TEXTO PARA DISCUSSÃO

2349

AVALIAÇÃO DE PRIORIZAÇÕES PARA EXPANSÃO DA REDE DE ACESSO À BANDA LARGA NOS MUNICÍPIOS BRASILEIROS

Alexandre Ywata de Carvalho
Mário Jorge Mendonça
José Jaime da Silva

ipea

2349
TEXTO PARA DISCUSSÃO
Brasília, novembro de 2017

AVALIAÇÃO DE PRIORIZAÇÕES PARA EXPANSÃO DA REDE DE ACESSO À BANDA LARGA NOS MUNICÍPIOS BRASILEIROS

Alexandre Ywata de Carvalho[1]
Mário Jorge Mendonça[2]
José Jaime da Silva[3]

1. Técnico de planejamento e pesquisa na Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea. *E-mail*: <alexandre.ywata@ipea.gov.br>.
2. Técnico de planejamento e pesquisa na Dirur do Ipea. *E-mail*: <mario.mendonca@ipea.gov.br>.
3. Assistente de pesquisa na Dirur e mestre em estudos populacionais da Escola Nacional de Ciências Estatísticas (Ence). *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, Interino
Rogério Boueri Miranda

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.

JEL: C21; O33.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO 7

2 CONTEXTUALIZAÇÃO DOS CRITÉRIOS DE PRIORIZAÇÃO DE ACESSO E CUSTO DE EXTENSÃO DA REDE 9

3 DETERMINAÇÃO DO CUSTO MÍNIMO DE EXTENSÃO DA REDE DE FIBRA 12

4 DIMENSIONAMENTO DO MERCADO POTENCIAL DE BANDA LARGA NO BRASIL 17

5 AVALIAÇÃO DOS EFEITOS DOS INVESTIMENTOS EM TELECOMUNICAÇÕES SOBRE O PIB 23

6 IMPLICAÇÕES DAS POLÍTICAS DE PRIORIZAÇÃO 31

REFERÊNCIAS 37

APÊNDICE A 38

SINOPSE

Este estudo tem como propósito apresentar um modelo de avaliação econômica da expansão da rede de banda larga para os municípios brasileiros não conectados. Isso é feito levando em consideração não apenas o retorno econômico dessa expansão, mas também o seu custo. O retorno é calculado a partir de três critérios de priorização: o tamanho da população, o índice de desenvolvimento humano (IDH) e o mercado potencial. Uma vez escolhido um desses critérios, estima-se o efeito sobre certas variáveis de interesse, como o efeito sobre o produto, o acesso à população mais pobre etc. Para determinar o mercado potencial, estimamos um modelo de dimensionamento de mercado como base na regressão logística, usando dados da Pesquisa Nacional por Amostra de Domicílio (Pnad) 2015 e do Censo Demográfico 2010, do Instituto Brasileiro de Geografia e Estatística (IBGE). O efeito da expansão da banda larga sobre o produto interno bruto (PIB) municipal foi obtido empregando o modelo de dados em painel. O custo de expansão da rede associado às escolhas priorizadas é gerado com base na solução da árvore geradora mínima (AGM), da teoria dos grafos. Observou-se que a priorização com base na escolha do mercado potencial é a alternativa que gera maior retorno econômico. Também os resultados obtidos com base no tamanho da população são próximos ao custo de expansão da rede, enquanto a priorização pelo IDH municipal é o critério que apresenta o pior desempenho, no tocante à população pobre e extremamente pobre.

Palavras-chave: banda larga; rede de fibra; teoria dos grafos; árvore geradora mínima; regressão logística.

ABSTRACT

This study aims to present a model for economic evaluation of the expansion of the broadband network for the Brazilian municipalities. This is done taking into account not only the economic return, but also the cost of expanding the network. Return is calculated from three criteria of prioritization: size of population, HDI, and potential market. Once one of these criteria is chosen, the effect on certain variables of interest, such as the effect on the product, access to the poorest population, etc., is estimated. To determine the potential demand of broadband, we estimated model combining two databases: PNAD 2015 and Census 2010. The effect of broadband on municipal GDP was obtained using the panel data model. To evaluate the cost to expand network

we the applied the solution of the minimum generating tree from the graph theory. It was observed that the criterion based on the potential market is the alternative that generates greater economic return. Also the results obtained based on the size of the population are close to this one, whereas the prioritization by the municipal HDI is the criterion that presents the worst performance that penalizes the poor and the extremely poor population.

Keywords: broadband; backhaul; graph theory; minimum spanning tree; logit regression; panel data.

1 INTRODUÇÃO

Embora no Brasil já se tenha a percepção da importância do acesso à banda larga, há muito ainda a ser feito. Diferentemente do que acontece com a telefonia fixa, na qual a universalização já foi alcançada, há um caminho longo a percorrer no que se refere à banda larga. No Brasil, dados da Agência Nacional de Telecomunicações (Anatel) mostram que a infraestrutura de banda larga é ainda precária. Existem ainda 2.325 municípios que não possuem *backhaul* de fibra, sendo que 58% se situam nas regiões Norte e Nordeste. Assim, 14% da população brasileira está desassistida por essa infraestrutura. Os 54% dos municípios sem fibra estão nas regiões Norte e Nordeste; sendo que, na região Sudeste, 23% dos municípios sem fibra se situam no estado de Minas Gerais. Mesmo nos municípios dotados de *backhaul* de fibra, a qualidade do serviço de banda larga fica aquém do desejado. Na média, nos 2.221 municípios que possuem *backhaul*, a velocidade da banda larga de fibra é de 5 Mbps.[1] Novamente, as regiões Norte e Nordeste possuem as mais baixas velocidades.

Tendo em vista a relevância dos serviços derivados do acesso à banda larga, sua massificação está inserida na agenda prioritária de política pública. No entanto, ocorre que, tendo em vista a dimensão territorial do país, o custo de expansão da rede e a limitação de recursos, sobretudo no momento atual de crise, existe restrição quanto a levar a cabo a ampliação do acesso à banda larga de modo pleno. Isso significa que devem existir critérios para priorização da expansão da rede de banda larga, nos quais devem ser confrontados tanto os benefícios decorrentes do aumento do acesso à rede como os custos de expansão. Em outras palavras, deve-se proceder a algum tipo de análise de custo-benefício, como meio de racionalizar o gasto de expansão do acesso aos serviços de banda larga. Portanto, embora o conhecimento do retorno econômico ou os benefícios derivados do acesso à rede de banda larga sejam importantes, o emprego apenas desses indicadores sem considerar os custos associados a essa expansão faz com que a avaliação de tal política seja incorreta, tanto na ótica do setor público quanto na do setor privado.

Este estudo tem como finalidade lançar luz na discussão sobre a análise de custo-benefício de expansão da banda larga no Brasil. Existem vários indicadores sociais que podem ser utilizados de modo a servir de base para avaliação de política pública, sendo difícil identificar um critério que seja superior aos demais. Portanto, a escolha de um deles se dará sempre em certa margem de subjetividade, exceto quando se tem um

1. O Brasil está entre os países com as mais baixas velocidades de banda larga.

objetivo específico. Optamos por analisar a expansão da banda larga com base em três critérios de priorização: o índice de desenvolvimento humano (IDH),[2] o tamanho da população e o mercado potencial. Para cada um deles, estima-se o custo de expansão do acesso à banda larga nos municípios ranqueados de acordo com um dos critérios de priorização. Além disso, para cada um desses critérios, examinamos o efeito sobre algumas variáveis de interesse, à medida que mais municípios priorizados vão sendo atendidos com acesso à fibra.

Existem diferentes alternativas que podem ser empregadas para aumentar a rede de banda larga, como ampliar a rede de transporte de alta capacidade, bem como possibilitar a expansão do serviço móvel de telefonia, da rede de acesso à banda larga fixa (SCM) e do *backhaul* de fibra etc. Neste estudo, abordaremos o problema do custo da ampliação da rede de banda larga, por meio da expansão do *backhaul* de fibra. O cálculo desse custo é feito por meio do emprego da teoria dos grafos. Com base nessa teoria e utilizando informações disponibilizadas pela Anatel, é possível construir uma rede de banda larga que conecta todos os municípios; sendo essa rede a de menor custo. Aqui, o custo é entendido como proporcional à distância entre as conexões. Uma vez que o custo mínimo tenha sido determinado, é possível associar qualquer critério de priorização para ampliação da rede de banda larga ao seu custo. Por exemplo, vamos supor que é meta fazer chegar a banda larga aos municípios de população mais pobre. Tendo ranqueado tais municípios, podemos identificar o custo em termos de adicional de fibra, para conectar os municípios nos quais essas populações habitam.

Esta pesquisa está estruturada da seguinte forma. A seção 2 descreve as linhas mestras da metodologia adotada neste estudo, para avaliação dos critérios de priorização. A seção 3 introduz uma breve digressão sobre a teoria dos grafos e apresenta a solução, com base nessa teoria, para a determinação do custo mínimo de expansão da rede de banda larga. Nas seções 4 e 5, descrevemos a metodologia e os resultados para determinação, respectivamente, do mercado potencial de banda larga e do efeito devido à expansão da banda larga sobre o produto interno bruto (PIB). Por fim, na seção 6, destacam-se as implicações sobre a economia, a população e o dimensionamento do mercado decorrentes das escolhas dos diferentes critérios de priorização.

2. Indicador social elaborado pela combinação de informações de expectativa de vida, escolaridade e renda *per capita*, ajustada pela paridade do poder de compra.

2 CONTEXTUALIZAÇÃO DOS CRITÉRIOS DE PRIORIZAÇÃO DE ACESSO E CUSTO DE EXTENSÃO DA REDE

Como mencionado na introdução deste estudo, decidimos por nortear a ampliação do acesso à rede de banda larga com base em três diferentes critérios, que são: o tamanho da população, o IDH e o mercado potencial. Para cada um desses critérios de priorização, verificamos o efeito derivado sobre algumas variáveis de interesse. Essas variáveis são as seguintes:

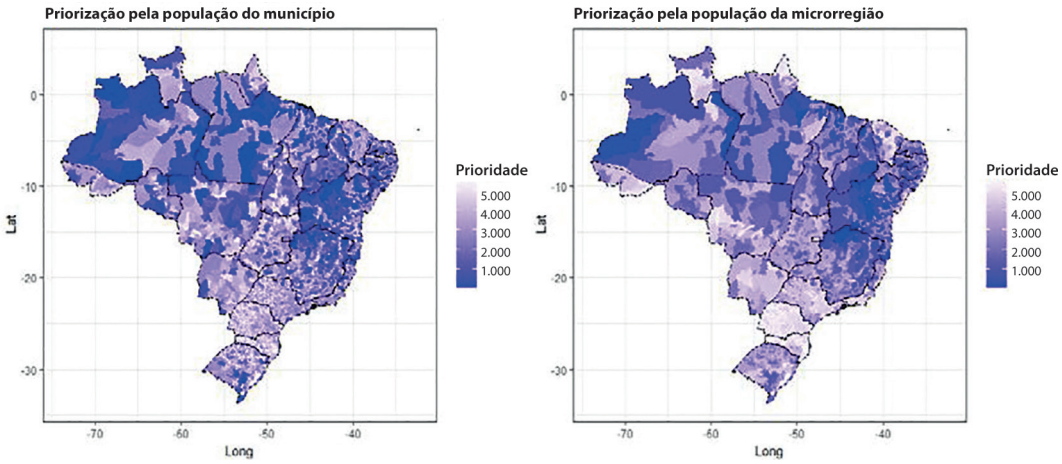
- população total atingida nos municípios com novo acesso;
- população pobre beneficiada. Estimativa feita para 2015, a partir dos dados do Atlas do Desenvolvimento Humano (ADH);
- população extremamente pobre beneficiada. Estimativa feita para 2015, a partir dos dados do ADH;
- população jovem (até 20 anos) beneficiada. Estimativa feita para 2015, a partir dos dados do ADH;
- mercado potencial atingido;
- PIB dos municípios atendidos; e
- impacto sobre o produto interno bruto (PIB do município multiplicado pela elasticidade média do *cluster*).

