais.

2 Revisão da literatura

O estudo do desenvolvimento econômico, em sua dimensão espacial, tem sido dominado por duas perspectivas teóricas contrastantes. A primeira, com um enfoque mais tradicional, sustenta que o desenvolvimento tende a ser um processo uniforme no espaço, desde que não haja barreiras à operação das forças de mercado: em uma economia nacional, completamente integrada, espera-se uma convergência de renda e desenvolvimento ao longo do tempo pelo movimento dos preços, salários e taxa de lucro. Diferenças nas rendas colocariam em movimento mecanismos de correção das disparidades que levam à convergência entre regiões de um espaço nacional.³ Por isso, nessa perspectiva, disparidades regionais não seriam persistentes, pois a sua ocorrência seria compensada pela operação do mecanismo de mercado, corrigindo os desequilíbrios regionais. Entretanto, como observado anteriormente, ainda que formulada sobre bases mais tradicio-

³ Para referências a esse respeito, ver Martin e Sunley (1998, p. 201). A possibilidade de divergência também pode ser deduzida dos modelos da "nova" teoria do crescimento, como o modelo AK, porém, por diferentes motivos teóricos daqueles apresentados neste estudo.

nais, a hipótese dos clubes de convergência tem ganhado respaldo tanto na literatura teórica⁴ quanto na empírica⁵. Em ambos os casos, o modelo tradicional de Solow poderia gerar a formação de *clusters* e a persistência de desigualdades de renda *per capita* e desenvolvimento mesmo no longo prazo.

De acordo com a perspectiva alternativa, sintetizada por Martin e Sunley (1998), não se pode esperar, teoricamente, que a convergência de renda e desenvolvimento entre regiões seja o caso mais provável, mesmo para longos períodos de tempo, dentro de um mesmo país. Para autores como Myrdal (1968), Kaldor (2013), Dixon e Thirlwall (1975), a divergência de renda e desenvolvimento entre as regiões é o caso mais provável. A explicação teórica baseia-se no princípio da causalção cumulativa, em que as forças de mercado, quando atuam livremente, tendem a reforçar as diferenças regionais. Economias de escala e de aglomeração levam a uma cumulativa concentração de investimentos em capital fixo, trabalho qualificado e alta produtividade em certas regiões-polo, em detrimento de outras regiões menos desenvolvidas. As regiões que inicialmente concentram os investimentos geram uma atratividade para capital fixo e capital humano qualificado, ausente nas regiões menos desenvolvidas, reforçando seu potencial econômico de forma acumulativa. Os ganhos de produtividade estão correlacionados com a acumulação de capital e os acréscimos nos salários e nos lucros da região.

Mesmo com a ação política de governos no sentido da desconcentração da produção e dos investimentos, a convergência é uma hipótese improvável de ocorrer.⁶ Setterfield (1997) desenvolve um modelo teórico na perspectiva de Kaldor (2013), mostrando a possibilidade teórica de divergência persistente entre países ou regiões.

Quando se considera o modelo de Solow como base teórica para a hipótese da convergência, questões metodológicas complicadas emergem, pois é necessário postular que os casos considerados já se encontram em seu estado estacionário (FERREIRA; CRUZ, 2008). A suposição teórica que fundamenta a hipótese de convergência é a de produtividade marginal decrescente do capital. De acordo com essa suposição, regiões com reduzido

⁴ Ver, especialmente, Galor (1996).

⁵ Ver, especialmente, Durlauf e Johnson (1995) e Quah (1996) para suporte empírico a essa hipótese.

⁶ Por exemplo, a migração é sempre seletiva no fator idade. Os jovens, em geral, migram para regiões mais desenvolvidas à custa de sua região de origem, que perde capacidade de trabalho com a migração. Sobre esse aspecto, após se analisarem vários estudos na área, Martin e Sunley (1998, p. 212) concluem: "*these analyses suggest that the migration of better-educated, highly skilled, and more enterprising workers is regionally disequilibrating, in that it benefits destination regions at the expense of the areas of origin*".

estoque de capital teriam altas taxas de retorno e, desse modo, desenvolver-se-iam mais rapidamente que as regiões ricas, que têm um estoque de capital mais elevado e retorno mais baixo. A falácia do argumento consiste em supor que uma alta taxa de retorno potencialmente considerada seja suficiente para efetivar uma alta taxa de acumulação de capital. De fato, é bem sabido, na teoria econômica, que uma alta taxa de retorno só poderá ser obtida depois que uma alta taxa de acumulação for concretizada pelos capitalistas e/ou pelo estado.⁷

A experiência tem mostrado que, em geral, as regiões mais desenvolvidas é que se tornam alvo dos investimentos de empresas e acumulação de capital fixo, em vista da mão de obra já qualificada, da infraestrutura de transporte e comunicações, das facilidades locais, etc. Ao final, a acumulação de capital fixo em certas áreas reforça os desequilíbrios regionais de uma forma acumulativa: quanto mais desenvolvida uma região, mais desenvolvida ela se torna e vice-versa.

Uma das ideias-chave dos trabalhos de Myrdal (1968), Kaldor (2013), Dixon e Thirlwall (1975) consiste na lei de Verdoorn. Na medida em que uma região obtém uma vantagem comparativa para seu crescimento, ela tende a mantê-la ou aumentá-la ao longo do tempo através dos retornos crescentes da indústria e de seus efeitos de encadeamento sobre o produto agregado. Segundo Kaldor (2013, p. 482-483):

The prevailing distribution of real income in the world — the comparative riches or poverty of nations, or regions — is largely to be explained, not by 'natural factors', but by unequal incidence of development in industrial activities.

Muitos trabalhos têm sido realizados para investigar a hipótese da convergência ou divergência no crescimento econômico. A hipótese da convergência tem recebido maior atenção dos pesquisadores. Ferreira e Cruz (2008), com uma base de dados municipal para os períodos de 1991 e 2000 e utilizando um modelo de **regressão em árvore** com um parâmetro de *threshold*, investigaram a hipótese de convergência de renda dos municípios brasileiros. O objetivo do trabalho foi verificar se estava havendo convergência na desigualdade entre os municípios e se as condições iniciais, medidas pelo Índice de Gini, geravam diferentes equilíbrios (clubes de convergência). Os autores constataram a ocorrência de seis clubes de convergência, nos quais a renda do trabalho foi a variável com maior influência na redução das desigualdades. As transferências governamentais e a escolaridade contribuíram pouco para esse processo.

Gligor e Ausloos (2011) investigaram a hipótese de convergência sob um enfoque multidimensional a partir de dados de séries temporais, consi-

⁷ Teoricamente, os capitalistas ganham o que gastam.

derando variáveis como PIB, PIB *per capita*, consumo e investimento para o continente europeu. Através da análise de *cluster* e correlação, os autores concluem que há evidência de formação de clubes de convergência dentro de uma mesma região da Europa. Para os autores, as aglomerações industriais e o crescimento são forças que se retroalimentam no sentido da formação dos clubes de convergência.

Marques (2009), empregando uma base de dados municipal para os anos 1990 e 1999, através de análise de variância e correlação, definindo *dummies* regionais, constatou que as três macrorregiões que constituem o Estado do Rio Grande do Sul diferem significativamente em termos de crescimento e desigualdade, com uma significativa correlação inversa entre ambas. Segundo o autor, os municípios situados na região nordeste do estado combinam mais alto crescimento com um grau de desigualdade substancialmente menor. Embora o período de uma década seja relativamente curto, o trabalho conclui que os resultados apontam para uma divergência entre as regiões do Rio Grande do Sul.

Moreira, Braga e Toyoshima (2010), empregando uma base de dados estadual para o período de 1996 a 2007, através de modelos dinâmicos de equações simultâneas para dados em painel, constataram a persistência de disparidades regionais no Brasil. Eles concluem que essa segmentação tem efeito desacelerador sobre as taxas de investimento produtivo a longo prazo e, portanto, sobre as taxas de crescimento econômico. Assim, a segmentação acaba perpetuando a incidência da pobreza por meio de um círculo vicioso ao longo do tempo, de modo que a coexistência da pobreza e desigualdade dificulta o crescimento econômico.

