

de Carvalho, Alexandre Ywata; de Mendonça, Mário Jorge Cardoso; da Silva, José Jaime

Working Paper

Avaliando o efeito dos investimentos em telecomunicações sobre o PIB

Texto para Discussão, No. 2336

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: de Carvalho, Alexandre Ywata; de Mendonça, Mário Jorge Cardoso; da Silva, José Jaime (2017) : Avaliando o efeito dos investimentos em telecomunicações sobre o PIB, Texto para Discussão, No. 2336, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/177552>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2336

**AVALIANDO O EFEITO
DOS INVESTIMENTOS EM
TELECOMUNICAÇÕES SOBRE O PIB**

TEXTO PARA DISCUSSÃO

**Alexandre Ywata de Carvalho
Mário Jorge Mendonça
José Jaime da Silva**



AVALIANDO O EFEITO DOS INVESTIMENTOS EM TELECOMUNICAÇÕES SOBRE O PIB¹

Alexandre Ywata de Carvalho²
Mário Jorge Mendonça³
José Jaime da Silva⁴

1. Os autores agradecem os valiosos comentários de Hélio Maurício M. da Fonseca, Felissa Sousa Alarcon, Luciano Chalita de Freitas e dos demais integrantes da equipe técnica da Agência Nacional de Telecomunicações (Anatel). Os autores são também muito agradecidos ao grupo de estatísticos da Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea, formado por Janaína Carvalho dos Santos, Fernando Henrique A. Esteves e Wesley de Jesus Silva, pelo trabalho de formatação e manutenção da base de dados. Em especial, registram-se agradecimentos ao conselheiro Igor Vilas Boas de Freitas da Anatel, por disponibilizar as informações acerca de infraestrutura de banda larga.

2. Diretor da Dirur/Ipea. *E-mail*: <alexandre.ywata@ipea.gov.br>.

3. Técnico de planejamento e pesquisa na Dirur/Ipea. *E-mail*: <mario.mendonca@ipea.gov.br>.

4. Assistente de pesquisa na Dirur/Ipea. *E-mail*: <jaime.silva@ipea.gov.br>.

Governo Federal

Ministério do Planejamento, Desenvolvimento e Gestão

Ministro Dyogo Henrique de Oliveira

ipea

**Instituto de Pesquisa
Econômica Aplicada**

Fundação pública vinculada ao Ministério do Planejamento, Desenvolvimento e Gestão, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Ernesto Lozardo

Diretor de Desenvolvimento Institucional

Rogério Boueri Miranda

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia

Alexandre de Ávila Gomide

Diretor de Estudos e Políticas Macroeconômicas

José Ronaldo de Castro Souza Júnior

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Alexandre Xavier Ywata de Carvalho

Diretor de Estudos e Políticas Setoriais de Inovação e Infraestrutura

João Alberto De Negri

Diretora de Estudos e Políticas Sociais

Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais

Sergio Augusto de Abreu e Lima Florencio Sobrinho

Assessora-chefe de Imprensa e Comunicação

Regina Alvarez

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

Texto para Discussão

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2017

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento, Desenvolvimento e Gestão.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 REVISÃO DA LITERATURA	9
3 APLICAÇÃO DA ANÁLISE DE AGRUPAMENTO PARA IDENTIFICAÇÃO DA TIPOLOGIA DOS MUNICÍPIOS	12
4 ESTIMAÇÃO DO EFEITO DOS INVESTIMENTOS EM BANDA LARGA SOBRE O PIB MUNICIPAL	28
5 COMENTÁRIOS FINAIS	45
REFERÊNCIAS	46

SINOPSE

O objetivo deste estudo é avaliar o efeito regional da expansão da banda larga no Brasil. Isso é feito por meio do emprego do modelo econométrico de dados em painel para a população de 5.564 municípios, no período 2007-2014. De modo a incorporar as diferenças regionais, usamos informação advinda da análise de agrupamentos que isola grupos homogêneos de municípios. O problema da causalidade reversa entre o acesso à banda larga e o produto interno bruto (PIB) é identificado com o emprego do método de variável instrumental (IV). Entre os resultados obtidos, podemos constatar a existência de efeitos regionalmente diferenciados da expansão da banda larga sobre o PIB municipal, de modo que o maior impacto da banda larga ocorre nos municípios de maior renda *per capita* e alta concentração urbana, seguidos daqueles municípios com forte concentração nos setores de serviço, comércio e construção. O impacto da banda larga é também expressivo nos municípios ligados à parcela mais dinâmica do setor agrícola. Nos municípios mais pobres, o efeito da expansão da banda larga também ocorre, embora com menor intensidade. Por fim, testamos a hipótese de que o efeito do acesso à banda larga sobre o produto é do tipo “U invertido”, que assinala que tal efeito se eleva até determinado nível de acesso, mas diminui após ultrapassá-lo. Os resultados econométricos obtidos não evidenciam que essa hipótese possa se negada.

Palavras-chave: banda larga; análise de agrupamento; K-médias; dados em painel; variável instrumental.

ABSTRACT

The aim of this study is to evaluate the effect of broadband on the GDP taking into account regional differences. This is done by using the econometric model of panel data based on database of 5564 municipalities from 2007 to 2014. In order to incorporate the regional differences we use information derived from cluster analysis that isolates homogeneous groups of municipalities. Simultaneous causality between broadband and GDP is identified through instrumental variable method. Regional differentiated effects of broadband among municipalities were observed. The greatest impact of broadband occurs on those municipalities with the highest income *per capita* and high urban concentration followed by ones with a strong concentration in the service, commerce and construction sectors. The impact of broadband is also significant in the municipalities

linked to the dynamic agricultural sector. In poorer municipalities the effect of broadband also occurs although with less intensity. Finally, we test the hypothesis that the effect of broadband on the output has an “inverted U” form that indicates that the effect rises to a certain level of access, but decreases after it is exceeded. The econometric results do not indicate that this hypothesis can be denied so that the impact is more intense in the intermediate levels of access of broadband.

Keywords: broadband; cluster analysis; K-means; panel data; instrumental variable.

1 INTRODUÇÃO

O setor de telecomunicação tem vivenciado uma profunda mudança, no que se refere à introdução de novas tecnologias. Inserido nesse contexto, aparece, em especial, a difusão dos serviços de banda larga, cuja tecnologia permite a transferência de dados em alta velocidade.¹ Para o público em geral, os efeitos mais perceptíveis da banda larga estão associados aos usos mais populares que dela derivam, como ampliação e melhor qualidade do acesso à internet, do serviço de TV a cabo etc. No entanto, os benefícios da banda larga vão muito além, alcançando os serviços de telefonia, redes de computadores, conexão por satélites etc. É desnecessário dizer que a expansão da banda larga tem implicações importantes sobre a atividade econômica, impactando o crescimento econômico, o nível de emprego, a produtividade etc. (Katz, 2012).

No Brasil, a regulação dos serviços de telecomunicação é feita pela Agência Nacional de Telecomunicações (Anatel), que objetiva promover a modernização e a ampliação do acesso à banda larga em todo o país. No entanto, para que isso seja efetivado, duas ações são necessárias. Primeiro, aprofundar substancialmente os investimentos em infraestrutura no setor de telecomunicações. Segundo, ter uma noção prévia de como se efetua o impacto decorrente da expansão da banda larga sobre a economia, pois os efeitos, como apontado na literatura, ocorrem de modo heterogêneo entre os setores da economia, entre regiões, além de outros fatores. Assim sendo, é imperativo impor como agenda de política pública um modelo de gestão que faça com que os investimentos realizados pelas operadoras promovam a instalação de infraestrutura de telecomunicações adequada para prover serviços de banda larga de qualidade e volume compatíveis às necessidades da economia brasileira.

1. O Brasil ainda não tem uma regulamentação que indique qual é a velocidade mínima para uma conexão ser considerada de banda larga. O critério significado tem sofrido as modificações com o tempo. Inicialmente, a denominação de banda larga era usada para definir qualquer conexão de internet acima da velocidade padrão dos *modems* analógicos. A velocidade que implica uma conexão de banda larga se desenvolveu com o passar dos anos. Dos 256 Kbps que proporcionavam as conexões *asymmetric digital subscriber line* (ADSL), foi-se passando para tecnologias que oferecem vários Mbps. Hoje, a ideia de banda larga engloba várias noções que transcendem a conexão em si.

Portanto, é necessário examinar como se dá a relação entre a expansão da banda larga e o seu efeito sobre o produto, considerando as disparidades existentes entre as regiões. Neste estudo, isso é feito pelo emprego de um modelo econométrico usando uma amostra de dados em painel dos municípios brasileiros no período 2007-2014. Com o propósito de contemplar as diferenças regionais, foi aplicada a análise de agrupamento (clusterização), visando isolar os grupos de municípios com características aproximadas. Tal procedimento é feito em uma etapa que antecedeu a estimação do modelo econométrico. O emprego da técnica de dados em painel permite, por sua vez, levar em consideração a heterogeneidade existente entre os municípios. O problema da causalidade reversa entre a banda larga e o produto, pouco investigado, abordado na literatura, é tratado com o emprego do método de variável instrumental (IV).

Os resultados obtidos para todos os modelos estimados mostram que o efeito da expansão da banda larga sobre o produto interno bruto (PIB) é positivo e significativo. Na média, a ampliação de 1% do acesso à banda larga acarreta um aumento de 0,077% no PIB. Deve-se ainda ressaltar a existência de efeito diferenciado entre os agrupamentos homogêneos de municípios. O maior impacto da banda larga ocorre nos municípios de maior renda *per capita* e alta concentração urbana. Logo em seguida, aparecem aqueles municípios em que a economia se concentra nos setores de serviço, comércio e construção. Em tais regiões, o efeito da expansão da banda larga pode alcançar um crescimento de 0,19% do PIB. Embora o impacto da expansão da banda larga nos municípios mais pobres seja menor que nos demais, percebe-se que o efeito sobre o PIB desses municípios é também positivo. Por fim, testamos a hipótese de que o efeito do acesso à banda larga sobre o produto é do tipo “U invertido”, no qual tal efeito se eleva até determinado nível de acesso, mas diminui após ultrapassá-lo. O resultado econométrico não atesta evidência de que se possa negar essa hipótese, de modo que o impacto é mais intenso nos níveis intermediários de acesso.

Este estudo está estruturado da seguinte forma. Na seção 2, apresentamos a revisão da literatura acerca do efeito da expansão dos serviços de banda larga sobre a economia. Também nessa seção, indicamos os principais pontos em que procuramos avançar na literatura. Na seção 3, aplicamos a análise de agrupamentos, de modo a identificar, assim, como contextualizar os grupos homogêneos de municípios. Os aspectos metodológicos do modelo econométrico, da base de dados e dos resultados obtidos são descritos na seção 4. Por fim, os comentários finais aparecem na seção 5.

2 REVISÃO DA LITERATURA

Conforme foi colocado na introdução, a literatura aponta que o avanço na disponibilidade dos serviços de banda larga, ao ampliar o acesso à tecnologia de informação e comunicação (TIC), tem gerado fortes benefícios econômicos na forma de efeitos diretos, assim como em externalidades positivas. Esses ganhos estão distribuídos na economia entre o aumento da produtividade, a criação de empresas, o aumento no emprego e o crescimento econômico.

Czernich *et al.* (2009), usando dados de 25 países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), concluíram que um aumento de 10% do acesso à banda larga implicou aumento entre 0,9% e 0,15% do crescimento do PIB *per capita*; enquanto Koutroumpis (2009), para uma amostra de 22 países da OCDE, chega à conclusão que um aumento na mesma magnitude no acesso implica 0,25% no crescimento do PIB *per capita*. Também Qiang e Rossotto (2009), com uma amostra de 66 países de alta renda, determinaram que um aumento de 10% no acesso implica 1,21% no crescimento do PIB *per capita*. Entretanto, conforme coloca Katz (2012), os impactos não ocorrem de modo homogêneo entre os distintos setores e, consequentemente, entre os países.

No que se refere a produtividade, emprego e produto, Katz (2012) destaca os pontos a seguir. No caso da produtividade total dos fatores, o efeito é mais sentido nas indústrias mais intensivas em tecnologia da informação (TI). Stiroh (2002) afirma que o emprego da tecnologia da informação aumentou substancialmente a produtividade da economia americana. O aumento na eficiência é também sentido com maior relevância em setores com alto custo de transação – por exemplo, o setor financeiro –, ou intensivos em trabalho, tal como o turismo. O efeito sobre o crescimento econômico parece estar ligado às “economias” que apresentam maior intensidade tecnológica. Em regiões menos desenvolvidas, a ampliação do acesso à rede pode levar à incorporação de processos de produção mais eficientes, tendo como consequência a substituição do fator trabalho pelo capital. O efeito de ampliação da rede sobre pequenas e médias empresas não é imediato, devido à necessidade de reestruturação das firmas para se adequarem à nova tecnologia. O efeito da ampliação da rede de banda larga parece ser mais sentido quando é combinado com o estímulo à inovação.

Além do que foi dito anteriormente, muitas questões importantes permeiam a discussão acerca do efeito da expansão dos serviços de banda larga sobre a economia. A partir do fato apontado, de que o impacto sobre o crescimento econômico é mais sentido em economias mais intensivas em tecnologia, pode-se questionar se é necessário existir um nível de “massa crítica” em que somente a partir do qual o efeito da ampliação da rede se faça sentir? Uma evidência nesse sentido aparece em Koutroumpis (2009), que mostra que o acesso à rede é maior para países com alta penetração. Outro ponto interessante é se existe um ponto de crítico a partir do qual o acesso à rede apresenta retornos decrescentes. De fato, Czernich *et al.* (2009) usam a técnica de variável *dummy* para checar essa hipótese. As *dummies* são indicadoras de 10% e 20% de acesso à rede de banda larga. Verificou-se que somente a *dummy* de 10% de acesso se mostrou estatisticamente significativa.

Para o caso brasileiro, alguns bons estudos também assinalam o efeito positivo que a banda larga tem sobre o PIB. Para o Brasil, Macedo e Carvalho (2010a), aplicando modelos econométricos de dados em painel, analisam a relação entre o aumento da densidade de acessos de banda larga por habitantes e o crescimento do PIB e também do PIB *per capita*, com base em dados desagregados por estados, para o período 2000-2008. Os resultados para os coeficientes estimados permitem concluir que, para cada ponto percentual (p.p.) de aumento da densidade de acessos de banda larga, haveria um aumento entre 0,053 p.p. e 0,11 p.p. do PIB *per capita*.

Em um estudo posterior, Macedo e Carvalho (2010b) aprofundam o estudo a respeito do impacto sobre o PIB e o PIB *per capita* decorrente do aumento da difusão da banda larga no Brasil, utilizando-se de sistemas de equações simultâneas de oferta e demanda. Os dados empregados foram os de Macedo e Carvalho (2010a). Os valores encontrados indicam que um aumento de 1 p.p. da densidade de acessos de banda larga por mil habitantes poderia ocasionar o crescimento do PIB entre 0,037 p.p. e 0,178 p.p. e do PIB *per capita* entre 0,196 p.p. e 0,359 p.p.