A completa caracterização dos procedimentos envolvidos na determinação da análise de custo e benefício, decorrentes da expansão do acesso à rede de banda larga de acordo com os critérios de priorização definidos, somente ficará completamente explícita com a leitura completa deste texto. No entanto, iremos aqui estabelecer as bases iniciais para esse entendimento. As etapas para a elaboração da análise são as seguintes.

1) Define-se a escolha de um critério. Por exemplo, opta-se por priorizar pelo tamanho da população.
2) Procede-se à ordenação dos municípios em ordem decrescente de número de habitantes, de modo que a priorização seja estabelecida pelo número de indivíduos a serem atendidos. Por exemplo, suponha-se que se deseje levar o acesso à banda larga a mais de 5 milhões de pessoas.
3) Verifica-se quais são os primeiros *N* municípios, cuja população somada atinge a meta estabelecida.

As figuras 1 e 2 mostram os municípios e as microrregiões[3] priorizadas pelo tamanho da população. A utilização da abrangência geográfica por microrregião permite estudar potenciais ganhos de aglomeração espacial na expansão da rede. Para os demais critérios, o procedimento é o mesmo.

FIGURA 1
Priorização pela população do município

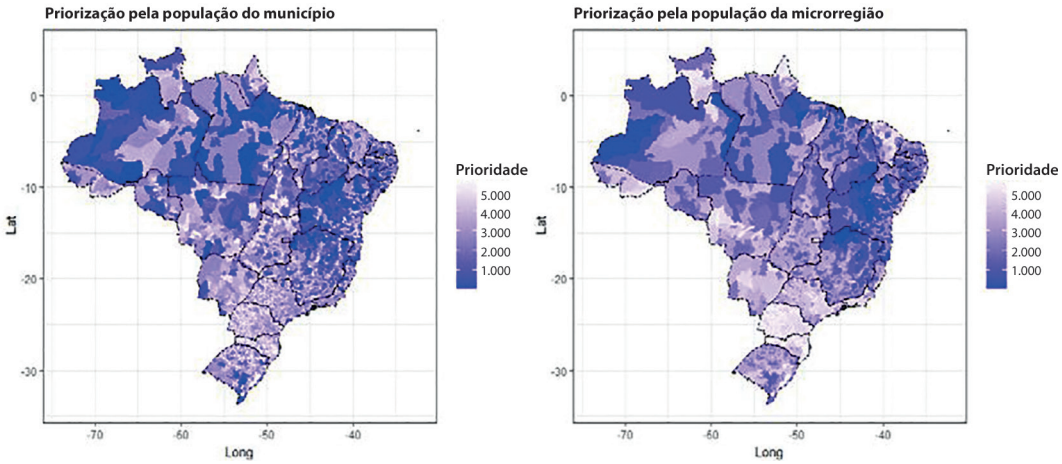


Elaboração dos autores.

Tendo escolhido o critério com que se deseja priorizar o acesso à rede, a tarefa a seguir diz respeito à estimativa do custo necessário para que os municípios priorizados sejam conectados. Poder-se-ia imaginar que uma solução seria conectar cada município à rede de fibra mais próxima. No entanto, conforme será demonstrado, isso é uma solução ótima quando o número de municípios envolvido é pequeno. À medida que esse número cresce, esta solução se afasta da solução ótima decorrente do modelo da árvore geradora mínima (AGM), que se configura como um tema de relevância da teoria dos grafos. Esse assunto será aprofundado na seção 3.

3. De acordo com a Constituição Federal (CF) de 1988, microrregião é um agrupamento de municípios limítrofes com o objetivo de integrar a organização, o planejamento e a execução de funções públicas de interesse comum, definidas por lei complementar estadual. No entanto, poucas são as microrregiões definidas dessa forma. Consequentemente, o termo é muito mais conhecido em função de seu uso prático pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que, para fins estatísticos e com base em similaridades econômicas e sociais, divide os diversos estados da Federação em microrregiões. No Brasil, existe um total de 558 microrregiões.

FIGURA 2
Priorização pela população da microrregião



Elaboração dos autores.

Vale também examinar as implicações da escolha dos critérios de priorização sobre as variáveis de interesse elencadas antes. Esse tipo de análise permite verificar se determinado critério de priorização é compatível com outras agendas de política pública. Por exemplo, a priorização pelo IDH pode levar a um menor efeito sobre o PIB. Não necessariamente a escolha pelo IDH deverá contemplar os mais pobres. Das variáveis de interesse listadas, cabe aprofundar os comentários com respeito ao mercado potencial e ao impacto sobre o PIB, visto que as demais são obtidas por leitura direta da base de dados. No que concerne ao mercado potencial e ao efeito sobre o PIB decorrente da expansão da banda larga, é necessário empreender todo um estudo para acessar essas grandezas.

O mercado potencial de banda larga é obtido por projeção com base no modelo de regressão logística, cuja variável dependente é a probabilidade de se ter acesso à internet por banda larga, calcada na Pesquisa Nacional por Amostra de Domicílio (Pnad) 2015, do Instituto Brasileiro de Geografia e Estatística (IBGE). A projeção do mercado de banda larga é feita usando os parâmetros do modelo estimado, mas com as "mesmas" variáveis preditoras da base do Censo Demográfico 2010 do IBGE. Projetamos, tendo em vista as características físicas e socioeconômicas dos domicílios, qual a demanda potencial (acessar ou não a internet) do domicílio não atendido por banda larga caso existisse conexão. Na seção 4, descreveremos todos os procedimentos envolvidos e apresentaremos as estimativas do mercado potencial. A outra variável de interesse não obtida diretamente dos dados é o efeito sobre o PIB decorrente do acesso à banda larga.

Obtivemos essa medida por meio do emprego do modelo econométrico de dados em painel dos municípios brasileiros. Tal modelo incorpora as diferenças regionais. A análise referente à estimação do efeito da banda larga sobre o PIB será tratada na seção 4. Por fim, na seção 6, procura-se demonstrar as implicações sobre as variáveis de interesse das decisões decorrentes da aplicação dos diferentes critérios de priorização.

3 DETERMINAÇÃO DO CUSTO MÍNIMO DE EXTENSÃO DA REDE DE FIBRA

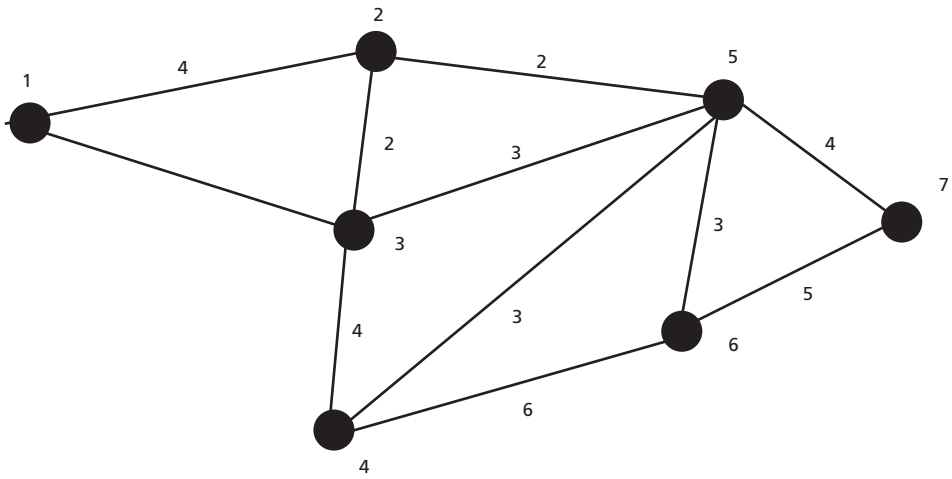
Suponha-se que existe um conjunto de municípios que não possui acesso à banda larga. Para sanar esse problema, é necessário estender a rede de fibra, de modo que todos os municípios fiquem conectados. A princípio, poder-se-ia pensar em ligar cada um deles à rede de fibra mais próxima. Contudo, tendo em vista o custo dessa conexão, esse seria um modo pouco econômico e ineficiente de solucionar o problema. Uma maneira mais inteligente de resolver essa questão se daria da seguinte forma: dado um conjunto de municípios em que cada um deles pode ser conectado, verificar se existe uma rede que conecta todos os municípios com uma quantidade mínima de fibra. Entende-se aqui a quantidade mínima como a menor distância. Desse modo, este problema pode ser modelado utilizando-se a teoria dos grafos (Harary, 1969; Lucchesi, 1979; Szwarcfiter, 1986).

A teoria dos grafos pode ser utilizada para resolver problemas de várias áreas do conhecimento aparentemente distintas, como emprego em rede de computadores, rede de transmissão de energia, rede de tráfego de veículos etc. Um grafo $G(V, E)$ é uma estrutura matemática constituída por um conjunto $V = \{v_1,..., v_n\}$ finito de n vértices e um conjunto $E = \{e_1,..., e_m\}$ de m arestas, que são pares não ordenados de elementos de V. Toda aresta é definida por dois vértices que são seus extremos. As arestas representam alguma relação para aquele conjunto de vértices, sendo que elas podem ou não ter direcionamento. As arestas podem ser ponderadas – ou seja, apresentar valores ligados a elas. Por exemplo, em uma situação em que os vértices representam os municípios, o peso ou valor não negativo de cada aresta pode significar a distância entre os vértices (municípios) que eles conectam. De cada vértice, pode sair mais de uma aresta. O grau de vértice v, $d(v)$ é o número de arestas que saem do vértice.

Na figura 3, mostramos a representação geométrica do grafo ponderado não orientado $G1(V1, E1)$, definido pelos conjuntos $V1=\{1,2,3,4,5,6,7\}$ e $E1=\{1,4,2,4,2,3,3,6,3,4,5\}$.

