

## AVALIANDO O EFEITO DA EXPANSÃO DA BANDA LARGA SOBRE O PIB

ALEXANDRE YWATA DE CARVALHO \*

MÁRIO JORGE MENDONÇA †

MARCUS GERARDUS LAVAGNOLE ‡

JOSÉ JAIME DA SILVA §

### Resumo

Este estudo tem como objetivo avaliar o efeito da banda larga no Brasil por meio de um modelo de dados em painel de 5.564 municípios no período de 2007 a 2014. Usamos a análise de agrupamento para controlar as diferenças regionais. Pode-se constatar que o efeito da banda larga sobre o PIB municipal é regionalmente diferenciado. O maior impacto incide nos municípios de maior renda *per capita* e alta concentração urbana. Testa-se ainda a hipótese de que o efeito da banda larga é do tipo “U invertido”. Os resultados não evidenciam que essa hipótese pode ser negada.

**Palavras-chave:** banda larga, análise de agrupamento, k-médias, dados em painel, variável instrumental.

### Abstract

This study aims to assess the effect of broadband in Brazil through a panel data model of 5,564 municipalities in the period 2007 to 2014. We use cluster analysis to control regional differences. It can be seen that the effect of broadband on municipal GDP is regionally differentiated. The greatest impact is on municipalities with higher *per capita* income and high urban concentration. We also test the hypothesis that the broadband effect is of the “inverted U” type. The results do not show that this hypothesis can be denied.

**Keywords:** broadband, cluster analysis, k-means, panel data, instrumental variable.

**JEL classification:** J43, O15, Q26.

**DOI:** <http://dx.doi.org/10.11606/1980-5330/ea143788>

---

\* Técnico de Planejamento e Pesquisa da Diretoria de Estudos Regionais e Urbanos (Dirur) do Instituto de Pesquisa Econômica Aplicada (Ipea). do IPEA. E-mail: alexandre.ywata@ipea.gov.br.

† Técnico de Planejamento e Pesquisa da Diretoria de Estudos Regionais e Urbanos (Dirur) do Ipea. E-mail: mario.mendonca@ipea.gov.br.

‡ Assistente de pesquisa da Diretoria de Estudos Regionais e Urbanos (Dirur) do (IPEA)

§ Assistente de pesquisa na Diretoria de Estudos Macroeconômicos (Dimac). E-mail: jaime.silva@ipea.gov.br.

## 1 Introdução

O setor de telecomunicações 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<sup>1</sup>. 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. (Czernich et al. 2009, Katz 2012, Koutroumpis 2009, Qiang & Rossotto 2009). A literatura aponta que, ao ampliar o acesso à tecnologia de informação e comunicação (TIC), o avanço na disponibilidade dos serviços de banda larga 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.

No Brasil, a regulação dos serviços de telecomunicação é feita pela Agência Nacional de Telecomunicações (Anatel), que para os próximos anos tem como objetivo promover a modernização e ampliação do acesso à banda larga. No entanto, para que isso seja efetivado, duas ações preliminares são necessárias. Primeiro, aprofundar substancialmente os investimentos em infraestrutura no setor de telecomunicações, tendo em vista os benefícios decorrentes. Segundo, dada a grande heterogeneidade econômica e social existente entre as localidades, não basta somente intensificar os investimentos, mas ter conhecimento de como se disseminam os efeitos da banda larga; pois, como aponta a literatura, o impacto da banda larga ocorre de modo heterogêneo entre as localidades<sup>2</sup>, considerando os fatores socioeconômicos.

Considerando essas duas ações, este estudo faz uso de um modelo econômico de dados em painel para os municípios brasileiros no período 2007-2014<sup>3</sup>. O emprego da técnica de dados em painel permite, por sua vez, levar em consideração a heterogeneidade existente entre os municípios. Com o propósito de contemplar as diferenças regionais foi aplicada a análise de agrupamento (clusterização), visando isolar os grupos de municípios com características aproximadas. Mais do que apenas isolar grupos heterogêneos de municípios, a análise de clusterização permite intuir como fatores como a escolaridade, renda e a participação dos setores econômicos predispõem o maior ou menor efeito da expansão da banda larga. Outro ponto também enfocado

---

<sup>1</sup>O Brasil ainda não tem uma regulamentação que indique qual é a velocidade mínima para uma conexão ser considerada de banda larga. O critério tem sofrido 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.

<sup>2</sup>Conforme dito, nossa base de dados é formada por informações referentes a municípios.

<sup>3</sup>Na verdade, podemos dizer que se trata da população dos municípios brasileiros, pois de um modo geral todos os municípios do Brasil foram contemplados.

neste artigo é o problema da causalidade reversa entre a banda larga e o produto. Ele é tratado com o emprego do método de variável instrumental, pois se por um lado a banda larga gera aumento da produtividade e portanto da produção, também é razoável admitir que produção requer maior necessidade dos serviços derivados da banda larga.

Os resultados obtidos mostram que o efeito da expansão da banda larga sobre o PIB municipal é positivo e significativo, mas distinto entre os agrupamentos regionais. Na média, a ampliação de 1% do acesso à banda larga acarreta um aumento de 0,077% no PIB municipal. O maior impacto da banda larga ocorre nos municípios de maior renda *per capita*, alta concentração urbana e alta escolaridade. Logo em seguida, aparecem aqueles municípios onde 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. O menor impacto da banda larga sobre o produto acontece nos municípios mais pobres, mas percebe-se que este efeito é ainda assim positivo e significativo. Testamos ainda a hipótese de que o efeito do acesso à banda larga sobre o produto municipal é do tipo “U invertido”, no qual o 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. Os resultados obtidos são condizentes não apenas com aqueles encontrados na literatura internacional, assim como com aqueles feitos para identificar o efeito da banda larga no Brasil, como Macedo & Carvalho (2010a,b) e Fonseca (2015), os quais acenam para uma relação positiva entre a expansão da banda larga e o crescimento econômico. Este estudo está estruturado da seguinte forma. Na seção 2, apresentamos uma breve revisão da literatura relacionada ao efeito da expansão da banda larga sobre a economia. Também nesta seção indicamos os principais pontos em que procuramos avançar. 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, base de dados e resultados obtidos são descritos na seção 4. Por fim, nos comentários finais, indicamos algumas sugestões de políticas públicas que visem expandir o acesso à banda larga no Brasil e que estão baseadas nos resultados alcançados neste artigo.

## 2 Breve Revisão da Literatura e Contribuições

Além dos efeitos da banda larga sobre a economia já apontados, outras questões importantes permeiam esse tema. Por exemplo, 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” 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 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.

Também a questão acerca do impacto diferenciado entre os setores da eco-

nomia é tema de análise. 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 intensivo 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.