Segundo esses autores, melhorar a equidade em um ambiente marcado pela alta heterogeneidade teria o efeito de acelerar a taxa de crescimento a longo prazo. A elevação da taxa de crescimento a longo prazo, por sua vez, atuaria sobre a redução da pobreza, levando a um maior nível de equidade e a um maior crescimento, engendrando um círculo virtuoso de desenvolvimento. Os autores ressaltam a ocorrência de grandes disparidades entre os estados situados nas regiões Norte e Nordeste vis-à-vis Região Sul do Brasil.

Na mesma perspectiva teórica de Myrdal (1968), Aryeetey, Owusu e Mensah (2009), utilizando o Gini da renda e três coeficientes de entropia generalizada, analisaram a existência de disparidades regionais em Gana. A base de dados leva em conta o Gini da renda e vários indicadores de desenvolvimento, como consumo de energia elétrica, acesso à escola, taxa de alfabetização, consumo de água potável (entre outros) para os períodos de 1991-2006 e 1970-2000. Os autores concluem que a persistente divergência entre as regiões do País (Norte e Sul) e as zonas (rural e urbana) tem

suas origens no processo de colonização. Essas condições, porém, foram reforçadas pelas estratégias de política que se seguiram. Por exemplo, o processo de substituição de importações e a expansão dos investimentos nas áreas de exportação (indústria de mineração) criaram infraestrutura e facilidades de transporte em áreas concentradas, reforçando as disparidades entre as regiões, em especial nas zonas portuárias, em detrimento de áreas menos desenvolvidas. Nessas áreas, criou-se um incentivo para novos investimentos e deslocamento de mão de obra qualificada, em detrimento das demais regiões.

Com um enfoque espacial, Silveira-Neto e Azzoni (2006) analisaram a evolução da renda *per capita* dos estados brasileiros no período 1985-2001. O estudo apresenta forte evidência de correlação espacial positiva (cada estado é parecido com os estados vizinhos em termos de renda *per capita*), fator que ajuda explicar a grande persistência das disparidades regionais do Brasil. Em particular, os autores identificaram dois *clusters* geográficos: um formado por estados de alta renda, situado no Sul-Sudeste, e outro formado por estados de baixa renda, situado no Norte-Nordeste. Segundo os autores, a formação desses dois “clubes” acelerou-se na década de 90, ocorrendo simultaneamente à intensificação da dependência espacial, medida pelo coeficiente de Moran. Ao final do período, apenas dois estados de renda média (Santa Catarina e Mato Grosso do Sul) mudaram seus *status* e migraram para o *cluster* de mais alta renda, sugerindo uma forte estabilidade na formação dos *clusters*.

Mossi *et al.* (2003) analisaram a dependência espacial dos estados brasileiros no período 1939-1998, com dados anuais de renda *per capita* estadual. Os autores constatam que o índice de Moran da renda *per capita* é positivo e estatisticamente significativo durante todo o período, intensificando-se após os anos 80. Assim, se a renda *per capita* é baixa em um estado, ela tende a ser baixa também em seus estados vizinhos e vice-versa. Em particular, dois *clusters* espaciais com alta persistência são identificados na análise: um *cluster* de baixa renda, situado na Região Nordeste, e um *cluster* de alta renda, situado na Região Sudeste, com baixa mobilidade entre eles, em que apenas Minas Gerais e Espírito Santo migraram para o *cluster* de alta renda. Ao analisar a dinâmica de transição, os autores concluem:

The results indicate a much higher persistence for the extreme classes of income than for the intermediary classes. The tests on neighborhood effects indicate that states with wealthier neighbors have a greater chance of moving upward on the income ladder (MOSSI *et al.*, 2003, p. 415-416).

Com o objetivo de testar a hipótese de convergência absoluta e condicional do PIB *per capita* entre os estados brasileiros no período de 1994 a

2010, Gazonato, Gomes e Reis (2014) utilizam uma metodologia em duas etapas. Na primeira, por meio da estatística *I* de Moran, foi avaliada a presença de dependência espacial do PIB *per capita* entre as regiões analisadas. Os resultados dessa estatística indicaram a presença de autocorrelação espacial nos dois anos analisados. Na segunda etapa, os autores utilizaram um modelo de convergência para verificar a existência de convergência absoluta e condicional. Os resultados deste modelo indicaram a presença de convergência no período analisado.

Com foco no capital humano, Gomes e Esperidião (2016) analisaram a hipótese da convergência absoluta e condicional do PIB *per capita* nas regiões brasileiras, no período 1995-2009. Visando uma maior homogeneidade na análise, os estados brasileiros foram divididos em quatro grupos e foi estimado um painel dinâmico. Em linhas gerais, seus resultados apontaram para a presença de convergência. No entanto, considerando-se especificamente o capital humano, a convergência condicional foi diferente entre os grupos utilizados, sinalizando para a persistência de desigualdades regionais.

Guimarães e Almeida (2017) analisaram a convergência de renda entre os anos 1999 e 2005 com modelos de econometria espacial e usando diferentes escalas de agregação geográfica. Seus resultados diferem de acordo com a escala geográfica considerada. Isso pode indicar dependência de escala na análise de convergência de renda no Brasil. Duas questões tornam-se relevantes quando se fazem estudos dessa natureza: a escolha da escala geográfica mais adequada e a falta de comparabilidade dos resultados quando são usadas diferentes escalas.

3 Metodologia e base de dados

3.1 Descrição dos dados

A escolha das variáveis foi inspirada especialmente em Cracolici, Cuffaro e Nijkamp (2010) e Aryeetey, Owusu e Mensah (2009), que apresentam uma síntese de outros trabalhos na mesma perspectiva e também utilizam vários indicadores sociais e econômicos que expressam a utilização de bens e serviços pelos agentes econômicos para medir a qualidade de vida e o bem-estar da população. Outras variáveis poderiam ser testadas, porém, a escolha foi também guiada pela facilidade de acesso e homogeneidade da fonte dos dados do Instituto de Pesquisa Econômica Aplicada (IPEA) (2013).

O período estudado vai de 1970 até 2000. Para o conjunto de variáveis utilizadas, ainda não é possível estender o estudo até 2010, pois algumas variáveis usadas não estão disponíveis, como o estoque de capital residencial rural e o estoque de capital residencial urbano. As variáveis empregadas no trabalho estão descritas no Quadro 1. Para facilitar a análise, essas variáveis podem ser agrupadas em seis dimensões do desenvolvimento: condições de moradia (krrpc, krupc, cerpc), desenvolvimento industrial (ceipc, pindpc), pobreza e desigualdade (theil, p0), desenvolvimento humano (idhl, idhe, idhr), produção agrícola (yappc, yatpc) e emprego (empurb, emprur).

Quadro 1

Descrição das variáveis utilizadas no trabalho

VARIÁVEL	DESCRIÇÃO
Krrpc	Capital residencial rural <i>per capita</i> — em milhares de reais, a preços do ano 2000
Krupc	Capital residencial urbano <i>per capita</i> — em milhares de reais, a preços do ano 2000
Cerpc	Consumo de energia elétrica residencial <i>per capita</i> — MWh
ceipc	Consumo de energia elétrica industrial <i>per capita</i> — MWh
theil	Desigualdade renda — índice L de Theil
p0	Pessoas pobres — %
idhl	Índice de Desenvolvimento Humano (IDH) — longevidade
idhe	Índice de Desenvolvimento Humano (IDH) — educação
idhr	Índice de Desenvolvimento Humano (IDH) — renda
yappc	Produção da lavoura permanente <i>per capita</i> — em milhares de reais, a preços do ano 2000
yatpc	Produção da lavoura temporária <i>per capita</i> — em milhares de reais, a preços do ano 2000
empurb	População urbana ocupada como proporção da população total
emprur	População rural ocupada como proporção da população total
pindpc	PIB industrial estadual <i>per capita</i> (preços básicos) — em milhares de reais, a preços do ano 2000

FONTE DOS DADOS BRUTOS: IPEA (2013).

NOTA: Por não existirem informações sobre todas as variáveis, em todos os períodos analisados, os Estados do Mato Grosso do Sul e do Tocantins foram retirados da análise. O estudo considera, portanto, 25 estados.