Esse grafo possui seis vértices, ilustrados por pontos, e onze arestas, ilustradas por linhas que ligam os vértices. Por exemplo, do vértice 1, saem duas arestas: uma que o conecta ao vértice 2 e possui valor 4; e outra que o conecta ao vértice 3, tendo valor 1. No nosso caso, os vértices representam os municípios e as arestas, as conexões entre os municípios. Quando as arestas têm direção, diz-se que o grafo é direcionado, também denominado dígrafo.

FIGURA 3
Grafo *G*1(*V*1,*E*1)



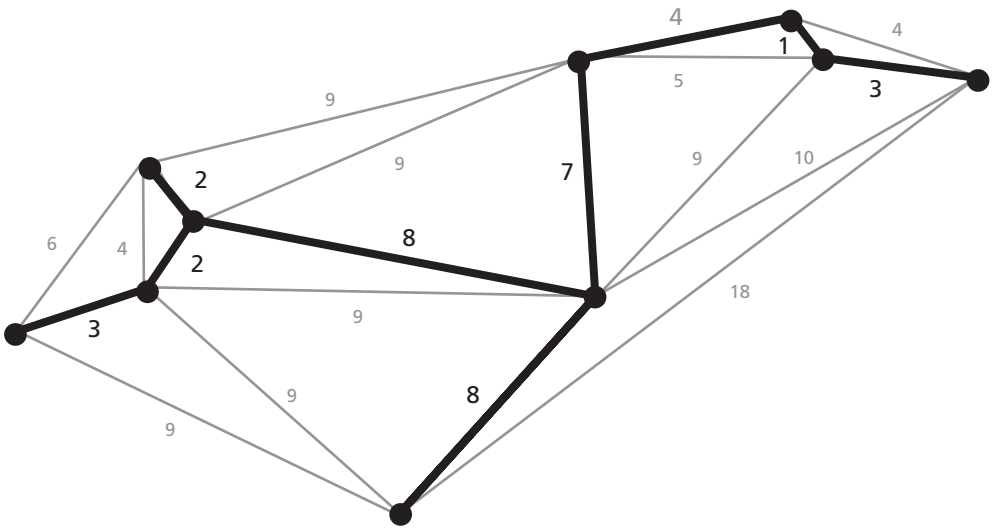
Elaboração dos autores.

Vejamos ainda três conceitos que devemos definir antes de avançar na solução do problema da rede mínima, que são os conceitos de regularidade, conexidade e ciclo. Um grafo é regular (de grau *k*, ou ainda *k*-regular) quando todos os seus vértices têm o mesmo grau (*k*). Um grafo conexo é aquele no qual existe um caminho ligando qualquer par de pontos. Por fim, um ciclo é um grafo conexo regular de grau 2. Visto isso, podemos caminhar novamente para a solução do problema original. Iremos partir da hipótese de que existe uma rede de fibra que conecta todos os municípios. Essa rede pode ser vista como um grafo $G(V,E)$ conexo, ponderado e não orientado. Sendo o grafo $G(V,E)$ conexo, então existe um subconjunto ou subgrafo $\overline{G}\left(V, \overline{E}\right) \subset G\left(V, E\right)$, conexo e acíclico, com $\overline{E} \subset E$. Diz-se então que tal subconjunto é uma árvore geradora de $G(V,E)$, porque possui o mesmo número de vértices G e conecta todos eles.

Vimos que a árvore geradora conecta todos os vértices (municípios) do grafo (rede) a um menor custo, visto que o número de conexões da árvore é menor ou igual ao número

de conexões do grafo dado por $\overline{E} \subset E$. Nosso problema é encontrar a árvore (rede) que irá gerar o menor custo, e a sua solução está na teoria dos grafos. Conhecido como o problema da árvore geradora mínima, ele consiste em encontrar, dado um grafo conexo e ponderado, uma árvore geradora de menor peso possível, em que o peso é dado pela soma dos pesos das arestas escolhidas. Existem algoritmos muito eficientes para resolver o problema da AGM; entre os quais, os algoritmos de Prim e Kruskal são comumente empregados (Cormen *et al.*, 2009). Esses algoritmos são vistos como de caráter guloso, pois em cada passo acumulam a aresta que parece mais localmente promissora, sem considerar o efeito global dessa escolha. A figura 4 ilustra um exemplo de uma árvore mínima geradora que no grafo está destacada em negrito. Note que todos os vértices, representados no grafo por pontos, estão conectados.

FIGURA 4
Árvore geradora mínima



Elaboração dos autores.

A solução do problema da árvore geradora mínima pode ser usada para obtermos a rede de custo mínimo, que conecta todos os municípios sem fibra. Nossa tarefa recai somente na elaboração de grafo que represente uma rede factível de municípios conectados. Isso pode ser feito com base nas seguintes informações:

1) distância de cada município até a rede de fibra mais próxima[4] (*backbone*) (Anatel);
2) distâncias entre os centroides dos municípios (IBGE, 2015);
3) um critério de vizinhança entre dos municípios; e
4) uma medida de ajuste da distância geográfica.

O grafo (rede) pode ser construído tomando como vértices os centroides dos municípios e os pontos de fibra mais próximos dos municípios, cada município possui o seu ponto de fibra mais próximo. As arestas (conexões da rede) são obtidas ligando cada município ao seu vizinho e conectando o centroide do município à rede de fibra mais próxima a esse município. O critério de vizinhança adotado aqui é que um município vizinho ao outro faz fronteira com esse. A escala usada para ajustar a distância se faz de acordo com o ajuste entre as estradas, tal que a distância utilizada é igual a 1,5 vezes a distância geográfica.

Veremos agora como podemos usar a solução da árvore geradora mínima para avaliar o custo mínimo de expansão da rede de banda larga, tendo em vista a priorização de diferentes objetivos de política pública. Conforme dito na introdução, foram escolhidos três critérios para a avaliação de priorização dos investimentos de ampliação da rede de banda larga. São eles: o tamanho da população; o IDH; e o dimensionamento do mercado potencial. De modo a entender como podemos usar a teoria dos grafos acíclicos, vejamos, por exemplo, o caso de priorizar com base no IDH. Pode ocorrer ao gestor de política imaginar que o acesso à internet seja uma maneira eficaz de reduzir a desigualdade regional com base nesse indicador social. Nesse caso, o gestor agiria ranqueando os municípios brasileiros em ordem crescente de nível de IDH. Uma vez feito isso, ele pode escolher priorizar os N primeiros municípios com os piores índices de escolaridade. O grafo desse problema é montado do mesmo modo como descrito para o problema geral, exceto pelo fato de que agora se devem incluir somente vértices representativos dos municípios priorizados e pontos de fibras mais próximos destes. Em termos formais, temos a seguinte situação:

$$\ddot{G}\left(\hat{V},\ddot{E}\right)\subset\hat{G}\left(\hat{V},\hat{E}\right)\subset G\left(V,E\right)\text{ com }\hat{V}<V\text{ e }\ddot{E}\leq\hat{E}<E.$$

Em que:

4. Quando não existe ponto de fibra dentro do município, considera-se essa distância como nula.

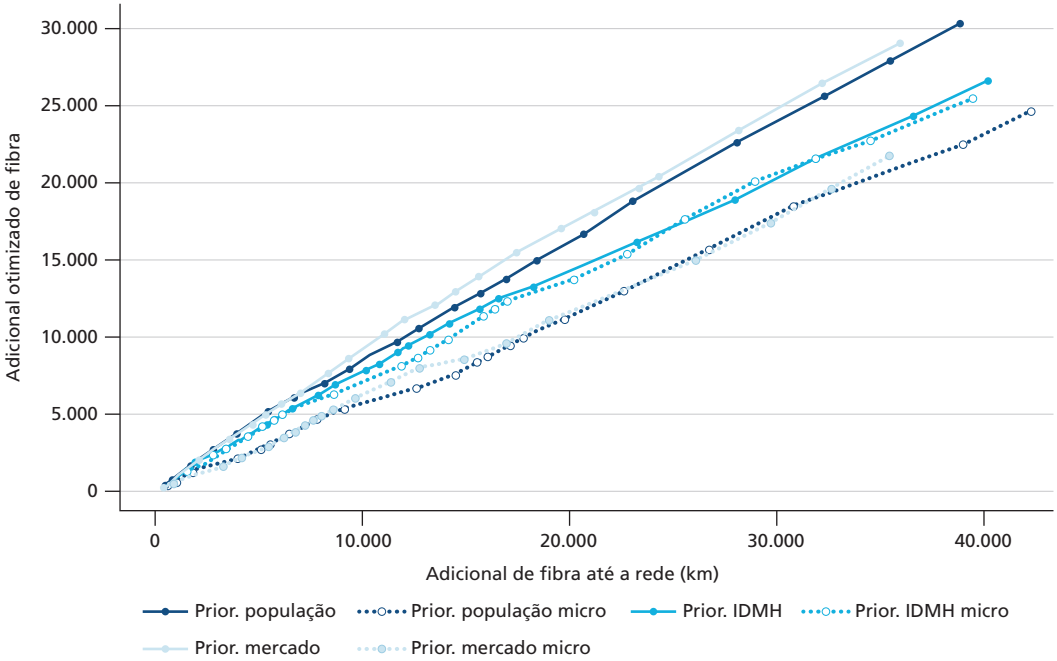
- G é o grafo de toda a rede original de municípios sem fibra;
- V é o conjunto de vértices da rede original;
- E é o conjunto de arestas da rede original;
- $\hat{G}$ é o grafo da rede de municípios priorizados; e
- $\ddot{G}$ é a árvore mínima geradora da rede de municípios priorizados.

O gráfico 1 relaciona o adicional ótimo de fibra com o adicional de fibra do centroide do município até a rede mais próxima para três diferentes critérios de priorização, nas abrangências geográficas de município e microrregião: menor IDH, maior população e maior mercado potencial. A estimação do mercado potencial de banda larga será tratada na seção 4. A leitura apropriada do gráfico deve ser feita da seguinte forma. Tomemos como base o critério de priorizar a expansão da rede com base no tamanho da população (Prior. população). Com base nessa opção, o gestor de política deseja ampliar a rede de banda larga para 250 municípios mais populosos, o que tornaria possível conectar cerca de mais 10 milhões de pessoas (apêndice). A quantidade de fibra ótima necessária para conectar esses 250 municípios estaria em torno de 13 mil quilômetros de fibra (figura 3).[5] De acordo com o gráfico 1, e tendo em vista o critério de priorizar o tamanho da população, 13 mil quilômetros de fibra ótima correspondem aproximadamente a 17 mil quilômetros, caso tais municípios fossem conectados com o ponto da rede de fibra mais próximo. Importante ressaltar que, para cada escolha do tamanho da população a ser priorizada e, portanto, do número de municípios contemplados, o grafo se altera e, assim, uma nova árvore geradora mínima deve ser criada.