Para o Brasil, a literatura também indica o efeito positivo que a banda larga tem sobre o PIB. Para o Brasil, Macedo & 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 do PIB *per capita*, com base em dados desagregados obtidos para o período 2000-2008. Os resultados para os coeficientes estimados permitem concluir que, para cada ponto percentual (pp) de aumento da densidade de acessos de banda larga, haveria um aumento entre 0,053 pp e 0,11 pp do PIB *per capita*. Em estudo posterior, Macedo & Carvalho (2010b) aprofundam o estudo 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 utilizados foram os de Macedo & Carvalho (2010a). Os valores encontrados indicam que o aumento de 1 pp da densidade de acessos de banda larga por mil habitantes poderia ocasionar o crescimento do PIB entre 0,037 pp e 0,178 pp e do PIB *per capita* entre 0,196 pp e 0,359 pp. 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 vai 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.

No que se refere às contribuições do corrente estudo, podemos citar as seguintes. Primeiro, os estudos foram produzidos a partir 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 delineamos, 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.

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. Diferentemente, nosso modelo econométrico foi estimado com base na metodologia de dados em painel usando uma base de dados composta de dados de 5.564 municípios brasileiros<sup>4</sup> para o período de 2007 a 2014. 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. Cabe ressaltar ainda que a questão sobre o efeito regionalmente diferenciado que a expansão da banda larga pode ter é objeto de estudo deste trabalho. Visando determinar o efeito da ampliação do acesso à banda larga, e tendo em vista as diferenças de impactos regionais, empregamos o método de *k-médias* para isolar os agrupamentos proximamente homogêneos de municípios.

Por fim, o problema da causalidade reversa também é analisado com robustez no trabalho corrente. Os resultados obtidos por grande parte dos estudos que visam determinar uma relação causal entre a expansão da banda larga e o crescimento econômico não consideram a possibilidade de existência de causalidade simultânea entre o PIB e a banda larga. 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 variáveis instrumentais (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.

### **3 Aplicação da Análise de Agrupamento para Identificação da Tipologia dos Municípios**

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 gestão de política pública e para 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 dentro do âmbito da análise de multivariados como análise de agrupamentos. Como definem Lattin et al. (2011), esta técnica envolve categorização, ou seja, dividir um grande conjunto de observações em conjuntos menores para que as observações dentro de 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 dentro de uma determinada área, mas separadas por regiões de densidade local baixa. Dito em outras palavras, os agrupamentos correspondem a uma modalidade de dados e o número de agrupamento ao número de modas em uma distribuição multimodal. Diversos enfoques podem ser empregados para a

---

<sup>4</sup>Esta base de dados concentra toda a população de municípios brasileiros, já que inclui todos os municípios do território nacional.

análise de agrupamentos. Dois deles são de uso mais comum, os métodos hierárquicos e os métodos de partição<sup>5</sup>. O resultado gerado pela aplicação de um método hierárquico é representado pela estrutura de árvore, no qual a solução de agrupamento  $k$  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 determinado de subgrupos, sendo que a solução de agrupamento  $k$  e a solução de agrupamento  $k + 1$  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 dentro dos grupos aumenta.

Os métodos hierárquicos geralmente enfocam 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 e 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 restem  $N$  agrupamentos com somente uma observação.

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

- 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  $D_{i,j}$ , dada por  $D_{i,j} = \frac{\|x_i - x_j\|^2}{2}$ ;
- 2) Passo 2: um novo pseudomunicípio é 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, com  $x_{i \cup j} = \frac{x_i + x_j}{2}$ ;
- 3) Passo 3: para todos os pares de municípios e pseudomunicípios, selecionar o par com menor “dissimilaridade” entre os vetores de variáveis selecionadas, considerando a métrica de Ward, dada por  $D_{i,j} = \frac{\|x_i - x_j\|^2}{\frac{1}{n_i} + \frac{1}{n_j}}$ , em que  $n_i$  e  $n_j$  são, respectivamente os números dos agrupamentos  $i$  e  $j$ ;
- 4) Passo 4: um novo pseudomunicípio é composto pela união dos pseudomunicípios com menor dissimilaridade, com  $x_{i \cup j} = \frac{n_i x_i + n_j x_j}{n_i + n_j}$ ;
- 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, dentro de um montante considerável de subjetividade, qual

<sup>5</sup>Uma descrição geral dos algoritmos de clusterização (hierárquicos e *k-means*) pode ser encontrada em Khattree & Naik (2000), Berry & Linoff (1997).

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, à 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 relação 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 número 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 dentro de cada grupo sejam relativamente similares. Para tal, precisamos encontrar um modo de medir a similaridade dentro do 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: escolhe-se, de início, uma partição dos dados em  $K$  agrupamentos. Diversos enfoques podem ser aplicados para a escolha dessa partição inicial. Uma 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 dentro do 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 dada por ESS<sup>6</sup>. 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.

---

<sup>6</sup>Error sum of square.

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 dentro de cada agrupamento pelo maior número de grupamentos). 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 & Harabasz (1974) definido da seguinte forma:

$$pseudo-F = \frac{tr(B/(K-1))}{tr(W/(N-K))}.$$

Note-se que esta 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 dentro dos agrupamentos  $W$ . Se a função  $F$  for monotonicamente crescente com o número de grupos  $K$ , os dados não apresentam uma partição “natural”. No caso da função  $F$  apresentar um ponto de máximo, então o número de conglomerados e a partição correspondem 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. Dessa forma, esta estatística estaria rejeitando a hipótese de igualdade de vetores de médias populacionais, tendo como resultado maior heterogeneidade entre os grupos.

A estatística *pseudo-t2* proposta por Duda & 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  $C_l$ , então o *pseudo-t2* é definido por:

$$PseudoT^2 = \frac{B_{il}}{[\sum_{j \in C_i} \|X_{ij} + \bar{X}_i\|^2 + \sum_{j \in C_l} \|X_{lj} + \bar{X}_l\|^2](n_i + n_l - 2)^{-1}},$$

em que  $B_{il}$  é a distância entre os grupos. O *pseudo-t2* utiliza a mesma lógica do *Pseudo-F*, buscando o valor máximo da estatística, onde 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 um, pois aquele agrupamento explica totalmente a variação do conjunto de dados. À medida que o processo de agrupamento acontece, o  $R^2$  vai decrescendo.

### 3.1 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 pelo *Atlas do Índice de Desenvolvimento Humano Municipal 2010*<sup>7</sup>. Podemos categorizar as variáveis do seguinte modo: indicadores de renda, longevidade, saúde, educação, emprego e características dos domicílios. Estas variáveis são formadas por 159 indicadores dos municípios retirados do *Atlas de Desenvolvimento Humano*. Contudo, vale ressaltar que antes da análise de agrupamentos foi feita uma análise de componentes principais, objetivando reduzir o número de variáveis utilizadas para a análise de agrupamentos. A lista de variáveis usadas na análise de agrupamentos aparece no Apêndice A.