Sinteticamente, as principais vantagens da metodologia proposta em relação à análise de regressão convencional são: (a) nenhuma relação assimétrica *a priori* é postulada entre as variáveis analisadas (não existe a distinção entre “explicada” e “explicativa”); (b) não se faz nenhuma hipótese sobre a distribuição de probabilidade dos dados para a determinação dos *clusters*; e (c) a alta correlação entre as variáveis é uma propriedade desejável, e, contrariamente, poderia implicar problema de multicolinearidade na

análise de regressão convencional, inflacionando a variância e covariâncias dos coeficientes. Esta última característica está presente na maioria dos indicadores econômicos e sociais, o que reforça a adequabilidade da metodologia proposta.

Esses dados foram organizados para o formato de matriz ($n \times p$), em que n é o número de indivíduos (casos), e p é o número de variáveis. Neste estudo, os casos são as unidades da Federação, e as variáveis, as diferentes dimensões do desenvolvimento econômico. A descrição completa das variáveis encontra-se no Quadro A.1. Todas as estimações e gráficos foram feitos no ambiente **R**. No total, quatro matrizes de dados foram organizadas para os seguintes momentos do tempo: 1970, 1980, 1991 e 2000.

A matriz de correlação amostral com os coeficientes de Pearson ($R_{14 \times 14}$) oferece uma indicação da interdependência entre as dimensões do desenvolvimento econômico. O objetivo da análise de correlação, preliminarmente, é detectar padrões de relacionamento entre as variáveis analisadas, que, em princípio, mudam pouco com o tempo. O Quadro A.2 apresenta o coeficiente de correlação amostral para cada par de variáveis do estudo nos anos de 1970, de 1980, de 1991 e de 2000. Espera-se, teoricamente, que certos padrões de relacionamento mantenham-se estáveis ao longo do tempo, tais como: (a) a relação negativa entre a percentagem de pobres e os IDHs de renda, de longevidade e de educação; e (b) a correlação positiva entre a produção industrial *per capita* e os indicadores de desenvolvimento e bem-estar.

A análise das matrizes de correlação permite que sejam extraídas algumas conclusões preliminares sobre o relacionamento entre as variáveis analisadas. A primeira conclusão é que a melhora nas condições de moradia, ao longo das quatro décadas, tanto urbanas quanto rurais, está altamente correlacionada com melhora nos IDHs (educação, renda e longevidade), cujo coeficiente de correlação é, em geral, maior que 0,70 e, em alguns casos, maior que 0,80. Ademais, as melhores condições de moradia estão fortemente relacionadas à redução da percentagem de pobres. Em relação à melhora nas condições de moradia, em especial, verifica-se que são as melhores condições de moradia do meio rural (e não as urbanas) que mais fortemente se relacionam com a redução do grau de desigualdade da renda.

A segunda conclusão é que a produção agrícola *per capita*, ao longo das quatro décadas, foi deixando de apresentar uma correlação importante com os indicadores de desenvolvimento humano (IDHs). Isso pode ser observado nas duas últimas colunas das matrizes de correlação. No ano de 1970, por exemplo, a produção agrícola *per capita* apresentava uma correlação muito alta com a redução da percentagem de pobres (-0,94) e tam-

bém com a melhora dos IDHs. Entretanto, nos anos 1980, 1991 e 2000, essa correlação praticamente deixou de existir. Essa mudança provavelmente tem relação com a redução de participação que a agropecuária teve na economia. Em termos agregados, a sua participação caiu pela metade no período, passando de 12% para 6%.

A terceira conclusão evidencia que, em contraste com a produção agrícola *per capita*, a produção industrial *per capita* manteve, ao longo dos anos, sua forte influência no sentido da redução da percentagem de pobres e do aumento dos IDHs. Além disso, a produção industrial *per capita*, ao longo dos anos, passou a ter importante influência no sentido da redução da desigualdade, medida pelo índice de Theil, algo que também não se observa no caso da produção agrícola. Essa redução da desigualdade pode estar relacionada com os ganhos de produtividade ocasionados pela participação da indústria e as correspondentes melhorias salariais. Essa interpretação encontra respaldo nos resultados obtidos por Ferreira e Cruz (2008) com relação à renda do trabalho e sua influência sobre a redução da desigualdade.

Por fim, com relação à utilização da mão de obra nas áreas urbana e rural da economia, constata-se um vivo contraste entre essas duas áreas e sua marcada influência sobre as condições de moradia, os IDHs, a pobreza e a desigualdade, especialmente em 1991 e em 2000. Em particular, a taxa de ocupação da mão de obra no meio urbano tem forte correlação positiva com a melhora nas condições de moradia e a elevação dos IDHs. Por outro lado, a utilização da mão de obra do meio rural tem forte correlação negativa.

A participação da indústria na economia pode ter tido um papel importante no desenvolvimento econômico, pois, no período analisado, a redução da porcentagem de pobres, da desigualdade de renda e o aumento do IDH foram relativamente mais significativos nos estados mais industrializados. Por ser tipicamente um setor em que a intensidade da demanda se manifesta nos crescentes ganhos de produtividade (e na qualificação dos trabalhadores) e escala da produção, esses resultados sugerem que a indústria continua tendo papel preponderante no desenvolvimento econômico do Brasil, algo que não se constata comumente na agricultura, considerada uma atividade sujeita a retornos decrescentes à escala (THIRLWALL, 2005).

3.2 Análise de *cluster*

O passo seguinte do trabalho consiste em verificar se existem diferentes grupos de estados em termos de desenvolvimento. Para isso, o primeiro passo consiste na separação inicial dos grupos. Caso existam diferentes

grupos, será preciso verificar se, ao longo do tempo, ocorrem mudanças em suas composições e, também, se eles se aproximam ou se distanciam. Há dois casos extremos que servem de guia para a análise: o caso em que todos os estados têm o mesmo grau de desenvolvimento e estão todos situados no mesmo *cluster* e o caso em que há grande dispersão em termos de desenvolvimento e há a formação de certo número de *clusters* hierárquicos. É importante verificar se houve mudança na composição desses *clusters* ao longo do período analisado ou se os estados inicialmente menos desenvolvidos continuam, ao longo do tempo, na mesma situação relativa. Para aferir a significância das diferenças das distâncias entre os grupos ao longo dos anos, é utilizado o teste de Wilcoxon.

Na identificação dos *clusters*, o passo inicial consiste na definição da medida de distância a ser utilizada. Nesse aspecto, a maioria dos autores emprega a medida de distância mais simples, a distância euclidiana. Para tanto, considere o vetor aleatório de medidas X_j definido por:

$$\mathbf{X}_j = [X_{1j} X_{2j} \dots X_{pj}]'; \quad j = 1, 2, \dots, n \quad (1)$$

onde X_{ij} representa o valor observado da variável i medida no elemento j . Como a medida de distância reflete dissimilaridade, quanto menor seu valor, mais similares serão os casos que estão sendo comparados. A distância euclidiana entre dois elementos amostrais X_l e X_k , $l \neq k$ é definida por:

$$d(X_l, X_k) = [(X_l - X_k)'(X_l - X_k)]^{\frac{1}{2}} \quad (2)$$

Os resultados do cálculo das distâncias foram organizados em uma matriz de distâncias, $D_{25 \times 25}$, uma vez que a análise é feita para 25 estados da Federação. A partir da obtenção de $D_{25 \times 25}$, é possível aplicar um critério de ligação entre os indivíduos para a formação do dendograma, que é um gráfico em que no eixo das abscissas aparecem os *clusters* e no eixo das ordenadas aparecem as distâncias (adimensionais) calculadas entre os elementos amostrais.

Por levar em consideração os vetores de médias amostrais e baseando-se nos princípios da análise de variância em cada passo do processo de aglomeração, o método de Ward tem recebido especial interesse. Mingoti (2007, cap. 6) oferece uma discussão detalhada e uma comparação desse método com os outros critérios de aglomeração. Neste trabalho, segue-se a abordagem de Saint-Arnaud e Bernard (2003), que também empregaram o critério de Ward para a obtenção do dendograma.

Para o cálculo das matrizes de distância, as variáveis foram normalizadas, haja vista que as unidades de medida poderiam influenciar os resultados (MINGOTI, 2007). A ideia subjacente à separação em grupos é de que a partição desejada ou esperada, resultante do cálculo das distâncias e do algoritmo de aglomeração, seja a que possua a maior heterogeneidade possível **entre** os grupos (*clusters*) formados e a maior homogeneidade possível **dentro** dos grupos.