Para os dois outros critérios adotados, ocorre igualmente uma grande economia em termos de fibra. Isso ocorre quando se compara a solução de ampliar a rede de banda larga com base na solução da árvore mínima geradora, com a opção de conectar cada município à rede de fibra mais próxima. Contudo, deve-se ter em mente que, se o número de municípios escolhido for pequeno, no caso menor ou igual a cinquenta, a solução gerada pelo método da árvore mínima é muito próxima àquela gerada por conectar o município ao ponto de fibra mais próximo.

5. Na figura 3, o gráfico na linha de cima e à esquerda da figura mostra a relação entre a população priorizada e a quantidade de fibra necessária aos municípios que comportam essa população.

GRÁFICO 1
Redução de extensão de fibra pela otimização
(Em km)



Elaboração dos autores.
Obs.: IDHM = índice de desenvolvimento humano municipal.

4 DIMENSIONAMENTO DO MERCADO POTENCIAL DE BANDA LARGA NO BRASIL

4.1 Metodologia

Como foi dito na introdução deste estudo, os números da Anatel mostram que há muito ainda a ser feito para que se alcance a massificação da banda larga no Brasil, pois há nitidamente uma carência pelo lado da oferta. Nesta seção, trataremos de estimar o dimensionamento do mercado potencial de banda larga no Brasil. Para esse fim, usamos como medida de dimensionamento do mercado de banda larga o acesso à internet que se dá por meio da banda larga. O mercado potencial é estimado com base no total predito de domicílios que poderiam acessar a internet por banda larga caso existisse essa opção. Os dados utilizados advêm de duas bases de dados, o Censo Demográfico 2010 e a Pnad 2015. As informações do primeiro são empregadas para realizar as projeções e as da Pnad 2015, para estimar um modelo que associa a variável dependente relacionada ao acesso à

banda larga com suas possíveis variáveis preditoras. Assim, o primeiro passo é a seleção de variáveis preditoras compatíveis com o Censo Demográfico 2010 e a Pnad 2015.

Para a escolha de variáveis, seguimos o estudo de Cadorna *et al.* (2007), cujo escopo do assunto e da metodologia, como será visto, tem muito em comum com este trabalho. Cadorna *et al.* (2007) estimam uma função de demanda por internet residencial na Áustria com foco em conexão banda larga, empregando o modelo de escolha discreta *nested logit.* Entretanto, diferentemente desse estudo que faz uso das características dos moradores como variáveis dependentes, o corrente trabalho utiliza as variáveis relacionadas às características físicas e socioeconômicas dos domicílios como variáveis preditivas, para que o domicílio esteja ou não conectado à banda larga. As variáveis preditoras escolhidas – e compatíveis com – para ambas as bases de dados, do Censo Demográfico 2010 e da Pnad 2015, foram as seguintes:[6]

Urbano: se o domicílio está localizado na área urbana (*dummy*);

TemTV: se o domicílio tem televisão (*dummy*);

TemTelFixo: se o domicílio tem telefone fixo convencional (*dummy*);

TemGeladeira: se o domicílio tem geladeira (*dummy*);

DomAlugado: se o domicílio é alugado (*dummy*);

Densidade: número de moradores por cômodo;

NumComodos: número de cômodos;

NumMoradores: número de moradores;

RendaDom: renda domiciliar de todas as pessoas; e

ValorAluguel: aluguel mensal pago no mês de referência.

Utilizou-se o Censo Demográfico 2010 para fazer a predição, por ele possuir informações dos domicílios que podem ser expandidas por meio dos pesos para todo o território nacional. No entanto, pelo fato de esse censo possuir informação limitada sobre a utilização da internet, o modelo econométrico para determinação do acesso à banda larga teve de ser estimado com base nos dados da Pnad 2015, que possui informações

6. Outras variáveis foram também inclusas na análise de regressão. Contudo, elas não apresentaram significância estatística.

mais atualizadas e detalhadas acerca da utilização desses serviços. A figura 5 mostra o nível de detalhamento da informação referente ao acesso à internet (banda larga) aferido pela Pnad 2015. Contudo, essa base não permite desagregação no nível de municípios, Unidades da Federação (UFs), regiões metropolitanas (RMs), áreas urbanas não metropolitanas e áreas rurais.

FIGURA 5
Informação do questionário da Pnad 2015

32.4	O acesso à internet que o(s) morador(es) tem(êm) no próprio domicílio é feito através de:		
a. *Sinal de rede celular 3G ou 4G*?	**02424**	2 ☐ Sim	4 ☐ Não
b. Conexão discada por linha telefônica (*como uma chamada telefônica comum*)	**02425**	2 ☐ Sim	4 ☐ Não
c. *Banda larga (ADSL, VDSL, cabo de tv por assinatura, cabo de fibra ótica, satélite ou algum tipo de rádio, como WI-FI e WIMAX*?	**02426** (siga 32a)	2 ☐ Sim	4 ☐ Não

Fonte: IBGE.

Para podermos usar adequadamente as variáveis do Censo Demográfico em um modelo cujos parâmetros são estimados a partir dos dados da Pnad, é necessário que as variáveis desta base estejam na mesma escala do censo – ou seja, sejam reescalonadas. O reescalonamento das variáveis da Pnad 2015 para a mesma escala do Censo Demográfico 2010 foi feito utilizando, para cada variável do modelo, transformação que aparece na equação (1). Esta equação é uma translação das variáveis da Pnad 2015. Chegamos a isso igualando as variáveis normalizadas de ambas as bases.

$$x_{2015} = \left(\left(x_{2010} - \bar{x}_{2010} \right) / \sigma_{2010} \right) * s_{2015} + \bar{x}_{2015} . \qquad (1)$$

Em que:

- x_{2015} é a variável *x* reescalonada da Pnad 2015;
- $\bar{x}_{2015}$ é a média da variável *x* da Pnad;
- s_{2015} é o desvio-padrão da variável *x* da Pnad;
- x_{2010} é a variável *x* do Censo Demográfico 2010;
- $\bar{x}_{2010}$ é a média da variável *x* do Censo Demográfico 2010; e
- σ_{2010} é o desvio-padrão da variável *x* do Censo Demográfico 2010.

O passo seguinte refere-se à estimação do modelo. A variável dependente Y_i é uma *dummy* definida da seguinte forma: $Y_i = 1$ se o domicílio *i* está conectado à internet banda larga – ou seja, se ele atende ao item c do questionário da Pnad 2015 – e $Y_i = 0$, caso contrário.

Pelo fato de a variável Y_i estar limitada entre 0 e 1, não podemos fazer uso do modelo de regressão linear tal como comumente é feito. Em casos como este, em que a variável dependente é dita qualitativa, devemos limitar o comportamento de *Y* por meio de uma função Φ, que expressa a probabilidade, no caso, de o domicílio ter conexão por banda larga. Assim, temos que:

$$\Pr(Y_i = 1 \mid X) = \Phi\left(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_N x_N + \varepsilon_i\right). \tag{2}$$

Desse modo, a probabilidade de o domicílio *i* estar conectado por banda larga, $\Pr(Y_i = 1 \mid X)$, é dada por uma função não linear limitada no intervalo [0,1], cujo argumento é expresso pela relação linear $\beta_0 + \beta_1 x_{1i} + \cdots + \beta_N x_N + \varepsilon_i$, na qual x_{ji} é uma variável preditiva *j* associada ao domicílio *i*, β_j é o parâmetro dessa variável, enquanto ε_i é o distúrbio da regressão não linear.

Neste estudo, usamos para representar Φ a função de distribuição logística[7] (Greene, 1993; Johnston e Dinardo, 1997; Maddala, 1983; Kutner, 2005; Stock e Watson, 2004), de modo que $\Phi = 1/1 + e^{-x}$. Assim temos que:

$$\Pr(Y_i = 1) = p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{i,1}+\beta_2 x_{2,i}+\cdots+\beta_N x_{i,N})}}. \tag{3}$$

A expressão (3) também pode ser escrita da seguinte forma:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{2,i} + \cdots + \beta_N x_{i,N} \tag{3'}$$

A estimação do modelo de regressão logística é feita a partir do emprego do estimador de máxima verossimilhança.

7. Poderíamos ter feito uso do modelo *probit*, em que a função Φ é representada por uma distribuição acumulada normal.

A etapa seguinte diz respeito à descrição dos procedimentos relativos ao dimensionamento do mercado de banda larga. Para isso, estimamos a equação (3), usando os dados da Pnad 2015 apenas para os domicílios *nos quais o acesso à banda larga está disponível*. Feito isso, utilizamos a regressão estimada para todos os domicílios do Censo Demográfico 2010, para fazer a predição da probabilidade de o domicílio usar internet. O dimensionamento do mercado de banda larga – isto é, o número estimado de domicílios que demandariam internet – é calculado por meio do somatório dessas probabilidades da seguinte forma:

$$\sum Pred.Internet * pesodom = \sum p_i * pesodom = nº\ domicílios\ . \qquad (4)$$

Em que:

- *Pred.Internet* é a probabilidade predita do domicílio demandar internet; e
- *Pesodom* é o peso do domicílio no Censo Demográfico 2010.

Para calcular o número de domicílios de uma área geográfica (municípios, região etc.), faz-se a restrição do somatório dentro da região que se deseja. Por fim, de modo a estimar o novo dimensionamento do mercado de banda larga, caso ocorra um aumento na penetração média desse serviço, usamos o seguinte procedimento. Determinamos a mudança no intercepto da regressão que levaria a um aumento dessa magnitude. Temos assim uma regressão modificada, que usamos novamente para predizer o novo dimensionamento do mercado, com os dados do Censo Demográfico 2010.

4.2 Resultados

Apresentamos os resultados da estimação do modelo proposto na seção 4.1 assim como o valor predito da dimensão do mercado potencial brasileiro da banda larga. Na tabela 1, são mostrados os resultados da regressão *logit* estimada com base nos dados da Pnad 2015, usando apenas as variáveis que se mostraram significativas.[8] Cabe destacar que, diferentemente da regressão linear, os coeficientes da regressão logística não podem ser diretamente interpretados na equação como o efeito marginal da variável regressora.[9] Na tabela 1, também aparece o R^2 de McFadden, cuja interpretação dessa estatística é

8. Ver nota de rodapé 2.

9. O efeito marginal no modelo *logit* $\partial p_i / \partial x_j = \Phi'(X\beta)\beta_{j_i}$ em que Φ' é a derivada de Φ' em relação à variável x_{ij}.

análoga ao R^2 da regressão linear.[10] Conforme antes exposto, o intuito do nosso modelo é o de realizar um exercício preditivo de dimensionamento do mercado de banda larga. Razão pela qual designamos por variável preditiva, e não por variável explicativa do modelo. Portanto, a regressão estimada deve ser entendida como uma regressão relacionada à forma reduzida do modelo, e não à forma estrutural.