Foram selecionadas variáveis em um total de 159 e, a partir disso, foi efetuada uma análise de componentes principais (ACP) (Lattin et al. 2011) para redução de dimensionalidade. A análise de componentes principais permite, por exemplo, acomodar situações nas quais diversas variáveis possuem alta correlação. Selecionamos um número de componentes tal que esses contabilizassem 99% da variabilidade das 159 variáveis originais<sup>8</sup>. Ao final, selecionamos 81 componentes – esses 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.

No Apêndice B é possível visualizar os seis *clusters* obtidos a partir do método *K-means* na forma de um mapa temático. 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*. Para facilitar a visualização e interpretação dos *clusters* (contíguos ou não), as figuras de 1 a 5 apresentam o diagrama de caixa<sup>9</sup> (boxplot) de variáveis agrupadas de acordo com determi-

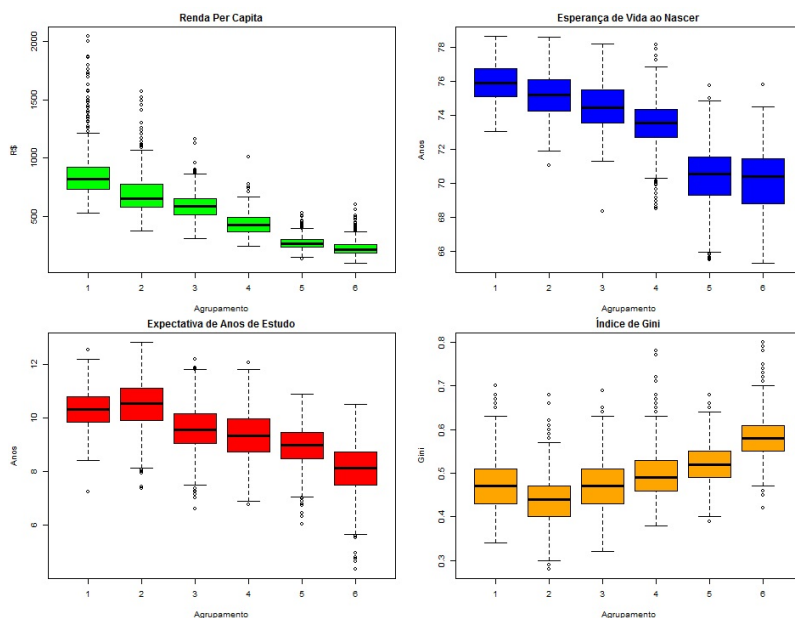
<sup>7</sup>Disponível em: <[http://www.atlasbrasil.org.br/2013/pto\\_atlasidhm/](http://www.atlasbrasil.org.br/2013/pto_atlasidhm/)>

<sup>8</sup>A análise de componentes principais (ACP) é uma técnica que converte um conjunto de variáveis possivelmente correlacionadas em um conjunto de vetores não linearmente correlacionados chamados de componentes principais. Esta transformação é feita de modo que o primeiro componente principal tem a maior variância possível, ou seja, ele é responsável pelo máximo de variabilidade nos dados, e cada componente seguinte, por sua vez, tem a máxima variância sob a restrição de ser ortogonal, isto é, não correlacionado com os componentes anteriores. Deste modo, não há redução no número de variáveis, pois ocorre apenas uma transformação das variáveis em entidades matemáticas que preservam o conteúdo genuíno da informação contido no conjunto original de variáveis.

<sup>9</sup>O diagrama de caixa é uma ferramenta gráfica para representar a variação de dados observados de uma variável numérica por meio de quartis. O boxplot também indica variabilidade fora do quartil superior e do quartil inferior.

nadas características socioeconômicas. Cada boxplot analisa quatro variáveis enquadradas dentro de um determinado aspecto socioeconômico, tal como escolaridade, renda, desigualdade, pobreza, longevidade, IDHM, condições de saneamento etc. Além disso, a Tabela 1 traz para cada *cluster* a média de algumas das variáveis socioeconômicas mais relevantes.

**Figura 1:** Boxplot da renda *per capita*, esperança de vida, esperança de anos de estudo e Índice de Gini

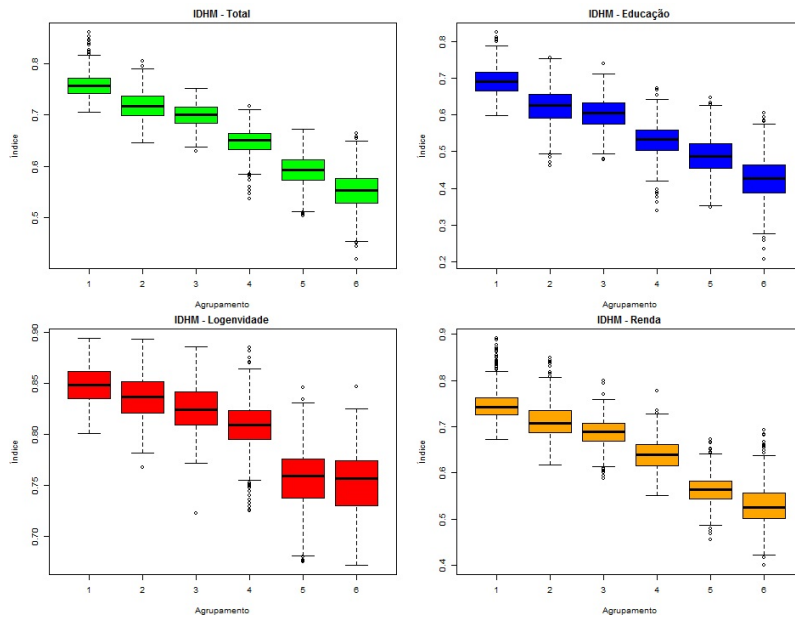


Fonte: Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/2013/>>.

Com base na análise dos diagramas de caixa que aparecem nas figuras de 1 a 5, podemos tecer as seguintes conjecturas acerca dos *clusters*:

