

de Carvalho, Alexandre Ywata; de Mendonça, Mário Jorge Cardoso; Lavagnole, Marcus Gerardus; Medrano, Luis Alberto Toscano

Working Paper

Avaliação do impacto econômico setorial da banda larga

Texto para Discussão, No. 2366

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: de Carvalho, Alexandre Ywata; de Mendonça, Mário Jorge Cardoso; Lavagnole, Marcus Gerardus; Medrano, Luis Alberto Toscano (2018) : Avaliação do impacto econômico setorial da banda larga, Texto para Discussão, No. 2366, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/177582>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

2366

**AVALIAÇÃO DO IMPACTO ECONÔMICO
SETORIAL DA BANDA LARGA**

TEXTO PARA DISCUSSÃO

**Alexandre Ywata de Carvalho
Mário Jorge Mendonça
Marcus Gerardus Lavagnole
Luis Alberto Medrano**



AVALIAÇÃO DO IMPACTO ECONÔMICO SETORIAL DA BANDA LARGA

Alexandre Ywata de Carvalho¹
Mário Jorge Mendonça²
Marcus Gerardus Lavagnole³
Luis Alberto Medrano⁴

1. Diretor da 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/Ipea. *E-mail*: <mario.mendonca@ipea.gov.br>.

3. Assistente de pesquisa na Diretoria de Estudos e Políticas Macroeconômicas (Dimac) do Ipea. *E-mail*: <marcus.lavagnole@ipea.gov.br>.

4. Assistente de pesquisa na Dimac/Ipea e professor da Universidade Federal Rural do Rio de Janeiro (UFRRJ). *E-mail*: <lmedrano10@msn.com>.

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

Fabiano Mezadre Pompermayer

Diretora de Estudos e Políticas Sociais

Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais

Ivan Tiago Machado Oliveira

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** 2018

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: O57; O47; L69.

SUMÁRIO

RESUMO

ABSTRACT

1 INTRODUÇÃO	7
2 REFERENCIAL TEÓRICO.....	9
3 REGRESSÃO MULTIVARIADA EM DADOS EM PAINEL	11
4 CLUSTERIZAÇÃO	16
5 RESULTADOS E COMENTÁRIOS FINAIS	22
REFERÊNCIAS	26

RESUMO

Embora exista concordância de que o efeito da expansão da banda larga sobre a economia, visto de um prisma agregado, seja positivo, não é voz unânime que esse impacto seja positivo sobre todos os setores. Por meio da aplicação de um modelo multivariado de dados em painel *two way* estimado por metodologia bayesiana, este estudo objetiva determinar o efeito da expansão da banda larga sobre os setores da economia. Com o propósito de contemplar as diferenças regionais, foi aplicada a análise de agrupamento (clusterização), visando isolar os grupos de municípios com características aproximadas. Os resultados ilustram o fato de que o impacto da banda larga não é positivo para todos os setores, sendo positivo nos setores mais dinâmicos e mais intensivos em tecnologia, como é o caso dos setores industrial e de serviços, e negativo onde isso não acontece, como na agricultura.

Palavras-chave: banda larga; modelo multivariado; dados em painel; clusterização; amostrador de Gibbs.

ABSTRACT

Although there is agreement that the effect of broadband expansion on the economy, from an general overview, is positive; not an unanimous voice arguments that this impact be positive on all sectors. By applying a multivariate two way panel data model and estimated by bayesian methodology, this study aims to determine the effect of broadband expansion on the sectors of the economy. In order to take into account the geographical, social and economic differences among municipalities cluster analysis was employed in order to isolate groups of municipalities with approximate characteristics. The results illustrate that the impact of broadband is not positive for all sectors, being positive in the most dynamic and technology-intensive sectors such as the industrial and service sectors; and negative where this does not happen as in agricultural sector.

Keywords: broadband; multivariate model; panel data; clustering; Gibbs sampler.

1 INTRODUÇÃO

Embora os estudos objetivando detectar o efeito da banda larga sobre o produto interno bruto (PIB) sejam baseados em metodologias e base de dados de caráter bastante diferenciado, de um modo geral, é consenso que a banda larga, ao ampliar o acesso às tecnologias de informação e comunicação (TICs), tem gerado efeito positivo sobre a economia. Esses ganhos estão distribuídos na economia entre o aumento da produtividade, a criação de empresas, o aumento no emprego e o crescimento econômico (Czernich *et al.*, 2009; Koutroumpis, 2009; Qiang e Rossotto, 2009). Também para o Brasil, as pesquisas se mostram em conformidade, indicando que o impacto da banda larga sobre o produto é significativo, tendo sinal positivo (Macedo e Carvalho, 2010a; 2010b; Carvalho, Mendonça e Silva, 2017; Fonseca, 2015).

Ocorre, contudo, que, embora o efeito da expansão da banda larga sobre o produto agregado seja positivo, esse impacto não se dá de modo homogêneo sobre a economia. Katz (2012) assinala que o efeito é mais sentido nos setores intensivos em tecnologia da informação (TI). Stiroh (2002) afirma que o emprego da tecnologia da informação aumentou substancialmente a produtividade da economia americana, sendo que o aumento na eficiência é também sentido com maior relevância em setores com alto custo de transação – por exemplo, o setor financeiro –, ou intensivos em trabalho, tal como o turismo. Assim, o efeito sobre o crescimento econômico parece estar ligado aos setores da economia 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.