Uma vez estando de posse dos resultados, a tarefa seguinte será a de estimar o mercado potencial conforme a metodologia descrita nesta seção. A predição para todo o território nacional é feita com os dados do Censo Demográfico 2010. Para municípios não existentes em 2010, foi feita uma imputação com base nos municípios dos quais os novos municípios faziam parte. Observou-se que os efeitos finais são residuais. A projeção para o mercado potencial de banda larga no Brasil chega a 32,6 milhões de domicílios. Os estados das regiões Norte e Nordeste (Pará, Amapá, Maranhão, Piauí, Amazonas e Tocantins) têm potencial de quase dobrar ou triplicar o número de domicílios, caso haja mais acesso à banda larga. As projeções do mercado potencial para todos os municípios e microrregiões do território brasileiro estão bem ilustradas nas figuras 6 e 7.

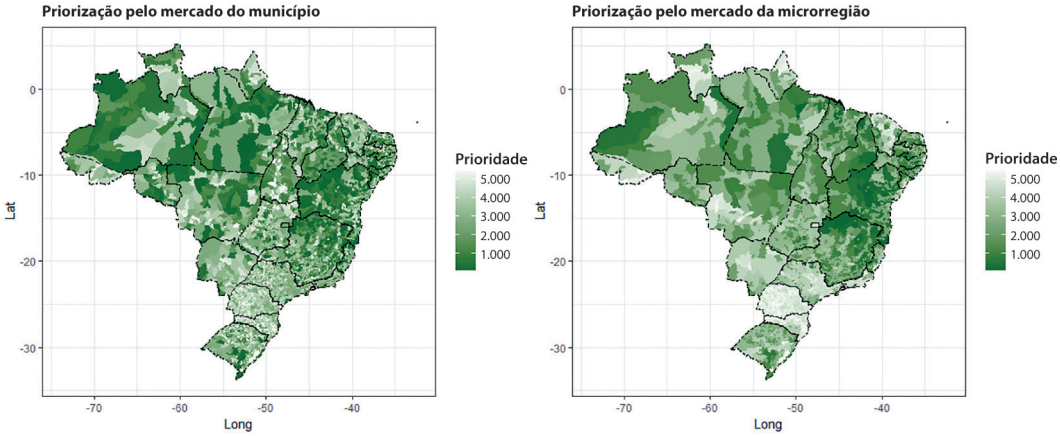
TABELA 1
Resultado da regressão logística

Regressores	Coeficiente	Erro-padrão	P-valor[1]
Intercepto	-4.385	0,233	0,000
Urbano	1.133	0.091	0,000
TemTV	0.459	0.115	0,000
TemTelFixo	1.071	0.024	0,000
TemGeladeira	0.625	0.182	0,000
DomAlugado	-0.164	0.056	0,001
Densidade	-0.377	0.104	0,000
NumComodos	0.116	0.011	0,000
NumMoradores	0.202	0.021	0,000
RendaDom	0.0002	0.000	0,000
ValorAluguel	0.001	0.000	0,000
R^2 de McFadden	0.193		

Elaboração dos autores.
Nota: [1] IC: 95%.

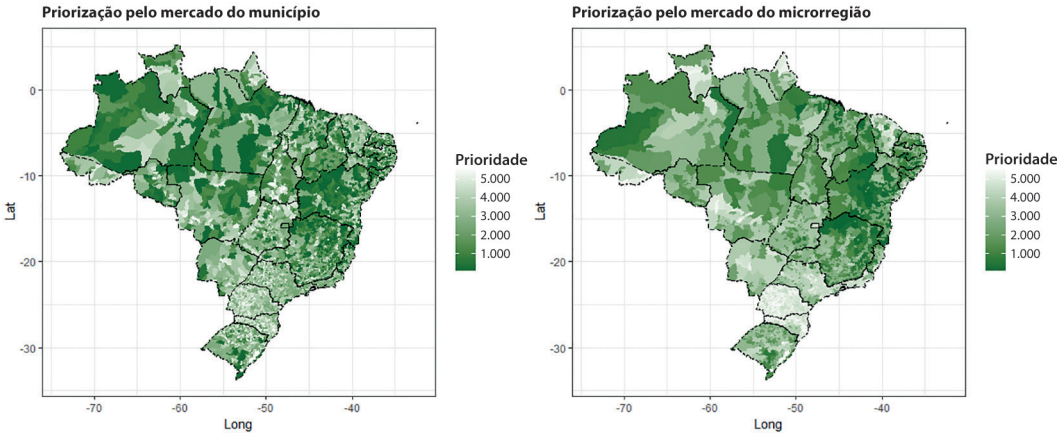
10. O R^2_{McF} de McFadden é definido como $R^2_{McF} = 1 - (\ln(L_M/\ln(L_0))$, em que L_M é o valor da função verossimilhança da regressão, enquanto L_0 é o valor dessa mesma função sem nenhum preditor.

FIGURA 6
Priorização pelo mercado do munícipio



Elaboração dos autores.

FIGURA 7
Priorização pelo mercado da microrregião



Elaboração dos autores.

5 AVALIAÇÃO DOS EFEITOS DOS INVESTIMENTOS EM TELECOMUNICAÇÕES SOBRE O PIB

5.1 Modelo econométrico

Uma vez tendo apresentado o modelo de custo de expansão da banda larga, nosso objetivo nesta seção é apresentar indicadores que visem captar o benefício ou retorno econômico da ampliação da banda larga. Uma das principais medidas para verificar o retorno sobre

a economia de um determinado tipo de política pública se dá pelo efeito que esta tem sobre o PIB. De modo a tornar efetiva essa proposta para a avaliação do impacto da banda larga sobre o produto, usaremos um modelo econométrico cujo objeto principal é determinar o efeito marginal da expansão da banda larga sobre o produto municipal real (PIBM). A literatura tem mostrado que a banda larga apresenta implicações importantes sobre a economia, sendo que seu efeito tem sido percebido sobre vários indicadores econômicos, como o PIB, o nível de emprego, a produtividade etc. (Katz, 2012).

A tarefa deste estudo recai em estimar um modelo econométrico capaz de definir a relação entre a expansão da banda larga e o PIB. Nossa base de dados é composta de dados em painel de 5.564 municípios brasileiros, entre 2007 e 2015. A equação de regressão então estima a relação linear entre o PIB municipal e a banda larga da seguinte forma:

$$\log(PIBM_{i,t}) = \alpha_i + \sum_{j=1}^{J} \phi_j \log(w_{i,t}) * Dclusteri_{i,j} + \sum_{k=1}^{K} \delta_k \log(x_{i,t,k}) + \varepsilon_{i,t} \qquad (5).$$

Em que:

- i, para i=1,...N, é o índice das unidades (municípios);
- t, para i=1,...,T, é o índice de tempo;
- k é o índice relativo à variável de controle $x_{i,t,k}$; e
- $w_{i,t}$ é a variável de política relacionada ao serviços de banda larga.

O modelo da equação (5) contempla ainda os efeitos da banda larga para subgrupos homogêneos de agrupamento de municípios homogêneos[11] (*clusters* j=1,...,J), por meio da variável $Dclusteri_{i,j}$, que 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.

Por hipótese, vamos assumir que a equação (5) é uma representação empírica da função de produção da economia e, portanto, devemos ter as medidas de capital, emprego, produto ou renda, além, naturalmente, de um indicador verossímil de acesso à banda larga.

11. 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 utilizando o método K-médias (K-means) (Lattin, Carrol e Green, 2011), que determinou a existência de seis *clusters*.

Naturalmente, entre a pesquisa teórica e a sua aplicabilidade, geralmente existe, por vezes, certa distância. Na prática econométrica, não estão disponíveis 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 das seguintes variáveis de controle, que são as receitas totais (*RECEITA*).[12]

Para representar o fator trabalho ou emprego, usamos a variável *vínculos*[13] da Relação Anual de Informações Sociais (Rais);[14] e para representar a variável T da equação (5), que é a nossa variável de política ($w_{_(i,t)}$), usamos a densidade de acessos à banda larga (*DEM_SCM*).[15] Incluímos também, como variáveis de controle, as transferências governamentais decorrentes do Programa Bolsa Família (PBF), as transferências governamentais decorrentes do programa de Benefício de Prestação Continuada (BPC) e os recursos advindos da Previdência Social, excetuando o BPC (*PREVI*). O parâmetro ϕ_j associado ao termo composto $\log(w_{i,t}) * Dclusteri_{i,j}$ define o efeito sobre o PIB nos municípios inseridos no agrupamento *j*, decorrente de um aumento dos serviços de banda larga.

Cabe ainda comentar sobre 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 em 1%, o PIB dos municípios do agrupamento J aumenta em ϕ %.

A base de dados deste estudo é composta pela população de 5.564 municípios brasileiros entre 2007 e 2015 e por informações oriundas de base de dados distintas, como dados dos censos demográficos, 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 Bolsa Família e Benefício de Prestação Continuada, cujas bases de dados

12. 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.
13. Emprego que abrange todos os vínculos formais (celetistas, estatutários, temporários, avulsos, entre outros).
14. 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).
15. Razão entre o número de acessos do Sistema de Comunicação Multimídia (SCM) por grupo de cem domicílios, de acordo com a Agência Nacional de Telecomunicações (Anatel).

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

É necessário que o nosso modelo de regressão linear não deixe despercebido o problema de causalidade reversa, que aparece quando uma das variáveis explicativas está correlacionada com o termo de erro da população (variáveis endógenas). Contudo, existe uma parte da variação da variável explicativa que não tem correlação com o termo de erro. Em outras palavras, a variável densidade de acesso de banda larga está correlacionada com o termo de erro que determina o PIB, tornando o estimador inconsistente. O método de variáveis instrumentais (IV) utiliza uma variável adicional para isolar a parte da variação da variável explicativa que não tem correlação com o termo de erro. 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 $(w_{i,t})$ $*Dclusteri_{i,j}$ deve ser instrumentalizada no primeiro estágio, conforme o modelo da equação (6):

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

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

O problema do modelo de variáveis instrumentais recai, sobretudo, na escolha correta dos instrumentos. Os critérios estatísticos que definem uma IV já foram descritos antes. No entanto, vale a pena discorrer de modo mais intuitivo,para 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 estão relacionadas à infraestrutura da banda larga. Usamos três conjuntos de IV:

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 a cada 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 atuam no município; e
- *ERBS*: números de estações de rádio base (agregado por todas as tecnologias, 2G, 3G e 4G) por município (Anatel).

Vejamos como é possível defender a hipótese de que esse conjunto de variáveis pode ser candidato a bons instrumentos. No caso da variável *DEM_STFC*, temos que o emprego 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 também interpretada como um insumo de produção, pelo fato de também ser um meio de troca de informação. De fato, Norton (1992) observou, em um estudo usando uma amostra de 47 países, no período 1957-1977, que a telefonia fixa teve efeito positivo, sendo também estatisticamente significativa. Norton (1992) conclui ainda que a infraestrutura de telefonia fixa tem impacto positivo sobre o crescimento. Não obstante a questão de que o referido estudo não endereça 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 e, no presente, 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.