A distância euclidiana e o método de Ward de aglomeração foram a base para a apresentação dos dendogramas para todos os anos considerados no estudo (CARVALHO; SANTANA; MENDES, 2006; SAINT-ARNAUD; BERNARD, 2003). O método de Ward procura, em cada passo da aglomeração, levar em conta estrutura de variabilidade dos dados, minimizando a perda de informação em cada passo do algoritmo (JOHNSON; WICHERN, 1998, p. 751). A linha que separa os *clusters* foi estabelecida de acordo com a sugestão de Pereira (2001), em que o maior salto no algoritmo de aglomeração determina a separação dos grupos. Os resultados da análise de *cluster* são apresentados e analisados na seção quatro. Para a validação dos *clusters* identificados, utilizou-se a estatística Λ de Wilks. Para detalhes sobre a aplicação da estatística Λ de Wilks, pode-se consultar Reis (2001, cap. 9). Todas as operações computacionais foram realizadas no ambiente **R** (R DEVELOPMENT CORE TEAM, 2017) utilizando-se o pacote **R Commander** (FOX, 2005).

3.3 Validação dos resultados

A aplicação de técnicas multivariadas requer que as p variáveis em estudo sejam correlacionadas em um grau mínimo (MINGOTI, 2007; PEREIRA, 2001). Em particular, a estatística Λ de Wilks depende da existência de correlações não nulas entre as variáveis analisadas (REIS, 2001, p. 169). Admitindo-se a suposição de que o vetor aleatório \mathbf{X} segue distribuição normal p -variada, pode-se testar a hipótese de que a matriz de correlação populacional é uma identidade, contrariamente à hipótese alternativa bicaudal. Na prática, a não rejeição da hipótese nula indica que as variáveis são independentes, portanto, inadequadas para a análise de correlação. A estatística de teste é dada por:

$$T = - \left\{ n - \frac{1}{6} (2p + 11) \right\} \left\{ \sum_{j=1}^p \ln(\hat{\lambda}_j) \right\} \quad (3)$$

onde T é a estatística de teste, n é o tamanho da amostra, p é o número de variáveis, $\hat{\lambda}_j$ são os autovalores estimados a partir da matriz de correlação amostral do vetor X . Para n suficientemente grande, a estatística T tem uma distribuição qui-quadrado com $\frac{1}{2} p(p-1)$ graus de liberdade.

Ademais, a separação em diferentes grupos de estados, que apresentam diferentes indicadores de desenvolvimento, pode ser testada quanto à sua significância estatística. Em especial, o pesquisador pode estar interessado em testar a hipótese nula de que não há diferença entre os grupos identificados. A estatística Λ de Wilks, a partir do quociente de verossimilhanças, não se encontra tabelada, mas pode ser aproximada por:

$$\left\{ \frac{(n-k) - p + 1}{p} \right\} \left(\frac{1 - \Lambda}{\Lambda} \right) \tag{4}$$

que segue aproximadamente uma distribuição F com $(p, n - k - p + 1)$ graus de liberdade, em que p é o número de variáveis, k é o número de grupos, e n , o número de indivíduos. A estatística Λ é obtida de acordo com a expressão

$$\Lambda = \prod_{j=1}^s (1 + \lambda_j)^{-1} \tag{5}$$

em que λ_j são os autovalores do produto de matrizes $(W^{-1}B)$ e s é o número de autovalores da matriz resultante, que é definida sob a condição de que $n \geq p + k$, onde n é o tamanho da amostra, p é o número de variáveis, e k é quantidade de grupos. A matriz W expressa os produtos cruzados entre os grupos e dentro dos grupos, definida por:

$$W = \sum_{j=1}^k \sum_{u=1}^{n_j} (X_{ju} - \bar{X}_j)(X_{ju} - \bar{X}_j)' \tag{6}$$

Por sua vez, a matriz B é definida como

$$B = \sum_{j=1}^k n_j (\bar{X}_j - \bar{X})(\bar{X}_j - \bar{X})' \tag{7}$$

em que $\bar{X}_{\sim j} = \frac{\sum_{u=1}^{n_j} X_{\sim ju}}{n_j}$ é o vetor de médias amostrais para o grupo j ,

$\bar{X}_{\sim} = \frac{\sum_{j=1}^k \sum_{u=1}^{n_j} X_{\sim ju}}{n}$ é o vetor de médias amostrais para a totalidade da amos-

tra, e $n = \sum_{j=1}^k n_j$ é o tamanho total da amostra. Nesse contexto, as hipóte-

ses nula e alternativa a serem testadas são as seguintes:

$$H_0 : \mu_{\sim 1} = \mu_{\sim 2} = \dots = \mu_{\sim k} \text{ e } H_1 : \mu_{\sim i} \neq \mu_{\sim j}; \quad i \neq j.$$

Visando validar a detecção dos *clusters* pelos critérios escolhidos, procedeu-se ao teste para diferença entre os grupos, dado pela estatística Λ de Wilks (equações 4 e 5). Os resultados são apresentados na Tabela 1. Em todos os casos, constata-se que os grupos diferem estatisticamente a 0,01 de significância.

Tabela 1

Resultados do teste para a diferença entre os grupos de estados do Brasil — 1970, 1980, 1991 e 2000

ANO	ESTATÍSTICA DE WILKS $\hat{\Lambda}$	ESTATÍSTICA F COM ($p, n - k - p + 1$) GRAUS DE LIBERDADE	P-VALOR DO TESTE
1970	0,0008074	(1)21,98	0,0000
1980	0,0001152	(1)11,82	0,0000
1991	0,0000500	(1)16,09	0,0000
2000	0,0179000	(1)39,19	0,0000

FONTE DOS DADOS BRUTOS: IPEA (2013).

(1) Estatisticamente significativo a 0,01 de significância.

A Tabela 2 apresenta os resultados do teste de Bartlett para a matriz de correlação populacional. Em todos os casos, a hipótese nula de ausência de correlação pode ser rejeitada a 0,01 de significância.

Sobre a hipótese de convergência ou divergência aludida anteriormente, alguma evidência pode ser obtida pelo teste de Wilcoxon (discutido em Siegel e Castellan (1988)) para a igualdade das distâncias e contra a hipótese alternativa específica de divergência ou convergência. Assim, com base nos dados deste estudo e levando-se em consideração as diferentes dimensões do desenvolvimento vistas anteriormente, pode-se verificar se a

distância entre os estados vem aumentando, configurando certa divergência, vem diminuindo, configurando algum tipo de convergência, ou se está estabilizada em termos de desenvolvimento econômico.

Tabela 2

Resultado do Teste de Bartlett para a matriz de correlação populacional em estados do Brasil — 1970, 1980, 1991 e 2000

ANO	ESTATÍSTICA \hat{T}	P-VALOR
1970	(1)573,05	0,0000
1980	(1)468,81	0,0000
1991	(1)488,84	0,0000
2000	(1)515,07	0,0000

FONTES DOS DADOS BRUTOS: IPEA (2013).

(1) Estatisticamente significativo a 0,01 de significância.

Seguindo o trabalho de Gligor e Ausloos (2011, p. 12-13), a partir do cálculo das quatro matrizes de distâncias, as distâncias euclidianas foram organizadas em vetores, e, então, foi utilizado o teste de Wilcoxon (para dados pareados) para se testar a hipótese de igualdade das distâncias ao longo dos anos. Os resultados dos testes de hipóteses encontram-se nas Tabelas 3 e 4. Esses resultados consideram a hipótese nula como igualdade das distâncias. Na hipótese alternativa, têm-se os resultados para o caso unicaudal à esquerda (divergência), como postulado teoricamente por Myrdal (1968) e Kaldor (2013).