Diante do que foi observado, pressupõe-se que as regiões mais desenvolvidas tecnologicamente se beneficiariam mais que as demais da expansão da rede de banda larga. Essa suposição é natural, visto que alguns estudos já evidenciam que o efeito desta expansão sobre as regiões ligadas ao setor agrícola é menor que em regiões onde predominam os setores de serviços e também naquelas com forte grau de industrialização. No entanto, também existe a percepção de que o efeito positivo da banda larga tem sido registrado com forte intensidade no setor agrícola (Forman, Goldfarb e Greenstein, 2005), assim como no meio rural (Larose *et al.*, 2011). Isso porque existem segmentos do setor agrícola que estão se tornando cada vez mais intensivos em tecnologia, como

é o caso do agronegócio, em que existe um entrelaçamento cada vez mais intenso da agricultura com a indústria. Pode-se ainda conjecturar que os residentes e as empresas das pequenas cidades e das áreas rurais em locais remotos, mas com acesso à internet, obtêm retornos marginais em comunicação mais elevados.

No Brasil, o governo, por meio da Agência Nacional de Telecomunicações (Anatel), regula as diretrizes básicas desse setor, sendo que serão levadas a cabo, nos próximos anos, várias iniciativas que redundarão na expansão e modernização da infraestrutura de telecomunicações do país. Tendo em vista que o efeito do avanço da banda larga sobre os diferentes setores da economia acontece de modo heterogêneo, é fundamental que tenhamos, de antemão, alguma previsibilidade de qual será o impacto da expansão da banda larga e como isso se efetuará.

Uma vez que a literatura atesta, como visto, que o efeito da banda larga sobre o PIB agregado no Brasil é positivo, é necessário que se investigue o impacto de sua expansão entre os diversos setores da economia. Dutz *et al.* (no prelo) elaboraram um estudo minucioso para o Brasil acerca dos efeitos da expansão da internet, um subproduto da banda larga, sobre os diversos ramos da atividade econômica, mostrando que os impactos sobre os diversos setores são fortemente heterogêneos, e podem, inclusive, atuar negativamente sobre o emprego e a produtividade, a depender do segmento da atividade econômica que se analise.

Este estudo objetiva determinar o efeito da expansão da banda larga sobre os setores da economia. Isso é feito com base no emprego de um modelo multivariado de dados em painel, estimado por meio da prática bayesiana (Koop, 2003; Lancaster, 2004; Greenberg, 2007; entre outros). Nosso modelo de dados em painel é do tipo *two way*, capaz de controlar tanto o efeito fixo entre as unidades como o efeito exógeno do tempo sobre as unidades. Nossa base de dados incorpora informação oriunda de várias fontes, e é composta de dados de 5.564 municípios brasileiros,¹ no período 2007-2014. A análise contempla o setor de serviços, o setor agrícola, o setor industrial e a administração pública. Com o propósito de contemplar as diferenças regionais, foi aplicada a análise de agrupamento (clusterização), visando isolar os grupos de municípios com

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

características aproximadas. Tal procedimento é feito em uma etapa que antecede a estimação, sendo que a informação advinda da clusterização é incorporada na análise econométrica com o intuito de identificar o efeito da banda larga nos *clusters*.

Os resultados ilustram o fato de que o impacto da banda larga não é positivo para todos os setores, impactando positivamente sobre o setor industrial e o de serviços, e atuando negativamente sobre a agricultura. No entanto, quando se considera a clusterização, visualizam-se situações nas quais a banda larga tem efeito positivo sobre a agricultura, assim como são identificados casos em que esse efeito é negativo sobre o setor industrial e o setor agrícola.

Este trabalho está estruturado da forma a seguir, além desta introdução. A seção 2 descreve o referencial teórico que embasa o modelo econométrico a ser estimado de acordo com a metodologia bayesiana para estimar o modelo multivariado de dados em painel *two way*, descrito na seção 3. A seção 4 aplica a análise de agrupamentos, de modo a identificar, assim como contextualizar, os grupos homogêneos de municípios. A seção 5 apresenta os resultados estimados e as principais conclusões do estudo.

2 REFERENCIAL TEÓRICO

Nosso modelo estrutural parte da ideia consagrada na literatura econômica de que o produto setorial da economia ($Y(s)$) pode ser modelado por uma função de produção cujos fatores de produção mais usados 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 e Sala-i-Martin, 1995). Inseridos nesses últimos, aparecem os serviços de telecomunicações, destacando-se aqueles ofertados pela tecnologia de banda larga (T). Denotando por O_Y os demais fatores que afetam a produção, temos que o produto do setor pode ser representado da seguinte forma:

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

Assim, nossa tarefa é estimar um sistema de equações cujas variáveis endógenas sejam os produtos setoriais da economia. Cada equação do modelo multivariado modela a variável $PIB(s)$, em que o índice s representa um dos quatro setores

da economia: o setor agrícola, o industrial, o de serviços e a administração pública. O sistema de equações é definido por um conjunto de quatro regressões e representado pela equação (2). Cada equação é estimada usando-se uma amostra de dados em painel dos municípios brasileiros, que contempla os efeitos da banda larga para subgrupos homogêneos (*clusters* $j=1, \dots, J$) de municípios, e é definida da seguinte forma:

$$\log(PIB_{i,t}(s)) = \alpha(s)_i + \lambda(s)_t + \sum_{j=1}^J \phi(s)_j \log(w_{i,t}) * Dcluster_{i,j} + \sum_{k=1}^K \delta(s)_k \log(x_{i,t,k}) + \varepsilon(s)_{i,t}. \quad (2)$$

Em que:

s = setor agrícola, industrial, de serviços e administração pública;

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

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

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

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

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

O parâmetro $\alpha(s)_i$ indica o efeito individual ou efeito específico referente ao município i na equação s capaz de contemplar a heterogeneidade existente entre os municípios. De modo que a equação (2) possa ser uma representação empírica da equação (1), devemos ter as medidas de capital, emprego, produto ou renda, além, naturalmente, de um indicador verossímil de acesso à banda larga. Naturalmente, entre a pesquisa teórica e a sua aplicabilidade existe, por vezes, certa distância. Na prática econométrica, é comum não se dispor de algumas pesquisas que aparecem no referencial teórico. Isso acontece por vários motivos. No caso corrente, isso é devido ao nível de desagregação da nossa base de dados. No Brasil, não existe medida de capital desagregada para municípios. Assim, procuramos controlar essa dificuldade por meio de variáveis de controle, que são as receitas totais (*RECEITA*).²