5.2 Resultados

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 2 apresenta as regressões para dados em painel – por exemplo, Baltagi (1995), Hsiao (1995), Hausman (1981), Greene (1993), *inter alia* – para a equação (5). As colunas (1) e (2) apresentam os resultados obtidos, respectivamente, pelos modelos de efeito aleatório (RE) e o efeito fixo (FE). Nesses dois modelos, assume-se que os regressores são exógenos – ou seja, inexiste 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.[16] 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) e (5), são mostrados os resultados obtidos pela aplicação do método de IV, visando corrigir o problema de endogeneidade ou causalidade reversa entre o PIB do município (PIBM) e a densidade de banda larga (*DEN_SCM_i, i=1,...6*). Usamos um conjunto de instrumentos na regressão da coluna (5) maior que aquele usado em relação à regressão que aparece na coluna (4). Vários testes de especificação foram aplicados sobre as regressões (4). Por economia, não mostramos os resultados do primeiro estágio, porque teríamos de mostrá-los para cada regressão da variável *DEM_SCM_I, I=1,...,6*.[17]

É 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*), elas 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 de 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.

16. 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).
17. As saídas econométricas podem ser obtidas mediante contato com os autores.

A variável *ERBS* não apresenta significância. Isso provavelmente é explicado pela existência de colinearidade dessa variável com outro instrumento.[18] 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 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 *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.

Por fim, passamos 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. Assim, tomando a coluna 4 da tabela 2, temos que, para cada 1% de crescimento na variável *EMPREGO*, há 0,10% de crescimento do PIB do município. No que concerne 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 a tabela 2, 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 – no agrupamento 1, que é composto por municípios de maior renda *per capita* e alta concentração

18. 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*.

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 apontado 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 e de menor renda *per capita*. O agrupamento 6 concentra a menor taxa de urbanização e elevada participação do setor agrícola em relação ao demais setores. Não obstante tudo isso, ambos os agrupamentos assinalam o efeito da banda larga sobre a economia.

TABELA 2
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)	–	–	–	–

(Continua)

(Continuação)

	RE (1)	FE (2)	IV-FE (3)	IV-FE (4)	IV-GMM[1] (5)
Teste F de Anderson-Rubin	–	–	355.53 (0.000)	–	–
Teste de sub--identificação	–	–	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

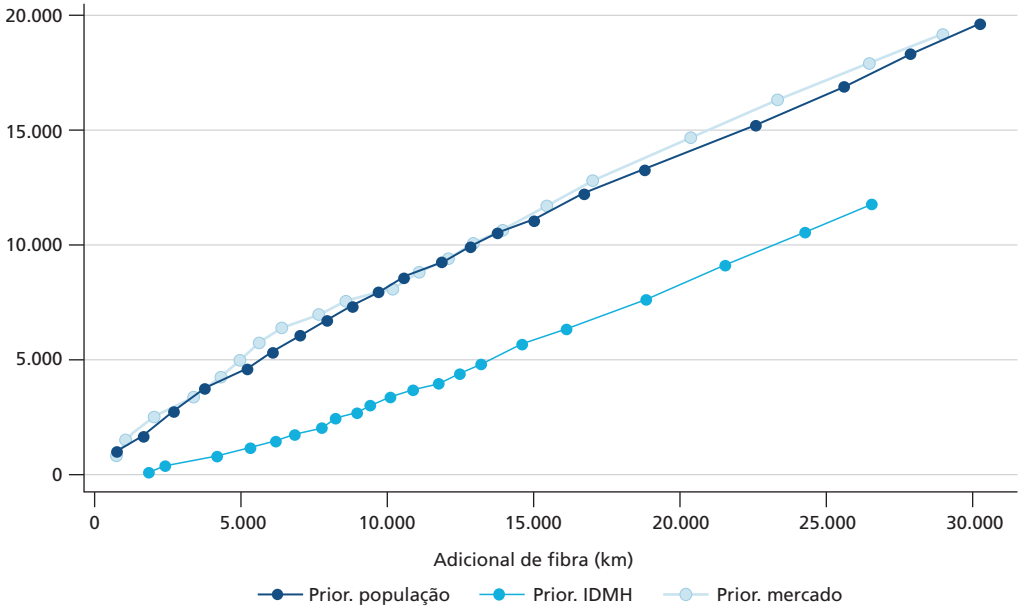
Notas:[1] Teste Wald para verificar a significância do regressor endógeno na equação estrutural.
[2] Testa a hipótese de "identificação fraca", que surge quando os instrumentos excluídos são correlacionados com regressores endógenos, mas apenas fracamente. A estatística F de Cragg-Donald está na parte 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, correlacionado com os regressores endógenos.
[4] Testa um subconjunto de condições de ortogonalidade – ou seja, 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, 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] GMM com correção de autocorrelação serial de ordem dois.
[9] Testa a hipótese de que as variáveis *3G* e *OUTRAS* são redundantes como instrumentos.
[10] Testa a hipótese de que as variáveis *OPERADORAS* E *ERBS* são redundantes.

6 IMPLICAÇÕES DAS POLÍTICAS DE PRIORIZAÇÃO

Conforme exposto na seção 2, uma vez tendo optado por um dos três critérios de priorização, devem-se examinar as implicações sobre as variáveis de interesse descritas na seção 2, que decorrem dessa escolha.[19] Os gráficos 2, 3, 4 e 5 mostram o efeito sobre as variáveis de interesse associadas a certos segmentos populacionais. Os resultados sugerem que a priorização pelo critério do IDH municipal não é a melhor alternativa, em termos de benefícios para a população pobre. Mais famílias de baixa renda podem ser atendidas por critérios como população ou mercado potencial. Os critérios de priorização a partir do mercado potencial ou da população total podem ser mais vantajosos em termos da população jovem atingida, população pobre e também extremamente pobre. Com exceção da população extremamente pobre, para os demais segmentos populacionais analisados, os critérios de mercado e tamanho da população geram resultados fortemente parecidos.

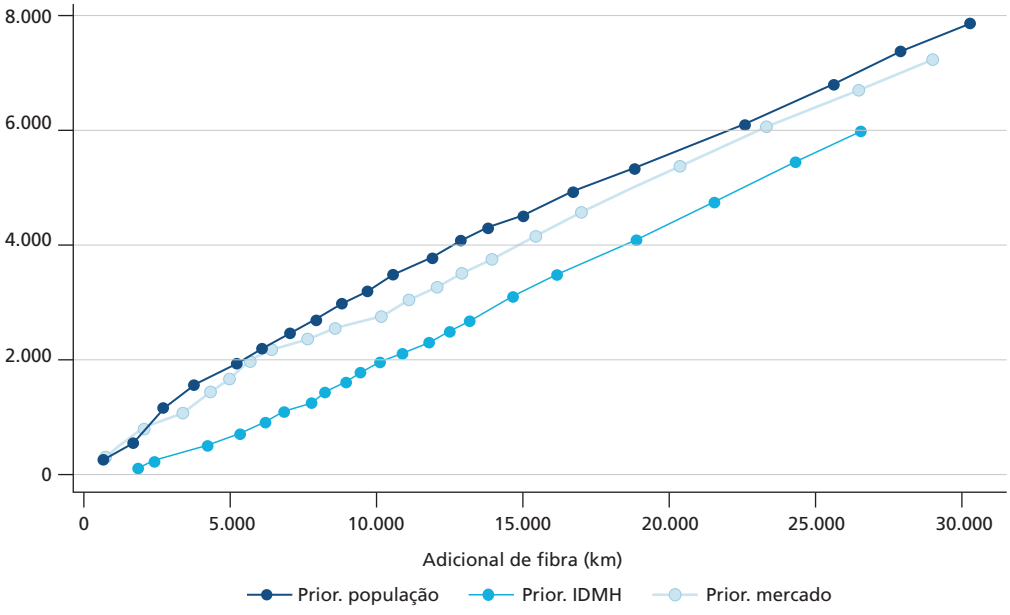
19. Nada impede que as variáveis de interesse pudessem ser tomadas como critério de priorização.

GRÁFICO 2
População total atingida
(Em milhares de habitantes)



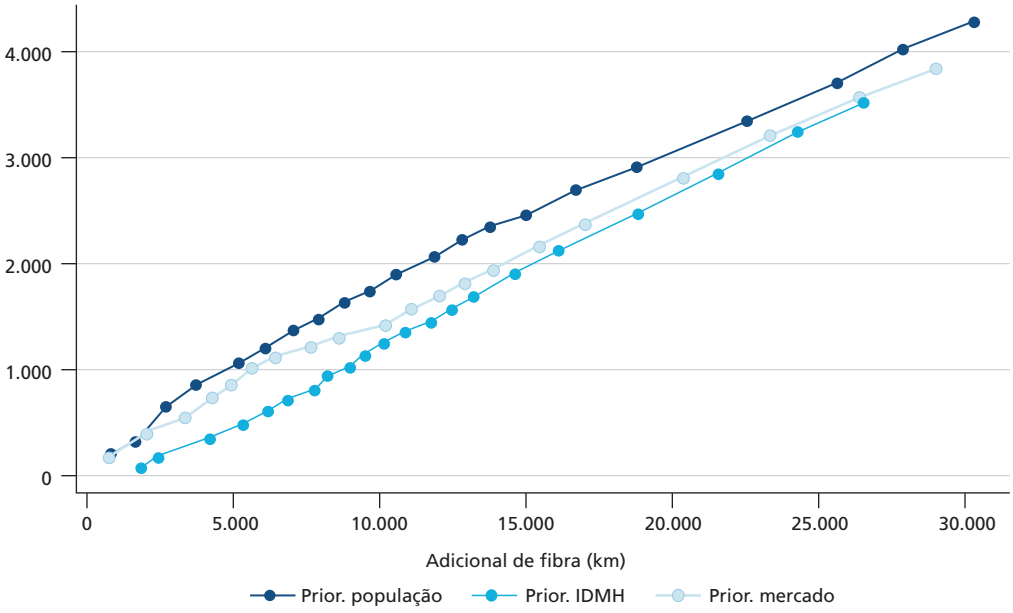
Elaboração dos autores.

