

Determinantes Robustos do Crescimento nos Estados Brasileiros: Uma Abordagem Bayesiana

Fábio Augusto Reis Gomes

Departamento de Economia, FEA-RP/USP

Gian Paulo Soave*

Departamento de Economia, UFBA

Resumo

Este artigo utiliza a metodologia de ponderação bayesiana de modelos para acomodar a incerteza sobre os potenciais determinantes do crescimento econômico nos estados brasileiros. O procedimento leva em conta a possível endogeneidade de algumas variáveis e calcula uma média ponderada de coeficientes entre inúmeros modelos, com pesos dados pela probabilidade posterior do modelo. Nossos resultados, baseados em dados em painel de 1992 a 2016 e um grande número de potenciais determinantes do crescimento econômico, sugerem que PIB inicial, qualidade de vida, educação, investimento privado, composição setorial e despesas em saúde pública são determinantes robustos do crescimento nos estados brasileiros.

Palavras-chaves: Ponderação de Modelos Bayesiana em Painel, Incerteza Sobre o Modelo, Crescimento Empírico, Determinantes Robustos do Crescimento Econômico.

Classificação JEL: C11, C21, C52, O10, O47, R11

Área ANPEC: Área 4 - Macroeconomia, Economia Monetária e Finanças

Abstract

This paper uses Bayesian Model Averaging methodology to accommodate uncertainty about the potential determinants of economic growth in the Brazilian states. The procedure takes into account the possible endogeneity of some variables and calculates a weighted average of coefficients among a substantial number models, with weights given by the posterior probability of the model. Our results, based on panel data from 1992 to 2016 and a large number of potential determinants of economic growth, suggest that initial GDP, standard of living, education, private investment, sectoral composition, and public health expenditures are robust determinants of growth in the Brazilian states.

Keywords: Panel-Based Bayesian Model Averaging, Model Uncertainty, Growth Empirics, Robust Growth Determinants

*Autor para correspondência. Email: gianps@ufba.br

1 Introdução

Os modelos neoclássicos de crescimento econômico apontam que países com os mesmos parâmetros estruturais de preferência e de tecnologia tenderiam para o mesmo nível de renda per capita de longo prazo (Solow 1956). Isso significa que neste grupo de países os mais pobres cresceriam mais rapidamente do que os mais ricos, o que é conhecido na literatura como convergência condicional de renda. Nesta perspectiva, o nível inicial da renda seria negativamente correlacionado com a taxa de crescimento da renda. Desde o trabalho seminal de Barro (1991), esta conjectura tem sido investigada por meio de regressões da taxa de crescimento da renda contra a renda inicial, levando-se em conta outros determinantes do crescimento econômico. Ao estudar 98 países no período de 1960 a 1985, o próprio Barro (1991) conclui que a taxa de crescimento do PIB per capita é negativamente correlacionada com o PIB inicial (de 1960). Além disso, os resultados de Barro (1991) sugerem que o crescimento é positivamente correlacionado com capital humano inicial (medido por taxas de matrículas) e com medidas de estabilidade política, e negativamente correlacionado com o percentual do consumo do governo no PIB e com medidas de distorções de mercado.

Como discutido por Sala-i-Martin (1997), a literatura empírica identificou um elevado número de variáveis parcialmente correlacionadas com a taxa de crescimento econômico por meio de regressões *à la* Barro (1991), o que inclui renda inicial, taxas de investimento, medidas de educação e indicadores de políticas. De fato, Sala-i-Martin (1997) documenta que mais de 60 variáveis são parcialmente correlacionadas com o crescimento econômico, enquanto Durlauf, Johnson e Temple (2005) identificam 140 potenciais determinantes do crescimento econômico. Cabe, então, a pergunta: quais seriam os determinantes robustos do crescimento econômico? Segundo Brock e Durlauf (2001), em seu estágio atual, a teoria de crescimento econômico é incapaz de fornecer um guia para a escolha de determinantes plausíveis do crescimento econômico, no sentido de que a teoria encontra-se em evolução – *open-ended*, na terminologia dos autores –, o que resulta no problema conhecido por “incerteza sobre o modelo”.¹