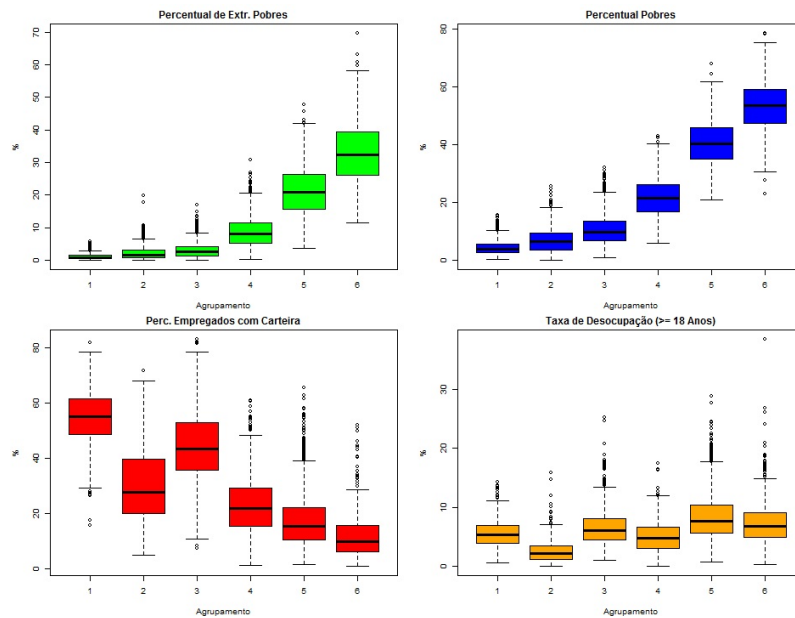
- Grupo 1: formado por municípios com maior renda *per capita*; o setor de serviços tem maior participação; melhores indicadores sociais como renda, desigualdade, expectativa de vida e escolaridade;
- Grupos 1 e 3: com alto percentual de população urbana;
- Grupo 2: com alto percentual de população rural, alta longevidade e com alta renda;
- Grupo 3: apresenta renda intermediária, mas apresenta grande percentual de trabalhadores em serviço, comércio e construção, escolaridade acima da média do país;
- 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;

**Figura 2:** Boxplot do IDHM (total e por categoria: educação, longevidade e renda)

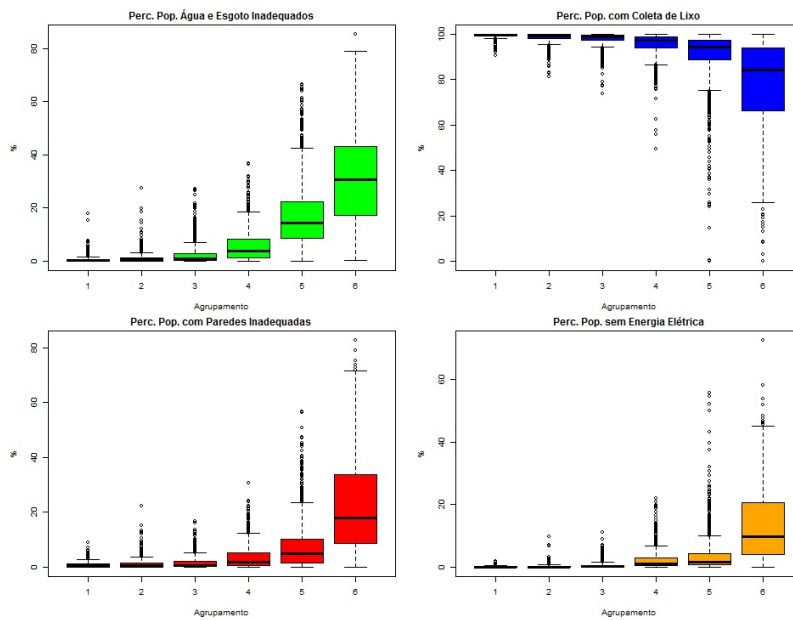


Fonte: Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/2013/>>.

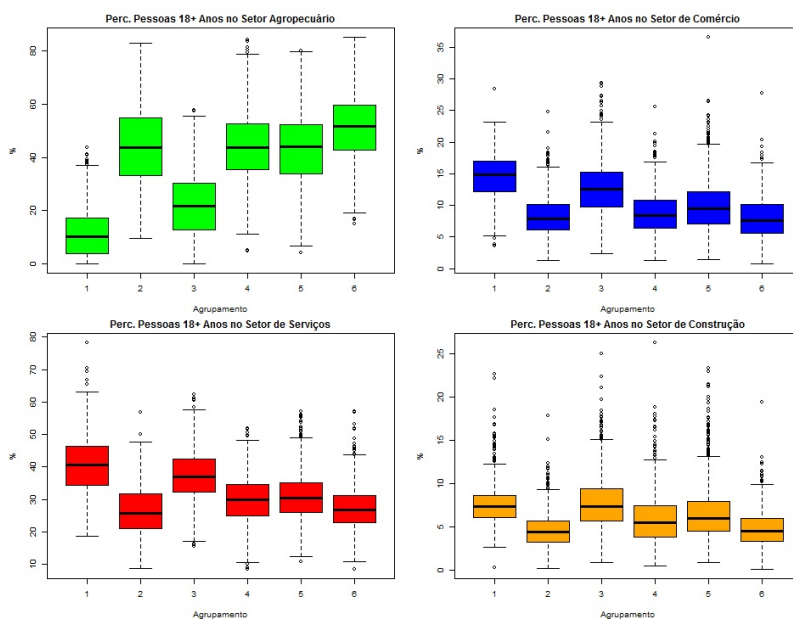
**Figura 3:** Boxplot da população extremamente pobre, população pobre, empregados com carteira assinada e taxa de desocupação



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

**Figura 4:** Boxplot das condições de saneamento básico

Fonte: Atlas do Desenvolvimento Humano no Brasil.

**Figura 5:** Dispersão da população entre os setores da economia.

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

Tabela 1: 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 agropecuário (%)	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 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
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
STFC* por habitante	7,543	3,9720	4,3780	2,2790	1,3590	0,9012	3.2824

Fonte: Instituto Brasileiro de Geografia e Estatística (2010), Relação Anual de Informações Sociais (Rais) de 2013 e Relatório Anual 2013 da Agência Nacional de Telecomunicações (Anatel).

Sistema de Telefonia Fixo Comutado.

- 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 2:** Indicadores do número de *clusters* não espaciais

Número <i>clusters</i>	CCC <sup>10</sup>	Pseudo- $t^2$	Pseudo-F	R <sup>2</sup> <i>clusters</i>	WSS <sup>11</sup>
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

Fonte: Elaborada pelos autores.

#### 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 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 dentro de um referencial teórico. Isso significa enquadrar a relação entre o PIB municipal e os investimentos em infraestrutura de banda larga dentro de 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, o 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 desses, podemos adicionar os diversos tipos de serviços de infraestrutura que contribuem, facilitam e dinamizam o processo produtivo (Barro & Sala-i Martin 1995). Inseridos nesses últimos aparecem os serviços de telecomunicações, destacando-se aqueles relacionados à 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, respectivamente,  $DT$  e  $ST$ . Naturalmente,  $DT$  depende do preço do serviço  $P_T$ , da renda  $Y$ , do nível do capital humano  $H$ , além de outras variáveis denotadas por  $I_D$  que 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<sup>12</sup>. Essas duas funções podem ser representadas de modo que

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

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

Sabendo que  $DT = ST$ , temos que

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

Substituindo a equação (4) em (1), temos que

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

Desconsiderando por hora  $H$ ,  $IF$ ,  $I_D$  e  $P_T$ , temos que

$$Y = F_2(N, K, T(Y), O_Y). \quad (1'')$$

Podemos entender de onde se origina o problema da causalidade reversa entre  $T$  e  $Y$ . Os serviços gerados pela banda larga  $T$  afetam o produto  $Y$ , mas este, por sua vez, também o afeta a demanda pelos serviços de banda larga. O modelo descrito pelas equações de (1) a (4) é próximo àquele que aparece em Roller & Waverman (2001), que 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, diferentemente, identificamos o efeito da banda larga sobre o produto com base no método de variáveis instrumentais (IV).