Também para o Brasil, Fonseca (2015) investigou a relação entre os investimentos em infraestrutura de telecomunicações e o crescimento econômico com base na metodologia de vetor autorregressivo (VAR) e na identificação de causalidade, aplicando o teste de Granger. A relação causal observada é na direção de que o crescimento das telecomunicações implica crescimento do produto. Importante registrar que a causalidade foi verificada tanto no curto quanto no longo prazo.

Um ponto que chama atenção quando se examina o estado da literatura referente ao efeito dos serviços de telecomunicação sobre os indicadores econômicos é que os resultados obtidos a partir dos diversos estudos apresentam uma grande discrepância. Podemos citar pelo menos cinco justificativas que indicam que os resultados não são comparáveis, assim como acreditamos ainda que tais estudos carecem de maior embasamento técnico, para que o efeito da expansão da banda larga sobre os indicadores econômicos possa ser fidedignamente estimado. Primeiro, os estudos foram produzidos em cima de bases de dados distintas. Segundo, não utilizaram o mesmo conjunto de variáveis de controle. Em terceiro lugar, é observado que grande parcela dos artigos nesta literatura não faz uso de técnicas capazes de controlar o efeito fixo inerente às unidades. Outro ponto importante é o problema da causalidade reversa entre o crescimento do produto e o acesso à rede de banda larga. Assim como a expansão da rede implica um efeito positivo sobre os indicadores econômicos pelos motivos que já delimitamos, por sua vez, o crescimento da economia e da eficiência também pode levar os agentes econômicos, as firmas e os consumidores a demandar maior acesso à rede.

Por fim, um ponto sumamente importante é que grande parte dos estudos usa uma amostra com dados fortemente agregados, cujas unidades são os estados ou os países. No nosso contexto, os resultados econométricos derivados a partir do uso de dados com tal nível de agregação dificilmente são estatisticamente fidedignos, na medida em que eles não permitem captar o efeito que subsiste da grande diferença entre as diversas localidades e que poderiam ser controlados usando dados de municípios ou área mínima comparada.

Nosso estudo avança na literatura com base em diversos pontos. Nossa base de dados é composta de dados de 5.564 municípios brasileiros² e se refere ao período 2007-2014. Nosso modelo econométrico foi estimado com base na metodologia de dados em painel. Esses dois aspectos metodológicos permitem tratar, no caso da base de dados, a grande dispersão quanto à natureza das unidades espaciais, enquanto a técnica de dados de painel permite controlar o efeito fixo inerente às unidades. Além disso, tal como é apontado na literatura, os resultados mostrados por grande parte dos estudos que visam determinar uma relação causal entre a expansão da banda larga e o crescimento econômico estão sujeitos a um possível viés de causalidade simultânea entre o PIB e a banda larga.

2. Essa base de dados concentra toda a população de municípios brasileiros, já que inclui todos os municípios do território nacional.

A causalidade simultânea implica que é necessário distinguir dois efeitos: *i*) o aumento do crescimento econômico, que é atribuível ao aumento da infraestrutura de telecomunicações e desenvolvimento de serviços; e *ii*) o aumento da demanda por serviços de telecomunicações, que é atribuível ao aumento do crescimento econômico – ou seja, a elasticidade-renda da demanda de telecomunicações. Neste estudo, usamos o método de IV, que endogeniza o acesso à banda larga em um modelo em que o PIB é determinado por uma função na qual o acesso à banda larga é visto como um fator de produção.

Cabe ressaltar que, para o objetivo de política pública que quer determinar o efeito da ampliação do acesso à banda larga tendo em vista as diferenças de impactos regionais, ponto comentado na seção introdutória, é necessário, a princípio, isolar os agrupamentos proximamente homogêneos de municípios. Neste estudo, isso é feito por meio do método de abordagem *K-médias*, oriundo da análise multivariada de dados.

3 APLICAÇÃO DA ANÁLISE DE AGRUPAMENTO PARA IDENTIFICAÇÃO DA TIPOLOGIA DOS MUNICÍPIOS

3.1 Metodologia

Devido à heterogeneidade existente entre os municípios brasileiros, faz-se necessário identificar os grupos homogêneos em termos de suas variáveis. Este ponto é fundamental para a gestão de política pública, o direcionamento eficiente dos investimentos em telecomunicações, na medida em que o impacto dos investimentos deve ser sentido de forma distinta entre os grupos selecionados. Com o objetivo de identificar os grupos homogêneos de municípios, faremos uso do que se conhece no âmbito da análise de multivariados como análise de agrupamentos. Como define Lattin, Carrol e Green (2011), esta técnica envolve categorização – ou seja, dividir um grande conjunto de observações em conjuntos menores, para que as observações em cada um dos conjuntos sejam relativamente similares, e as observações entre os diferentes conjuntos sejam relativamente dissimilares.

Encontrar agrupamentos que ocorrem naturalmente requer que exista um grupo de observações com alta densidade local – isto é, muitas observações em uma determinada área –, mas separados por regiões de densidade local baixa. Dito em outras

palavras, os agrupamentos correspondem a uma modalidade de dados e o número de grupamento, ao número de modas em uma distribuição multimodal. Diversas abordagens podem ser empregadas para a análise de grupamentos. Duas destas são de uso mais comum: os métodos hierárquicos e os métodos de partição.³ O resultado gerado pela aplicação de um método hierárquico é representado pela estrutura de árvore, no qual a solução de k grupamento é formada pela junção de dois agrupamentos da solução de agrupamento $k+1$. No caso dos métodos de partição, as observações são separadas em um número determinados de subgrupos, sendo que a solução de k agrupamento e a solução de $k+1$ agrupamento não são necessariamente aninhadas. Quando passa de um número k de grupos para um número $k+1$, o nível de similaridade decresce. Ou seja, a variação entre os grupos diminui e a variação nos grupos aumenta.

Os métodos hierárquicos geralmente abordam a análise de dados por intermédio de dois procedimentos distintos. O procedimento aglomerativo, de baixo para cima, começando com cada observação em um agrupamento separado, unindo os agrupamentos a cada etapa do processo, até que reste um único agrupamento com as N observações totais da amostra. O procedimento divisivo, de cima para baixo, inicia-se com todas as observações em um único agrupamento, dividindo o agrupamento em dois a cada etapa do processo, até que reste N agrupamentos com somente uma observação.

O algoritmo do método hierárquico aglomerativo segue os seguintes passos, considerando que cada município constitui um *cluster* de tamanho 1.

1. Passo 1: para todos os pares de municípios i e j , o par com menor *dissimilaridade* entre os vetores de variáveis escolhidas é combinado e passa a constituir um novo município (pseudomunicípio). Esta seleção é feita de acordo com a seguinte métrica $D_{i,j}$:
$$D_{i,j} = \frac{\|x_i - x_j\|^2}{2}.$$
2. Passo 2: novo pseudomunicípio composto pela união do par de municípios com menor dissimilaridade. Apenas um pseudomunicípio é feito a cada passo, e, devido à propriedade de hierarquia, uma vez juntos, não se separam nos estágios subsequentes: $x_{i \cup j} = \frac{x_i + x_j}{2}.$
3. Passo 3: para todos os pares de municípios e pseudomunicípios, selecionar o par com menor *dissimilaridade* entre os vetores de variáveis selecionadas, considerando

3. Descrição geral dos algoritmos de clusterização (hierárquicos e *K-means*): Khattree e Naik (2000) e Berry e Linoff (1997).

a métrica de Ward: $D_{i,j} = \frac{\|x_i - x_j\|^2}{\frac{1}{n_i} + \frac{1}{n_j}}$, em que: n_i e n_j são, respectivamente os números dos agrupamentos i e j .

4. Passo 4: novo pseudomunicípio composto pela união dos pseudomunicípios com menor dissimilaridade: $x_{i \cup j} = \frac{n_i x_i + n_j x_j}{n_i + n_j}$.

5. Passo 5: retorne ao passo 1, até que reste somente um agrupamento.

O agrupamento aglomerativo não fornece uma resposta definitiva à questão de como escolher o número de agrupamentos. No entanto, por meio do *dendograma*, que é uma representação gráfica de uma hierarquia de solução de agrupamentos aninhados (uma solução de um agrupamento, solução de dois agrupamentos e assim por diante, até uma solução de N agrupamentos), é possível dizer, em um montante considerável de subjetividade, qual o melhor agrupamento de dados. Isso é feito verificando as distâncias relativas em relação às quais o número de agrupamentos não se altera. Ou seja, a medida que o algoritmo de agrupamento avança, a similaridade vai decrescendo entre os conglomerados. Logo, a distância entre eles aumenta, criando “pontos de salto” relativamente grandes em no que concerne às demais distâncias. O momento ideal de parar o algoritmo é exatamente nestes pontos de salto, e o número de conglomerados final será o daquele ponto – a constituição final dos grupos também. Caso a função apresente mais de um ponto de salto, é possível determinar uma região de prováveis números de grupos.

O método de partição tem por objetivo repartir a amostra em um determinado número K de grupos não sobrepostos, de maneira que os objetos em cada grupo sejam relativamente similares. Para tal, precisamos encontrar um modo de medir a similaridade no grupo e a diferença entre os grupos, para que possamos comparar as duas partições. Também é necessário encontrar a melhor dessas partições, pelo menos uma localmente ótima, quando não for factível encontrar a solução globalmente ótima.

Usaremos neste estudo um método de partição conhecido como *K-means* (Hartigan, 1975), em que número K de grupos é definido antecipadamente. O algoritmo é simples e eficiente do ponto de vista computacional. Entretanto, esse método é propenso a encontrar apenas soluções localmente ótimas, pois se baseia em um procedimento heurístico que realiza melhorias locais para uma partição inicial, até que melhorias posteriores não sejam mais possíveis. A descrição do algoritmo *K-means* pode ser feita da forma a seguir.

1. Passo 1: escolha-se, de início, uma partição dos dados em K agrupamentos. Diversas abordagens podem ser aplicadas para a escolha dessa partição inicial. Outra variante importante do método K -means é aquela que começa com um conjunto inicial de centroides de semente K , designando cada objeto para o ponto de semente mais próximo. Em alguns casos, os pontos de semente são simplesmente os pontos reais da amostra, de modo que sejam amplamente dispersos; contudo, na maioria dos casos, os pontos de semente são escolhidos aleatoriamente no conjunto de dados.
2. Passo 2: calcule o centroide para cada agrupamento C , $x_{C(i)}$.
3. Passo 3: calcule a soma das distâncias ao quadrado de cada objeto ao seu centroide do agrupamento – isto é, o quadrado da soma dos erros da partição ESS.⁴ Essa é a medida que desejamos minimizar: $ESS = \sum_{i=1}^n (x_i - \bar{x}_{C(i)})' (x_i - \bar{x}_{C(i)})$
4. Passo 4: torne a relacionar cada objeto i a um outro agrupamento, cujo centroide é o mais próximo de cada objeto, minimizando a distância ESS. Este procedimento é repetido até que ao final da etapa os elementos do agrupamento permaneçam sem alterações para outros agrupamentos. Assim, o processo convergiu localmente; caso contrário, retorne ao passo 2 com uma nova partição.

O algoritmo K -means encontra uma solução de agrupamentos para um dado valor de K ; compete ao analista decidir qual valor de K resulta na melhor solução de agrupamento. A resposta consiste em testar diversos valores de K e depois decidir a melhor solução, conforme o objetivo do estudo. Isso envolve uma análise de custo-benefício entre uma solução mais simples (simplicidade de um menor número de agrupamentos) e sua adequação (redução da heterogeneidade em cada agrupamento pelo maior número de agrupamentos). O critério ESS não é útil, pois vai diminuir à medida que o número de partições aumenta, gerando soluções com um número maior de grupos. Um indicador capaz de conjugar essas duas coisas é a estatística *pseudo-F* usada por Calinski e Harabsz (1974), definida da seguinte forma: $pseudo-F = \frac{tr(B)/(K-1)}{tr(W)/(N-K)}$.

Note-se que essa estatística é similar ao teste- F , porque é efetivamente a razão entre a soma média dos quadrados entre os grupos, B , e a soma média dos quadrados nos agrupamentos, W . Se a função F for monotonicamente crescente com o número de grupos K , os dados não apresentam uma partição “natural”. No caso de a função F apresentar um ponto de máximo, então o número de conglomerados e a partição cor-

4. Error sum of square.

respondem à partição “ideal” do conjunto de dados. A ideia central é buscar o maior valor do *pseudo-F*, que está relacionado com a menor probabilidade de significância do teste. Desta forma, esta estatística estaria rejeitando a hipótese de igualdade de vetores de médias populacionais, tendo como resultado maior heterogeneidade entre os grupos.

A estatística o *pseudo-t2* proposta por Duda e Hart (1973), também é importante para indicar o número de *clusters*. Esta estatística é calculada em cada passo do algoritmo de clusterização. Se em um passo do algoritmo o *cluster* C_k é a união de C_i e C_p , então o *pseudo-t2* é definido por:
$$PseudoT^2 = \frac{B_{il}}{[\sum_{j \in C_i} \|x_{ij} + \bar{x}_i\|^2 + \sum_{j \in C_p} \|x_{ij} + \bar{x}_i\|^2]^{(n_i + n_p - 2)^{-1}}}$$

Em que B_{il} é a distância entre os grupos. O *pseudo-t2* utiliza a mesma lógica do *Pseudo-F*, buscando o valor máximo, e o número K de grupos corresponde aos grupos do passo em que o algoritmo atinge este valor máximo. As duas principais diferenças em relação à estatística F são que o *pseudo-t2* não é aleatório na alocação, devido aos critérios de agrupamento, e que o *pseudo-t2* faz um teste de comparações dos vetores das médias dos dois grupos que se uniram. O R^2 também é uma estatística utilizada para a escolha do número de *clusters*. O R^2 mostra a explicação da variabilidade. Logo, quando cada indivíduo é considerado um grupo, no início do algoritmo, o R^2 é igual a 1, pois aquele agrupamento explica totalmente a variação do conjunto de dados. À medida que o processo de agrupamento acontece, o R^2 vai decrescendo.

3.2 Resultados da clusterização

Tendo descrito na subseção anterior a ideia geral acerca da análise de agrupamento e de alguns métodos que permitem sua aplicação, passaremos para a aplicação empírica do método, como meio para a construção dos grupos de municípios homogêneos a partir da nossa amostra. De início, devemos ter em mente que amostras diferentes de variáveis podem levar a grupos homogêneos diferentes. Por outro lado, a nossa experiência mostra que, devido à grande correlação em geral entre os indicadores socioeconômicos, a utilização de indicadores diferentes não incorre em agrupamentos significativamente diferentes. Neste trabalho, as variáveis utilizadas correspondem às características socioeconômicas dos municípios, a partir dos indicadores disponibilizados para o Atlas do Índice de Desenvolvimento Humano Municipal (IDHM) 2010.⁵ Podemos categorizar

5. Disponível em: <goo.gl/kSlbnP>.

as variáveis do seguinte modo: indicadores de renda, longevidade, saúde, educação, emprego e características dos domicílios.