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

Para representar o fator trabalho ou emprego, usamos a variável vínculos³ da Relação Anual de Informações Sociais (Rais)⁴ (*EMPREGO*), e para representar a variável *T* da equação (1), que é a nossa variável de política ($w_{i,t}$) da equação (2), usamos a densidade de acessos à banda larga (*DEM_SCM*).⁵ Incluímos também, como variáveis de controle, as transferências governamentais decorrentes do Programa Bolsa Família (*PBF*), do Benefício de Prestação Continuada (*BPC*) e os recursos advindos da Previdência Social (*PREVT*), excetuando-se o BPC. O parâmetro ϕ_j , associado ao termo composto $\log(w_{i,t}) * Dcluster_{i,j}$, define o efeito sobre o PIB nos municípios inseridos no agrupamento *j* decorrente de um aumento dos serviços de banda larga.

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

A base de dados do presente estudo é composta de dados de 5.564 municípios brasileiros, no período 2007-2015, e é formada por informações oriundas de bases 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 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/MTE.

3 REGRESSÃO MULTIVARIADA EM DADOS EM PAINEL

Nesta seção, descrevemos a metodologia utilizada para estimar o modelo multivariado de dados em painel. Partimos da prática de inferência bayesiana, que traz vantagens importantes em termos da capacidade de obter “estimativas mais confiantes na presença

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

4. 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).

5. Razão entre o número de acessos do Sistema de Comunicação Multimídia (SCM) por grupo de 100 domicílios. Fonte: Anatel. Disponível em: <www.anatel.gov.br>.

de pequenas amostras com um espaço de parâmetros de alta dimensão” (Gelman *et al.*, 2003, p. 696). O modelo de dados em painel conjuga séries transversais e séries temporais dentro de uma mesma estrutura (Hsiao, 1995; Baltagi, 1995; Arellano, 2003). A análise dos dados longitudinais aqui apresentados nos permite examinar não apenas o efeito fixo como também o efeito temporal exógeno, ambos inerentes às unidades individuais. O modelo univariado para dados em painel foi amplamente tratado no paradigma bayesiano. Laird e Ware (1982) e Chib e Carlin (1999) estão entre as principais referências. Boas exposições também aparecem em Koop (2003), Lancaster (2004) e Greenberg (2007).

Na perspectiva frequentista, o modelo de regressão multivariada em dados em painel foi estudado por muitos autores. Baltagi (1981; 1995), Baltagi e Li (1992) e Kinal e Lahiri (1990; 1993) mostram que uma estimativa consistente pode ser obtida usando-se 3SLS; Cornwell, Schmidt e Wyhowski (1992) empregam a abordagem GMM, enquanto o emprego da abordagem bayesiana no tratamento de modelos multivariados aparece em Greenberg (2007), Gamerman e Moreira (2004) e Gamerman e Lopes (2006).

3.1 Modelo multivariado de dados agrupados (*pooling*)

Seja $y_{it} = (y_{it1}, \dots, y_{itQ})'$ um vetor $Q \times 1$ das variáveis endógenas e $u_{it} = (u_{it1}, \dots, u_{itQ})'$ o vetor $Q \times 1$ dos termos de distúrbio com $u_{it} \sim (0, \Sigma_u)$, onde Σ_u é uma matriz $Q \times Q$. Seja, ainda, $x_{itk} = (x_{itk1}, \dots, x_{itkQ})'$ com $k = 1, \dots, K$ um vetor de variáveis exógenas. Cada k está associado ao efeito da k -ésima variável exógena sobre a variável endógena Q . Os índices i e t estão associados, respectivamente, à unidade espacial e ao tempo, de modo que $i = 1, \dots, N$ e $t = 1, \dots, T$. Precisamos de algumas anotações adicionais para alocar os dados da amostra de forma mais compacta. As informações relativas às variáveis endógenas de cada unidade i podem ser colocadas em uma matriz Y_i de modo que

$$Y_i = \begin{bmatrix} (y_{i11}, \dots, y_{i1Q}) \\ \vdots \\ (y_{iT1}, \dots, y_{iTQ}) \end{bmatrix} = \begin{bmatrix} y'_{i1} \\ \vdots \\ y'_{iT} \end{bmatrix},$$

em que Y_i é uma matriz de ordem $T \times Q$.

Podemos agrupar as informações de todas as unidades individuais em uma matriz Y de ordem $TN \times Q$, de maneira que Y pode ser definida da seguinte forma:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}.$$

A informação relacionada às variáveis exógenas e aos distúrbios podem ser também representadas, respectivamente, pelas matrizes X e E , construídas de modo semelhante ao que foi feito para a matriz Y . Assim, o modelo de dados agrupados pode ser expresso da seguinte forma:

$$Y = XB^* + U, \quad (3)$$

em que Y e U são matrizes de ordem $NT \times Q$ e X é uma matriz $NT \times K$. Iremos assumir que U tem uma distribuição matriz normal, de modo que $U \sim MN(0, \Sigma_u \otimes I_{TN})$.