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:

<sup>12</sup>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.

$$\log(\text{PIBM}_{i,t}) = \alpha_i + \sum_{j=1}^J \phi_j \log(w_{i,t}) * \text{Dcluster}_{i,j} \quad (5)$$

$$+ \sum_{k=1}^K \delta_k \log(x_{i,t,k}) + \varepsilon_{i,t}$$

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

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

O parâmetro  $\alpha_i$  indica o efeito individual ou efeito específico referente ao município  $i$  capaz de contemplar a heterogeneidade existente entre os municípios. Para que a equação (5) possa ser uma representação empírica da equação (1), devemos ter as medidas de capital, emprego, produto ou renda, além de um indicador verossímil de acesso à banda larga. Naturalmente, entre a pesquisa teórica e a sua aplicabilidade geralmente existe uma certa distância. Na prática econométrica é comum não dispor de algumas variáveis que aparecem no referencial teórico. Isso acontece por vários motivos. No caso corrente, isso é devido ao grau 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*)<sup>13</sup>.

Para representar o fator trabalho ou emprego, usamos a variável vínculos<sup>14</sup> da Rais.<sup>15</sup> 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 (*DEN\_SCM*)<sup>16</sup>. Incluímos também, como variáveis de controle, as transferências governamentais decorrentes do Programa Bolsa Família (PBF), das transferências governamentais decorrentes do programa de Benefícios de Prestação Continuada (BPC) e os recursos advindos da Previdência Social, excetuando o BPC (Previ). O parâmetro  $\phi_j$  associado ao termo composto  $\log(w_{i,t}) * \text{Dcluster}_{i,j}$  define o efeito sobre o PIB nos municípios inseridos no agrupamento  $j$  decorrente de um aumento dos serviços de banda larga.

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  $\log(w_{i,t}) * \text{Dcluster}_{i,j}$  deve ser instrumentalizada no primeiro estágio conforme o modelo da equação (6).

<sup>13</sup>Esta variável foi criada envolvendo fontes de recursos federais, estaduais e municipais.

<sup>14</sup>Emprego que abrange todos os vínculos formais (celetistas, estatutários, temporários, avulsos, entre outros).

<sup>15</sup>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).

<sup>16</sup>Razão entre o número de acessos do Sistema de Comunicação Multimídia (SCM) por grupo de 100 domicílios.



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

onde  $l$  é o índice relativo às variáveis instrumentais ( $VINST$ ).

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

A base de dados do presente 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 dos censos demográficos do IBGE, 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 PBF e de BPC, cujas bases de dados são administradas pelo Ministério do Desenvolvimento Social e Agrícola (MDS). As informações sobre o emprego são oriundas da Rais. Para obter o valor em termos reais usamos como deflator o IPCA<sup>17</sup>.

#### 4.1 Resultados econométricos

Nossa estratégia econométrica é levada a cabo perfazendo um estudo comparativo entre alguns modelos econométricos e 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 3. Antes de comentarmos os resultados, cabe esclarecer alguns pontos. A escolha do período amostral de 2007 a 2015 procura considerar a maior disponibilidade de informação possível. A diferença no número de observações entre as regressões tem a ver com o fato de que nem todas as variáveis possuem observações disponíveis para todo o período amostral. Portanto, o número de observações está condicionado àquela variável cuja dimensão temporal é menor.

A coluna (1) mostra os resultados obtidos por OLS com intuito apenas de calcular a estatística VIF<sup>18</sup> (Judge 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 é verificável. 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 a estatística VIF calculada com base no método de mínimos quadrados ordinários (OLS)<sup>19</sup> 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

<sup>17</sup>Índice de preço ao consumidor amplo (IBGE 2010).

<sup>18</sup>Variation Inflation Factor (VIF). Um valor abaixo de 5.00 indica a não ocorrência de problema de multicolinearidade.

<sup>19</sup>Ordinary least square.

VIF tem uma propensão a estar acima do verdadeiro valor que inflaciona a variância de cada coeficiente quando se emprega 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 (RE) e o efeito fixo (FE). Por hipótese, assumimos nesses dois modelos que os regressores são exógenos, ou seja, que inexistente correlação entre as variáveis explicativas e o erro idiossincrático  $\varepsilon_{i,t}$ , mas não que 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<sup>20</sup>. 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 3 são mostrados os resultados obtidos pela aplicação do método de variáveis instrumentais, visando corrigir o problema de endogeneidade ou causalidade reversa entre o produto do município (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 seção 4.1 foram aplicados sobre as regressões (IV). Por economia, são mostrados apenas os resultados obtidos a partir do modelo estimado da coluna 4. Os mesmos 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 (IV) recai, sobretudo, na escolha correta dos instrumentos. Os critérios estatísticos que definem uma variável instrumental já foram descritos anteriormente. No entanto, vale a pena discorrer de modo mais intuitivo, de forma 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 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 aquelas relacionadas à infraestrutura da banda larga. Usamos três conjuntos de variáveis instrumentais:

$$IV\_1 = [DEN\_STFC, 521\_KBPS].$$

$$IV\_2 = [IV\_1, 3G, OUTRAS].$$

---

<sup>20</sup>A não rejeição da hipótese nula do teste de Breusch-Pagan indicaria que o modelo deve ser estimado usando uma regressão OLS com dados agrupados (POLS ou Pooling OLS).

IV\_3 = [IV\_2, OPERADORAS, ERBS].

em que:

*DEN\_STFC*: densidade de acessos por telefonia fixa por 100 domicílios;

*521\_KBPS*: número de acessos na faixa de velocidade entre 521 Kbps e 2 Mbps 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.

**Fonte:** Dados diretamente fornecidos pela Anatel.

Vejamos como podemos defender a hipótese de que esse conjunto de variáveis podem ser candidatos a bons instrumentos. Vejamos o caso da variável *DEN\_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 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. De mesma forma, justificamos o uso das demais variáveis como instrumentos apelando para o fato de que elas têm efeito indireto sobre o PIB devido ao seu efeito direto sobre a banda larga.

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 4 porque teríamos que mostrá-los para cada regressão da variável *DEN\_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 3<sup>21</sup>. Interessante notar que os coeficientes são significativos e apresentam os sinais “esperados”. No caso da variável *DEN\_STFC*, temos a correlação positiva entre essa variável e *DEN\_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

<sup>21</sup>Os resultados econômicos podem ser obtidos mediante contato com os autores.

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 do preço em decorrência do aumento da competição, seja devido à maior disponibilidade do serviço.