GRÁFICO 3
População pobre atingida
(Em milhares de habitantes)



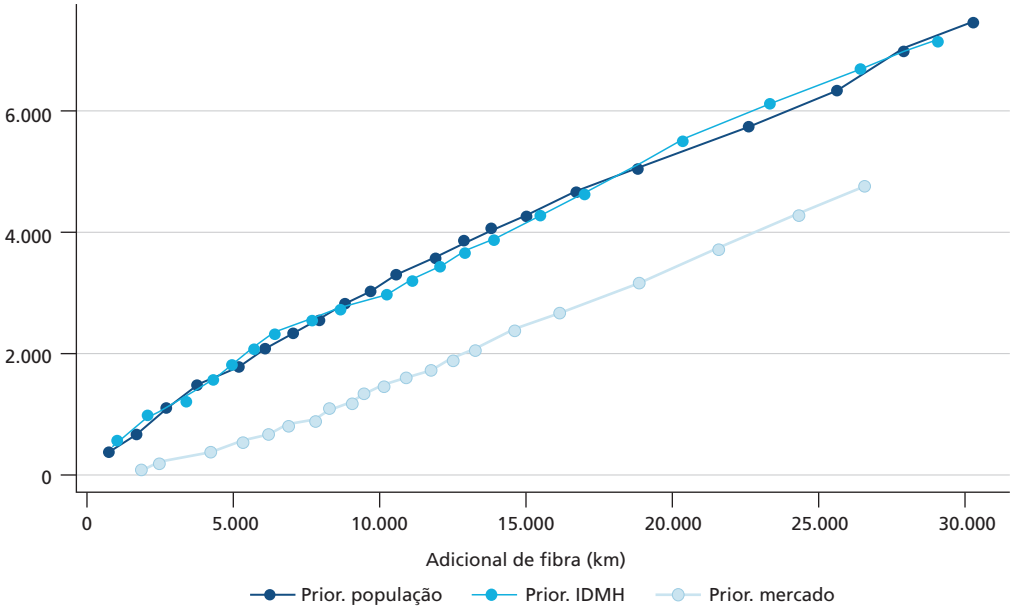
Elaboração dos autores.

GRÁFICO 4
População extremamente pobre atingida
(Em milhares de habitantes)



Elaboração dos autores.

GRÁFICO 5
População com até 20 anos atingida
(Em milhares de habitantes)

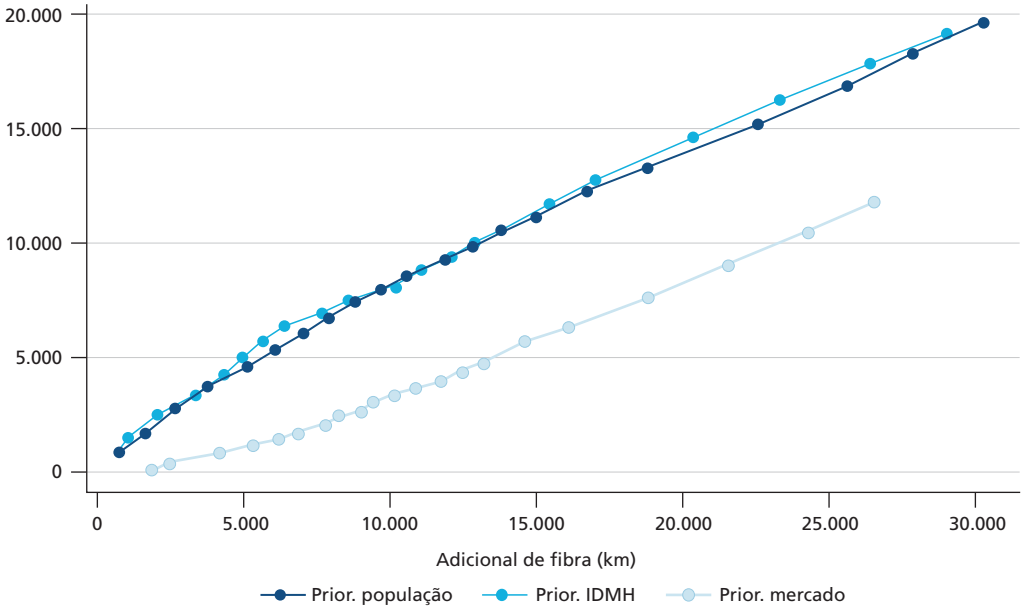


Elaboração dos autores.

Os gráficos 6, 7, 8 e 9, na linha inferior, ilustram o impacto sobre a economia. Eles mostram quanto do produto total está associado a determinada escolha, assim como o efeito marginal sobre o PIB, em ambos os casos, para os municípios priorizados. Observa-se que o critério de priorizar com base na escolha do mercado potencial é o que gera maior retorno econômico, seguido pelo critério com base no tamanho da população. Novamente, o IDH é o critério de priorização que gera o menor retorno. Portanto, uma boa escolha de política pública deve ser aquela que contempla a priorização com base no mercado potencial, pois é aquela que gera o maior ganho econômico, além de atender aos diferentes segmentos da população. Isso poderia gerar algum tipo de sinergia no qual as empresas podem estender futuramente as redes para municípios vizinhos, dado que já estão localizadas em municípios de maior mercado.

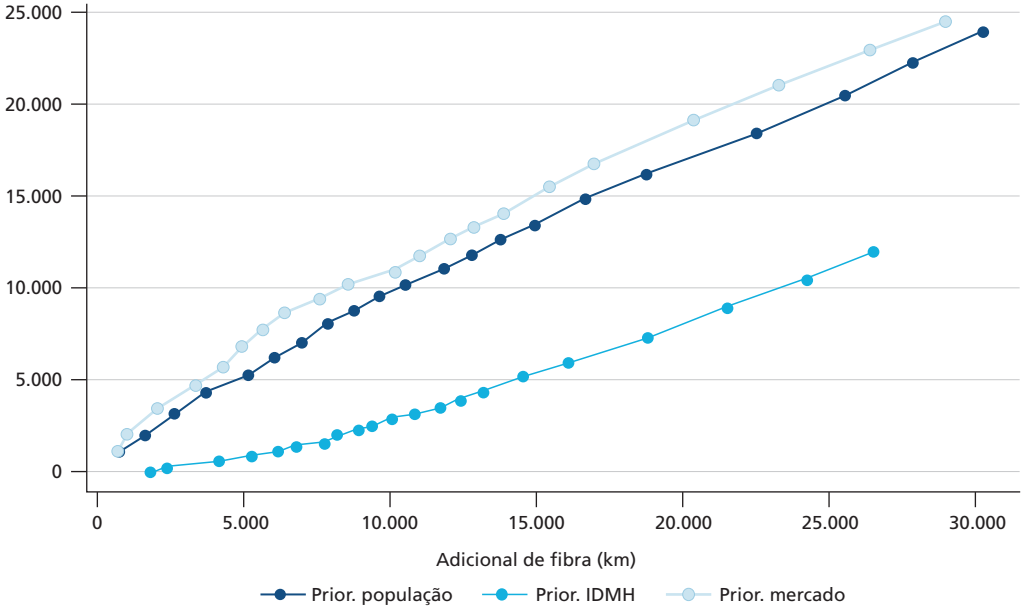
Deve-se ressaltar que a opção de priorização pelo IDH foi aquela que se mostrou a menos promissora. Estudamos ainda o caso no qual a abrangência geográfica ocorre por microrregião. Na maioria dos casos, os resultados não se alteram significativamente. Apenas no caso da população extremamente pobre, o IDH mostra um resultado mais favorável.

GRÁFICO 6
População total atingida
(Em milhares de habitantes)



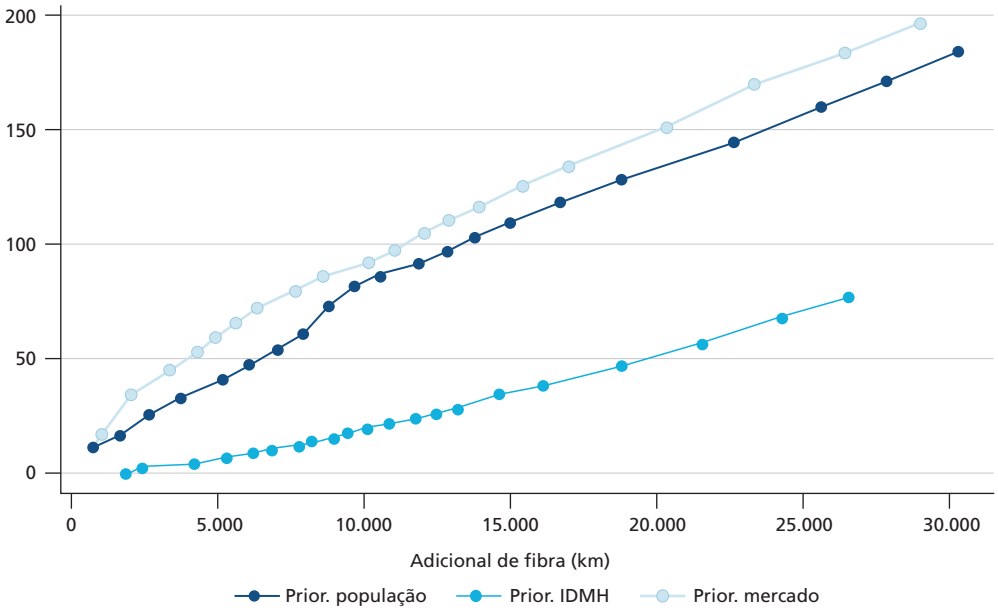
Elaboração dos autores.

GRÁFICO 7
Mercado: potencial atingido
(Em milhares de domicílios)



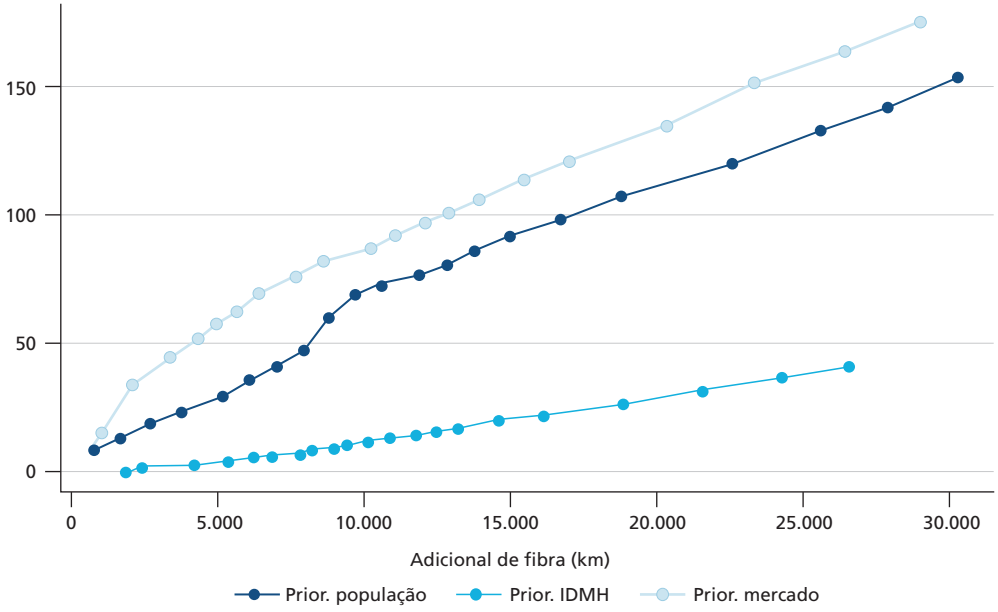
Elaboração dos autores.