3.2 Modelo painel *two way* multivariado

O modelo *two way* que contempla o efeito fixo com efeito temporal pode ser introduzido expandindo-se o modelo de dados agrupados (*pooling*). Seja o vetor $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iQ})'$, que para uma determinada unidade espacial i representa o efeito fixo sobre essa unidade espacial em cada uma das equações. Para que o modelo multivariado comporte o efeito fixo, precisamos definir uma matriz de efeitos individuais $\alpha = [\alpha_{ij}]_{N \times Q}$, em que α_{ij} é o efeito fixo sobre a unidade i na equação j . Podemos então considerar essa matriz como sendo um conjunto a mais de parâmetros. A inclusão da matriz de efeito fixo α na equação (3) pode ser feita tal como aparece na equação (4).

$$Y = XB^* + (I_N \otimes i_T)\alpha + (i_N \otimes I_T)\lambda + U, \quad (4)$$

em que i_T e i_N são, respectivamente, de ordem $T \times 1$ e $N \times 1$, com todos os elementos iguais a um, e I_T é a matriz identidade de ordem $T \times T$. O modelo multivariado pode ser acomodado de modo mais compacto, de forma que

$$Y = \begin{bmatrix} I_N \otimes i_T & i_N \otimes I_T & X \end{bmatrix} \begin{bmatrix} \alpha \\ \lambda \\ B^* \end{bmatrix} + U$$

Ou

$$Y = ZB + U, \tag{4'}$$

em que $\begin{bmatrix} I_N \otimes i_T & i_N \otimes I_T & X \end{bmatrix} = Z$, com Z de ordem $TN \times (N + T + K)$ e $\begin{bmatrix} \alpha \\ \lambda \\ B^* \end{bmatrix} = B$, onde B é de ordem $(N + T + K) \times Q$.

3.3 Inferência bayesiana

Consideramos estimar o modelo da equação (4) com base na abordagem bayesiana. De acordo com o teorema de Bayes, a distribuição *a posteriori* para os parâmetros dessa equação é definido da seguinte forma:

$$P(B, \Sigma_u | Y) \propto P(Y | B, \Sigma_u) P(B, \Sigma_u),$$

em que $P(Y | B, \Sigma_u)$ é a função de verossimilhança e $P(B, \Sigma_u)$ é a distribuição *a priori*. Assumimos aqui a hipótese de que a função de verossimilhança é uma matriz normal (*MN*) (Press, 1989), de modo que

$$P(Y | B, \Sigma_u) \propto |\Sigma_u|^{-Q/2} \exp \left[-\frac{1}{2} \text{tr} \left(\Sigma_u^{-1} (Y - ZB)' (Y - ZB) \right) \right].$$

No caso da distribuição *a priori* para (B, Σ_u) , assume-se que a distribuição conjunta para esses dois parâmetros é da forma matriz normal Wishart invertida (Press, 1989), tal que $(B, \Sigma_u) \sim \text{MNIW}(B_0, H_0, \nu_0, S_0)$, com densidade $P(B, \Sigma_u) \propto P(B | \Sigma_u) P(\Sigma_u)$, onde

$$P(B | \Sigma_u) \propto |H_0|^{-Q/2} |\Sigma_u|^{-Q/2} \exp \left[-\frac{1}{2} \text{tr} \left(\Sigma_u^{-1} (B - B_0)' H_0^{-1} (B - B_0) \right) \right]$$

$$P(\Sigma_u) \propto |\nu_0 S_0|^{-Q/2} |\Sigma_u|^{-(\nu_0+Q+1/2)} \exp\left[-\frac{1}{2} \text{tr}(\Sigma_u^{-1} \nu_0 S_0)\right].$$

Aqui, H_0 e S_0 são matrizes definidas positivas, sendo que $\nu_0 > 0$ é um número inteiro. Em outras palavras, condicional a Σ_u , $B | \Sigma_u \sim N(B_0, \Sigma_u \otimes H_0^{-1})$, enquanto a distribuição marginal para Σ_u^{-1} é uma Wishart invertida denotada por $\Sigma \sim IW(\nu_0, S_0)$ e com valor esperado $E[\Sigma] = S_0 / (\nu_0 - (Q+1))$.

Pelo fato de o núcleo da função de verossimilhança apresentar uma distribuição normal multivariada, as distribuições *a priori* especificadas anteriormente são denominadas *a priori* conjugadas; dessa forma, as distribuições *a posteriori* pertencerão às mesmas famílias de distribuições. Logo, as distribuições condicionais completas serão $B | \Sigma_u, Y \sim N(B_T, \Sigma_u \otimes H_T^{-1})$ e $\Sigma_u | Y \sim IW(S_T, \nu_T)$. Então, temos que:

$$H_T = H_0 + Z'Z; \tag{5}$$

$$\nu_T = \nu_0 + T; \tag{6}$$

$$S_T = \frac{\nu_0}{\nu_T} S_0 + \frac{T}{\nu_T} \hat{\Sigma} + \frac{1}{\nu_T} (\hat{B} - B_0)' H_0 H_T Z' Z (\hat{B} - B_0); \tag{7}$$

$$B_T = H_T^{-1} (H_0 B_0 + Z' Z \hat{B}). \tag{8}$$

Todas as distribuições condicionais assumidas são conhecidas, podendo ser simuladas pelo algoritmo de amostragem de Gibbs (Gamerman e Lopes, 2006; Dagpunar, 2007; Greenberg, 2007). Conhecido o resultado da (i-1)-ésima amostragem, a *i*-ésima amostragem é obtida com base nos passos a seguir.

Passo 1: $\Sigma_u^{(g)} \sim IW(S_T, \nu_T)$.

Passo 2: $B^{(g)} | \Sigma_u^{(g-1)} \sim N(B_T^{(g-1)}, \Sigma^{(g-1)} \otimes H_T^{-1})$.

Esse processo deve ser repetido um número muito grande de vezes, descartando os resultados das primeiras N iterações, sendo que os resultados das simulações restantes podem ser usados para a inferência. Informação sobre a convergência pode ser

acessada pelo monitoramento da correlação serial das amostragens, pelo cálculo do desvio-padrão dos parâmetros estimados baseados nos resultados das amostragens e por meio do diagnóstico de Gelman e Rubin (1992).