A variável *ERBS* não apresenta significância. Isso provavelmente é explicado pela existência de colinearidade dessa variável com outro instrumento<sup>22</sup>. 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 em relação 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 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 Sogo rejeita a hipótese de fraqueza dos instrumentos, enquanto o teste de Wu-Hausman rejeita a hipótese nula de que a variável *DEN\_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 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 3, 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 *DEN\_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 a Tabela 3, 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 com o esperado, 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 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 uma 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

---

<sup>22</sup>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*.

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 SCM<sup>23</sup> 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 percentis  $q(10)$ <sup>24</sup>,  $q(25)$ ,  $q(50)$ ,  $q(75)$ ,  $q(90)$  e  $q(99)$ , calculados para a variável ACS\_SCM. Por exemplo, a variável ACM\_SCM\_25 indica o número de acessos SCM maior  $q(10)$  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 dentro das faixas intermediárias de acesso.

## 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 municipal e 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. 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<sup>25</sup>. Além disso, as regiões Norte e Nordeste também apresentam as mais baixas velocidades de banda larga, o que naturalmente implica menor qualidade. Assim, o menor efeito da banda larga sobre a economia dos municípios das regiões Norte e Nordeste possivelmente seja explicado pela baixa eficiência desse serviço. Portanto, como proposta de política pública, fica a sugestão de considerar um programa que minimize as disparidades regionais de acesso à banda larga; não somente em termos do acesso, mas também de qualidade. Não incorpo-

<sup>23</sup>Sistema de comunicação multimídia.

<sup>24</sup> $q(10)$ : percentis de 10%.

<sup>25</sup>Não possuem *backhaul* de fibra.

**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)	11,302 (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)	-	-	-	-
Teste F de Anderson-Rubin <sup>1</sup>	-	-	-	1960,81 (0,000)	-	-
Teste F de Stock-Yogo <sup>2</sup>	-	-	-	24591,68 (16,85)	-	1916,98 (0,000)
Teste de sub-identificação <sup>3</sup>	-	-	-	21239,68 (0,000)	-	-
Teste Qui2 de Sargan/Hansen <sup>4</sup>	-	-	-	1649,79 (0,000)	-	-
Teste F de Wu-Hausman <sup>5</sup>	-	-	-	575,71 (0,000)	605,842 (0,000)	34,958 (0,000)
R <sup>2</sup> <sup>8</sup>	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 <sup>6</sup>	-	-	-	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 <sup>7</sup>	-	-	-	-	-	-
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

Fonte: Elaborada pelos autores.

Notas: <sup>1</sup>: Teste Wald para verificar a significância do regressor endógeno na equação estrutural. <sup>2</sup>: 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. <sup>3</sup>: Teste LM para verificar a hipótese de que se os instrumentos excluídos são “relevantes”, isto é, correlacionado com os regressores endógenos. <sup>4</sup>: Testa um subconjunto de condições de ortogonalidade, isto é, é um teste da exogeneidade de um ou mais instrumentos. <sup>5</sup>: Teste de endogeneidade: sob a hipótese nula o regressor endógeno é tratado como se fosse exógeno. <sup>6</sup>: 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. <sup>7</sup>: 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. <sup>8</sup>: Considera-se o R<sup>2</sup> “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 <sup>1</sup> (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)	-0,080 (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)	0,084 (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 <sup>2</sup>	-	-	6800,00 (0,000)	-	-
Teste de Redundância <sup>3</sup>	-	-	-	820,95 (0,000)	822,46 (0,000)
R2	0,88	0,76	-	-	-
NOBS	44511	44511	32644	32037	32037

Notas: <sup>1</sup>: GMM com correção de autocorrelação serial de ordem dois. <sup>2</sup>: testa a hipótese de que as variáveis 3G e *OUTRAS* são redundantes como instrumentos. <sup>3</sup>: 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 <sup>1</sup>	-	-	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 de Wu-Hausman	-	-	-	-	49,987 (0,000)	47,842 (0,000)
Teste de Redundância <sup>2</sup>	-	-	-	-	61000,00 (0,000)	-
Teste de Redundância <sup>3</sup>	-	-	-	-	-	4206,00 (0,000)
R2	0,75	0,76	0,87	-	-	-
NOBS	44098	44098	44098	44098	32540	31935

Notas: <sup>1</sup>: *ACS\_SCM\_10*: percentil de 10%  $q(10)$  da variável. <sup>2</sup>: testa a hipótese de que as variáveis *3G* e *OUTRAS* são redundantes como instrumentos. <sup>3</sup>: testa a hipótese de que as variáveis *OPERADORAS* e *ERBS* são redundantes.



rar esse ponto em uma agenda prioritária de política pública pode levar ao aprofundamento da desigualdade entre as regiões do Brasil.

Também 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.

Podemos usar os resultados descritos nos parágrafos anteriores para embasar algumas sugestões de políticas públicas que visem regular os investimentos em banda larga no Brasil. A incorporação das informações advindas do exercício de clusterização permitiu identificar que o impacto econômico ocorre de forma diferenciada entre os agrupamentos de municípios homogêneos. Com base nisso, deve-se, numa primeira instância, partilhar os investimentos de modo que o maior peso incida sobre os *clusters* onde o efeito decorrente da expansão da banda larga é maior. O menor peso deve estar associado aos *clusters* cujo efeito econômico da banda larga é menor. Num segundo momento, a dispersão dos investimentos dentro dos *clusters* deve se dar de modo a priorizar os municípios que possuem a economia concentrada em setores nos quais a banda larga tem maior impacto econômico e, concomitantemente, deve procurar fazer que a distribuição dos acessos à banda larga se concentre em torno dos níveis médios de acessos de maneira a considerar os resultados da Tabela 3.

## Referências Bibliográficas

- Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, New Jersey: Wiley.
- Barro, R. J. & Sala-i Martin, X. (1995), *Economic growth*, New York: McGraw-Hill.
- Berry, M. & Linoff, G. (1997), *Data Mining techniques*, New York: Wiley.
- Calinski, T. & Harabasz, J. (1974), ‘A dendrite method for cluster analysis’, *Communications in Statistics* 3, 1–27.
- Czernich, N., Falck, O., Kretschmer, T. & Woessmann, L. (2009), *Broadband infrastructure and economic growth*, CESifo.
- Duda, R. O. & Hart, P. E. (1973), *Pattern classification and scene analysis*, New York: Wiley.
- Fonseca, H. M. M. (2015), *Ensaio em teoria de contratos aplicada no setor de telecomunicações com foco nos investimentos*, PhD thesis, Tese (Doutorado) –, Brasília: Universidade de Brasília.
- Hartigan, J. A. (1975), *Clustering algorithms*, New Jersey: Wiley.
- IBGE (2010), *Censo Demográfico 2010*, Rio de Janeiro: IBGE. Instituto Brasileiro de Geografia e Estatística.
- Judge, G., Hill, R. C., Griffiths, W. E., Lutkepohl, H. & Lee, T. C. (1982), *Introduction to the theory and practice of econometrics*, New York: Wiley.