4 Resultados e discussão

Considerando-se as diferentes dimensões do desenvolvimento contempladas neste trabalho, pode-se dizer que o Brasil ficou mais desigual no período de 1970 a 2000, comparando-se o grupo de estados formado pelas Regiões Norte e Nordeste com o grupo de estados do Sudeste, Sul e Centro-Oeste. Todavia, o desenvolvimento econômico tornou-se mais igualitário entre os estados dentro desses dois grupos: pode-se dizer que houve convergência dentro dos *clusters*, e divergência entre os dois *clusters* de desenvolvimento, que, ao longo dos anos, tornaram-se mais definidos e distantes em termos de desenvolvimento. Considerando, conforme Pereira (2001), que o maior salto no eixo das ordenadas determina a separação dos grupos⁸, o Brasil, em 1970, tinha dois grupos de estados com desenvolvi-

⁸ Os dendogramas das Figuras 1, 2, 3 e 4 mostram, nos eixos das ordenadas, a distância euclidiana multidimensional entre os grupos de unidades da Federação (distância conside-

mento relativamente similar, em 1980, tinha três grupos, em 1991, dois grupos, e, em 2000, também, dois grupos.

O Quadro A.1 apresenta os valores médios das 14 variáveis consideradas por *cluster* e por período. Essas informações são úteis para acompanhar a evolução das distâncias (diferenças) entre todas as unidades estaduais analisadas, verificar as diferenças entre os *clusters* formados e acompanhar a evolução dos níveis de desenvolvimento desses *clusters*. Analisando-se mais detalhadamente os grupos formados ao longo do tempo, a Figura 1 mostra a posição dos estados brasileiros com relação às variáveis medidas em 1970. Constata-se que as distâncias entre os *clusters* não eram substanciais, entretanto, eram suficientes para gerar, pelo menos, dois grupos de estados, em que os mais dessemelhantes em relação aos demais estão situados no *cluster* mais próximo da origem. Observa-se que a distância que separa esse *cluster* do outro está próxima de 30.

Figura 1

Dendograma dos resultados da aglomeração de estados do Brasil — 1970



O primeiro *cluster* (designado G1-1970) era formado pelos Estados de São Paulo, Rio de Janeiro e Distrito Federal, e o segundo (o G2-1970) era formado pelos demais estados brasileiros. Dentro do segundo grupo havia dois subgrupos: o primeiro formado pelos demais estados do Sudeste (Minas Gerais e Espírito Santo) e os estados do Sul, do Centro-Oeste (exceto Distrito Federal) e do Norte (exceto Acre). O segundo subgrupo era formado

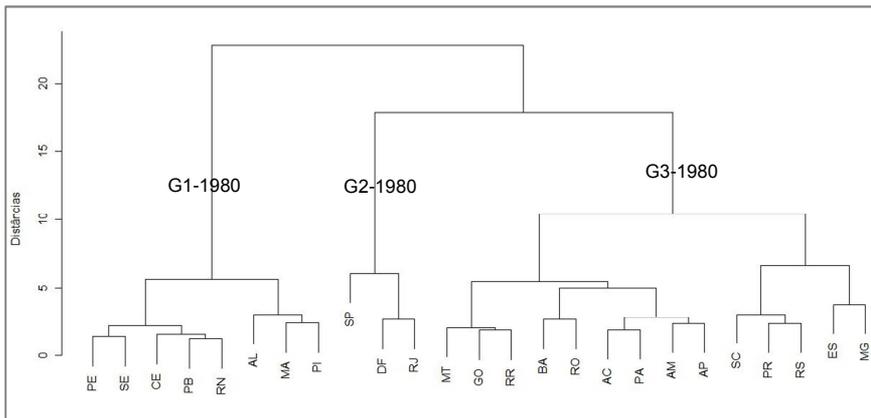
rando as 14 variáveis usadas como indicadoras de desenvolvimento). Nos eixos das abscissas, os dendogramas mostram as unidades da Federação e os grupos a que elas pertencem. A separação dos grupos foi validada pela estatística Λ de Wilks, cujos resultados aparecem na Tabela 1.

pelos estados do Nordeste (mais Acre). Observa-se que há grande homogeneidade dentro dos *clusters*, pois as distâncias que os separam, medidas no eixo das ordenadas, são muito pequenas se comparadas com as distâncias que separam as aglomerações (saltos). Esse é um resultado desejável do ponto de vista dos princípios da aglomeração.

Ao longo do tempo, houve uma divisão no grupo dois (G2-1970), formando três *clusters*, como se observa na Figura 2. Assim, em 1980, o primeiro grupo (G1-1980) passou a ser formado pelos estados do Nordeste (exceto Bahia), o segundo grupo (G2-1980) ficou com a mesma composição do primeiro grupo do período anterior, ou seja, com os Estados de São Paulo, Rio de Janeiro e o Distrito Federal, e o terceiro grupo (G3-1980) passou a ser composto pelos demais estados do Sudeste (Minas Gerais e Espírito Santo), mais os estados do Sul, do Centro-Oeste (exceto Distrito Federal), do Norte e o Estado da Bahia.

Figura 2

Dendograma dos resultados da aglomeração de estados do Brasil — 1980



Com o tempo, o terceiro grupo de 1980 foi se dividindo com os estados da Região Norte e o Estado da Bahia convergindo para o *cluster* formado pelos estados do Nordeste (G1-1991; G1-2000) e os demais estados convergindo para o *cluster* formado pelos Estados de São Paulo, do Rio de Janeiro e o Distrito Federal (G2-1991 e G2-2000).

O resultado disso pode ser visto tanto na Figura 3, referente a 1991, quanto na Figura 4, referente a 2000, em que os mesmos grupos são constituídos pelos mesmos estados da Federação. Nessas duas figuras percebe-se a persistência de dois grupos distintos de estados em termos de desenvolvimento: o grupo mais desenvolvido, formado pelos estados das Regiões

Sudeste, Sul e Centro-Oeste, e o grupo menos desenvolvido, composto pelos estados da Região Norte e Nordeste.

Observando os valores no eixo das ordenadas desses dois grupos, nos anos de 1991 e 2000, percebe-se que a distância entre eles aumentou, pois estava ao redor de 30 em 1991, ao passo que, em 2000, ela aumentou para aproximadamente 35. Por outro lado, as distâncias dentro dos grupos diminuíram de um ano para o outro.

Figura 3

Dendrograma dos resultados da aglomeração de estados do Brasil — 1991

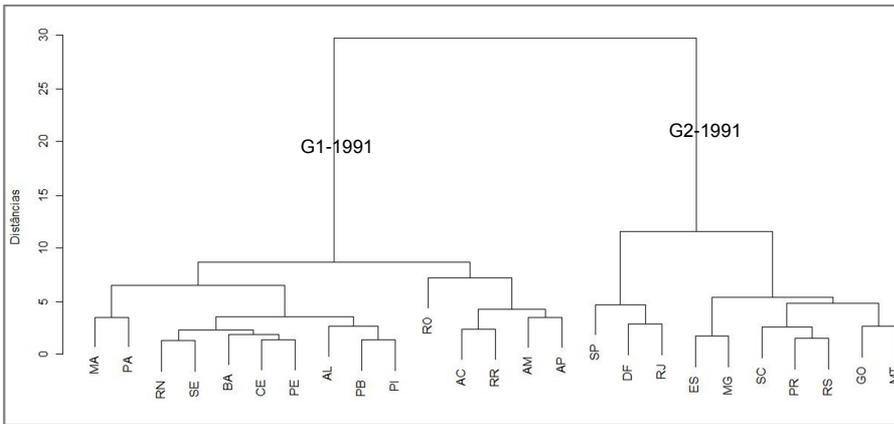
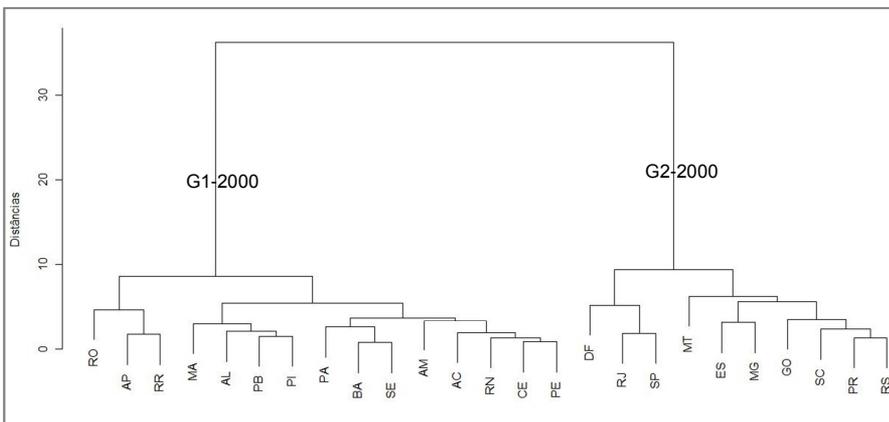


Figura 4

Dendrograma dos resultados da aglomeração de estados do Brasil — 2000



Avaliando-se a evolução dos agrupamentos formados ao longo do tempo, em linhas gerais, pode-se dizer que dois grupos ficaram estáveis em situações opostas: por um lado, o grupo formado pelos Estados de São Paulo, Rio de Janeiro e o Distrito Federal e, por outro lado, o grupo formado pelos estados do Nordeste. Entre os dois flutuaram os demais estados e, ao longo do tempo, os estados do Norte convergiram para o grupo do Nordeste, enquanto que os demais estados convergiram para o primeiro grupo.