Foram selecionadas as variáveis, em um total de 159, e, a partir disso, foi efetuada uma análise de componentes principais (Lattin, Carrol e Green, 2011) para redução de dimensionalidade. A análise de componentes principais permite, por exemplo, acomodar situações nas quais diversas variáveis possuem alta correlação entre elas. Selecionamos um número de componentes, tais que estes contabilizassem 99% da variabilidade das 159 variáveis originais. Ao final, selecionamos 81 componentes – estes foram então utilizados como variáveis para as análises de agrupamentos; portanto, reduziu-se o número de variáveis pela metade.

Utilizando os critérios para seleção dos números de agrupamentos – principalmente, a estatística *pseudo-t2* –, chegamos a quatro configurações para as tipologias de municípios, usando o método de clusterização hierárquica (número de *clusters* = 6, 23, 26 e 29). Chegou-se até esses números de *cluster* pelos maiores valores da variação da estatística *pseudo-t2*, como pode ser visto na tabela 1. Para esses mesmos números de agrupamentos, rodamos o algoritmo de *clusters* via *K-means*. Para fins de nossa análise de subgrupos, vamos usar seis *clusters*. Em etapas posteriores, pode-se investigar a utilização de análises de subgrupos com números maiores de agrupamentos. Porém, conforme assinalamos, isso pode trazer também diversas desvantagens.

Para facilitar a visualização e a interpretação dos *clusters* – contíguos ou não –, a figura 1 mostra os seis *clusters* obtidos a partir do método *K-means*. A variável que identifica o *cluster* foi remarcada, de forma que os municípios correspondentes ao *cluster* 1 são aqueles com maior renda *per capita*; os municípios correspondentes ao último *cluster* são aqueles com menor renda *per capita*. Vejamos algumas observações acerca dos diferentes grupos:

- grupos 1: formados por municípios com maior renda *per capita*;
- grupos 1 e 3: com alto percentual de população urbana;
- grupo 2: com alto percentual de população rural, alta longevidade e alta renda;
- grupo 3: apresenta renda intermediária, mas apresenta grande percentual de trabalhadores em serviço, comércio e construção;

- grupo 5: com alta taxa de desocupação para população maior ou igual a 18 anos de idade;
- grupos 2, 4, 5 e 6 (rurais): com baixa taxa de empregados com carteira; e
- grupo 6: com baixa renda, alto percentual da população morando em domicílios com água e esgoto inadequados e sem acesso à energia elétrica e com alta fecundidade.

TABELA 1
Indicadores do número de *cluster* não espaciais

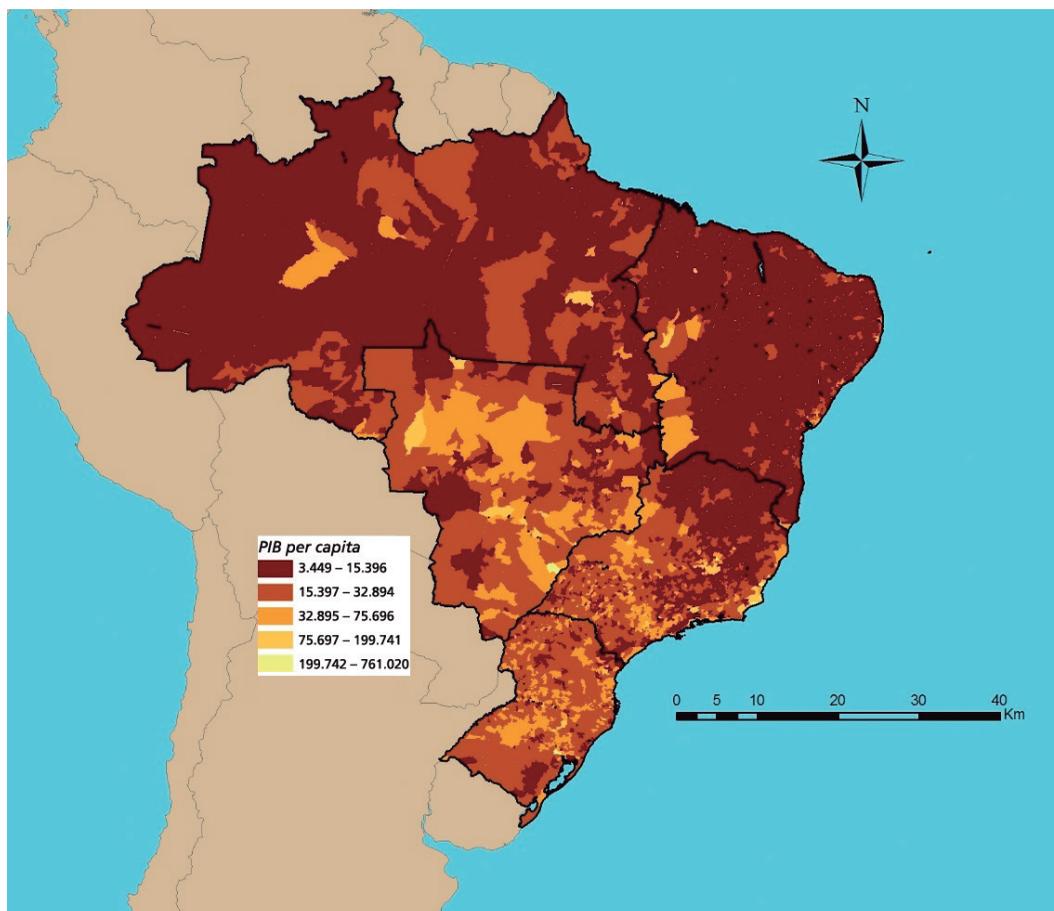
Número de <i>clusters</i>	<i>Cubic clustering criterium</i> (CCC)	Pseudo-t ²	Pseudo-F	R ² clusters	<i>Within sum of squares</i> (WSS)
1	0	2476	875712	0.00	875712
2	-9	515	2476	0.31	605927
3	-13	280	1632	0.37	551838
4	-19	268	1216	0.40	528809
5	-17	248	1015	0.42	506136
6	-14	161	890	0.44	486423
7	-11	145	805	0.46	468507
8	-10	141	729	0.48	456425
9	-10	110	667	0.49	446721
10	-9	76	613	0.50	439158

Elaboração dos autores.

3.3 Contextualização dos agrupamentos espaciais

Vale apontar alguns dados descritivos sobre a distribuição espacial da população e do PIB *per capita* do Brasil.

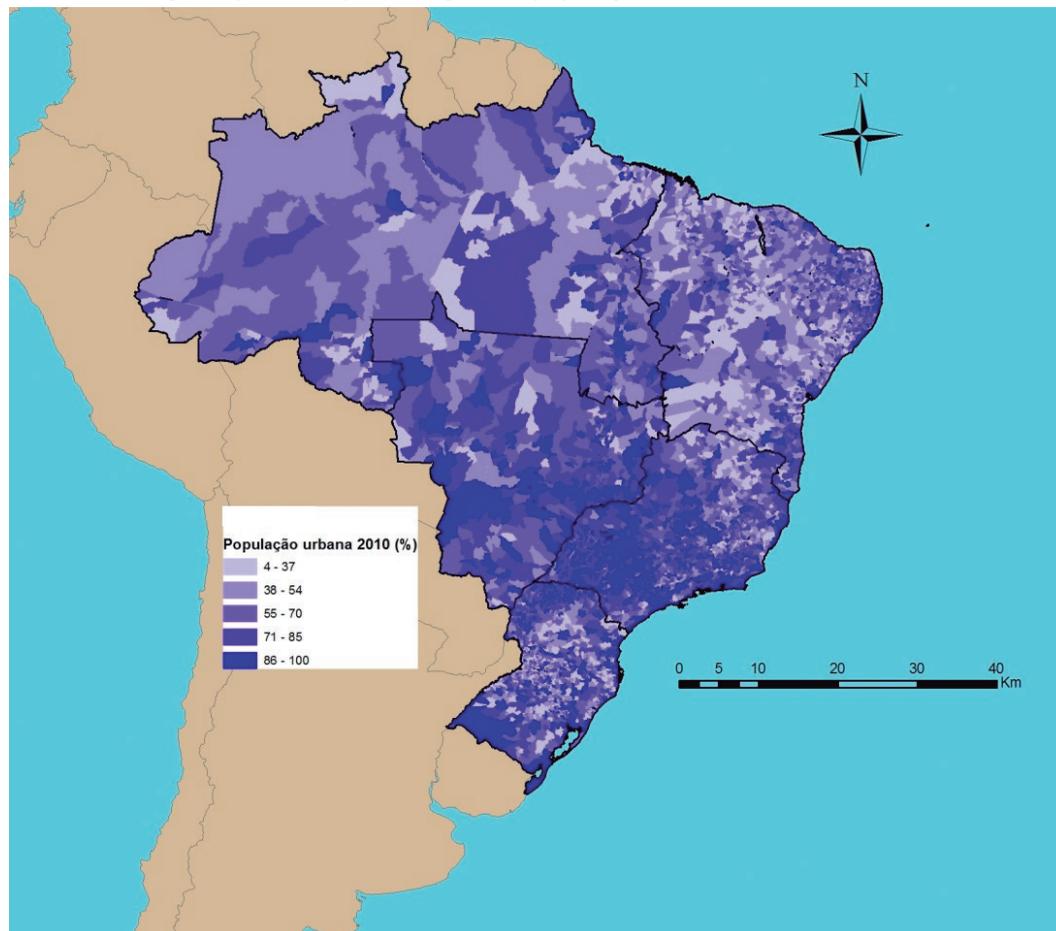
FIGURA 1
Distribuição espacial do PIB municipal *per capita* – Brasil (2013)



Fonte: Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/2013/>>.

No intuito de mostrar as diferenças regionais na economia do país, é mostrado na figura 1 um mapa com a distribuição do PIB *per capita* (dados de 2013). As cinco categorias apresentadas foram feitas pelo método de quebras naturais (algoritmo de Jenks). É possível observar os contrastes entre as grandes regiões. A maioria dos municípios das regiões Norte e Nordeste está na categoria de menor PIB *per capita* (de R\$ 3.449,00 até R\$ 15.396,00). Os municípios da mesorregião do Jequitinhonha e do Norte de Minas Gerais também estão nesta categoria. Enquanto a maior parte dos municípios da região com economia mais dinâmica (Sudeste, Centro-Oeste e Sul) apresenta a maioria dos municípios com PIB *per capita* nas categorias intermediárias e maiores (acima de R\$ 15.397,00, mas abaixo de R\$ 75.600,00). Os municípios com PIB muito alto (acima de R\$ 199.700,00) estão nas proximidades das regiões metropolitanas das grandes regiões de PIB *per capita* maior.

FIGURA 2
Distribuição espacial da porcentagem da população urbana – Brasil (2010)



Fonte: Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/2013/>>.

A taxa de urbanização mostra a densidade de pessoas morando em áreas urbanas. As atividades econômicas que utilizam serviços de comunicação são atividades tipicamente urbanas e do setor de serviços. Também vale ressaltar que a renda do trabalho no setor de serviços resulta em maiores salários, que atraem as pessoas para a área urbana dos municípios. Na figura 2, é possível observar as diferenças na distribuição da taxa de urbanização⁶ entre as grandes regiões do país. A maioria dos municípios das regiões Norte e Nordeste apresenta taxa de urbanização de até 54% da população. Enquanto a maioria dos municípios cuja taxa de urbanização está acima de 70% está nas regiões

6. População residente em área urbana dividida pela população total do município.

Sudeste, Centro-Oeste e Sul. Também é possível observar alguns contrastes internos, como nos municípios ao norte de Mato Grosso e de Minas Gerais, que têm menor taxa de urbanização.

3.4 Estatística descritiva dos *clusters*

O *cluster* 1 tem 758 municípios distribuídos em sua maioria no estado de São Paulo (345). Neste *cluster* específico, o PIB *per capita* é, em média, de R\$ 32.810,00. Estes municípios apresentam uma parcela grande da população morando em áreas urbanas (89% da população mora nas áreas urbanas). O setor que tem maior participação no valor adicionado do PIB é o setor de serviços, com 47% do PIB. Este setor também é o que mais emprega, dado que 40% do trabalho formal das pessoas com 18 anos de idade ou mais está neste setor, em relação a este total da população empregada. A taxa de desocupação é de 5% – para pessoas de 18 anos ou mais. Quanto aos indicadores sociais, pode-se afirmar que 10% da população com mais de 25 anos de idade tem ensino superior completo. A esperança de vida ao nascer é de 75 anos de idade e a mortalidade infantil – até 5 anos de idade – é de quatorze para mil nascidos vivos. A razão de dependência mostra a razão entre a população potencialmente ativa (entre 15 e 59 anos) e a população dependente (menores de 15 anos e maior que 60 anos). Quanto maior o valor, maior a proporção da população que deve ser dependente da população ativa. O *cluster* 1 mostrou valores baixos em relação aos outros *clusters*. A proporção de pessoas abaixo da linha de pobreza também apresentou valor de 4,4%, e em média 10% dos domicílios são beneficiários do Programa Bolsa Família (PBF). Apenas 0,7% das pessoas deste *cluster* moram em domicílios com abastecimento de água e esgotamento sanitário inadequados.

O *cluster* 2 tem 743 municípios distribuídos em sua maioria nos estados do Rio Grande do Sul (271), de Santa Catarina (130) e de São Paulo (100). Neste *cluster*, o PIB *per capita* é R\$ 23.610,00. A renda *per capita* média é de R\$ 688,00. Sendo que mesmo os municípios com menor PIB ainda apresentam uma renda *per capita* acima da linha de pobreza do PBF (R\$ 170,00). A população vive tanto na área urbana quanto na rural; este dado fica, respectivamente, em 54% e 46% da população. A maior participação no PIB dos municípios deste *cluster* é no setor agropecuário (35%), que absorve 44% da força de trabalho formal. Ou seja, os municípios deste *cluster* são voltados para a economia agrícola. A taxa de desemprego é de apenas 2% da população com 18 anos idade ou mais. Quanto aos indicadores sociais, a esperança de vida ao nascer é de 75 anos, a mortalidade infantil – até 5 anos de idade – é de quinze para cada mil nascidos vivos e a porcentagem de pobres fica em 7%, abaixo dos valores nacionais.