4 CLUSTERIZAÇÃO

Devido à heterogeneidade existente entre os municípios brasileiros, faz-se necessário identificar os grupos homogêneos em termos de suas variáveis. Este ponto é fundamental para a gestão de política pública e o direcionamento eficiente dos investimentos em telecomunicações, na medida em que o impacto dos investimentos deve ser sentido de forma distinta entre os grupos selecionados. Com o objetivo de identificar os grupos homogêneos de municípios, faremos uso do que se conhece no âmbito da análise de multivariados como análise de agrupamentos. Como definem Lattin, Carroll e Green (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 exige que exista um grupo de observações com alta densidade local, isto é, muitas observações dentro de uma determinada área, mas separados por regiões de densidade local baixa. Em outras palavras, os agrupamentos correspondem a uma modalidade de dados, e o número de agrupamento ao número de modas, a uma distribuição multimodal. Diversos métodos podem ser empregados para a análise de agrupamentos. Dois deles são de uso mais comum: os métodos hierárquicos e os métodos de partição.⁶ O resultado gerado pela aplicação de um método hierárquico é representado pela estrutura de árvore, no qual a solução de k agrupamento 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 k agrupamento e a solução de $k+1$ agrupamento não são necessariamente aninhadas. Quando passa de um número k de grupos para um número $k+1$, o nível de similaridade decresce. Ou seja, a variação entre os grupos diminui e a variação dentro dos grupos aumenta.

6. Descrição geral dos algoritmos de clusterização (hierárquicos e *k-means*): Hastie, Tibshirani e Friedman (1997), Khattree e Naik (2000), Berry e Linoff (1997) e Alpaydin (2004).

Os métodos hierárquicos geralmente abordam a análise de dados por intermédio de dois procedimentos distintos. O procedimento aglomerativo, de baixo para cima, começa com cada observação em um agrupamento separado, unindo-se 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, se inicia com todas as observações em um único agrupamento, e dividindo-se 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-se que cada município constitui um *cluster* de tamanho 1.

- Passo 1: para todos os pares de municípios i e j , o par com menor “dissimilaridade” entre os vetores de variáveis escolhidas é combinado e passa a constituir um novo município (pseudomunicípio). Esta seleção é feita de acordo com a seguinte métrica $D_{i,j}$.

$$D_{i,j} = \frac{\|x_i - x_j\|^2}{2}$$

- Passo 2: novo pseudomunicípio, composto pela união do par de municípios com menor dissimilaridade. Apenas um pseudomunicípio é feito a cada passo, e, devido à propriedade de hierarquia, uma vez juntos, os pseudomunicípios não se separam nos estágios subsequentes.

$$x_{i \cup j} = \frac{x_i + x_j}{2}$$

- Passo 3: para todos os pares de municípios e pseudomunicípios, deve-se selecionar o par com menor “dissimilaridade” entre os vetores de variáveis selecionadas, considerando-se a métrica de Ward.

$$D_{i,j} = \frac{\|x_i - x_j\|^2}{\frac{1}{n_i} + \frac{1}{n_j}},$$

em que n_i e n_j são, respectivamente, os números dos agrupamentos i e j .

- Passo 4: novo pseudomunicípio, composto pela união dos pseudomunicípios com menor dissimilaridade:

$$x_{i \cup j} = \frac{n_i x_i + n_j x_j}{n_i + n_j}.$$

- Passo 5: retorna-se 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, através 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 o melhor agrupamento de dados. Isso é feito verificando-se 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 números de grupos.

O método de partição tem por objetivo repartir a amostra em num 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 o 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 soluções localmente ótimas apenas, 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.

Passo 1: escolhe-se de início uma partição dos dados em K agrupamentos. Diversas abordagens podem ser aplicadas para a escolha dessa partição inicial. Outra variante importante do método *K-means* é aquela que começa com um conjunto inicial de centroides de semente K designando cada objeto para o ponto de semente mais próximo. Em alguns casos, os pontos de semente são simplesmente os pontos reais da amostra, de modo que sejam amplamente dispersos; contudo, na maioria dos casos, os pontos de semente são escolhidos aleatoriamente dentro do conjunto de dados.

Passo 2: calcula-se o centroide para cada agrupamento C , $\bar{x}_{C(i)}$.

Passo 3: calcula-se a soma das distâncias ao quadrado de cada objeto ao seu centroide do agrupamento, isto é, o quadrado da soma dos erros da partição, *ESS*. Essa é a medida que desejamos minimizar:

$$ESS = \sum_{i=1}^n (x_i - \bar{x}_{C(i)})' (x_i - \bar{x}_{C(i)}).$$

- Passo 4: relaciona-se novamente cada objeto i a um outro agrupamento cujo centroide é o mais próximo de cada objeto, minimizando-se 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, retorna-se ao passo 2, com uma nova partição.

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

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

Note-se que 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 de a 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 o *pseudo-t2* proposta por Duda e Hart (1973) também é importante para indicar o número de *clusters*. Esta estatística é calculada em cada passo do algoritmo de clusterização. Se em um passo do algoritmo o *cluster* C_k é a união de C_i e C_p , então o *pseudo-t2* é definido por:

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

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

4.1 Resultados da clusterização

Tendo descrito na seção anterior a ideia geral acerca da análise de agrupamento e de alguns métodos que permitem sua aplicação, passaremos para a aplicação empírica do método como meio para a construção dos grupos de municípios homogêneos a partir

da nossa amostra. De início, devemos ter em mente que amostras diferentes de variáveis podem levar a grupos homogêneos diferentes. Por outro lado, a nossa experiência mostra que, devido à grande correlação em geral entre os indicadores socioeconômicos, a utilização de indicadores diferentes não incorre em agrupamentos significativamente diferentes. Neste trabalho, as variáveis utilizadas correspondem às características socioeconômicas dos municípios, a partir dos indicadores disponibilizados para o *Atlas do IDHM 2010*. Podemos categorizar as variáveis do seguinte modo: indicadores de renda, longevidade, saúde, educação, emprego e características dos domicílios.