A seguir, nas Tabelas 3 e 4, são apresentados os resultados para as diferenças das distâncias entre os estados brasileiros ao longo dos anos considerados (1970, 1980, 1991 e 2000). Como observado anteriormente, foi utilizado o teste não paramétrico de Wilcoxon para dados pareados, uma vez que não depende das suposições de normalidade e homogeneidade das variâncias. Na Tabela 3 aparecem os resultados para o teste da hipótese de igualdade contra a hipótese bicaudal, e, na Tabela 4, os resultados referem-se ao teste de igualdade contra a hipótese unicaudal à esquerda, que indica divergência ao longo dos anos.

Tabela 3

Resultado para a diferença das distâncias entre os estados do Brasil — 1970, 1980, 1991 e 2000

ANO	MEDIANA
1970	3,91 ^A
1980	4,43 ^B
1991	4,70 ^B
2000	4,77 ^B

FONTE DOS DADOS BRUTOS: IPEA (2013).

NOTA: Os valores das medianas com os mesmos sobrescritos não diferem a 10% de significância, e os com sobrescritos distintos diferem estatisticamente a 10% de significância.

Tabela 4

Resultado para a diferença das distâncias entre os estados do Brasil — 1970, 1980, 1991 e 2000

ANO	MEDIANA
1970	3,91 ^A
1980	4,43 ^B
1991	4,70 ^B
2000	4,77 ^B

FONTE DOS DADOS BRUTOS: IPEA (2013).

NOTA: Os valores das medianas com os mesmos sobrescritos não diferem a 10% de significância, e os com sobrescritos distintos diferem estatisticamente a 10% de significância.

Os resultados apresentados acima permitem constatar que: primeiro, as distâncias entre os estados brasileiros, em termos de desenvolvimento

econômico, aumentaram ao longo do tempo: esse aumento é estatisticamente significativo. Segundo, como indica a Tabela 4, a hipótese divergência vem ganhando respaldo ao longo do tempo, como referida por Myrdal (1968) e Kaldor (2013), uma vez que a hipótese nula de igualdade foi rejeitada em todos os casos, nas comparações múltiplas, em favor da hipótese alternativa de divergência no desenvolvimento (unicaudal à direita).

Outra constatação, corroborada pela análise de *cluster* anteriormente, é que, no ano de 1970, a distância entre os estados era menor quando comparada com a configuração do ano 2000. Embora não existam diferenças significativas entre as distâncias dos anos 1980, 1991 e 2000, o comportamento da mediana parece indicar claramente o seu aumento ao longo do tempo. De 1991, em termos de mediana, as distâncias aumentaram 1,49%. Porém, de 1980 para 1991, esse aumento foi de 6,09%, que é sensivelmente maior que o período posterior.

Em síntese, os resultados encontrados são consistentes com as evidências apresentadas por Mossi *et al.* (2003) e Silveira-Neto e Azzoni (2006). Para esses autores, em vista da forte dependência espacial, ao final dos anos 90, o Brasil constituiu-se de duas regiões bem definidas em termos de renda *per capita*: Sul-Sudeste e Norte-Nordeste. Os resultados deste trabalho reforçam essas conclusões, com a inclusão dos estados da Região Centro-Oeste no *cluster* de estados das Regiões Sul e Sudeste. Outra diferença deste estudo é que ele não ficou restrito ao âmbito da renda *per capita*, incorporando mais variáveis e dimensões do desenvolvimento.

Vários são os fatores que podem causar a desigualdade entre os dois diferentes *clusters* de regiões identificados neste estudo Sul-Sudeste-Centro-Oeste e Norte-Nordeste. Entre eles podem se levantados os seguintes: diferenças no grau de industrialização, diferenças de capital humano, infraestrutura, entre outros. No entanto, a metodologia usada neste estudo possibilita uma análise multivariada, mas não permite que se estabeleçam relações de causalidade.

5 Conclusões

O principal objetivo do trabalho foi realizar uma análise exploratória de uma base de dados para a economia brasileira em âmbito estadual, levando-se em conta múltiplos critérios. No total, 14 variáveis que expressam seis dimensões do desenvolvimento econômico, como desenvolvimento industrial, produtividade agrícola, capital humano, pobreza, desigualdade e condições de moradia foram consideradas na análise.

A partir dos resultados da análise de *cluster*, foi possível constatar que o País está diminuindo a heterogeneidade *dentro* de dois grupos distintos

de estados. Há, porém, um aumento das distâncias entre esses grupos de estados. Após quatro décadas de desenvolvimento econômico, apenas dois grupos de estados configuram-se no Brasil: as macrorregiões Sul-Sudeste-Centro-Oeste e a Norte-Nordeste. Há evidência de que ambas, internamente, estão convergindo em várias dimensões do desenvolvimento, porém, enquanto grupo, as duas macrorregiões divergem entre si. É importante notar que essas diferenças são estatisticamente significativas, configurando um processo sistemático ao longo do tempo e não casual.

Nota-se que, a despeito da relativa similaridade entre os grupos existente no início da década de 70, as diferenças acentuaram-se ao longo dos últimos 40 anos de política econômica (a mediana das distâncias aumentou ao longo do tempo). Essa conclusão deriva do teste para a diferença das distâncias euclidianas, obtidas a partir da matriz de dados padronizada para cada ano analisado.

Assim, ao longo do tempo, as diferenças entre os grupos foram se acumulando, culminando na formação de apenas dois grupos distintos entre si. Os resultados dos testes para diferença das médias permitem concluir, com relação às condições iniciais, que a distância entre os estados brasileiros, tendo por base o ano de 1970, aumentou sistematicamente, apesar de esse aumento não ser estatisticamente significativo quando são comparados os anos 1980, 1991 e 2000.

A estatística Λ de Wilks, empregada para testar a significância das diferenças entre os grupos, indicou a formação sistemática de grupos diferentes ao longo dos anos. Em particular, apesar da migração dos estados de alguns *clusters* iniciais para outros mais desenvolvidos, as diferenças entre os grupos são estatisticamente significativas, mesmo quando se considera a formação de apenas dois *clusters* (duas macrorregiões, Sul-Sudeste-Centro-Oeste e Norte-Nordeste).

Um trabalho futuro poderá investigar os principais fatores históricos, políticos e institucionais que levaram o País a esse resultado. Nesse contexto, será possível questionar se estão sendo efetivas as políticas públicas e os programas sociais implantados para alavancar o desenvolvimento de áreas deprimidas e para reduzir as desigualdades regionais.

Embora o *cluster* de estados das Regiões Norte e Nordeste tenha sido o principal alvo de políticas de redução da pobreza e de melhorias do desenvolvimento no Brasil desde as décadas de 1960 e 1970, com a metodologia usada neste trabalho não foi possível afirmar se essas políticas foram ou não efetivas. Para tanto, seria necessário usar um método que possibilitasse acompanhar a trajetória das regiões sem a presença das referidas políticas.