GRÁFICO 8
PIB municipal: total atingido
(Em R$ bilhões)



Elaboração dos autores.

GRÁFICO 9
Impacto sobre o PIB
(Em R$ milhões)



Elaboração dos autores.

REFERÊNCIAS

BALTAGI, B. H. **Econometric analysis of panel data**. Hoboken: John Wiley & Sons, 1995.

CADORNA, M. *et al.* Demand estimation and market definition for broadband internet services. **Journal of Regulatory Economics**, v. 35, n. 1, p. 70-95, 2009.

CORMEN, T. *et al.* **Introduction to algorithms**. Cambridge, MA: The MIT Press, 2009.

GREENE, W. **Econometric analysis**. Upper Saddle River: Prentice Hall, 1993.

HARARY, F. **Graph theory**. Boston: Addison-Wesley, 1969.

HSIAO, C. **Analysis of panel data**. Cambridge, UK: Cambridge University Press, 1995.

JOHNSTON, J.; DINARDO, J. **Econometric methods**. 4. ed. New York: McGraw-Hill, 1997.

KATZ, R. **The impact of broadband on the economy**. Geneva: Regulatory & Market Environment/ITU, 2012. Broadband Series.

KUTNER, M. H. **Applied linear statistical models**. New York: McGraw-Hill/Irwin, 2005.

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

LUCCHESI, C. L. Introdução à teoria dos grafos. *In*: COLÓQUIO BRASILEIRO DE MATEMÁTICA DO INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA (IMPA), 12., 1979, Rio de Janeiro. **Anais...** Rio de Janeiro: Impa, 1979.

MADDALA, G. S. **Limited dependent and qualitative variables in econometrics**. Cambridge, UK: Cambridge University Press, 1983.

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

STOCK, J. H.; WATSON, M. W. **Econometria**. São Paulo: Adison Wesley, 2004.

SZWARCFITER, J. L. **Grafos e algoritmos computacionais**. 2. ed. Rio de Janeiro: Editora Campus Ltda., 1986.

BIBLIOGRAFIA COMPLEMENTAR

HAUSMAN, J.; TAYLOR, W. Panel data and unbservable individual effects. **Econometrica**, v. 49, n. 6, p. 1377-1398, 1981.

KATZ, R. **The economic and social impact of telecommunications output**: empirical evidence in the US and Europe. Brussels: ETNO, 2008. ETNO Regulatory Economics Task Force.

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

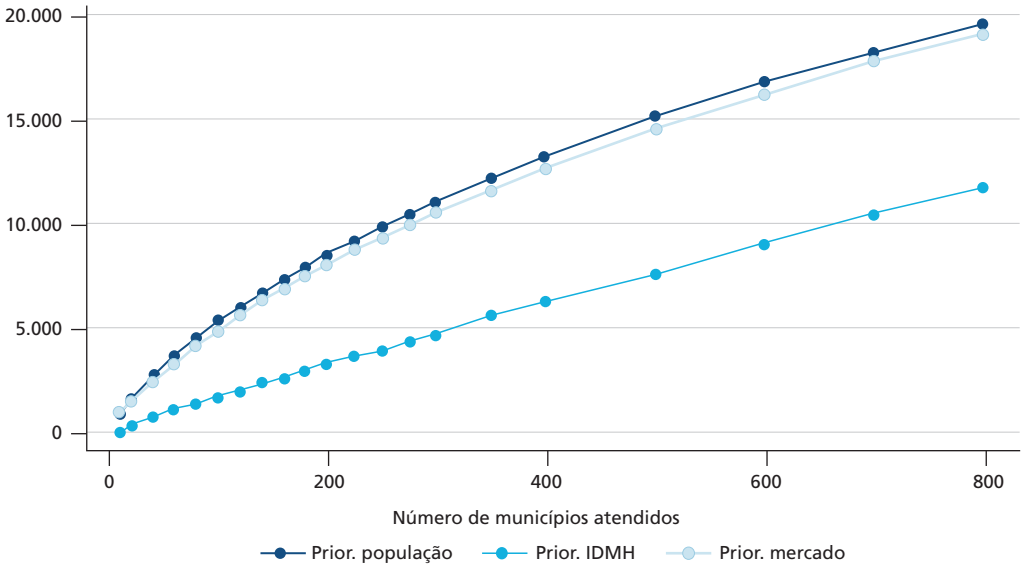
MOREIRA, M. J. A conditional likelihood test for structural models. **Econometrica**, v. 71, n. 4, p. 1027-1048, 2003.

QIANG, C.; ROSSOTTO, C. M. **In Information and Communications for Development 2009**: extending reach and increasing impact. Washington: World Bank, 2009. p. 35-50.

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

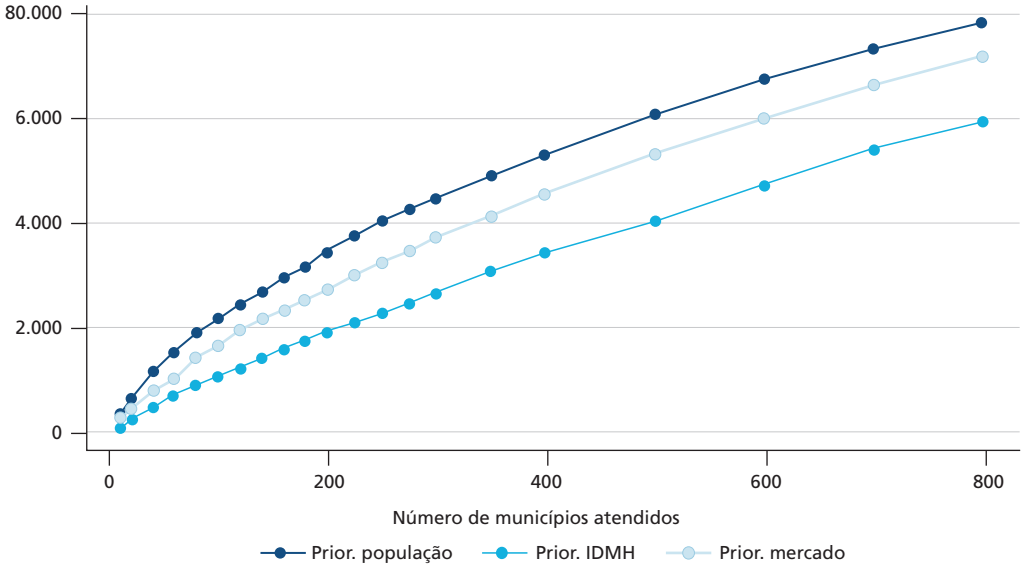
APÊNDICE

GRAFICO A.1
População total atingida
(Em milhares de habitantes)



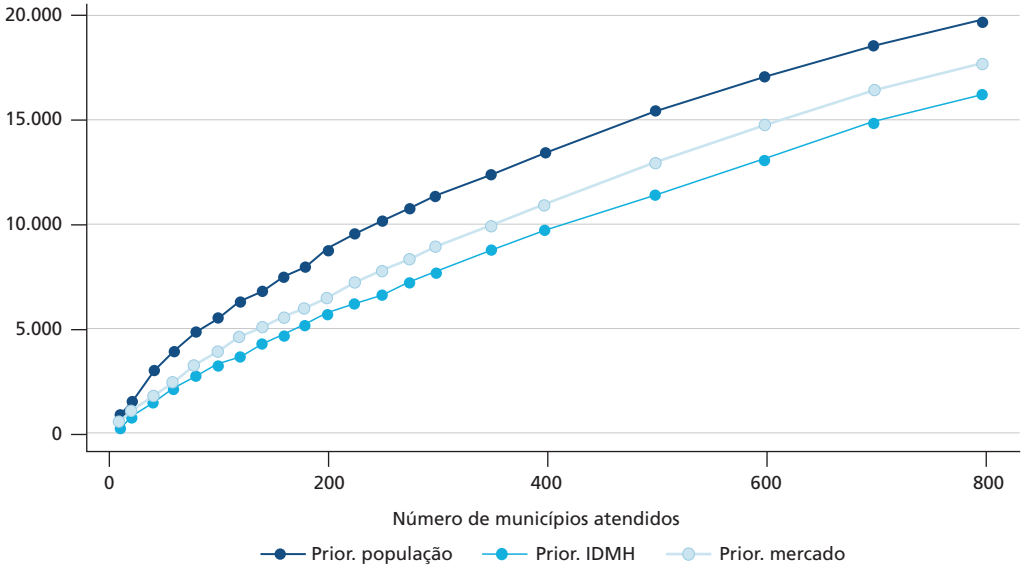
Elaboração dos autores.

GRAFICO A.2
População pobre atingida
(Em milhares de habitantes)



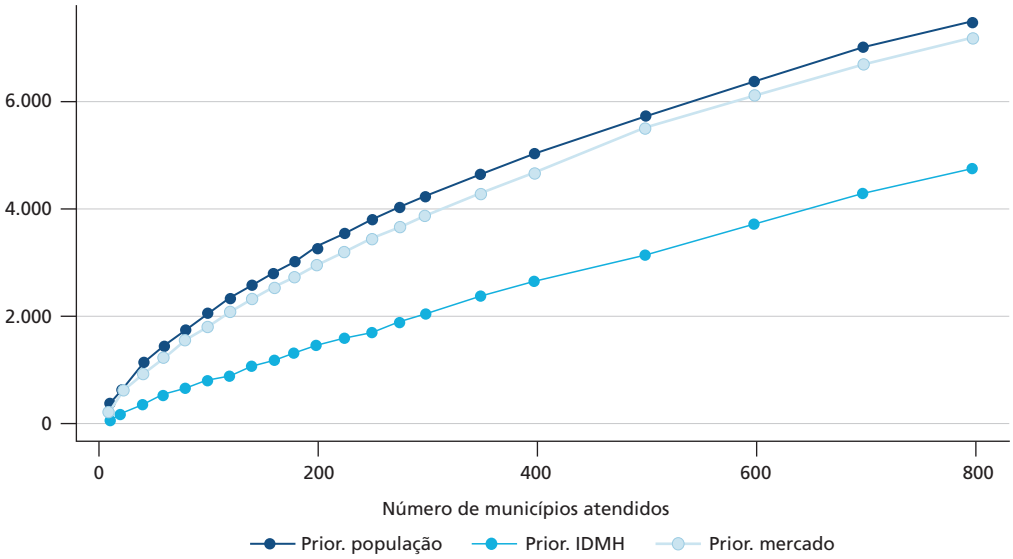
Elaboração dos autores.

GRAFICO A.3
População extremamente pobre atingida: total
(Em milhares de habitantes)



Elaboração dos autores.

GRAFICO A.4
População com até 20 anos atingida
(Em milhares de habitantes)



Elaboração dos autores.

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
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

Composto em adobe garamond pro 12/16 (texto)
Frutiger 67 bold condensed (títulos, gráficos e tabelas)
Brasília-DF

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

BRASIL
GOVERNO FEDERAL

ISSN 1415-4765