Diante dessa situação, a literatura tem se apoiado em procedimentos empíricos para selecionar os determinantes robustos do crescimento econômico. Inclusive, são utilizados métodos estatísticos capazes de lidar com bases de dados com grande número de variáveis. Por exemplo, Levine e Renelt (1992) usam *extreme-bounds tests* e concluem que poucas ou nenhuma variáveis podem ser consideradas determinantes robustos do crescimento. Por sua vez, Sala-i-Martin (1997) argumenta que um método plausível para a análise dos determinantes robustos do crescimento econômico deve acomodar os potenciais diferentes modelos, sendo que uma mesma variável pode fazer parte de vários deles. O autor propõe computar a distribuição do parâmetro associado à variável de interesse, levando em conta um conjunto de modelos distintos. O método proposto por Sala-i-Martin baseia-se num método frequentista de ponderação de modelos, no qual o parâmetro de interesse é computado como uma média ponderada de seus valores estimados para cada um dos modelos, cujo peso depende de uma razão de verossimilhanças. Assim, Sala-i-Martin (1997) estima diversos modelos e combina os coeficientes estimados para cada regressor, concluindo que variáveis relacionadas a região, políticas, religião, distorções e performance do mercado, setores da economia, abertura econômica, tipo de organização econômica e antigas colônias espanholas são robustas.

A análise dos determinantes do crescimento econômico encontra uma série de dificuldades discutidas por Durlauf (2001). Primeiro, dentre as muitas variáveis investigadas, algumas são claramente endógenas, mas, em larga medida, isso é ignorado nos estudos. Segundo, o elevado número de regressores acarreta um elevado número de especificações e, assim, as estimativas dos coeficientes e dos erros-padrão podem ser muito voláteis. Terceiro, o uso de muitos regressores

1. Tal incerteza pode se repercutir de duas maneiras: (i) incerteza sobre a teoria, em que há desacordo entre economistas sobre os fundamentos teóricos e, (ii) incerteza sobre a especificação, em que há incerteza sobre como a teoria pode ser traduzida sob a forma de um modelo econométrico. Ver Steel (2017) para uma discussão detalhada.

potencializa o problema de multicolinearidade. De fato, é esperado que certos determinantes do crescimento sejam colineares.

Neste contexto, Durlauf (2001) aponta a abordagem Bayesiana de ponderação de modelos como sendo promissora. Nesta abordagem, ao invés de se assumir que o pesquisador esteja lidando com o “verdadeiro” modelo empírico com probabilidade igual a um, todos os possíveis modelos candidatos são estimados e a probabilidade posterior de cada um deles é calculada. Feito isso, constrói-se uma média ponderada dos coeficientes dos regressores, cujos pesos são as probabilidades posteriores dos modelos que geraram tais coeficientes, o que é conhecido como ponderação bayesiana de modelos (*Bayesian Model Averaging*, daqui em diante, BMA). É importante destacar que na abordagem bayesiana são acomodadas incertezas com relação tanto aos modelos quanto aos parâmetros.²

Por todas essas razões, neste trabalho adotamos a abordagem BMA. De fato, essa abordagem tem ainda outro importante benefício. Sua flexibilidade nos permite acomodar a possível endogeneidade de parte dos regressores.³ Mais especificamente, empregamos um método que pode ser visto como a versão bayesiana do estimador de mínimos quadrados de dois estágios (2SLS). Com isso, acomodamos a possibilidade de alguns dos regressores serem endógenos devido à existência de simultaneidade ou causalidade reversa, como proposto em Leon-Gonzalez e Montolio (2015). Finalmente, todas as propriedades atrativas da abordagem BMA tornaram-na uma ferramenta popular nos anos recentes para se estimar os determinantes robustos das taxas de crescimento econômico tanto em regressões de países em cortes transversais quanto em painéis (Fernández, Ley e Steel 2001; Durlauf, Kourtellos e Tan 2008; Moral-Benito 2012; Ciccone e Jarocinski 2010; Leon-Gonzalez e Montolio 2015; Leon-Gonzalez e Vinayagathan 2015). Neste sentido, estamos aproximando a literatura nacional da abordagem adotada na recente literatura internacional, sendo ainda mais rigorosos do que certos trabalhos no que diz respeito ao tratamento do problema de endogeneidade.