Apêndice

Quadro A.1

Médias das variáveis e número de unidades da Federação por *cluster* — 1970, 1980, 1991 e 2000

1970															
Clusters	Krrpc	krupc	cerpc	ceipc	theil	p0	idhl	idhe	idhr	yappc	yatpc	empurb	emprur	pindpc	N° UFs
G1-1970	1,647	8,309	0,210	0,307	0,497	37,897	0,473	0,636	0,843	0,030	1,327	0,289	0,037	1,479	3
G2-1970	0,904	3,725	0,044	0,073	0,497	80,214	0,418	0,423	0,275	0,067	0,636	0,118	0,174	0,355	22
TODOS	0,993	4,275	0,064	0,101	0,497	75,136	0,425	0,449	0,343	0,063	0,719	0,138	0,158	0,490	25
1980															
Clusters	Krrpc	krupc	cerpc	ceipc	theil	p0	idhl	idhe	idhr	yappc	yatpc	empurb	emprur	pindpc	N° UFs
G1-1980	0,645	3,678	0,089	0,189	0,615	69,474	0,467	0,394	0,438	0,069	0,256	0,152	0,156	0,628	8
G2-1980	2,232	10,136	0,572	0,776	0,580	16,363	0,549	0,699	0,963	0,069	0,135	0,364	0,027	3,451	3
G3-1980	1,130	4,808	0,182	0,371	0,571	43,481	0,543	0,556	0,801	0,108	0,467	0,198	0,135	1,739	14
TODOS	1,107	5,086	0,199	0,361	0,586	48,544	0,520	0,521	0,704	0,091	0,360	0,203	0,129	1,589	25
1991															
Clusters	Krrpc	krupc	cerpc	ceipc	theil	p0	idhl	idhe	idhr	yappc	yatpc	empurb	emprur	pindpc	N° UFs
G1-1991	0,614	3,426	0,200	0,363	0,718	58,960	0,607	0,645	0,588	0,116	0,275	0,206	0,118	0,992	15
G2-1991	1,455	6,827	0,397	0,611	0,635	30,381	0,698	0,797	0,699	0,111	0,391	0,327	0,077	2,387	10
TODOS	0,950	4,786	0,278	0,462	0,685	47,528	0,643	0,706	0,633	0,114	0,321	0,254	0,101	1,550	25
2000															
Clusters	Krrpc	krupc	cerpc	ceipc	theil	p0	idhl	idhe	idhr	yappc	yatpc	empurb	emprur	pindpc	N° UFs
G1-2000	0,931	4,531	0,304	0,383	0,748	52,074	0,678	0,783	0,628	0,045	0,109	0,241	0,101	0,994	15
G2-2000	3,098	9,806	0,568	0,744	0,653	22,153	0,757	0,886	0,752	0,086	0,371	0,353	0,064	2,542	10
TODOS	1,798	6,641	0,410	0,527	0,710	40,106	0,710	0,824	0,678	0,062	0,214	0,286	0,086	1,613	25

Quadro A.2

Matrizes de correlação — 1970, 1980, 1991 e 2000

1970	<i>krrpc</i>	<i>krupc</i>	<i>cerpc</i>	<i>ceipc</i>	<i>theil</i>	<i>p0</i>	<i>idhl</i>	<i>idhe</i>	<i>idhr</i>	<i>yappc</i>	<i>yatpc</i>	<i>empurb</i>	<i>emprur</i>	<i>pindpc</i>
<i>krrpc</i>	1,00													
<i>krupc</i>	0,73	1,00												
<i>cerpc</i>	0,73	0,85	1,00											
<i>ceipc</i>	0,46	0,67	0,67	1,00										
<i>theil</i>	-0,27	0,18	-0,03	0,23	1,00									
<i>p0</i>	-0,88	-0,86	-0,89	-0,54	0,19	1,00								
<i>idhl</i>	0,71	0,69	0,53	0,38	-0,12	-0,57	1,00							
<i>idhe</i>	0,87	0,85	0,76	0,50	-0,08	-0,84	0,83	1,00						
<i>idhr</i>	0,87	0,88	0,92	0,57	-0,08	-0,98	0,59	0,84	1,00					
<i>yappc</i>	-0,51	-0,27	-0,29	-0,11	0,47	0,38	-0,50	-0,41	-0,36	1,00				
<i>yatpc</i>	0,81	0,78	0,85	0,52	-0,11	-0,94	0,42	0,67	0,96	-0,36	1,00			
<i>empurb</i>	0,73	0,88	0,93	0,60	0,17	-0,89	0,47	0,76	0,94	-0,19	0,88	1,00		
<i>emprur</i>	-0,60	-0,73	-0,86	-0,47	-0,18	0,79	-0,31	-0,67	-0,82	0,07	-0,75	-0,93	1,00	
<i>pindpc</i>	0,65	0,59	0,77	0,82	-0,10	-0,65	0,43	0,57	0,65	-0,18	0,60	0,62	-0,54	1,00
1980	<i>krrpc</i>	<i>krupc</i>	<i>cerpc</i>	<i>ceipc</i>	<i>theil</i>	<i>p0</i>	<i>idhl</i>	<i>idhe</i>	<i>idhr</i>	<i>yappc</i>	<i>yatpc</i>	<i>empurb</i>	<i>emprur</i>	<i>pindpc</i>
<i>krrpc</i>	1,00													
<i>krupc</i>	0,83	1,00												
<i>cerpc</i>	0,68	0,84	1,00											
<i>ceipc</i>	0,68	0,62	0,33	1,00										
<i>theil</i>	-0,38	0,06	-0,08	-0,17	1,00									
<i>p0</i>	-0,86	-0,83	-0,82	-0,49	0,31	1,00								
<i>idhl</i>	0,65	0,51	0,46	0,35	-0,47	-0,78	1,00							
<i>idhe</i>	0,84	0,79	0,76	0,45	-0,27	-0,96	0,81	1,00						
<i>idhr</i>	0,74	0,72	0,68	0,46	-0,21	-0,95	0,79	0,92	1,00					
<i>yappc</i>	0,07	0,08	-0,05	0,47	0,16	-0,04	0,11	0,01	0,12	1,00				
<i>yatpc</i>	0,21	0,01	-0,15	0,01	-0,15	-0,24	0,41	0,22	0,38	-0,03	1,00			
<i>empurb</i>	0,83	0,96	0,88	0,54	0,03	-0,88	0,50	0,85	0,79	0,00	0,06	1,00		
<i>emprur</i>	-0,67	-0,84	-0,82	-0,46	-0,09	0,74	-0,26	-0,72	-0,63	0,09	0,23	-0,90	1,00	
<i>pindpc</i>	0,84	0,67	0,53	0,70	-0,41	-0,75	0,56	0,70	0,65	0,16	0,15	0,67	-0,52	1,00
1991	<i>krrpc</i>	<i>krupc</i>	<i>cerpc</i>	<i>ceipc</i>	<i>theil</i>	<i>p0</i>	<i>idhl</i>	<i>idhe</i>	<i>idhr</i>	<i>yappc</i>	<i>yatpc</i>	<i>empurb</i>	<i>emprur</i>	<i>pindpc</i>
<i>krrpc</i>	1,00													
<i>krupc</i>	0,86	1,00												
<i>cerpc</i>	0,84	0,88	1,00											
<i>ceipc</i>	0,48	0,29	0,17	1,00										
<i>theil</i>	-0,63	-0,28	-0,51	-0,43	1,00									
<i>p0</i>	-0,79	-0,75	-0,93	-0,12	0,63	1,00								
<i>idhl</i>	0,76	0,73	0,83	0,16	-0,60	-0,91	1,00							
<i>idhe</i>	0,77	0,74	0,89	0,16	-0,58	-0,96	0,92	1,00						
<i>idhr</i>	0,78	0,80	0,94	0,06	-0,48	-0,98	0,87	0,94	1,00					
<i>yappc</i>	0,02	-0,06	0,00	0,01	-0,06	-0,05	0,04	0,13	0,03	1,00				
<i>yatpc</i>	-0,01	-0,12	-0,02	-0,19	-0,27	-0,16	0,13	0,12	0,08	0,37	1,00			
<i>empurb</i>	0,80	0,90	0,91	0,07	-0,44	-0,88	0,82	0,84	0,89	-0,03	0,04	1,00		
<i>emprur</i>	-0,58	-0,67	-0,75	0,05	0,23	0,67	-0,56	-0,61	-0,72	0,21	0,35	-0,80	1,00	
<i>pindpc</i>	0,58	0,64	0,67	0,40	-0,47	-0,60	0,61	0,58	0,56	-0,01	-0,08	0,58	-0,44	1,00

(continua)