Foram selecionadas as variáveis, em um total de 159, e a partir disso foi efetuada uma análise de componentes principais (Lattin, Carroll e Green, 2011) para redução de dimensionalidade. A análise de componentes principais permite, por exemplo, acomodar situações nas quais diversas variáveis possuem alta correlação entre elas. Selecionamos um número de componentes tais que esses contabilizassem 99% da variabilidade das 159 variáveis originais. 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 *clusters* 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.

A variável que identifica o *cluster* foi remarcada, de forma que os municípios correspondentes ao *cluster* 1 são aqueles com maior renda *per capita*; os municípios correspondentes ao último *cluster* são aqueles com menor renda *per capita*. Vejamos algumas observações acerca dos diferentes grupos.

- Grupos 1: formados por municípios com maior renda *per capita*.
- Grupos 1 e 3: são os grupos 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.
- Grupo 5: com alta taxa de desocupação para população maior ou igual a 18 anos.
- Grupos 2, 4, 5 e 6 (rurais): com baixa taxa de empregados com carteira.
- Grupo 6: com baixa renda, alto percentual da população morando em domicílios com água e esgoto inadequados e sem acesso à energia elétrica, e com alta fecundidade.

TABELA 1
Indicadores do número de *clusters* não espaciais

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

Elaboração dos autores.

5 RESULTADOS E COMENTÁRIOS FINAIS

Nesta seção, estimamos o modelo econométrico painel *two way* multivariado para determinar o efeito da expansão da banda larga sobre o PIB setorial, com base na metodologia bayesiana descrita na seção 3. Pelo fato de adotar o método de dados em painel com efeito fixo e controle exógeno do tempo (modelo *two way*), nossa abordagem permite considerar a heterogeneidade existente entre as unidades espaciais (municípios), assim como controlar o efeito exógeno do tempo. Os resultados são mostrados nas tabelas 2 e 3. Na tabela 3, o modelo incorpora informação decorrente da análise de agrupamentos (clusterização), em que se procura identificar as diferenças entre agrupamentos homogêneos de municípios (*clusters*), objeto de estudo da seção 4.

Nas tabelas 2 e 4, os resultados dos coeficientes estimados para cada regressor estão associados a uma determinada equação (coluna 1). Nas colunas 2 e 4, temos os

intervalos de confiança de 2,5% e 97,5%, enquanto a *média* é mostrada na coluna 3. As variáveis endógenas da coluna 1 são os valores adicionados do setor agrícola (*AGRO*), do setor industrial (*IND*), do setor de serviços (*SERV*) e a administração pública (*ADM*). O modelo estimado é do tipo *log-log*. Assim, o coeficiente estimado representa a elasticidade do PIB municipal com relação à variável explicativa. Inicialmente, vejamos os resultados mostrados na tabela 2 referentes ao modelo sem identificação de agrupamentos. A tabela 3 mostra o efeito exógeno do tempo para esse modelo.

No que se refere às variáveis de controle, destaca-se que *EMPREGO*, variável relacionada ao emprego formal, apresenta significância estatística e sinal esperado para todos os setores. Vejamos agora o efeito da banda larga (*DEM_SCM*), nossa variável de política, sobre o PIB setorial. Muito embora a literatura aponte para um efeito positivo da expansão da banda larga sobre o PIB agregado, conforme visto na introdução deste estudo, isso não acontece quando discriminamos tal efeito sobre os setores da economia. Nota-se que o efeito da expansão da banda larga é negativo sobre o setor agrícola e sobre a administração pública, enquanto é positivo para os setores industrial e de serviços. Tais resultados se mostram em conformidade com a literatura, na medida em que ilustram o fato de que o impacto da banda larga não é positivo entre os setores, sendo positivo nos setores mais dinâmicos e mais intensivos em tecnologia, e negativo onde isso não acontece.

Cabe, no entanto, indagar se os resultados obtidos na tabela 2 permanecem iguais caso se incorpore na análise o efeito dos agrupamentos de municípios com características homogêneas. Em outras palavras, é possível que, apesar de o efeito agregado da banda larga sobre o setor agrícola ser negativo, isso possa se reverter em determinadas áreas agrícolas que, por hipótese, sejam mais intensivas em tecnologia (Forman, Goldfarb e Greenstein, 2005; Larose *et al.*, 2011). De modo contrário, pode ser o caso de que existam certos agrupamentos de municípios onde o efeito da banda larga sobre o setor industrial seja negativo. De modo a verificar esse fato, incorporamos no modelo seis variáveis *dummies* que identificam os agrupamentos homogêneos com base na análise de *clusters* implementada na seção 4. Os resultados são apresentados na tabela 4, sendo que a tabela 5 mostra o efeito exógeno temporal para esse modelo.

Conforme se observa, existe certa diferenciação tanto quanto à dimensão como quanto ao sinal do efeito da banda larga entre os agrupamentos, considerando-se o mesmo setor. Por exemplo, no *cluster* 1, o efeito da banda larga sobre o setor industrial é

negativo. No caso do setor agrícola, temos agora que o impacto da banda larga é inócuo no *cluster 3*, enquanto é positivo no *cluster 4*. Nesse agrupamento, a expansão da banda larga não gera nenhum impacto sobre o setor de serviços. Nos demais casos, os resultados permanecem em concordância com aqueles já mostrados na tabela 2.