Nossa amostra é composta por um painel de estados brasileiros, além do distrito federal, de 1992 a 2016. A dificuldade comum a estudos intra-país é a indisponibilidade de variáveis comumente empregadas nas análises entre países. De todo modo, foi possível construir diversos potenciais regressores para as 27 unidades federativas brasileiras, o que inclui um indicador de qualidade de política macroeconômica, taxas de investimento pública e privada, medidas de capital humano, estrutura setorial, gastos de governo desagregados, condições demográficas e geográficas, além de fatores associados ao padrão de vida.⁴ Por fim, levamos em conta, ainda, a relação entre crescimento econômico e saúde no sentido de Nelson-Phelps, considerando o possível efeito causal de algumas *proxies* de condições de saúde sobre o crescimento.

Finalmente, os resultados empíricos sugerem que o crescimento econômico nos estados brasileiros está associado ao PIB inicial – havendo evidência de convergência condicional –, padrão e qualidade de vida, educação, investimento privado, composição setorial e despesas em saúde pública. Ainda, há evidência clara de que o PIB inicial, o investimento privado, um índice de

2. Sala-i-Martin (1997) ponderou modelos com base na abordagem clássica, sem levar em conta a própria incerteza sobre o modelo – que é formalmente explicitada em sua probabilidade posterior –, o que tende a resultar em inferência “muito otimista” (Hjort e Claeskens 2003). Além disso, uma limitação da abordagem clássica é que seus resultados são derivados da teoria assintótica, o que significa que tanto a amostra quanto o espaço de modelos precisam ser suficientemente grandes para que a inferência seja válida. Modelos omitidos, portanto, implicam em estimações imprecisas dos referidos pesos, o que afeta a distribuição dos parâmetros. Finalmente, deve-se salientar que os resultados obtidos por Sala-i-Martin (1997) dependem crucialmente de sua escolha arbitrária de pesos. Em contrapartida, na abordagem Bayesiana os pesos são derivados da regra de Bayes a partir da probabilidade posterior de cada modelo, após observarmos os dados. Por sua vez, a influência do pesquisador, se existir, é expressa formalmente por meio da escolha de priors.

3. Veja Koop, Leon-Gonzalez e Strachan (2012) e Lenkoski, Eicher e Raftery (2014).

4. A abordagem utilizada para construir esses regressores é similar àquela de Leon-Gonzalez e Montolio (2004), que emprega a abordagem BMA para analisar os determinantes robustos do crescimento econômico nas regiões espanholas.

políticas e o gasto como proporção do PIB são variáveis endógenas. Assim, estudos prévios que empregaram tais variáveis explicativas sem levar em conta o problema da endogeneidade devem ser vistos com cautela.

O artigo está organizado da seguinte forma. A seção 2 apresenta uma breve revisão da literatura. Na seção 3 é apresentada a estratégia econométrica, embora a abordagem BMA seja detalhada na seção 4. A seção 5 descreve o conjunto de variáveis, os critérios de seleção das mesmas e a análise de componentes principais efetuada para comprimir o espaço de modelos e mitigar o problema de multicolinearidade. A seção 6 apresenta os resultados da abordagem BMA. As conclusões são sumarizadas na Seção 7.

2 A Literatura de Crescimento Empírico

A literatura empírica tem buscado identificar os determinantes do crescimento econômico a partir de variáveis sugeridas pelos modelos teóricos. Assim, se empregam variáveis explicativas que são contrapartidas empíricas das variáveis presentes nos modelos teóricos. Por exemplo, modelos neoclássicos *à la* Solow (1956) sugerem variáveis como a renda inicial, o investimento privado, o crescimento populacional e o progresso tecnológico. No entanto, em conformidade com Mankiw, Romer e Weil (1992), também são empregadas variáveis relacionadas ao nível de eficiência dos trabalhadores, o que inclui fatores idiossincráticos das regiões sob estudo relacionados às dotações, ao clima e às instituições.