É possível observar que apenas 5% das pessoas com mais de 25 anos têm ensino superior completo. A razão de dependência é 42 para os municípios deste *cluster*, ficando com o segundo melhor resultado. Apenas 1,2% das pessoas deste *cluster* moram em domicílios com abastecimento de água e esgotamento sanitário inadequados.

O *cluster* 3 tem 1.161 municípios distribuídos em sua maioria nos estados de Minas Gerais (310), de São Paulo (179) e do Paraná (121). Neste *cluster*, o PIB *per capita* é de R\$ 24.870,00. Um pouco mais de 80% da população dos municípios deste *cluster* mora em áreas urbanas. A maior parte de seu PIB é formada pelo setor de serviços. A maioria dos trabalhadores deste *cluster* está empregada no setor de serviços também. A taxa de desocupação é de 6,5%. Com relação a indicadores sociais, pode-se destacar que apenas 6% das pessoas com mais de 25 anos de idade têm ensino superior completo e sua esperança de vida ao nascer é de 74 anos, em média. A mortalidade infantil é de dezessete para cada mil nascidos vivos. A proporção de pobres é de 10% da população, e 21% dos domicílios são beneficiados com o PBF. Apenas 2,2% das pessoas moram em domicílios com abastecimento de água e esgotamento sanitário inadequados.

O *cluster* 4 é formado por 859 municípios distribuídos em sua maioria no estado de Minas Gerais (292). O PIB *per capita* é, em média, de R\$ 17.070,00. A população está bem distribuída em áreas rurais e urbanas; a porcentagem da população que mora na área urbana é 54%. A economia é voltada para os setores agrícola e da administração pública; são estes setores os que mais contribuem para a formação do PIB, com 29% e 33%, respectivamente. Contudo, a maioria dos empregos está no setor agropecuário. A taxa de desocupação é 5%. Os indicadores sociais apontam que a população tem esperança de vida ao nascer de 73 anos de idade e 4% dos adultos – com 25 anos ou mais – têm ensino superior completo. A mortalidade infantil é próxima de vinte para cada mil nascidos vivos, e a razão de dependência é 51. O percentual de pobres é 21% – um valor acima da média nacional –, e 36% dos domicílios são beneficiários do PBF. Em média, 6% das pessoas não têm água e esgoto adequados em seus domicílios.

O *cluster* 5 é o que engloba o maior número de municípios, 1.555, distribuídos pelos estados da Bahia (346) e por outros estados da região Nordeste. O PIB *per capita* é de R\$ 8.261,00, e a renda *per capita* de alguns municípios fica abaixo da linha de pobreza (R\$ 170,00). Metade da população mora em áreas urbanas, em média. O setor da administração pública tem a maior participação no PIB dos municípios deste *cluster* (48%), e em segundo lugar fica o setor de serviços. Contudo, é o setor agropecuário que emprega a maior parte da mão de obra adulta (43%), enquanto o setor de serviços emprega 30% da mão de obra. A taxa de desocupação é de 8%.

Quanto aos indicadores sociais, pode-se observar que a esperança de vida ao nascer é de 70 anos de idade e a mortalidade infantil é de 28 para cada mil nascidos vivos. A razão de dependência é 57, e apenas 3,3% da população adulta tem ensino superior completo. A proporção de pobres é de 40% da população, e 36% dos domicílios são beneficiários do PBF. Uma parcela considerável da população (17%) mora em domicílios com abastecimento de água e esgoto inadequados.

O *cluster* 6 é formado por 488 municípios distribuídos nos estados das regiões Norte, como Pará (91) e Amazonas (59), e Nordeste, como Maranhão (122). O PIB *per capita* é de R\$ 8.500,00, em média. E a renda *per capita* é de R\$ 229,00, com alguns municípios apresentando renda *per capita* abaixo da linha de pobreza (R\$ 170,00). A maior parte da população vive em áreas rurais (55%). O PIB destes municípios recebe maior contribuição do setor da administração pública (48%) e do setor agropecuário (25%). Contudo, é este último setor que emprega 50% da mão de obra adulta. A taxa de desemprego deste *cluster* é de 7%. Quanto aos indicadores sociais, a esperança de vida ao nascer é de 70 anos de idade e a mortalidade infantil é de trinta para cada mil nascidos vivos. Apenas 3% da população adulta (25 anos ou mais) tem ensino superior completo. A razão de dependência é 68, o maior valor entre os *clusters*. A proporção de pobres é de 52% da população, e 53% dos domicílios são beneficiados pelo PBF. Quanto às condições sanitárias, 31% das pessoas moram em domicílios com água e esgoto inadequados. A tabela 2 faz um resumo destes valores.

TABELA 2
Média dos indicadores

Indicador/grupo	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Brasil
PIB <i>per capita</i> (R\$)	32.710	23.610	24.870	17.070	8.261	8.500	18.488
Taxa de urbanização (%)	89,88	54,320	80,49	54,050	54,560	45,000	63,8
Participação do setor agrícola (%)	18,92	35,00	19,46	29,57	14,14	25,45	22,05
Participação da indústria (%)	25,56	12,70	21,06	10,08	11,19	06,29	14,8
Participação do setor de serviços (%)	47,81	29,80	36,54	26,74	26,10	19,70	31,27
Participação da administração pública (%)	16,42	22,51	23,35	33,62	48,57	48,56	33,14
Emprego – agropecuária (%)	11,99	44,78	22,08	44,23	43,13	50,76	35,55
Emprego – serviços (%)	40,59	26,20	37,17	29,99	30,76	27,90	32,46
Emprego – comércio (%)	14,56	8,169	12,78	8,75	9,93	8,05	10,58
Taxa de desocupação (%)	5,50	2,49	6,44	5,02	8,34	7,50	6,19
Ensino superior (25 anos ou mais) (%)	10,9	5,5	6,5	4,643	3,2	3,0	5,4
Esperança de vida (em anos)	75,8	75,2	74,5	73,4	70,3	70,0	73,0
Mortalidade infantil (para cada mil nascidos vivos)	14,9	15,6	17,6	19,9	28,6	30,1	21,5

(Continua)

(Continuação)

Indicador/grupo	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Brasil
Razão de dependência	42,2	45,0	46,6	51,5	57,1	68,9	51,4
Proporção de pobres (%)	4,41	7,14	10,63	21,97	40,76	52,96	23,20
PBF por domicílio (%)	10,6	16,63	21,81	36,37	56,18	53,90	34,26
Condições sanitárias (%)	0,71	1,2	2,28	6,0	17,05	31,63	9,201
Sistema de telefonia fixo comutado (STFC) por habitante	7.543	3.9720	4.3780	2.2790	1.3590	0.9012	3.2824

Fontes: IBGE (2010), Relação Anual de Informações Sociais (Rais) de 2013 e *Relatório Anual 2013* da Agência Nacional de Telecomunicações (Anatel).

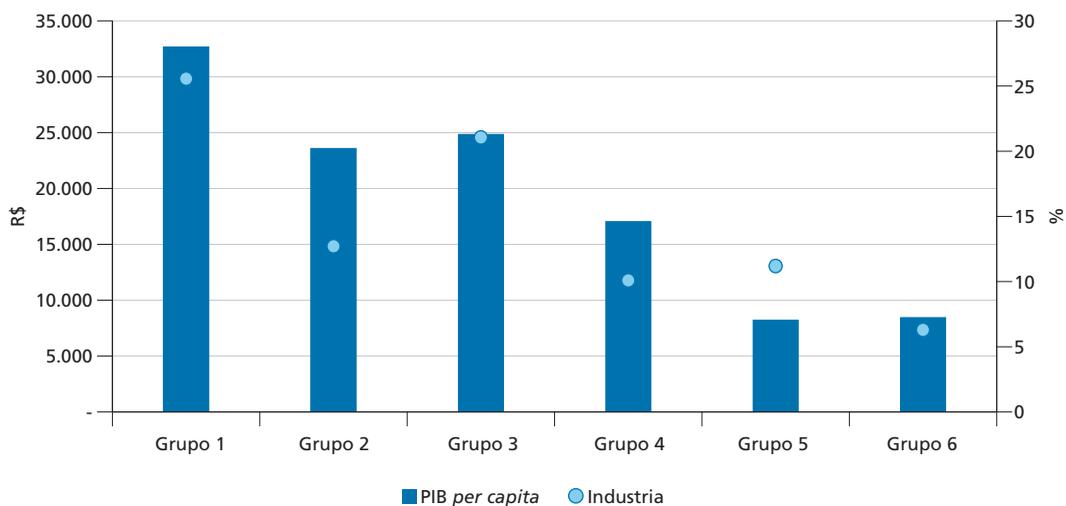
Quanto às características de algumas variáveis dos grupos, pode-se observar que, além das diferenças no PIB *per capita* entre os agrupamentos 1 e 2 e os demais agrupamentos, o emprego formal medido pelos dados da Relação Anual de Informações Sociais (Rais) apresenta destaque nos grupos 1 e 3, com maior número de vínculos, enquanto os demais grupos apresentam poucos vínculos de emprego formal. Uma possível explicação para este resultado são as altas taxas de emprego informal em algumas regiões. Um indicador relacionado com a divisão por agrupamentos é a participação da indústria no PIB, dado que a indústria pode gerar empregos que dinamizam a economia local e impulsionam o crescimento econômico. Os grupos 1 e 3 apresentam maiores rendas e maior participação da indústria (entre 20% e 25%), enquanto os grupos 4, 5 e 6, com menor renda, apresentam a participação da indústria entre 7% e 12%.

Como se pode observar no gráfico 2, os grupos que apresentam maior taxa de urbanização – definida como a porcentagem da população que vive na área urbana do município – são aqueles que têm maior PIB *per capita*. Além do mais, a *proxy* de acesso à internet se mostra maior nestes agrupamentos. O grupo 3, que apresenta renda intermediária, tem alta taxa de urbanização (acima de 80%) e, também, tem maior acesso à internet. Logo, existe uma relação fraca do acesso à internet com a urbanização. Por outro lado, os grupos com maior renda – e renda intermediária – também apresentam maior acesso à internet.

A partir da visualização da figura 3, é possível observar que o grupo 6 é formado por municípios localizados, predominantemente, na região Norte, enquanto os municípios que formam o grupo 5 estão localizados na região Nordeste. Nestes dois casos, a análise de *cluster* coincidiu bastante com a divisão por grandes regiões do Brasil feita pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Contudo, os outros grupos não obedecem a mesma distribuição espacial. O grupo 1 está localizado principalmente

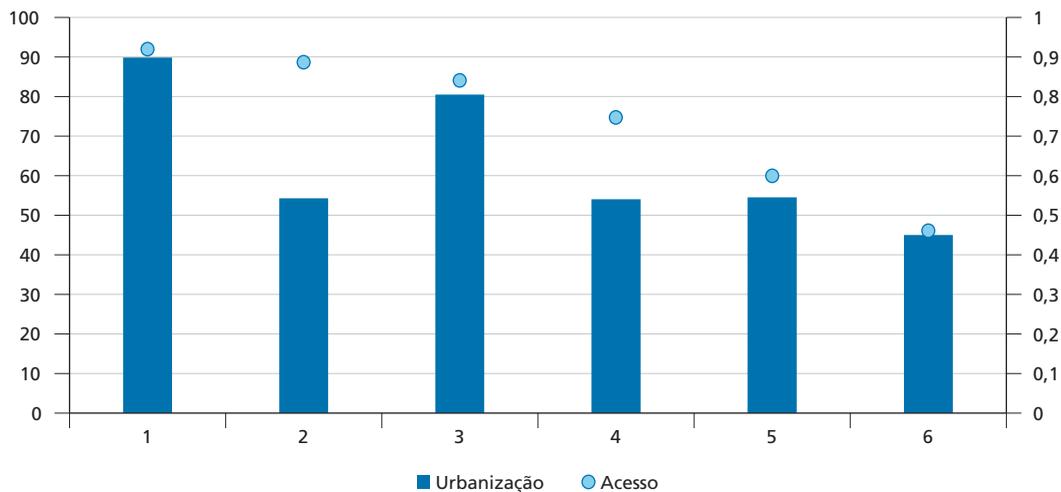
no estado de São Paulo, mas com alguns municípios em outros estados e regiões vizinhas. Os outros grupos estão distribuídos nas regiões Centro-Oeste, Sudeste e Sul de forma desconcentrada.

GRÁFICO 1
PIB per capita e acesso à internet – grupos



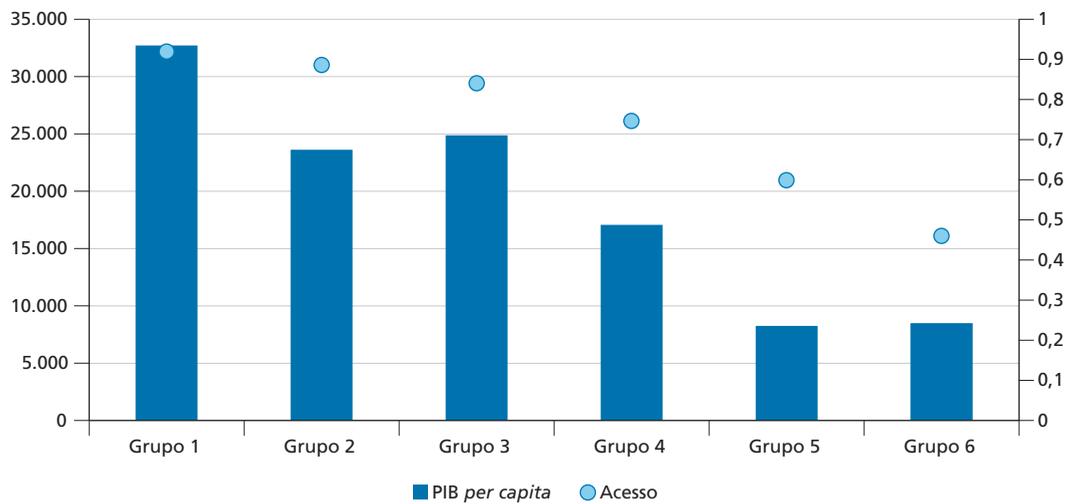
Fontes: IBGE (2010) e Relatório Anual 2013 da Anatel.