TABELA 2
Modelo multivariado para os PIBs regionais

Regressor	Equação (1)	2,5% (2)	Média (3)	97,5% (4)
EMPREGO	AGRO	0.0444	0.0534	0.0644
	IND	0.2710	0.2846	0.2987
	SERV	0.1133	0.1186	0.1241
	ADMP	0.0550	0.0577	0.0602
PBF	AGRO	0.0273	0.0417	0.0559
	IND	-0.0404	-0.0209	-0.0018
	SERV	0.0619	0.0700	0.0779
	ADMP	0.0771	0.0807	0.0844
DEN_SCM	AGRO	-0.0304	-0.0263	-0.0220
	IND	0.0297	0.0361	0.0427
	SERV	0.0080	0.0105	0.0126
	ADMP	-0.0056	-0.0045	-0.0035
RECEITAS	AGRO	-0.0037	0.0028	0.0093
	IND	0.0301	0.0400	0.0491
	SERV	0.0247	0.0285	0.0321
	ADMP	0.0193	0.0210	0.0225
BPC	AGRO	0.0020	0.0043	0.0066
	IND	-0.0025	0.0007	0.0042
	SERV	0.0001	0.0014	0.0026
	ADMP	0.0002	0.0008	0.0014
PREVI	AGRO	0.0109	0.0173	0.0230
	IND	-0.0001	0.0075	0.0169
	SERV	0.0050	0.0083	0.0116
	ADMP	0.0065	0.0080	0.0098

Elaboração dos autores.

Obs.: Os destaques (em cinza) indicam que a variável é não significativa.

TABELA 3
Efeito temporal

	T1	T2	T3	T4	T5	T6	T7	T8
AGRO	8.820	8.884	8.862	8.817	8.889	8.742	8.822	8.773
IND	6.986	6.945	6.990	7.089	7.050	6.973	6.928	6.921
SERV	8.274	8.279	8.289	8.281	8.275	8.296	8.312	8.316
ADMP	8.753	8.778	8.788	8.782	8.767	8.749	8.771	8.746

Elaboração dos autores.

TABELA 4
Modelo multivariado para os PIBs regionais (com clusterização)

Regressor	Equação (1)	2,5% (2)	Média (3)	97,5% (4)
EMPREGO	AGRO	0.0439	0.0537	0.0631
	IND	0.2654	0.2795	0.2929
	SERV	0.1134	0.1185	0.1239
	ADMP	0.0535	0.0559	0.0586
PBF	AGRO	0.0334	0.0472	0.0629
	IND	-0.0433	-0.0228	-0.0038
	SERV	0.0540	0.0623	0.0699
	ADMP	0.0744	0.0784	0.0819
DEN_SCM_1	AGRO	-0.0582	-0.0518	-0.0453
	IND	-0.0313	-0.0215	-0.0120
	SERV	0.0072	0.0107	0.0147
	ADMP	-0.0123	-0.0104	-0.0089
DEN_SCM_2	AGRO	-0.0292	-0.0198	-0.0112
	IND	0.0363	0.0481	0.0615
	SERV	-0.0119	-0.0074	-0.0025
	ADMP	-0.0010	0.0012	0.0035
DEN_SCM_3	AGRO	-0.0047	0.0006	0.0063
	IND	0.0203	0.0284	0.0363
	SERV	0.0013	0.0044	0.0076
	ADMP	-0.0127	-0.0112	-0.0097
DEN_SCM_4	AGRO	0.0113	0.0172	0.0227
	IND	0.0116	0.0186	0.0268
	SERV	-0.0057	-0.0027	0.0003
	ADMP	-0.0095	-0.0081	-0.0068
DEN_SCM_5	AGRO	-0.0739	-0.0686	-0.0630
	IND	0.0422	0.0490	0.0567
	SERV	0.0204	0.0231	0.0260
	ADMP	-0.0069	-0.0055	-0.0043

(Continua)

(Continuação)

Regressor	Equação (1)	2,5% (2)	Média (3)	97,5% (4)
DEN_SCM_6	AGRO	-0.0402	-0.0349	-0.0297
	IND	0.0302	0.0378	0.0461
	SERV	0.0108	0.0136	0.0166
	ADMP	-0.0001	0.0014	0.0026
RECEITAS	AGRO	0.0000	0.0061	0.0133
	IND	0.0309	0.0403	0.0495
	SERV	0.0239	0.0278	0.0313
	ADMP	0.0194	0.0211	0.0228
BPC	AGRO	0.0009	0.0033	0.0054
	IND	-0.0023	0.0008	0.0041
	SERV	0.0003	0.0016	0.0027
	ADMP	0.0005	0.0010	0.0017
PREVI	AGRO	0.0083	0.0144	0.0200
	IND	0.0010	0.0093	0.0177
	SERV	0.0056	0.0091	0.0125
	ADMP	0.0071	0.0085	0.0102

Elaboração dos autores.

Obs.: Os destaques (em cinza) indicam que a variável é não significativa.

TABELA 5
Efeito temporal (com clusterização)

	T1	T2	T3	T4	T5	T6	T7	T8
AGRO	8.722	8.783	8.761	8.716	8.786	8.638	8.719	8.669
IND	7.055	7.019	7.069	7.173	7.152	7.078	7.035	7.030
SERV	8.390	8.397	8.408	8.401	8.398	8.421	8.437	8.440
ADMP	8.796	8.823	8.834	8.829	8.817	8.800	8.822	8.797

Elaboração dos autores.

REFERÊNCIAS

ALPAYDIN, E. **Introduction to machine learning**. 1. ed. Cambridge: The MIT Press, 2004.

ARELLANO, M. **Panel data econometrics**. Oxford University Press, 2003.

BALTAGI, B. H. Simultaneous equations with error components. **Journal of Econometrics**, v. 17, p. 189-200, Nov. 1981.