Uma vez definido o conjunto de variáveis explicativas, é utilizada a regressão da taxa de crescimento da renda contra esses potenciais determinantes, como segue:⁵

$$\mathbf{g} = \alpha + \beta' \mathbf{x} + \varepsilon \quad (1)$$

em que \mathbf{g} é um vetor com taxas de crescimento das regiões sob estudo, no caso, os estados brasileiros; α é um vetor de constantes; \mathbf{x} é um vetor com variáveis explicativas cujos coeficientes estão contidos no vetor β . Como discutido, há uma dificuldade em determinar quais variáveis devem ser incluídas em \mathbf{x} , já que as teorias de crescimento não são suficientemente explícitas sobre quais variáveis pertencem ao “verdadeiro” modelo (Brock e Durlauf 2001; Sala-i-Martin, Doppelhofer e Miller 2004; Durlauf, Johnson e Temple 2005). Assim, geralmente a lista de variáveis contidas em \mathbf{x} é bastante ampla.

Outra questão importante na especificação do modelo (1) é a existência de um efeito fixo por região.⁶ De fato, um ramo da literatura argumenta que fatores idiossincráticos não observáveis afetam o crescimento econômico dos países e regiões, o que motiva o uso de modelos em painel com efeito fixo (Islam 1995; Caselli, Esquivel e Lefort 1996; Moral-Benito 2012). É importante destacar que a omissão de um efeito fixo relevante torna viesados os estimadores dos coeficientes β , uma vez que tal efeito seria correlacionado com variáveis explicativas empregadas (Moral-Benito (2012)).

Ainda, como discutido por Durlauf (2001), Brock e Durlauf (2001) e Leon-Gonzalez e Montolio (2015), parte das variáveis explicativas comumente usadas são potencialmente endógenas, um problema frequentemente ignorado pela literatura empírica. Como discutido na Seção 3, nossa metodologia econométrica leva em conta este problema de endogeneidade.

Dentre os potenciais determinantes do crescimento econômico, um dos mais investigados no caso brasileiro é o gasto do governo.⁷ Por exemplo, usando técnicas de séries temporais,

5. Veja, por exemplo, os trabalhos seminais de Kormendi e Meguire (1985), Barro (1991) Mankiw, Romer e Weil (1992)

6. Por efeito fixo entenda-se um componente específico de cada país ou região. Assim, não estamos nos referindo ao estimador de efeito fixo.

7. Veja Rocha e Giuberti (2007) para uma breve revisão da literatura nacional.

Rodrigues e Teixeira (2010) investigam qual é o gasto público – federal, estadual ou municipal – mais relacionado à taxa de crescimento econômico do Brasil. Os resultados dos autores indicam que o investimento é o gasto mais relevante e a esfera estadual é a que apresenta maior capacidade de impulsionar o crescimento econômico. Por sua vez, Divino e Silva Junior (2012) analisam uma *cross-section* de municípios brasileiros, investigando o efeito da composição dos gastos públicos – corrente e de capital – sobre o crescimento da renda per capita dos municípios. Os resultados obtidos indicam que municípios com renda menor do que a linha de pobreza têm uma necessidade maior de gastos correntes do que os que têm renda maior do que essa linha.

Outros estudos sobre o impacto dos gastos do governo se baseiam em métodos em painel para analisar os estados brasileiros. Rocha e Giuberti (2007) investigam quais componentes do gasto público tem maior influência sobre o crescimento econômico dos estados brasileiros, concluindo que os gastos de capital estimulam o crescimento econômico enquanto os gastos correntes teriam tal efeito somente até o limite de 61% da despesa orçamentária. Silva e Scatolin (2012) também analisam os efeitos dos gastos públicos sobre o crescimento dos estados brasileiros e concluem que aumentos nos gastos em infraestrutura como proporção do PIB contribuem para o crescimento econômico. Finalmente, Bertussi, Capeluppi e Ellery (2011) investigam a relevância dos gastos públicos em transporte, concluindo que tais gastos tem efeito positivo sobre o desempenho econômico de longo prazo dos estados brasileiros. Este efeito é maior nas regiões Norte, Nordeste e Centro-Oeste do país. É importante destacar que todos esses estudos empregaram metodologias econométricas frequentistas e apenas Silva e Scatolin (2012) utilizam um estimador de variáveis instrumentais (GMM).