GRÁFICO 2
Taxa de urbanização e acesso à internet – grupos (2010)
(Em %)



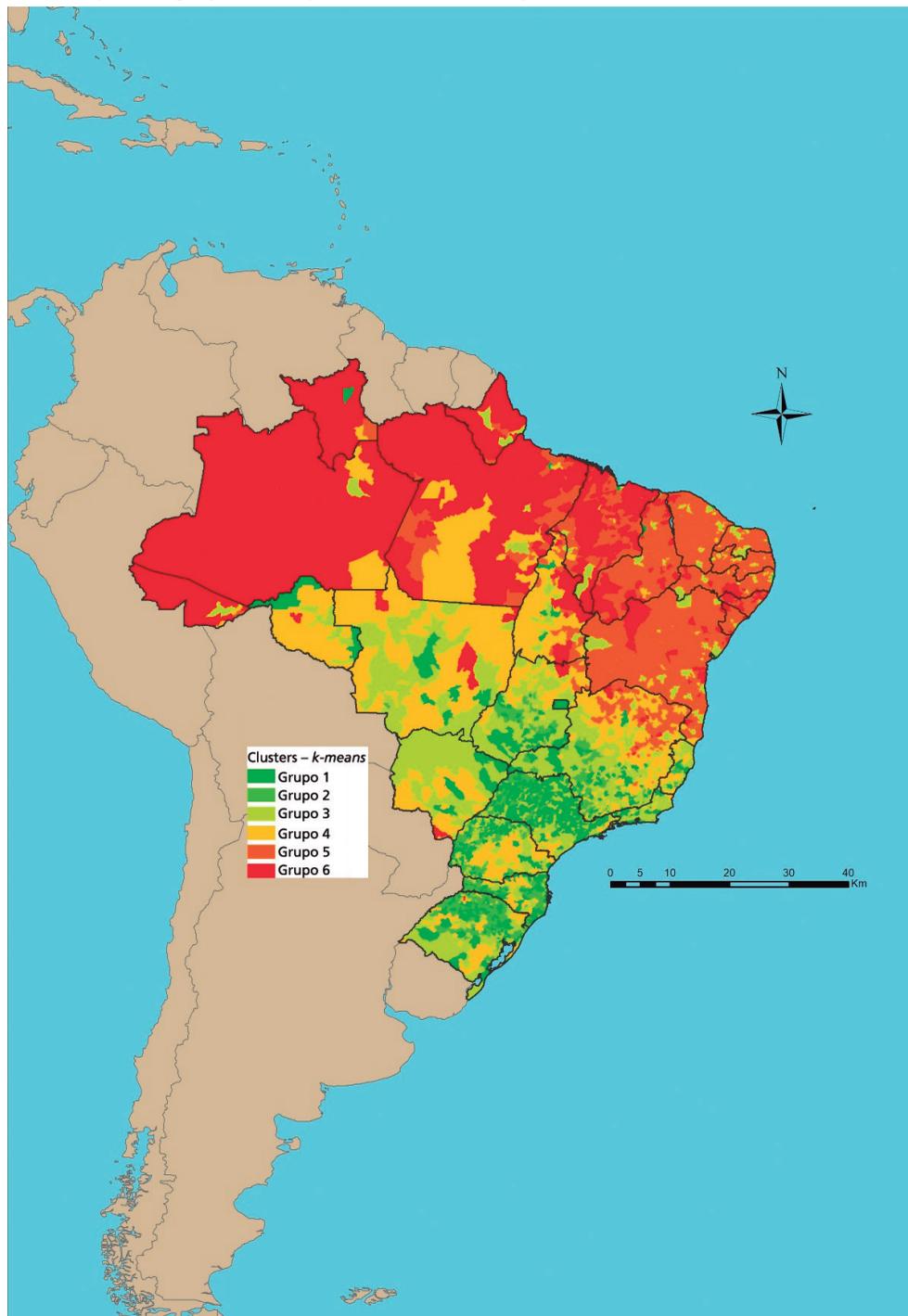
Fontes: IBGE (2010) e Relatório Anual 2013 da Anatel.

GRÁFICO 3
Taxa de urbanização e acesso à internet – grupos
(Em %)



Fontes: IBGE (2010) e Relatório Anual 2013 da Anatel.

FIGURA 3
Mapa do agrupamento para os seis *clusters* pelo método *K-means*



Fontes: Malha Digital dos Municípios e Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/2013/>>.

4 ESTIMAÇÃO DO EFEITO DOS INVESTIMENTOS EM BANDA LARGA SOBRE O PIB MUNICIPAL

Nossa tarefa nesta seção concerne à estimação econométrica do efeito dos investimentos em telecomunicações sobre o PIB municipal. Essa estimação é feita com base em uma análise de regressão linear, capaz de contemplar a heterogeneidade existente entre os municípios e, também, de levar em conta as diferenças entre agrupamentos homogêneos de municípios (*clusters*), determinados na seção 3, com base na metodologia de análise multivariada. No entanto, antes de passarmos para a análise econométrica, é necessário contextualizar a questão do impacto dos investimentos em telecomunicação em um referencial teórico. Isso significa enquadrar a relação entre o PIB municipal e os investimentos em infraestrutura de banda larga em um modelo econômico, de modo que, a partir disso, possamos ter uma ideia prévia das variáveis de controle do modelo, além de saber tratar com maior eficácia o problema da endogeneidade ou causalidade reversa; ponto já comentado neste trabalho, que significa, em última instância, recuperar o modelo estrutural por meio da forma reduzida. Portanto, temos que postular um modelo econômico ou estrutural que servirá de ponto de apoio para a nossa análise.

Nosso modelo estrutural parte da ideia consagrada na literatura econômica de que o produto de uma economia (Y) pode ser modelado por uma função de produção cujos fatores de produção mais citados são o capital (K) e a força de trabalho (N). Além destes, podemos adicionar os diversos tipos de serviços de infraestrutura que contribuem para o processo produtivo, bem como facilitam e dinamizam este (Barro e Sala-i-Martini, 1995). Inseridos nestes últimos, aparecem os serviços de telecomunicações, destacando-se aqueles ofertados pela tecnologia de banda larga (T). Denotando por O_Y os demais fatores que afetam a produção, temos que o produto pode ser representado da seguinte forma:

$$Y = F_1(N, K, T, O_Y). \quad (1)$$

Como qualquer insumo, a variável T deriva da interação das funções de demanda e oferta de serviços de banda larga: DT e ST , respectivamente. Naturalmente, DT depende do preço do serviço P_T , da renda Y , do nível do capital humano H , além de outras variáveis que, denotadas por I_D , influenciam a demanda. Da mesma forma, a oferta de banda larga, ST , é determinada naturalmente também pelo preço P_T .

assim como os demais fatores relacionados à infraestrutura IF , que contribuem para disponibilizar o acesso à banda larga.⁷ Essas duas funções podem ser representadas de modo que

$$DT = DT(P_T, Y, H, I_D) \quad (2)$$

$$ST = D(P_T, IF). \quad (3)$$

Sabendo que $DT = ST$, temos que

$$T = T(P_T, H, Y, IF, I_D), \quad (4)$$

substituindo a equação (4) em (1), de forma que

$$Y = F_1(N, K, T(P_T, H, Y, IF, I_D), O_Y). \quad (1')$$

Dessa forma, podemos entender de onde se origina o problema da causalidade reversa entre T e Y .

O modelo descrito pelas equações de (1) a (4) segue próximo àquele que aparece em Roller e Waverman (2001). De modo semelhante ao que fazemos neste estudo, Roller e Waverman (2001) tratam o problema da endogeneidade dos serviços de telecomunicação especificando as equações de oferta e demanda, que, por sua vez, são estimadas em conjunto com a função de produção, por meio de um sistema de equações simultâneas. Nós, por outro lado, apenas identificamos o efeito da banda larga sobre o produto usando o método de IV, permitindo tratar o efeito causal da infraestrutura de banda larga conforme discutido anteriormente. Isso significa que não identificamos as equações de demanda e oferta de banda larga. Deve-se, contudo, ter em perspectiva que a infraestrutura instalada para disponibilizar o acesso à banda larga não se altera no curto ou médio prazo, devido ao tempo necessário à maturação dos investimentos. Dessa forma, é razoável admitir a hipótese de que a curva de oferta de banda larga é relativamente fixa no curto prazo.

7. Como será visto adiante, um exemplo de serviço de infraestrutura que facilita o acesso à banda larga, além de outros, é a estrutura já implantada de telefonia fixa.

Uma vez tendo apresentado o modelo econômico, passamos para a implementação do modelo econométrico, cujo objeto principal é determinar o efeito da expansão da banda larga sobre o produto municipal (PIBM). A equação de regressão para o PIBM, a ser estimada usando uma amostra de dados em painel dos municípios brasileiros e que contempla os efeitos da banda larga para subgrupos homogêneos (*clusters* $j=1, \dots, J$) de municípios, é definida da seguinte forma:

$$\log(\text{PIBM}_{i,t}) = \alpha_i + \sum_{j=1}^J \phi_j \log(w_{i,t}) * Dcluster_{i,j} + \sum_{k=1}^K \delta_k \log(x_{i,t,k}) + \varepsilon_{i,t} \quad (5)$$

Em que:

i , para $i = 1, \dots, N$, é o índice das unidades (municípios);

t , para $t = 1, \dots, T$, é o índice de tempo;

k é o índice relativo à variável de controle $x_{i,t,k}$;

$w_{i,t}$ é a variável de política relacionada ao serviços de banda larga; e

$Dcluster_{i,j}$ é a variável *dummy* que indica se o município i pertence ao agrupamento j .

O parâmetro α_i indica o efeito individual ou efeito específico referente ao município i , capaz de contemplar a heterogeneidade existente entre os municípios. De modo que a equação (5) possa ser uma representação empírica da equação (1), devemos ter as medidas de capital, emprego, produto ou renda, além, naturalmente, de um indicador verossímil de acesso à banda larga. Naturalmente, entre a pesquisa teórica e a sua aplicabilidade, geralmente existe, por vezes, certa distância. Na prática econométrica, é comum não se dispor de algumas que aparecem no referencial teórico. Isso acontece por vários motivos. No caso corrente, isso é devido ao nível de desagregação da nossa base de dados. No Brasil, não existe medida de capital desagregada para municípios. Assim, procuramos controlar essa dificuldade por meio de variáveis de controle que são as receitas totais (*RECEITA*).⁸

8. Essa variável foi criada com a intenção de se chegar ao que seria a "renda disponível" dos municípios, envolvendo fontes de recursos federais, estaduais e municipais.

Para representar o fator trabalho ou emprego, usamos a variável vínculos⁹ da Rais.¹⁰ Para representar a variável T da equação (1), que é a nossa variável de política ($w_{i,t}$) da equação (5), usamos a densidade de acessos à banda larga (DEM_SCM).¹¹ Incluímos, também, como variáveis de controle, as transferências governamentais decorrentes do PBF, as transferências governamentais decorrentes do programa Benefício de Prestação Continuada (BPC) e os recursos advindos da Previdência Social, excetuando o BPC (Previ). O parâmetro ϕ_j associado ao termo composto $\log(w_{i,t}) * Dcluster_{i,j}$ define o efeito sobre o PIB nos municípios inseridos no agrupamento j , decorrente de um aumento dos serviços de banda larga.

11 Tendo em mente a existência de causalidade reversa entre a variável de política e o PIB, usaremos aqui o método de dois estágios (2SLS) para controlar essa suposta endogeneidade. Desse modo, cada variável $\log(w_{i,t}) * Dcluster_{i,j}$ deve ser instrumentalizada no primeiro estágio, conforme o modelo da equação (6):

$$\log(w_{i,t}) Dcluster_{i,j} = \alpha_{i,j} + \sum_l^L \beta_{j,l} \log(VINST_{i,t,l}) + \eta_{i,t,j} . \quad (6)$$

Dessa forma, l é o índice relativo às variáveis instrumentais ($VINST$).

Cabe ainda um comentário de como interpretar os parâmetros da forma funcional log-log que adotamos. Por exemplo, o parâmetro ϕ é a elasticidade do PIB em relação à variável DEM_SCM (densidade de acessos de banda larga). Assim, quando DEM_SCM_1 varia 1%, o PIB dos municípios do agrupamento J aumenta em ϕ %.

A base de dados deste estudo é composta de dados de 5.564 municípios brasileiros, entre 2007 e 2015, e formada por informações oriundas de base de dados distintas, como dados do Censo Demográfico do IBGE, os registros administrativos dos municípios etc. As variáveis de infraestrutura de banda larga foram cedidas pela Anatel. Estão também disponíveis variáveis de transferências para municípios, incluindo aquelas de programas sociais, como o PBF e de o BPC, cujas bases de dados são administradas

9. Emprego que abrange todos os vínculos formais (celetistas, estatutários, temporários, avulsos, entre outros).

10. A Relação Anual de Informações Sociais (Rais) é um registro administrativo sobre o mercado de trabalho formal brasileiro mantido pelo Ministério do Trabalho e Emprego (MTE)

11. Razão entre o número de acessos do Sistema de Comunicação Multimídia (SCM) por grupo de cem domicílios, de acordo com a Anatel.

pelo Ministério do Desenvolvimento Social (MDS). As informações sobre o emprego são oriundas da Rais.

4.1 Modelo de dados em painel e testes de especificação

Conforme foi visto, a equação (5) requer que no procedimento de estimação pelo menos três fatores sejam controlados. Primeiro, a existência de causalidade reversa entre o PIBM e os serviços de telecomunicações. Segundo, o efeito da variável de política diferenciado por agrupamento. Em terceiro lugar, a heterogeneidade entre as unidades espaciais ou municípios. Como forma de controlar a heterogeneidade intrínseca a cada unidade individual, e tendo em vista que a nossa base de dados é composta de informação que variam no tempo e espacialmente, faremos uso do modelo de dados em painel, que é uma técnica que permite um melhor tratamento em dados que não são puramente de corte ou uma série temporal. Além disso, como discorre Baltagi (1995), a metodologia de dados em painel traz consigo várias vantagens, tais como a incorporação de maior informação amostral, mais variabilidade, menor colinearidade, maior grau de liberdade e mais eficiência. O método de dados em painel se aplica a uma situação típica, que ocorre em painéis de microdados coletados anualmente, em que é comum acontecer que a amostra de dados possua dimensão temporal (T) curta, mas com um grande número de unidades de corte (N). Também, por usar as mesmas unidades de corte transversal, podemos estudar com maior acuidade a influência das mudanças no tempo.