Katz, R. L. (2012), *The economic and social impact of telecommunications output: empirical evidence in the US and Europe*, Brussels.

Khattree, R. & Naik, D. N. (2000), *Multivariate data reduction and discrimination with SAS software*, Cary, NC: SAS.

Koutroumpis, P. (2009), 'The Economic Impact of Broadband on Growth: A Simultaneous Approach', *Telecommunications Policy* **33**(9), 471–485.

Lattin, J., Carrol, D. & Green, P. (2011), *Análise de dados multivariados*, São Paulo: Cengage.

Macedo, H. R. & Carvalho, A. X. Y. (2010a), 'Aumento do acesso à internet em banda larga no Brasil e sua possível relação com o crescimento econômico: uma análise de dados em painel'.

Macedo, H. R. & Carvalho, A. X. Y. (2010b), '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'.

Norton, S. W. (1992), 'Transaction Costs, Telecommunications, and the Microeconomics of Macroeconomic Growth', *Economic Development and Cultural Change* **40**, 175–196.

Qiang, C. & Rossotto, C. (2009), Economic Impacts of Broadband, in 'In: Information and Communications for Development 2009: Extending Reach and Increasing Impact', Washington, D.C.: World Bank, p. 35-50.

Roller, L. & Waverman, L. (2001), 'Telecommunications and Infrastructure and Economic Development: A Simultaneous Approach', *American Economic Review* **91**(4), 909–923.

Stiroh, K. (2002), 'Information technology and the U. S. productivity revival', *American Economic Review* **92**(5), 1559–1576.

## Apêndice A

Tabela 6: Lista de variáveis

Nome da Variável	Eixos
perc_mulheres_10 a 14_tiveram_filhos	Social
perc_mulheres_15 a 17_tiveram_filhos	Social
perc_população_10 a 14 anos	Social
perc_população_11 a 13 anos	Social
perc_população_11 a 14 anos	Social
perc_população_12 a 14 anos	Social
perc_população_15 a 17 anos	Social
perc_população_15 a 17 anos 2	Social
perc_população_15 a 24 anos	Social
perc_população_15 anos ou mais	Social
perc_população_16 a 18 anos	Social
perc_população_18 a 20 anos	Social
perc_população_18 a 24 anos	Social
perc_população_18 anos ou mais	Social
perc_população_18 anos ou mais 2	Social
perc_população_19 a 21 anos	Social
perc_população_1 a 3 anos	Social
perc_população_25 anos ou mais	Social
perc_população_4 anos	Social
perc_população_5 anos	Social
perc_população_65 anos ou mais	Social
perc_população_6 a 10 anos	Social
perc_população_6 a 17 anos	Social
perc_população_6 anos	Social
perc_população_até 1 ano	Social
perc_população_residente_feminina	Social
perc_população_residente_masculina	Social
perc_população_rural	Social
perc_população_urbana	Social
taxa_envelhecimento	Social
taxa_fecundidade	Social
expec_anos_estudo	Social
perc_0 a 3_na_escola	Social
perc_11 a 13 anos_fund_quase completo	Social
perc_11 a 14_na_escola	Social
perc_12 a 14 anos_fund_quase completo	Social
perc_15 a 17_anos_no_fundamental	Social
perc_15 a 17_na_escola	Social
perc_15 a 17 anos_fund_completo	Social
perc_15 a 17 anos_médio_1 ano_atraso	Social
perc_15 a 17 anos_médio_2 anos ou mais_atraso	Social
perc_15 a 17 anos_médio_sem_atraso	Social

**Tabela 6:** Lista de variáveis (continuação)

<b>Nome da Variável</b>	<b>Eixos</b>
perc_15 a 17 anos_no_superior	Social
perc_16 a 18 anos_fund_completo	Social
perc_18 a 20 anos_médio_completo	Social
perc_18 a 24_anos_no_fundamental	Social
perc_18 a 24_na_escola	Social
perc_18 a 24 anos_fund_completo	Social
perc_18 a 24 anos_médio_completo	Social
perc_18 a 24 anos_no_médio	Social
perc_18 anos ou mais_fund_completo	Social
perc_18 anos ou mais_médio_completo	Social
perc_19 a 21 anos_médio_completo	Social
perc_25 a 29_na_escola	Social
perc_25 anos ou mais_fund_completo	Social
perc_25 anos ou mais_médio_completo	Social
perc_25 anos ou mais_superior_completo	Social
perc_4 a 5_na_escola	Social
perc_4 a 5 anos_no_fundamental	Social
perc_4 a 6_na_escola	Social
perc_5 a 6_na_escola	Social
perc_6 a 14 anos_fundam_1 ano_atraso	Social
perc_6 a 14 anos_fundam_2 anos ou mais_atraso	Social
perc_6 a 14 anos_fundam_sem_atraso	Social
perc_6 a 14 anos_na_escola	Social
perc_6 a 14 anos_no_médio	Social
perc_6 a 17 anos_básico_1 ano_atraso	Social
perc_6 a 17 anos_básico_2 anos ou mais_atraso	Social
perc_6 a 17 anos_básico_sem_atraso	Social
perc_6 a 17 anos_na_escola	Social
perc_6 anos_na_escola	Social
perc_crianças_em_dom_sem_fund_completo	Social
taxa_analf_11 a 14 anos	Social
taxa_analf_15 a 17 anos	Social
taxa_analf_15 ou mais	Social
taxa_analf_18 a 24 anos	Social
taxa_analf_18 ou mais	Social
taxa_analf_25 a 29 anos	Social
taxa_analf_25 ou mais	Social
taxa_freq_líquida_básico	Social
taxa_freq_líquida_fundam	Social
taxa_freq_líquida_médio	Social
taxa_freq_líquida_pré-escola	Social
taxa_freq_líquida_superior	Social
perc_pop_dom_com_coleta_lixo	Social
perc_pop_dom_densidade_maior_2	Social
perc_pop_dom_energia_életrica	Social
perc_pop_em_dom_água_encanada	Social