3 Espaço de Modelos em Regressões de Crescimento em Painel sob Endogeneidade

De modo geral, adaptando-se ao caso de painel e admitindo-se que alguns dos regressores são potencialmente endógenos, a equação (1) pode ser escrita como

$$g_{i,t} = \gamma' \mathbf{y}_{i,t} + \beta' \mathbf{x}_{i,t} + \zeta_t + \eta_i + u_{i,t} \quad (2)$$

em que $g_{i,t} : (1 \times 1)$ é a taxa de crescimento da região ou país i no período t , $\mathbf{x}_{i,t} : (k_{1,j} \times 1)$ são regressores exógenos no período t , $\mathbf{y}_{i,t} : (m \times 1)$ é um conjunto de variáveis potencialmente endógenas, γ e β são vetores de coeficientes compatíveis, $\zeta_t : (1 \times 1)$ é um efeito comum temporal – modelado por meio de *dummies* de anos –, $\eta_i : (1 \times 1)$ diz respeito à heterogeneidade individual invariante no tempo da região i , com $i = 1, \dots, N$. Por sua vez, j varia de 1 a J , sendo J o número total de modelos estimados. Assim, j indica um modelo específico dentro do espaço de modelos \mathcal{M} tal que $k_{1,j}$ denota a possibilidade de o número de regressores exógenos variar entre os modelos estimados.

Assumimos a possibilidade de termos um subconjunto de regressores endógenos. Como a principal fonte de endogeneidade no presente contexto é a simultaneidade, em conformidade com Leon-Gonzalez e Montolio (2015), representamos a equação (2) por um modelo de equações simultâneas (SEM) com dinâmica em uma estrutura de painel.

Antes de considerar a representação SEM e suas propriedades econométricas, discutimos problemas inerentes aos modelos em painel. Neste contexto, é sabido que deve-se acomodar o chamado efeito fixo, especialmente no caso deste trabalho devido às especificidades de cada estado brasileiro. Como este efeito fixo é potencialmente correlacionado com alguns dos regressores no modelo (2), sua omissão acarreta estimativas viesadas. Leon-Gonzalez e Montolio (2015) mostram que aplicar a transformação *forward orthogonal deviation* (FOD) ao modelo (2) equivale a integrar para fora da densidade posterior o efeito fixo, quando assumida uma priori não

informativa no sentido bayesiano. A transformação FOD é preferida à primeira-diferenciação dado que ela não introduz correlação serial no termo de erro $u_{i,t}$. Finalmente, a transformação FOD é baseada na seguinte fórmula:

$$u_{i,t}^* = \sqrt{\frac{T-t}{T-t+1}} \left[u_{i,t} - \frac{1}{T-t} \sum_{t+1}^T u_{i,j} \right] \quad (3)$$

Assim, sob a transformação FOD, o modelo (2) pode ser reescrito como segue:

$$g_{i,t}^* = \gamma' y_{i,t}^* + \beta' x_{i,t}^* + u_{i,t}^* \quad (4)$$

em que $t = 1, \dots, T-1$, já que uma observação é perdida. Além disso, sob o modelo (4) temos $Var(u_{i,t}) = \sigma^2 I_T$ e, assim, $Var(u_{i,t}^*) = \sigma^2 I_{T-1}$, sendo livre de correlação serial.