O modelo de dados em painel parte da seguinte especificação:

$$y_{it} = \alpha + x_{it}'\beta + \mu_{it}, \quad (7)$$

aqui, y_{it} é a variável dependente para a unidade i no período t , x_{it} é a de dimensão, $K \times 1$ é a observação relativa às K variáveis explicativas, β é o vetor de parâmetros $K \times 1$ relacionado às variáveis explicativas, enquanto μ_{it} é o componente de erro. O modelo básico de dados em painel tem por hipótese o componente de erro ou erro de composição μ_{it} , no qual $\mu_{it} = u_i + \varepsilon_t$, em que u_i é o componente individual de erro ligado às unidades. E em que $E(u_i^2) = \sigma_u^2$, enquanto ε_t é um distúrbio estocástico ou erro idiossincrático, por hipótese, não correlacionado com x_{it} , de modo que $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Sabendo que $i = 1, \dots, N$ e $t = 1, \dots, T$, e tomando o componente

de erro específico u_i das unidades como fixo, o modelo pode ser reescrito na forma vetorial do seguinte modo:

$$y = \alpha i_{NT} + X\beta + v = Z\delta + \mu. \quad (8)$$

$$\mu = Z_u u + \varepsilon. \quad (9)$$

Em que y é de dimensão $NT \times 1$, X é uma matriz $NT \times K$, $Z = [i_{NT}, X]$, $\delta = (\alpha, \beta)$ e i_{NT} é o vetor de unidade de ordem NT , $Z_u = I_N \otimes i_T$ tal que \otimes é o produto de Kronecker. Neste caso, podemos estimar o vetor de parâmetros β multiplicando a equação (8) pela matriz de desvios Q , $Q = I_{N \times T} - (I_N \otimes i_T i_T' / T)$. Visto que $Q i_{NT} = 0$ ficamos com,

$$Q i_{NT} = 0. \quad (10)$$

A partir dessa expressão, a estimação dos parâmetros da regressão por mínimos quadrados ordinários (MQO – em inglês, *ordinary least squares* – OLS), que gera o chamado estimador de efeito “fixo” b_w , em que $b_w = (X' Q X)^{-1} X' Q y$, produz um estimador consistente para os parâmetros da regressão. A aplicação da matriz Q sobre (8) faz com que fiquemos com:

$$y_{it} - \bar{y} = \beta(x_{it} - \bar{x}) + \varepsilon_{it} - \bar{\varepsilon}. \quad (11)$$

Por outro lado, assumindo que u_i é aleatório, o vetor de distúrbio μ não contempla mais a hipótese clássica de homocedasticidade, já que sua matriz de covariância é definida por $V = \sigma_\varepsilon^2 I_{N \times T} + \sigma_u^2 (I_N \otimes i_T i_T')$. Neste caso, é possível mostrar que o método de mínimos quadrados generalizados (MQG) proporciona estimadores consistentes e eficientes na classe de estimadores lineares não tendenciosos para os parâmetros da regressão. Fazendo $V = I \otimes \Omega$, em que $\Omega = \sigma_\varepsilon^2 I_N + \sigma_u^2 i_T i_T'$, temos que $V^{-\frac{1}{2}} = I \otimes \Omega^{-\frac{1}{2}}$, de modo que há a equação $\Omega^{-\frac{1}{2}} = I_N - \frac{\theta}{T} i i'$, na qual $\theta = 1 - \sigma_\varepsilon / (T \sigma_u + \sigma_\varepsilon)^{1/2}$. A aplicação da matriz $\Omega^{-\frac{1}{2}}$ para a estimação MQG sobre (8) faz com que essa equação seja reescrita da seguinte forma:

$$y_{it} - \theta \bar{y} = \beta(x_{it} - \theta \bar{x}) + \varepsilon_{it} - \theta \bar{\varepsilon}. \quad (12)$$

Esse estimador é conhecido na literatura como estimador de efeito aleatório (RE).

A questão de qual dos dois estimadores deve ser empregado é investigada por meio da aplicação do teste de Hausman e Taylor (1981). O teste de Hausman é entendido a partir do resultado obtido por Mundlak (1978), que mostrou que o estimador de efeito fixo é um estimador não viesado e de variância mínima, de β consistente e eficiente quando $E(\alpha | X) = 0$. Hausman (1978) observa que, sob a hipótese nula de ortogonalidade entre α e os regressores, tanto o estimador de efeito fixo (FE) quanto o estimador de MQG de efeito aleatório são consistentes; porém, b_w é ineficiente. Sob a hipótese alternativa, $E(\alpha | X) \neq 0$, b_w é consistente, enquanto o estimador MQG não apresenta essa propriedade. Portanto, sob a hipótese nula, os dois estimadores não devem diferir sistematicamente (Greene, 1993). Assim, a aplicação do teste de Hausman tem como objetivo indicar se existe correlação entre os regressores e o componente individual.

A aplicação do modelo de dados em painel pode ser facilmente estendida para o caso no qual existe variável explicativa endógena (Baltagi 1995; Hsiao, 1995; Hausman, 1978; Greene, 1993; entre outros). De modo análogo ao que ocorre em outras situações, o problema de endogeneidade pode ser tratado por meio dos métodos tradicionais, como o de variável instrumental, o método de momentos generalizados (GMM) etc. Antes de apresentar os resultados econométricos, faremos um breve interlúdio acerca dos testes de especificação que serão aplicados nas regressões estimadas pelo método IV.

Como se sabe, para que a aplicação do método IV seja adequada, faz-se necessário que os instrumentos sejam “bons instrumentos”, no sentido de serem relevantes, bem como válidos. No jargão econométrico, isso significa que as variáveis usadas como instrumentos devem ser correlacionadas com os regressores endógenos e ao mesmo tempo ortogonais ao distúrbio da regressão. Neste sentido, são aplicados o teste de subidentificação¹² (Cragg e Donald, 1993; Kleibergen e Paap, 2006),¹³ o teste

12. Para que uma equação seja identificada no modelo IV, tanto a *condição de ordem* ($L > K$) – em que L é o número de variáveis instrumentais e K , o número de regressores –, como a *condição de posto* devem ser preenchidas. Esta última assinala que $QXZ = E(X'Z)$ deve ser de posto completo, em que $X = (X_1, X_2) = (\text{Endógenas}, \text{Exógenas})$ é a matriz de regressores, enquanto $Z = (Z_1, Z_2) = (\text{Exclusos}, \text{Inclusos})$ é a matriz dos instrumentos, sendo $Z_2 = X_2$. Quando isso não ocorre, dizemos que o modelo é subidentificado ou não identificado. Pode-se testar a condição de posto por meio do teste de Cragg e Donald (1993), cuja não rejeição da hipótese nula sugere que o modelo é subidentificado.

13. Na presença de heteroscedasticidade e autocorrelação residual, a estatística de Cragg e Donald (1993) não mais permanece válida. Em tais circunstâncias, o teste de Kleibergen e Paap (2006) deve ser usado como alternativa ao teste de Cragg e Donald (1993), quando as circunstâncias assim exigirem.

de sobreidentificação de Sargan-Hansen,¹⁴ além do teste de Stock-Yogo (Stock e Yogo, 2005), para verificar se os instrumentos são fracos.¹⁵

A pesquisa teórica econométrica mais recente sobre o método de variáveis instrumentais tem enfatizado muito a questão dos instrumentos fracos (Stock, Wright e Yogo, 2002; Moreira, 2003; Stock e Yogo, 2005). Quando os instrumentos são fracos, dois problemas sérios ocorrem na estimação por dois estágios (2SLS). O primeiro é a questão do viés. Embora o método 2SLS seja consistente,¹⁶ as estimativas são sempre viesadas para pequenas amostras. Segundo, quando os instrumentos são fracos, o erro-padrão estimado se torna muito pequeno. Neste caso, o intervalo de confiança é não fidedigno, pois, junto ao fato de o ponto médio deste estimador ser viesado, temos que o intervalo de confiança se torna pequeno. Isso tudo faz com que o procedimento de teste de hipótese na estimação por 2SLS se fragilize, no caso da ocorrência de instrumentos fracos. Como Staiger e Stock (1997) mostraram, o problema de instrumentos fracos pode ocorrer mesmo se no primeiro estágio os testes são significativos nos níveis convencionais (5% ou 10%) e para grandes amostras.

Vários métodos são sugeridos na literatura para testar a hipótese de instrumentos fracos. Uma estatística comumente usada é o R^2 do primeiro estágio com os instrumentos dito “inclusos” (Bound, Jaeger e Baker, 1995).¹⁷ Contudo, quando existe mais de uma variável endógena na regressão, esse indicador pode não ser mais válido. Shea (1997) propôs uma estatística conhecida como “ R^2 parcial”, que capta a correlação entre os instrumentos. Quando existe somente uma variável endógena, as duas medidas de R^2 são equivalentes. Outra regra aplicada quando se tem apenas um regressor

14. A independência do instrumento com relação ao distúrbio somente pode ser acessada se, e somente se, houver uma “abundância” de instrumentos – isto é, se a equação é sobreidentificada. Isto se dá quando a condição de ordem é satisfeita na desigualdade: o número de instrumentos excluídos é superior ao regressores endógenos. O teste de Sargan/Hansen é usado para testar a hipótese de sobreidentificação. Sob a hipótese nula, os instrumentos são válidos – ou seja, não correlacionados com o distúrbio. Sob ainda esta hipótese, a estatística de teste tem distribuição qui-quadrado com L-K restrições sobreidentificadas.

15. Instrumentos que explicam pouco a variação da variável explicativa endógena (fraca correlação entre Z e X) são considerados instrumentos fracos. O teste de Stock-Yogo (Stock-Yogo, 2005) é calculado com base na estatística F de Cragg e Donald (1993). Sob a hipótese nula, o estimador é fracamente identificado, no sentido de que o viés verificado é inaceitavelmente grande.

16. De modo que certamente o valor médio do estimador 2SLS se aproxima do verdadeiro valor do parâmetro, quando o tamanho da amostra converge para o infinito.

17. Alternativamente, isso também pode ser expresso como um teste F da significância conjunta dos instrumentos excluídos Z1.

endógeno é checar o valor da estatística F no primeiro estágio. Nesse caso, um valor menor que dez pode ser uma indicação de que os instrumentos são fracos.

Alternativamente, Stock e Yogo (2005) sugerem um teste no qual a hipótese nula é que o viés do estimador 2SLS seja menor que uma fração (digamos, 10%) do estimador MQO.¹⁸ De acordo com Murray (2006), a mais recente abordagem em relação ao problema de teste de hipótese com instrumentos fracos e um único regressor endógeno é o *teste da razão de verossimilhança condicional*, desenvolvido por Moreira (2003) e aperfeiçoado por Andrews, Moreira e Stock (2006). O teste de Moreira (2003) supera as distorções encontradas nos testes convencionais, ajustando os valores críticos do teste de hipótese de acordo com cada amostra, de modo que o intervalo de confiança corrigido gere um nível de significância correto. Assim, seus valores críticos “são condicionados” nos dados à disposição, e não constantes.

4.2 Resultados econométricos

Nossa estratégia econométrica é levada a cabo perfazendo um estudo comparativo entre alguns modelos econométricos, verificando se existe ganho de eficiência à medida que introduzimos modelos mais complexos. A tabela 3 apresenta as regressões para a equação (5) que não contemplam as diferenças entre agrupamentos, enquanto na tabela 4 essas diferenças são consideradas. Inicialmente, passemos à análise da tabela 1. A coluna (1) mostra os resultados obtidos por MQO, com o intuito apenas de calcular a estatística VIF¹⁹ (Lutkepohl *et al*, 1982), cuja finalidade é checar a ocorrência de multicolinearidade. Como se observa pelo valor médio da VIF, não existe indicação de que isso se verifica. Pode-se contrapor a ideia da utilidade dessa informação na medida em que o nosso método principal recai no modelo de dados em painel, sendo que a estatística VIF é calculada com base no modelo MQO para os dados agrupados. Assim, o verdadeiro valor pelo qual a variância dos coeficientes estimados estaria sendo inflacionada não está acessível quando se aplica o método de dados em painel. De fato, isso é verdade. Contudo, o viés obtido pela estatística VIF tem uma propensão a estar acima do verdadeiro valor que inflaciona a variância de cada coeficiente, quando se emprega

18. O teste de Stock-Yogo (Stock e Yogo, 2005) não mais recai na estatística F , que comumente aparece na prática econométrica, mas em uma generalização da estatística de Cragg e Donald (1993). Os valores críticos deste teste são obtidos a partir de uma distribuição não padrão.

19. *Variation inflation factor* (VIF). Um valor abaixo de 5.00 indica a não ocorrência de problema de multicolinearidade.

a metodologia de dados em painel, porque essa técnica tem a vantagem de diminuir a colinearidade entre os regressores (Baltagi, 1995).

As duas colunas seguintes apresentam os resultados obtidos por dados em painel para efeito aleatório e efeito fixo. Por hipótese, assumimos nesses dois modelos que os regressores são exógenos – ou seja, inexistência de correlação entre as variáveis explicativas e o erro idiossincrático ε_{it} , mas não que se possa haver correlação entre os regressores e o componente individual de erro. O teste de Pagan e Breusch não suporta a hipótese nula de que a variância do efeito individual é zero. Portanto, a variância dos resíduos reflete diferenças individuais.²⁰ Este fato nos permite avançar na análise do modelo de dados em painel, sugerindo aplicação do teste de Hausman, que, por sua vez, indica a rejeição da hipótese nula de não correlação entre o componente individual e o conjunto dos regressores. Essa correlação aparece principalmente nas variáveis *VÍNCULOS*, *RECEITA* e *PBF*, devido à diferença entre os valores gerados pelos estimadores RE e FE.

Nas colunas 4, 5 e 6 da tabela 1, são mostrados os resultados obtidos pela aplicação do método de variáveis instrumentais, visando corrigir o problema de endogeneidade ou causalidade reversa entre o PIBM e a densidade de banda larga (*DEN_SCM*). Usamos um conjunto de instrumentos na regressão da coluna 5 maior que aquele usado em relação à regressão que aparece na coluna 4. Na coluna 6, são mostrados os resultados gerados pela aplicação do método de primeira diferença (FD). Os testes de especificação descritos na subseção 4.1 foram aplicados sobre as regressões IV. Por economia, são mostrados apenas os resultados obtidos a partir do modelo estimado da coluna 4. Esses testes foram aplicados nos modelos das colunas 5 e 6 e indicaram que os resultados se mantêm qualitativamente inalterados em relação ao modelo da coluna 4.

O problema do modelo de variáveis instrumentais recai, sobretudo, na escolha correta dos instrumentos. Os critérios estatísticos que definem uma IV já foram descritos anteriormente. No entanto, vale a pena discorrer de modo mais intuitivo, de modo a justificar a escolha do nosso conjunto de instrumentos. Apresentar correlação com a variável endógena e não correlacionada com o distúrbio são os critérios básicos

20. A não rejeição da hipótese nula do teste de Breusch-Pagan indicaria que o modelo deve ser estimado usando uma regressão de mínimos quadrados ordinários (MQO) com dados agrupados (POLS ou *pooling* OLS).

para que uma variável possa ser empregada como instrumento. No entanto, essas duas condições podem ser expressas de forma mais intuitiva, dizendo que a variável instrumental tem efeito sobre a variável endógena, mas não tem influência direta sobre a variável dependente da regressão. Sendo assim, testaremos como variáveis instrumentais variáveis relacionadas à infraestrutura da banda larga. Usamos três conjuntos de variáveis instrumentais:

$IV_1 = [DEN_STFC, 521_KBPS].$

$IV_2 = [IV_1, 3G, OUTRAS].$

$IV_3 = [IV_2, OPERADORAS, ERBS].$

Em que:

DEN_STFC: densidade de acessos por telefonia fixa por cem domicílios;

521_KBPS: número de acessos na faixa de velocidade entre 521kbps e 2Mbps de banda larga por DDD;

3G: número de acessos por tecnologia 3G por município;

OUTRAS: número de acessos por outras tecnologias além de 2G e 3G;

OPERADORAS: número de operadoras de telefonia móvel que operam no município; e

ERBS: números de estações de rádio base – agregado por todas as tecnologias: 2G, 3G e 4G – por município.