**Tabela 6:** Lista de variáveis (continuação)

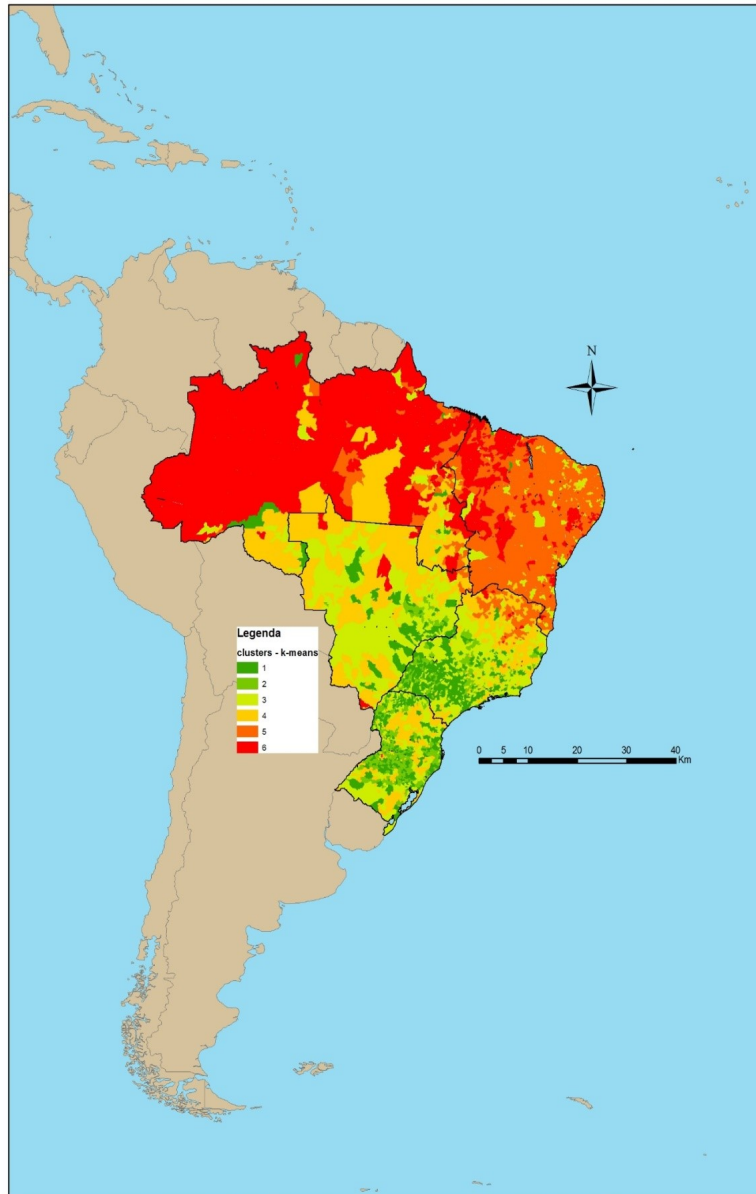
Nome da Variável	Eixos
perc_pop_em_dom_banheiro_e_água_encanada	Social
perc_pop_total_em_dom_particulares_permanentes	Social
perc_pessoas_dom_água_esgoto_inadequados	Social
perc_pessoas_dom_paredes_inadequadas	Social
perc_pessoas_em_dom_sem_energia_életrica	Social
índice_gini	Economia
índice_theil_1	Economia
perc_crianças_extrem_pobres	Economia
perc_crianças_pobres	Economia
perc_crianças_vulneráveis_à_pobreza	Economia
perc_extremamente_pobres	Economia
perc_mães_chefes_familia_sem_fund_completo_filhos_menos_15	Economia
perc_pessoas_15_a_24_nem_estudo_nem_trabalho_vuln_à_pobreza	Economia
perc_pessoas_18_ou_mais_sem_fund_completo_ocup_informal	Economia
perc_pessoas_dom_vuln_à_pobreza_sem_fund_completo	Economia
perc_pessoas_em_dom_vuln_à_pobreza_com_idosos_dependentes	Economia
perc_pessoas_vuln_à_pobreza_gasta_mais_de_1_hora_até_trabalho	Economia
perc_pobres	Economia
perc_vulneráveis_à_pobreza	Economia
razão_10_mais_ricos_a_40_mais_pobres	Economia
razão_20_mais_ricos_a_40_mais_pobres	Economia
razão_dependência	Economia
esperança_de_vida_ao_nascer	Social
mort_até_5_anos	Social
mort_infantil	Social
prob_sobrev_até_40_anos	Social
prob_sobrev_até_60_anos	Social
grau_formalizacao_ocupados_18_anos_ou_mais	Economia
perc_PEA_10_a_14_anos	Economia
perc_PEA_10_anos_ou_mais	Economia
perc_PEA_15_a_17_anos	Economia
perc_PEA_18_ou_mais_anos	Economia
perc_empregadores_18_anos_ou_mais	Economia
perc_empregados_cart_assinada_18_anos_ou_mais	Economia
perc_ocupados_com_renda_até_1_sm_18_anos_ou_mais	Economia
perc_ocupados_com_renda_até_2_sm_18_anos_ou_mais	Economia
perc_ocupados_com_renda_até_3_sm_18_anos_ou_mais	Economia
perc_ocupados_com_renda_até_5_sm_18_ano_sou_mais	Economia
perc_ocupados_fund_completo_18_anos_ou_mais	Economia
perc_ocupados_indústria_transf_18_anos_ou_mais	Economia
perc_ocupados_médio_completo_18_anos_ou_mais	Economia
perc_ocupados_no_SIUP_18_anos_ou_mais	Economia
perc_ocupados_sem_renda_18_anos_ou_mais	Economia
perc_ocupados_superior_completo_18_anos_ou_mais	Economia
perc_renda_vindo_de_trabalho	Economia
perc_setor_agropecuário_18_anos_ou_mais	Economia

**Tabela 6:** Lista de variáveis (continuação)

Nome da Variável	Eixos
perc_setor_comércio_18 anos ou mais	Economia
perc_setor_construção_18 anos ou mais	Economia
perc_setor_extrativo_mineral_18 anos ou mais	Economia
perc_setor_público_18 anos ou mais	Economia
perc_setor_serviços_18 anos ou mais	Economia
perc_trabs_conta_própria_18 anos ou mais	Economia
perc_trabs_sem_carteira_18 anos ou mais	Economia
renda_per_capita_décimo_mais_rico	Economia
renda_per_capita_média_extrem_pobres	Economia
renda_per_capita_média_pobres	Economia
renda_per_capita_média_vulnerável_à_pobreza	Economia
renda_per_capita_quinto_mais_rico	Economia
renda_per_capita_quinto1_mais_pobre	Economia
renda_per_capita_quinto2_mais_pobre	Economia
renda_per_capita_quinto3_mais_pobre	Economia
renda_per_capita_quinto4_mais_pobre	Economia
renda_per_capita_sem_nulos	Economia
rendimento_média_ocupados_18 anos ou mais	Economia
taxa_desocupação_10 a 14 anos	Economia
taxa_desocupação_10 anos ou mais	Economia
taxa_desocupação_15 a 17 anos	Economia
taxa_desocupação_18 a 24 anos	Economia
taxa_desocupação_18 anos ou mais	Economia
taxa_desocupação_25 a 29 anos	Economia
renda_per_capita	Economia

Apêndice B

**Figura 6:** Mapa do agrupamento para os seis *clusters* pelo método *k-means*



Fonte: Elaboração pelos autores.