No entanto, é importante notar que sob tal formulação, a presença do PIB inicial em y_i repercute-se numa correlação entre $y_{i,t}^*$ e $u_{i,t}^*$. Para lidar com tal endogeneidade, em conformidade com Leon-Gonzalez e Montolio (2015) e Leon-Gonzalez e Vinayagathan (2015), utilizamos uma abordagem bayesiana análoga ao estimador 2SLS. Assumindo-se uma equação auxiliar para $y_{i,t}^*$, a representação na forma SEM para (4) é dada por:

$$\begin{aligned} g_{i,t}^* &= \gamma' y_{i,t}^* + \beta' x_{i,t}^* + u_{i,t}^* \\ y_{i,t}^* &= \Pi_x x_{i,t}^* + \Pi_z z_{i,t} + \varepsilon_{i,t}^* \end{aligned} \quad (5)$$

em que $z_{i,t} : (k_{2,j} \times 1)$ é um vetor de instrumentos predeterminados, $u_{i,t}^*$ e $\varepsilon_{j,s}^*$ são normalmente distribuídos com $E(u_{i,t}^*) = E(\varepsilon_{j,s}^*) = 0$ e $E(u_{i,t}^*, \varepsilon_{j,s}^*) = 0, \forall i, j, t, s$. Tanto $x_{i,t}$ como $z_{i,t}$ são tratados como exógenos, implicando

$$E \left[x_{i,t}^* \begin{pmatrix} u_{i,t}^* \\ \varepsilon_{i,t}^* \end{pmatrix}' \right] = 0 \quad \text{e} \quad E \left[z_{i,t} \begin{pmatrix} u_{i,t}^* \\ \varepsilon_{i,t}^* \end{pmatrix}' \right] = 0$$

Por sua vez, $y_{i,t}^*$ é instrumentalizado utilizando-se o nível e defasagens de $y_{i,t}$, que são variáveis predeterminadas. Já que não há uma clara indicação quanto a escolha do número de defasagens a serem incluídas no conjunto de instrumentos, restringe-se a uma única defasagem em consonância com a recente literatura.⁸ De modo similar à Leon-Gonzalez e Montolio (2015), a aplicação empírica deste trabalho permite que as dimensões dos parâmetros ($\gamma, \beta, \Pi_x, \Pi_z$) possam divergir no espaço de modelos⁹, que inclui todos modelos nos quais a restrição de exata identificação ou de sobre-identificação ($k_{2,j} \geq m$) é satisfeita. Assim, sob tal configuração, o espaço de modelos neste trabalho inclui modelos que se diferenciam nos seguintes aspectos:

1. Conjunto de instrumentos: há incerteza quanto a quais variáveis em $z_i \in Z^*$, que é o conjunto de potenciais instrumentos, deveriam ser incluídas no modelo. Z^* é formado por dois subconjuntos disjuntos, $Z^* = Z_1^* \cup Z_2^*$, em que Z_1^* são variáveis que podem ser incluídas no modelo tanto em x_i quanto em z_i . Neste trabalho, a única variável em Z_1^* é a taxa de crescimento da população e as variáveis em Z_2^* são as defasagens de todos os potenciais instrumentos;
2. Variáveis em x_i : incerteza neste nível significa incerteza sobre os parâmetros do vetor β , o que se traduz em incerteza sobre os determinantes do crescimento econômico. Note que x_i é um subconjunto de $Z_1^* \cup X^*$, em que X^* é o conjunto de todos os potenciais regressores que não são admitidos como instrumentos;

8. Roodman (2009) mostra que elevar o número de defasagens pode acarretar um viés positivo nas estimativas. No contexto de BMA, Leon-Gonzalez e Montolio (2015) evidenciam que as defasagens mais próximas tendem a ter probabilidades posterior mais elevadas.

9. Para simplificar a notação, suprimiu-se o índice j .

3. Exogeneidade: em alguns casos a covariância entre u_i e ε_i pode ser zero, implicando que o correspondente regressor em y_i é exógeno;
4. Restrições de zero no vetor γ : um coeficiente zero em γ implica que a variável correspondente não possui impacto em g_i .

Como mostram Koop, Leon-Gonzalez e Strachan (2012), o espaço de modelos neste contexto consiste de $2^{2^m} N^A$ modelos, em que N^A é dado por:

$$N^A = \sum_{j=m}^{k_2^T} 2^{k_1^T + k_2^T - j} C_j^{k_2^T}$$

em que k_1^T é o número total de potenciais regressores em X^* , k_2^T é o número total de potenciais instrumentos e $C_j^{k_2^T}$ é um número combinatório com parâmetros j e k_2^T .