Vejamos como podemos defender a hipótese de que esse conjunto de variáveis pode ser candidato a bom instrumento. Vejamos o caso da variável *DEM_STFC*. Naturalmente, o uso dessa variável para instrumentalizar a banda larga não está isento de crítica, na medida em que a telefonia fixa, tal como a banda larga, pode ser igualmente vista como um insumo de produção e porque também é um meio de troca de informação como a tecnologia de banda larga. De fato, Norton (1992) observou, em um estudo feito com base em uma amostra de 47 países para o período 1957-1977, que a variável de telefonia fixa apresentou sinal positivo, sendo estatisticamente significativa. Norton (1992) conclui que a infraestrutura de telefonia fixa tem impacto positivo sobre o crescimento. Não obstante a questão de que o referido estudo não trata a contento a questão da causalidade reversa, vale a colocação de que o impacto direto da telefonia fixa sobre o produto, embora tenha sido significativo no passado, já não seria

tão relevante no presente e se dissipou, enquanto os efeitos diretos da expansão da banda larga são inúmeros, indo muito além da simples troca de informação. Nesse caso, a estrutura de telefonia fixa atualmente serviria de ponte para expansão dos serviços de banda larga, enquanto esta última teria implicações substanciais sobre a economia.

Nas colunas 4, 5 e 6 da tabela 3, são mostrados os resultados do primeiro estágio das regressões. Por economia, não mostramos esses mesmos resultados na tabela 3, porque teríamos que mostrá-los para cada regressão da variável DEM_SCM_I , $I=1, \dots, 6$. No entanto, podemos adiantar que os resultados para o primeiro estágio são qualitativamente semelhantes aos obtidos para as regressões IV da tabela 1.²¹

Interessante notar que os sinais dos coeficientes são significativos e apresentam os sinais “esperados”. No caso da variável DEM_STFC , temos a correlação positiva entre essa variável e DEM_SCM , o que está em conformidade com a ideia já colocada acerca da estrutura de telefonia fixa servir como suporte para ampliação do serviço de banda larga. No que diz respeito às variáveis relacionadas à velocidade de transmissão de dados (512_KBPS) e à tecnologia ($3G$ e $OUTRAS$), estas apresentam sinal positivo em conformidade com a literatura sobre demanda de banda larga. O sinal positivo para o coeficiente da variável $OPERADORAS$ pode estar associado ao fato que quanto maior a quantidade de empresas de telecomunicação atuando em um município, maior o acesso à banda larga; seja por razão de preço em decorrência do aumento da competição, seja devido à maior facilidade e disponibilidade do serviço.

A variável $ERBS$ não apresenta significância. Isso provavelmente é explicado pela existência de colinearidade dessa variável com outro instrumento.²² O teste de redundância indica a relevância dos instrumentos, assumindo a hipótese nula de que a eficiência assintótica da estimativa não é melhor quando se expande o conjunto de variáveis instrumentais. Testamos as hipóteses de que o conjunto de instrumentos IV_2 é redundante em relação ao conjunto IV_1 e de que o conjunto IV_3 é redundante no que concerne ao conjunto IV_2 . Em ambos os casos, a hipótese nula de redundância foi rejeitada. No entanto, o ganho de eficiência é pequeno quando se

21. Os resultados econométricos podem ser obtidos mediante contato com os autores.

22. A estatística VIF para a regressão do primeiro estágio mostrou que $ERBS$ é a variável que apresenta maior inflação da variância após 512_KBPS . Deduz-se que a variável $ERBS$ não detém informação relevante que já não esteja contida na variável 512_KBPS .

emprega o conjunto IV_3. O que é compatível com o que foi comentado acerca da não significância da variável *ERBS* no primeiro estágio da regressão. No que se refere aos demais testes de especificação, temos que o teste de Stock e Yogo (2005) rejeita a hipótese de fraqueza dos instrumentos, enquanto o teste de Wu-Hausman rejeita a hipótese nula de que a variável *DEM_SCM* é exógena. O teste de subidentificação não permite rejeitar a hipótese de que o conjunto de instrumentos não seja correlacionado com a variável endógena.

Passamos, por fim, para a análise dos coeficientes estimados para a regressão principal. No que se refere às variáveis de controle, temos que *EMPREGO*, variável relacionada ao emprego formal, apresenta significância estatística e sinal esperado em todos os modelos. O modelo que estimamos é do tipo log-log. Assim, o coeficiente estimado representa a elasticidade do PIB municipal com relação à variável explicativa. Desse modo, tomando a coluna 4 da tabela 1, temos que, para cada 1% de crescimento na variável *EMPREGO*, há 0,10% de crescimento do PIB do município. No que se refere às demais variáveis de controle, se fosse o caso, caberia maior investigação. Cabe destacar que, *em todas as regressões*, a variável *DEM_SCM* mostra sinal positivo e significância estatística. Isso demonstra que a expansão dos serviços de banda larga tem efeito positivo sobre o PIB municipal.

Considerando-se a tabela 4, observa-se que, embora a elasticidade do PIB municipal com relação à banda larga seja positiva, existe diferença de valor dessa medida entre os agrupamentos. O maior impacto ocorre em conformidade ao esperado e acontece no agrupamento 1, que é aquele composto por municípios de maior renda *per capita* e alta concentração urbana. O segundo maior efeito está vinculado ao agrupamento de municípios 3, cuja economia se concentra nos setores de serviço, comércio e construção, que a literatura tem apontando como sendo setores que se beneficiam fortemente da introdução da banda larga. Interessante notar que o efeito sobre o agrupamento 2 também é expressivo. Esta região é marcada por alta participação da população rural (45%), forte participação do setor agrícola na economia e alta renda *per capita*. Aqui, portanto, se insere o setor agrícola de maior dinamismo. Os agrupamentos 5 e 6 são os mais pobres, de menor renda *per capita*. O agrupamento 6 é o que concentra menor taxa de urbanização e elevada participação do setor agrícola em relação aos demais setores. Não obstante tudo isso, ambos os agrupamentos assinalam o efeito da banda larga sobre a economia.

De modo a validar os resultados obtidos e apresentados nas tabelas 3 e 4, usaremos outra variável *proxy* denotada por banda *ACS_SCM*, definida como o número de acessos do sistema de comunicação multimídia (SCM) por habitante. Usaremos essa variável para testar a hipótese de que o efeito da banda larga sobre o PIB cresce até determinado nível, mas arrefece ao ultrapassá-lo. Em outras palavras, a relação que descreve o impacto da banda larga sobre o produto é do tipo “U invertido”. Para isso, criamos variáveis para determinar seis faixas de acesso com base nos quartis $q(10)$,²³ $q(25)$, $q(50)$, $q(75)$, $q(90)$ e $q(99)$, calculados para a variável *ACS_SCM*. Por exemplo, a variável *ACM_SCM_25* indica o número de acessos SCM maior que $q(10)$ e menor ou igual a $q(25)$. Os resultados da regressão são mostrados na tabela 5. Tomando por base os resultados da coluna 6 dessa tabela, vemos que o efeito da banda é positivo para todas as faixas de acesso. Contudo, tais resultados parecem corroborar a hipótese de que o maior impacto acontece nas faixas intermediárias de acesso.

TABELA 3
Regressões: variável dependente – PIB municipal

Regressores	OLS (1)	RE (2)	FE (3)	IV-FE (4)	IV-FE (5)	IV-FD (6)
CTE	1.677 (0.000)	4.372 (0.000)	10.143 (0.000)	11.271 (0.000)	11302 (0.000)	0.045 (0.000)
EMPREGO	0.684 (0.000)	0.172 (0.000)	0.084 (0.000)	0.103 (0.000)	0.107 (0.000)	0.037 (0.000)
RECEITA	0.230 (0.000)	0.068 (0.000)	-0.007 (0.000)	-0.003 (0.124)	-0.003 (0.178)	0.004 (0.000)
PBF	0.036 (0.002)	0.172 (0.000)	-0.004 (0.000)	-0.001 (0.949)	-0.004 (0.521)	-0.003 (0.679)
PREVI	0.048 (0.000)	0.002 (0.000)	-0.030 (0.000)	-0.018 (0.000)	-0.018 (0.000)	0.001 (0.695)
PBC	-0.049 (0.002)	0.004 (0.000)	-0.004 (0.304)	-0.022 (0.000)	-0.024 (0.004)	-0.001 (0.604)
DEN_SCM	0.077 (0.000)	0.068 (0.000)	0.067 (0.000)	0.077 (0.000)	0.079 (0.000)	0.021 (0.000)
VIF Média	2.63	–	–	–	–	–
Teste de Breusch-Pagan	–	82824 (0.000)	–	–	–	–
Teste de Hausman	–	-22323 (0.000)	–	–	–	–

(Continua)

23. $q(10)$: quartil de 10%

(Continuação)

Regressores	OLS (1)	RE (2)	FE (3)	IV-FE (4)	IV-FE (5)	IV-FD (6)
Teste F de Anderson-Rubin ¹	–	–	–	1960.81 (0.000)	–	–
Teste F de Stock-Yogo ²	–	–	–	24591.68 (16.85)	–	1916.98 (0.000)
Teste de sub-identificação ³	–	–	–	21239.68 (0.000)	–	–
Teste Qui2 de Sargan/Hansen ⁴	–	–	–	1649.79 (0.000)	–	–
Teste F de Wu-Hausman ⁵	–	–	–	575.71 (0.000)	605.842 (0.000)	34.958 (0.000)
R2 ⁸	0.90	0.88	0.74	–	–	–
DEN_STFC 1_ESTÁGIO	–	–	–	0.198 (0.000)	0.215 (0.000)	0.404 (0.000)
512_KBPS 1_ESTÁGIO	–	–	–	0.625 (0.000)	0.623 (0.000)	0.462 (0.000)
Teste de Redundância ⁶	–	–	–	10000 (0.000)	–	–
3G 1_ESTÁGIO*	–	–	–	0.348 (0.000)	0.353 (0.000)	0.086 (0.000)
OUTRAS 1_ESTÁGIO*	–	–	–	0.444 (0.000)	0.423 (0.000)	0.257 (0.000)
Teste de Redundância ⁷	–	–	–	–	–	–
OPERADORAS 1_ESTÁGIO*	–	–	–	–	0.127 (0.000)	0.84 (0.017)
ERBS 1_ESTÁGIO*	–	–	–	–	0.012 (0.612)	0.047 (0.104)
NOBS	44511	44511	44511	32646	32037	26484

Elaboração dos autores.

Notas: ¹ Teste Wald para verificar a significância do regressor endógeno na equação estrutural.² Testa a hipótese de "identificação fraca", que surge quando os instrumentos excluídos são correlacionados com regressores endógenos, mas apenas fracamente. A estatística F de Cragg e Donald (1993) está na parte superior da célula. Valor entre parênteses: máximo de 5% de viés relativo da variável instrumental.³ Teste LM para verificar a hipótese de que os instrumentos excluídos são "relevantes" – isto é, correlacionados com os regressores endógenos.⁴ Testa um subconjunto de condições de ortogonalidade – isto é, é um teste da exogeneidade de um ou mais instrumentos.⁵ Teste de endogeneidade: sob a hipótese nula o regressor endógeno é tratado como se fosse exógeno.⁶ Testa a hipótese de que os dois próximos instrumentos excluídos na coluna são redundantes – ou seja, se a eficiência assintótica da estimativa não é melhorada ao usá-los.⁷ Testa a hipótese de que os dois próximos instrumentos excluídos na coluna são redundantes – ou seja, se a eficiência assintótica da estimativa não é melhorada ao usá-los.⁸ Considera-se o R2 "overall" no modelo de dados em painel.

TABELA 4
Regressões com *clusters*: variável dependente – PIB municipal

	RE (1)	FE (2)	IV-FE (3)	IV-FE (4)	IV-GMM ¹ (5)
CTE	4.723 (0.000)	10.170 (0.000)	11.171 (0.000)	11.221 (0.000)	0.983 (0.000)
EMPREGO	0.359 (0.000)	0.122 (0.000)	0.101 (0.000)	0.104 (0.000)	0.103 (0.000)
RECEITA	0.067 (0.000)	-0.006 (0.000)	-0.002 (0.235)	-0.002 (0.264)	-0.002 (0.399)
PBF	0.188 (0.000)	-0080 (0.000)	0.003 (0.604)	-0.001 (0.877)	-0.002 (0.729)
BPC	0.004 (0.000)	-0.004 (0.000)	-0.019 (0.021)	-0.002 (0.012)	-0.017 (0.000)
PREVI	0.034 (0.000)	-0.030 (0.000)	-0.016 (0.000)	-0.016 (0.001)	-0.012 (0.017)
DEN_SCM_1	0.136 (0.000)	0.083 (0.000)	0.100 (0.000)	0.099 (0.000)	0.099 (0.000)
DEN_SCM_2	0.072 (0.000)	0.063 (0.000)	0.084 (0.000)	0.084 (0.000)	0.084 (0.000)
DEN_SCM_3	0.089 (0.000)	0.076 (0.000)	0.094 (0.000)	0.095 (0.000)	0.094 (0.000)
DEN_SCM_4	0.063 (0.000)	0.069 (0.000)	0.084 (0.000)	0.085 (0.000)	0.084 (0.000)
DEN_SCM_5	0.048 (0.000)	0.057 (0.000)	0.062 (0.000)	0.063 (0.000)	0.064 (0.000)
DEN_SCM_6	0.043 (0.000)	0.069 (0.000)	0.076 (0.000)	0.078 (0.000)	0.078 (0.000)
Teste de Hausman	-20947.34 (0.000)	–	–	–	–
Teste de Breusch-Pagan	81910.61 (0.000)	–	–	–	–
Teste F de Anderson-Rubin	–	–	355.53 (0.000)	–	–
Teste de subidentificação	–	–	21199.28 (0.000)	–	–
Teste Qui2 de Sargan/Hansen	–	–	1719.21 (0.000)	–	–
Teste F de Wu-Hausman	–	–	740.49 (0.000)	738.59 (0.000)	713.32 (0.000)
Teste de Redundância ²	–	–	6800.00 (0.000)	–	–
Teste de Redundância ³	–	–	–	820.95 (0.000)	822.46 (0.000)
R2	0.88	0.76	–	–	–
NOBS	44511	44511	32644	32037	32037

Elaboração dos autores.

Notas: ¹ GMM com correção de autocorrelação serial de ordem dois.

² Testa a hipótese de que as variáveis 3G e OUTRAS são redundantes como instrumentos.

³ Testa a hipótese de que as variáveis OPERADORAS e ERBS são redundantes.