Quadro A.2

Matrizes de correlação — 1970, 1980, 1991 e 2000

2000	<i>krpc</i>	<i>krupc</i>	<i>cerpc</i>	<i>ceipc</i>	<i>theil</i>	<i>p0</i>	<i>idhl</i>	<i>idhe</i>	<i>idhr</i>	<i>yappc</i>	<i>yatpc</i>	<i>empurb</i>	<i>emprur</i>	<i>pindpc</i>
<i>krpc</i>	1,00													
<i>krupc</i>	0,95	1,00												
<i>cerpc</i>	0,85	0,82	1,00											
<i>ceipc</i>	0,29	0,33	0,15	1,00										
<i>theil</i>	-0,50	-0,38	-0,62	-0,33	1,00									
<i>p0</i>	-0,85	-0,80	-0,92	-0,23	0,79	1,00								
<i>idhl</i>	0,72	0,69	0,77	0,29	-0,67	-0,87	1,00							
<i>idhe</i>	0,77	0,72	0,90	0,18	-0,71	-0,93	0,84	1,00						
<i>idhr</i>	0,91	0,88	0,96	0,18	-0,64	-0,97	0,84	0,92	1,00					
<i>yappc</i>	0,08	0,13	0,04	0,37	-0,36	-0,25	0,15	0,17	0,17	1,00				
<i>yatpc</i>	0,10	0,01	0,16	-0,02	-0,29	-0,32	0,31	0,23	0,24	-0,03	1,00			
<i>empurb</i>	0,87	0,85	0,93	0,20	-0,66	-0,96	0,83	0,90	0,96	0,13	0,29	1,00		
<i>emprur</i>	-0,56	-0,58	-0,79	0,07	0,22	0,60	-0,53	-0,67	-0,69	0,27	-0,02	-0,72	1,00	
<i>pindpc</i>	0,54	0,56	0,62	0,50	-0,51	-0,61	0,63	0,56	0,58	0,16	0,04	0,57	-0,37	1,00

Referências

ARYEETEEY, E.; OWUSU, G.; MENSAH, E. J. **An Analysis of Poverty and Regional Inequalities in Ghana**. New Delhi: GDN, 2009. (Working Paper Series, n. 27).

BRIDA, J. G. *et al.* An Alternative View of the Convergence Issue of Growth Empirics. **Growth and Change**, Lexington, v. 42, n. 3, p. 320-350, 2011.

CARVALHO, D. F.; SANTANA, A. C.; MENDES, F. A. T. Análise de *cluster* da indústria de móveis de madeira do Pará. **Novos Cadernos NAEA**, Belém, v. 9, n. 2, p. 25-54, 2006.

CRACOLICI, M. F.; CUFFARO, M.; NIJKAMP, P. The measurement of economic, social and environmental performance of countries: a novel approach. **Social Indicators Research**, [S.l.], v. 95, p. 339-356, 2010.

DIXON, R.; THIRLWALL, A. P. A model of regional growth-rate differences on kaldorian lines. **Oxford Economic Papers (New Series)**, Oxford, v. 27, n. 2, p. 201-214, 1975.

DURLAUF, N. S.; JOHNSON, P. Multiple regimes and cross country growth behavior. **Journal of Applied Econometrics**, Chichester, v. 10, n. 4, p. 365-384, 1995.

FERREIRA, R. T.; CRUZ, M. S. Clubes de convergência na desigualdade de renda nos municípios brasileiros. In: ENCONTRO NACIONAL DE ECONOMIA, 36., 2008, Salvador, BA. **Anais...** Niterói: Anpec, 2008. Disponível em:

<<http://www.anpec.org.br/encontro2008/artigos/200807171159380-.pdf>>.

Acesso em: 30 out. 2017.

FOX, J. The R Commander: a basic statistics graphical user interface to R. **Journal of Statistical Software**, [S.l.], v. 14, n. 9, p. 1-42, 2005.

GALOR, O. Convergence? Inference from Theoretical Models. **Economic Journal**, Cambridge, v. 106, n. 437, p. 1056-1069, 1996.

GAZONATO, M. C.; GOMES, A. L.; REIS, R. R. de G. Desigualdade de Renda no Brasil: uma análise da convergência da renda per capita dos estados para o período de 1994 e 2010. **Revista de Economia**, Curitiba, v. 40, n. 2, p. 81-104, 2014.

GLIGOR, M.; AUSLOOS, M. **Convergence and cluster structures in EU area according to fluctuations in macroeconomic series**. 2011. Disponível em:

<<http://arxiv.org/ftp/arxiv/papers/0805/0805.3071.pdf>>.

Acesso em: 20 out. 2011.

GOMES, R. R.; ESPERIDIÃO, F. Convergência de renda: uma análise em painel para as regiões brasileiras no período 1995-2009. **Ensaios FEE**, Porto Alegre, v. 37, n. 1, p. 115-144, 2016.

GREMAUD, A. P.; VASCONCELLOS, M. A. S.; TONETO JÚNIOR, R. **Economia Brasileira Contemporânea**. São Paulo: Atlas, 2010.

GUIMARÃES, P. M.; ALMEIDA, E. A análise de convergência de renda no Brasil e o problema de escala espacial. **Ensaios FEE**, Porto Alegre, v. 37, n. 4, p. 899-924, 2017.

HOBijn, B.; FRANCES, P. H. Are living standards converging? **Structural Change and Economic Dynamics**, Amsterdam, v. 12, p. 171-200, 2001.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). **Ipeadata**. 2013. Disponível em: <www.ipeadata.gov.br>. Acesso em: 17 out. 2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice Hall, 1998.

KALDOR, N. The case for regional policies. **Scottish Journal of Political Economy**, Harlow, v. 60, n. 5, p. 481-491, 2013.

LUCAS, R. On the mechanics of economic development. **Journal of Monetary Economics**, Amsterdam, v. 22, p. 3-42, 1988.

MARQUES, A. M. Crescimento e grau de desigualdade no Rio Grande do Sul: uma lição a partir dos anos 1990. **Estudos do Cepe**, Santa Cruz do Sul, n. 30, p. 6-29, 2009.

MARTIN, R.; SUNLEY, P. Slow Convergence? New Endogenous Growth Theory and Regional Development. **Economic Geography**, Worcester, v. 74, n. 3, p. 201-227, 1998.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: UFMG, 2007.

MOREIRA, R. C.; BRAGA, M. J.; TOYOSHIMA, S. H. Crescimento e desigualdade: prosperidade versus armadilhas da pobreza no desenvolvimento econômico dos estados brasileiros. In: ENCONTRO NACIONAL DE ECONOMIA, 38., 2010, Salvador, BA. **Anais...** Niterói: Anpec, 2010. Disponível em: <www.anpec.org.br/encontro2010>. Acesso em: 5 jan. 2012.

MOSSI, M. B. *et al.* Growth dynamics and space in Brazil. **International Regional Science Review**, Philadelphia, v. 26, n. 3, p. 393-418, 2003.

MYRDAL, G. **Teoria Econômica e Regiões Subdesenvolvidas**. Rio de Janeiro: Saga, 1968.

PEREIRA, J. C. R. **Análise de dados qualitativos**. São Paulo: Edusp, 2001.

QUAH, D. T. Empirics for economic growth and convergence. **European Economic Review**, Amsterdam, v. 40, n. 6, p. 1353-1375, 1996.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna, 2017. Disponível em: <<http://www.R-project.org/>>. Acesso em: 30 out. 2017.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa: Silabo, 2001.

SAINT-ARNAUD, S.; BERNARD, P. Convergence or Resilience? A Hierarchical Cluster Analysis of the Welfare Regimes in Advanced Economies. **Current Sociology**, London, v. 51, n. 5, p. 499-527, 2003.

SETTERFIELD, M. 'History versus equilibrium' and the theory of economic growth. **Cambridge Journal of Economics**, London, v. 21, p. 365-378, 1997.

SIEGEL, S.; CASTELLAN, N. J. **Nonparametric Statistics for the Behavioral Sciences**. New York: McGraw-Hill, 1988.

SILVEIRA-NETO, R.; AZZONI, C. R. Location and regional income disparity dynamics: the Brazilian case. **Papers in Regional Science**, Urbana, Ill., v. 85, n. 4, p. 599-613, 2006.

THIRLWALL, A. P. **A natureza do crescimento econômico**. Brasília, DF: IPEA, 2005.

WORLD BANK. **World Development Indicators**. Washington, D.C.: World Bank, 2001.