4 Ponderação Bayesiana de Modelos em Painel

Seja $p(M_j|Y)$ a probabilidade posterior do modelo M_j , $j = 1, \dots, J$. Esta probabilidade nos dá informação sobre a plausibilidade de cada modelo e, pela teoria de Bayes, é definida por:

$$p(M_j|Y) = \frac{f(Y|M_j)\pi(M_j)}{\sum_{r=1}^J f(Y|M_r)\pi(M_r)} \quad (6)$$

em que Y representa os dados observados, $f(Y|M_j)$ é a verossimilhança marginal do modelo M_j e $\pi(M_j)$ é a probabilidade *a priori* de que M_j seja o verdadeiro modelo. Ainda, a verossimilhança marginal é dada por:

$$f(Y|M_j) = \int f(Y|\theta^j, M_j)\pi(\theta^j|M_j)d\theta^j \quad (7)$$

em que θ^j é um vetor de parâmetros desconhecidos associados ao modelo M_j , cuja *prior* é dada por $\pi(\theta^j|M_j)$. Finalmente, $f(Y|\theta^j, M_j)$ é a verossimilhança de M_j .

Se soubéssemos *a priori* que a probabilidade posterior de um certo modelo é igual um, não haveria incerteza e bastaria estimar tal modelo. No entanto, como há incerteza sobre os modelos – não se espera que $p(M_j|Y)$ seja igual a 1 para qualquer j –, uma possível abordagem seria justamente calcular uma média ponderada de todos os modelos em que os pesos são proporcionais à probabilidade posterior de cada modelo. Esta abordagem é conhecida como BMA e, de fato, do ponto de vista da literatura de crescimento econômico, existe um grande número de potenciais modelos, cada um com alguma probabilidade de ser o “verdadeiro” (Moral-Benito 2012).

Assim, na abordagem BMA, a inferência sobre o vetor de coeficientes θ é baseada na densidade posterior dos modelos, levando-se em conta todos os modelos com probabilidade positiva. A distribuição posterior para θ é dada por:

$$f(\theta|Y) = \sum_{j=1}^J f(\theta^j|M_j, Y)p(M_j|Y) \quad (8)$$

Neste contexto, a média posterior de θ é tão somente a média ponderada da média posterior de cada modelo, ponderada pela probabilidade posterior de cada modelo, como segue:

$$E(\theta|Y) = \sum_{j=1}^J E(\theta^j|Y, M_j)p(M_j|Y) \quad (9)$$

De modo similar, a variância de θ é computada por meio da fórmula:

$$Var(\theta|Y) = \sum_{j=1}^J Var(\theta^j|Y, M_j)p(M_j|Y) + \sum_{j=1}^J (E(\theta^j|Y, M_j) - E(\theta|Y))^2 p(M_j|Y) \quad (10)$$

Fica evidente que a incerteza sobre θ é acomodada, considerando-se tanto a variância do parâmetro em cada modelo, como a variância do parâmetro entre os modelos.

4.1 Estimação e Priors

Na abordagem bayesiana deve-se escolher priors para os modelos ($p(M_j)$) e para os parâmetros ($p(\theta|M_j)$). Com relação aos modelos, assumimos uma priori uniforme dada por:

$$p(M_1) = \dots = p(M_J) = \frac{1}{J}$$

Portanto, *a priori* cada modelo é considerado igualmente provável. Priors similares são assumidas para os instrumentos e as restrições de exogeneidade. As priors para os parâmetros seguem a configuração proposta por Leon-Gonzalez e Montolio (2015). Intuitivamente, os parâmetros na forma SEM são normalmente distribuídos, com variância ponderada pelo número de estados e tendo distribuição Gama Inversa.