_____. **Econometric analysis of panel data**. New Jersey: John Wiley & Sons, 1995.

- BALTAGI, B. H.; LI, Q. A note on the estimation of simultaneous equations with error components. **Econometric Theory**, v. 8, n. 1, p. 113-119, 1992.
- BARRO, R. J.; SALA-I-MARTIN, X. **Economic growth**. McGraw-Hill, 1995.
- BERRY, M.; LINOFF, G. **Data mining techniques: for marketing, sales, and customer relationship**. Wiley Computer Publishing, 1997.
- CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics**, v. 3, p. 1-27, 1974.
- CARVALHO, A. Y.; MENDONÇA, M. J.; SILVA, J. J. **Avaliando o efeito dos investimentos em telecomunicações sobre o PIB**. Brasília: Ipea, 2017. (Texto para Discussão, n. 2336).
- CHIB, S.; CARLIN, B. P. On MCMC sampling in hierarchical longitudinal models. **Statistics and Computing**, v. 9, p. 17-26, 1999.
- CORNWELL, C.; SCHMIDT, P.; WYHOWSKI, D. Simultaneous equations and panel data. **Journal of Econometrics**, v. 51, p. 151-181, 1992.
- CZERNICH, N. *et al.* Broadband infrastructure and economic growth. **CESifo**, 2009. (Working Paper, n. 2861).
- DAGPUNAR, J. S. **Simulation and Monte Carlo: with applications in finance and MCMC**. Wiley, 2007.
- DUDA, R. O.; HART, P. E. **Pattern classification and scene analysis**. New York: Wiley, 1973.
- DUTZ, M. A. *et al.* Economy-wide and sectoral impacts on workers of Brazil's internet rollout. *In: Technology adoption and inclusive growth: impacts of digital technologies on skills and jobs of Latin American workers*. No prelo.
- FONSECA, H. M. M. **Ensaio em teoria de contratos aplicada no setor de telecomunicações com foco nos investimentos**. 2015. Tese (Doutorado) – Universidade de Brasília, Brasília, 2015.
- FORMAN, C.; GOLDFARB, A.; GREENSTEIN, S. How did location affect adoption of the commercial internet? Global village, urban density, and identity composition. **Journal of Urban Economics**, v. 58, p. 389-420, 2005.
- GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for bayesian inference**. Chapman & Hall, 2006.
- GAMERMAN, D.; MOREIRA, A. R. B. Multivariate spatial regression model. **Journal of Multivariate Analysis**, v. 91, p. 262-281, 2004.

- GELMAN, A. *et al.* **Bayesian data analysis**. 1. ed. New York: Chapman & Hall, 2003.
- GREENBERG, E. **Introduction to bayesian econometrics**. Cambridge University Press, 2007.
- GELMAN, A.; RUBIN, D. R. Inference from iterative simulation using multiple sequences (with discussion). **Statist. Sci.**, v. 7, n. 4, p. 457-511, 1992.
- HARTIGAN, J. A. **Clustering algorithms**. New Jersey: John Wiley & Sons, 1975.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. Springer Series in Statistics, 1997.
- HSIAO, C. **Analysis of panel data**. New York: Cambridge University Press, 1995.
- KATZ, R. L. **The economic and social impact of telecommunications output: empirical evidence in the US and Europe**. [s.l.], 2012.
- KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS software**. Cary, NC: SAS Institute Inc, 2000.
- KINAL, T.; LAHIRI, K. A computational algorithm for multiple equation models with panel data. **Economics Letters**, v. 34, p. 143-146, 1990.
- _____. On the estimation of simultaneous-equations error-components with an application to a model of developing country foreign trade. **Journal of Applied Econometrics**, v. 8, n. 1, p. 81-92, 1993.
- KOOP, G. **Bayesian econometrics**. John Wiley-Interscience, 2003.
- KOUTROUMPIS, P. The economic impact of broadband on growth: a simultaneous approach. **Telecommunications Policy**, v. 33, n. 9, p. 471-485, 2009.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, v. 38, n. 4, p. 963-974, 1982.
- LANCASTER, T. **Introduction to modern bayesian econometrics**. 1. ed. New Jersey: Wiley-Blackwell, 2004.
- LAROSE, R. *et al.* The impact of rural broadband development: lessons from a natural field experiment. **Government Information Quarterly**, v. 28, p. 91-100, 2011.
- LATTIN, J.; CARROLL, D.; GREEN, P. **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011.
- MACEDO, H. R.; CARVALHO, A. X. Y. **Aumento do acesso à internet em banda larga no Brasil e sua possível relação com o crescimento econômico: uma análise de dados em painel**. Rio de Janeiro: Ipea, 2010a. (Texto para Discussão, n. 1494).

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

PRESS, S. J. **Bayesian statistics**: principles, models and applications. Wiley, 1989.

QIANG, C.; ROSSOTTO, C. M. Economic impacts of broadband. *In: Information and communications for development 2009*: extending reach and increasing impact. Washington: World Bank, 2009.

STIROH, K. Information technology and the U. S. productivity revival: what do the industry data say? **American Economic Review**, v. 92, n. 5, p. 1559-1576, 2002.

BIBLIOGRAFIA COMPLEMENTAR

CHIB, S.; GREENBERG, E. Understanding the metropolis-hasting algorithm. **The American Statistician**, v. 49, n. 4, p. 327-335, 1995.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2010**. Rio de Janeiro: IBGE, 2010.

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

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

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

Ana Clara Escórcio Xavier

Camilla de Miranda Mariath Gomes

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Hislla Suellen Moreira Ramalho (estagiária)

Lilian de Lima Gonçalves (estagiária)

Lynda Luanne Almeida Duarte (estagiária)

Luiz Gustavo Campos de Araújo Souza (estagiário)

Editoração

Aeromilson Trajano de Mesquita

Bernar José Vieira

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Herllyson da Silva Souza

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Capa

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

Projeto Gráfico

Renato Rodrigues Bueno

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

Livraria Ipea

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

70076-900 – Brasília – DF

Tel.: (61) 2026-5336

Correio eletrônico: livraria@ipea.gov.br

Missão do Ipea

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

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
**PLANEJAMENTO,
DESENVOLVIMENTO E GESTÃO**



ISSN 1415-4765