TABELA 5
Regressões por faixa de acesso: variável dependente – PIB municipal

	RE (1)	FE (2)	RE (3)	FE (4)	IV (5)	IV (6)
CTE	5.288 (0.000)	10.568 (0.000)	5.299 (0.000)	10.557 (0.000)	11.562 (0.000)	11.607 (0.000)
EMPREGO	0.382 (0.000)	0.127 (0.000)	0.376 (0.000)	0.125 (0.000)	0.108 (0.000)	0.111 (0.000)
RECEITA	0.067 (0.000)	-0.006 (0.000)	0.070 (0.000)	-0.003 (0.16)	-0.004 (0.003)	-0.004 (0.036)
PBF	0.166 (0.000)	0.078 (0.000)	0.165 (0.000)	0.075 (0.000)	0.015 (0.009)	0.012 (0.043)
BPC	0.044 (0.000)	-0.004 (0.000)	0.005 (0.000)	-0.003 (0.012)	-0.002 (0.001)	-0.003 (0.000)
PREVI	0.037 (0.000)	-0.028 (0.000)	0.040 (0.000)	-0.025 (0.000)	-0.022 (0.017)	-0.022 (0.000)
ACS_SCM	0.068 (0.000)	0.067 (0.000)	–	–	–	–
ACS_SCM_10 ¹	–	–	0.071 (0.000)	0.070 (0.000)	0.070 (0.000)	0.071 (0.000)
ACS_SCM_25	–	–	0.076 (0.000)	0.072 (0.000)	0.075 (0.000)	0.077 (0.000)
ACS_SCM_50	–	–	0.095 (0.000)	0.089 (0.000)	0.092 (0.000)	0.093 (0.000)
ACS_SCM_75	–	–	0.100 (0.000)	0.095 (0.000)	0.103 (0.000)	0.106 (0.000)
ACS_SCM_90	–	–	0.109 (0.000)	0.102 (0.000)	0.123 (0.000)	0.130 (0.000)
ACS_SCM_99	–	–	0.163 (0.000)	0.127 (0.000)	0.113 (0.000)	0.080 (0.000)
Teste de Hausman	-21677.09 (0.000)	–	-21766.26 (0.000)	–	–	–
Teste de Breusch-Pagan	82069.55 (0.000)	–	82304.59 (0.000)	–	–	–
Teste F de Anderson-Rubin	–	–	–	–	261.17 (0.000)	174.28 (0.000)
Teste de subidentificação	–	–	–	–	3917.93 (0.000)	4206.61 (0.000)
Teste Qui2 de Sargan/Hansen	–	–	–	–	1326.73 (0.000)	1404.96 (0.000)
Teste F de Wu-Hausman	–	–	–	–	49.987 (0.000)	47.842 (0.000)
Teste de Redundância ²	–	–	–	–	61000.00 (0.000)	–
Teste de Redundância ³	–	–	–	–	–	4206.00 (0.000)
R2	0.75	0.76	0.87	–	–	–
NOBS	44098	44098	44098	44098	32540	31935

Elaboração dos autores.

Notas: ¹ *ACS_SCM_10*: quartil de 10% $q(10)$ da variável.

² Testa a hipótese de que as variáveis *3G* e *OUTRAS* são redundantes como instrumentos.

³ Testa a hipótese de que as variáveis *OPERADORAS* e *ERBS* são redundantes.

5 COMENTÁRIOS FINAIS

Este estudo procurou avaliar o impacto da expansão da banda larga no Brasil. Isso foi feito com base em diversos modelos de dados em painel para a população de municípios brasileiros, para o período 2007-2014. Pode-se constatar que, de fato, a banda larga tem efeito positivo sobre o crescimento do PIB, sendo que esse efeito ocorre de modo diferenciado regionalmente. Tais resultados estão em conformidade com a literatura, inclusive com aquela referente aos estudos feitos para o Brasil. Nossa pesquisa também identificou que o efeito da expansão da banda larga acontece de modo distinto entre agrupamentos de municípios com características similares.

O efeito mais expressivo da expansão do acesso à banda larga ocorre nos municípios de maior renda *per capita* com alta concentração urbana. Municípios com forte concentração nos setores de serviço, comércio e construção se beneficiam fortemente da banda larga. O impacto da banda larga é também expressivo nos municípios ligados à parcela mais dinâmica do setor agrícola. Nos municípios mais pobres, o efeito da expansão da banda larga, embora seja positivo, acontece com menor intensidade.

Por fim, testamos a hipótese de que o efeito do acesso à banda larga sobre o produto é do tipo “U invertido”. Essa hipótese assinala que tal efeito é crescente até determinado nível de acesso, mas diminui após ultrapassá-lo. Os resultados econométricos não evidenciam que essa hipótese possa ser negada.

Observamos ainda que o menor impacto econômico recai sobre os *clusters* 5 e 6, cujos municípios estão na sua maior parte situados nas regiões Norte e Nordeste. No entanto, deve-se ter em perspectiva que, segundo dados da Anatel, 58% dos municípios dessas regiões possuem infraestrutura de banda larga bastante precária.²⁴ Além disso, as regiões Norte e Nordeste também apresentam as mais baixas velocidades de banda larga, o que naturalmente implica uma menor qualidade. Assim, o menor efeito da banda larga sobre a economia das regiões Norte e Nordeste possivelmente seja explicado pela baixa eficiência desse serviço. Portanto, como proposta de política pública, fica a sugestão de se considerar um programa que minimize as disparidades regionais de acesso à banda larga; não somente em termos do acesso, mas também de qualidade. Não incorporar esse ponto em uma agenda prioritária de política pública pode levar ao aprofundamento da desigualdade entre as regiões do Brasil.

24. Não possuem *backhaul* de fibra.

REFERÊNCIAS

- ANDREWS, D. W. K.; MOREIRA, M. J.; STOCK, J. H. Optimal two-sided invariant similar tests for instrumental variables regression. **Econometrica**, v. 74, n. 3, p. 715-52, 2006.
- BALTAGI, B. H. **Econometric Analysis of Panel Data**. Hoboken: John Wiley & Sons, 1995.
- BARRO, R. J.; SALA-I-MARTIN, X. **Economic growth**. New York: McGraw-Hill, 1995.
- BERRY, M; LINOFF, G. **Data mining techniques**. Hoboken: Wiley Computer Publishing, 1997.
- BOUND, J.; JAEGER, D. A.; BAKER, R. Problems with instrumental variables estimation when the correlation between instruments and endogenous variables is weak. **Journal of the America Statistical Association**, v. 90, n. 430, p. 443-450, Jun. 1995.
- CALISNKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics**, v. 3, n. 1, p. 1-27, 1974.
- CRAGG, J. G.; DONALD, S. G. Testing identifiability and specification in instrumental variables models. **Econometric Theory**, v. 9, n. 2, p. 222-240, Jun. 1993.
- CZERNICH, N. *et al.* **Broadband infrastructure and economic growth**. München: CESifo, 2009. (Working Paper, n. 286).
- DUDA, R. O.; HART, P. E. **Pattern classification and scene analysis**. New York: Wiley, 1973.
- FONSECA, H. M. M. **Ensaio em teoria de contratos aplicada no setor de telecomunicações com foco nos investimentos**. 2015. Tese (Doutorado) – Universidade de Brasília, Brasília, 2015.
- GREENE, W. **Econometric analysis**. New York: Prentice Hall, 1993.
- HARTIGAN, J. A. **Clustering algorithms**. Hoboken: John Wiley & Sons, 1975.
- HAUSMAN, J. A. Specification test in econometrics. **Econometrica**, v. 46, n. 6, p. 1251-1271, 1978.
- HAUSMAN, J. A.; TAYLOR, W. Panel data and unobservable individual effects. **Econometrica**, v. 49, n. 6, p. 1377-1398, 1981.
- HSIAO, C. **Analysis of panel data**. New York: Cambridge University Press, 1995.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2010**. Rio de Janeiro: IBGE, 2010.
- KATZ, R. L. **The economic and social impact of telecommunications output: empirical evidence in the US and Europe**. ETNO Regulatory Economics Task Force. Brussels: ETNO, 2012. Power Point.

KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS software**. Cary: SAS Institute Inc., 2000. 558 p.

KLEIBERGEN, F.; PAAP, R. Generalized reduced rank tests using the singular value decomposition. **Journal of Econometrics**, v. 127, n. 1, p. 97-126, 2006.

KOUTROUMPIS, P. The economic impact of broadband on growth: a simultaneous approach. **Telecommunications Policy**, v. 33, n. 9, p. 471-485, 2009.

LATTIN, J.; CARROL, D.; GREEN, P. **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011.

LUTKEPOHL, H. *et al.* **Theory and Practice of Econometrics**. 2. ed. New York: John Wiley & Sons, 1982.

MACEDO, H. R.; CARVALHO, A. X. Y. **Aumento do acesso à internet em banda larga no Brasil e sua possível relação com o crescimento econômico: uma análise de dados em painel**. Brasília: Ipea, 2010a. (Texto para Discussão, n. 1494).

_____. **Aumento da penetração do serviço de acesso à internet em banda larga e seu possível impacto econômico: análise através de sistema de equações simultâneas de oferta e demanda**. Brasília: Ipea, 2010b. (Texto para Discussão, n. 1495).

MOREIRA, M. J. A conditional likelihood test for structural models. **Econometrica**, v. 71, n. 4, p. 1027-1048, 2003.

MUNDLAK, Y. On the pooling of time series and cross section data. **Econometrica**, v. 46, n. 1, p. 69-85, Jan. 1978.

MURRAY, M. P. Avoiding invalid instruments and coping with weak instruments. **Journal of Economic Perspectives**, v. 20, n. 4, p. 111-132, 2006.

NORTON, S. W. Transaction costs, telecommunications, and the microeconomics of macroeconomic growth. **Economic Development and Cultural Change**, v. 41, n. 1, p. 175-196, Oct. 1992.

QIANG, C.; ROSSOTTO, C. Economic impacts of broadband. *In*: WORLD BANK. **Information and Communications for Development 2009: extending reach and increasing impact**. Washington: World Bank, 2009.

ROLLER, L.; WAVERMAN, L. Telecommunications and infrastructure and economic development: a simultaneous approach. **American Economic Review**, v. 91, n. 4, p. 909-923, 2001.

SHEA, J. Instrument relevance in multivariate linear models: a simple measure. **Review of Economics & Statistics**, v. 79, n. 2, p. 348-352, 1997.

STAIGER, D.; STOCK, J. H. Instrumental Variables Regression with Weak Instruments. **Econometrica**, v. 65, n. 3, p. 557-586, 1997.

STIROH, K. Information technology and the U. S. productivity revival. **American Economic Review**, v. 92, n. 5, p. 1559-1576, 2002.

STOCK, J. H.; YOGO, M. Testing for weak instruments in linear IV regression. In: ANDREWS, D. W. K. (Ed.). **Identification and inference for econometric models**. New York: Cambridge University Press, 2005. p. 80-108.

STOCK, J. H.; WRIGHT, J. H.; YOGO, M. A survey of weak instruments and weak identification in generalized method of moments. **Journal of Business and Economic Statistics**, v. 20, p. 518-529, 2002.

BIBLIOGRAFIA COMPLEMENTAR

CRANDALL, R.; JACKSON, C.; SINGER, H. **The effect of ubiquitous broadband adoption on investment, jobs and the U. S. economy**. Washington: Criterion Economics, 2003.

CRANDALL, R.; LEHR, W.; LITAN, R. **The effects of broadband deployment on output and employment**. Washington: Brookings Institution, 2007.

FORMAN, C.; GOLDFARB, A.; GREENSTEIN, S. The internet and local wages: a puzzle. **American Economic Review**, v. 102, n. 1, p. 556-575, 2012.

FORNFIELD, M.; DELAUNAY, G.; ELIXMANN, D. **The impacts of broadband on growth and productivity**. Germany: Micus, 2008.

FRIEDERISZICK, H.; GRAJEK, M.; ROLLER, L.-H. **Analyzing the relationship between regulation and investment in the telecom sector**. Berlin: ESMI, 2008. p. 108-201. (ESMI White Paper).

GREENSTEIN, S.; MCDEVITT, R. **The broadband bonus: accounting for broadband internet's impact on U.S. GDP**. Cambridge, MA: NBER, 2009. (Working Paper, n. 14758).

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. Berlin: Springer, 1997. (Springer Series in Statistics).

HENRIKSEN, A. L. **A competição no mercado de banda larga no brasil: uma análise de possíveis determinantes da penetração do serviço de acesso à internet em banda larga em municípios brasileiros**. 2012. Dissertação (Mestrado) – Universidade de Brasília, Brasília, 2012.

ITU – INTERNATIONAL TELECOMMUNICATION UNION. **The impact of broadband on the economy**. Paris: ITU, 2012.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed. Upper Saddle River: Prentice-Hall, 1999, 815 p.

JUDGE, G. *et al.* **Introduction to the theory and practice of econometrics**. New York: Wiley, 1982.

LAROSE, R.; STROVER, S.; GREGG, J. L. The impact of rural broadband development: lessons from a natural field experiment. **Government Information Quarterly**, v. 28, n. 1, p. 911-100, Jan. 2011.

LEE, S.; BROWN, J. Examining broadband adoption factors: an empirical analysis. **Info**, v. 10, n. 1, 2008.

MACEDO, H. R.; CARVALHO, A. X. Y. **Análise de possíveis determinantes da penetração do serviço de acesso à internet em banda larga nos municípios brasileiros**. Brasília: Ipea, 2010. (Texto para Discussão, n. 1503).

THOMPSON, H.; GARBACZ, C. Broadband impacts on state GDP: direct and indirect impacts. *In*: INTERNATIONAL TELECOMMUNICATIONS SOCIETY, 17., 2008, Canada. **Annals...** Canada: ITS, 2008.

Ipea – Instituto de Pesquisa Econômica Aplicada

Assessoria de Imprensa e Comunicação

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Everson da Silva Moura

Leonardo Moreira Vallejo

Revisão

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Marcelo Araujo de Sales Aguiar

Marco Aurélio Dias Pires

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Alessandra Farias da Silva (estagiária)

Ana Clara Escórcio Xavier (estagiário)

Hislla Suellen Moreira Ramalho (estagiário)

Lilian de Lima Gonçalves (estagiária)

Lynda Luanne Almeida Duarte (estagiário)

Luiz Gustavo Campos de Araújo Souza (estagiário)

Paulo Ubiratan Araujo Sobrinho (estagiário)

Editoração

Bernar José Vieira

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Herllyson da Silva Souza

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Capa

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

Projeto Gráfico

Renato Rodrigues Bueno

The manuscripts in languages other than Portuguese published herein have not been proofread.

Livraria Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF

Tel.: (61) 2026-5336

Correio eletrônico: livraria@ipea.gov.br

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
**PLANEJAMENTO,
DESENVOLVIMENTO E GESTÃO**



ISSN 1415-4765