Há dois grandes desafios computacionais para a abordagem BMA. Um diz respeito à verossimilhança marginal (7). Como não é possível resolver aquela integral de modo analítico, é necessário o uso intenso de métodos numéricos. O segundo desafio se deve ao número extremamente elevado de modelos. Para lidar com tais problemas, adota-se um algoritmo baseado em Cadeias de Markov com Saltos Reversíveis (RJMCMC) desenvolvido por Koop, Leon-Gonzalez e Strachan (2012) e Leon-Gonzalez e Montolio (2015). Tal algoritmo é capaz de amostrar valores para os parâmetros θ^j e o modelo M_j de maneira iterativa. Dados valores iniciais arbitrários para (θ^j, M_j) , após um número suficiente de iterações, os valores gerados podem ser usados para se construir uma distribuição empírica da posterior de (θ^j, M_j) . Então, esta distribuição é usada para se calcular quantidades de interesse, como a probabilidade posterior do modelo e intervalos de confiabilidade. A convergência deste algoritmo é verificada por meio da *Total Visited Probability* (TVP), que é uma estimativa da proporção de massa total de probabilidade que é visitada pelo algoritmo.¹⁰

Finalmente, ressaltamos que alguns trabalhos empíricos recentes tem buscado reduzir o número de potenciais regressores. As razões são: (i) reduzir os custos computacionais; (ii) diminuir o risco de problemas severos de multicolinearidade; e (iii) reduzir o problema causado pela inclusão de mais de uma *proxy* para um mesmo fator (veja, por exemplo Ciccone e Jarocinski 2010; Moral-Benito 2012). Tais reduções são baseadas em outros trabalhos empíricos, que são utilizados como informação *a priori* para a exclusão de certas variáveis. Entretanto, devido à escassez de trabalhos sobre o caso brasileiro, decidiu-se incluir todas as potenciais variáveis explicativas disponíveis, testando-se a sensibilidade dos resultados obtidos numa segunda etapa.

5 Base de Dados

A despeito da literatura ter documentado a existência de uma lista enorme de potenciais regressores em análises *cross-countries*, para casos de regiões de um dado país, o número de variáveis disponível é naturalmente reduzido, já que muitas tornam-se inapropriadas ou não estão disponíveis.¹¹ Esta diferença torna-se um problema especialmente para o caso dos estados

10. Ver Leon-Gonzalez e Montolio (2015) para maiores detalhes sobre o TVP.

11. Ver Durlauf, Johnson e Temple (2005) para uma extensa revisão da literatura de crescimento. Neste artigo, os autores identificam 43 teorias distintas de crescimento.

brasileiros no qual, por exemplo, não há *proxies* apropriadas para a qualidade das instituições, as imperfeições de mercado tais como poder de mercado das firmas e preço do investimento. Entretanto, uma vez que utilizamos dados em painel, podemos levar em conta alguns destes problemas acomodando o efeito fixo do estado. Contudo, isso nos requer restringir os potenciais regressores àqueles para os quais há variação temporal no período sob análise.

Mesmo diante de tais empecilhos, tomamos o modelo de Solow aumentado como referência inicial para a seleção de variáveis. Quando possível, consideramos também variáveis derivadas de modelos de crescimento endógeno, como medidas referentes à saúde da população. Vale ressaltar que, em alguns aspectos, é possível aprofundar ainda mais a investigação comparativamente à literatura *cross-country* em virtude da disponibilidade de dados desagregados para gasto do governo. Isso nos permite analisar se os diferentes tipos de gastos públicos afetam o crescimento econômico dos estados brasileiros, como tem sido estudado na literatura nacional (Rodrigues e Teixeira 2010; Divino e Silva Junior 2012; Rocha e Giuberti 2007; Silva e Scatolin 2012; Bertussi, Capeluppi e Ellery 2011).

No contexto de painel, uma prática comum na análise dos determinantes das taxas de crescimento de longo prazo é empregar médias das variáveis a cada cinco ou dez anos, com vistas a suavizar os efeitos dos ciclos econômicos. Como as observações sobrepostas não são utilizadas para evitar problemas de autocorrelação dos resíduos, essa prática acarreta uma grande redução no número de observações. Para maximizar o tamanho da amostra, adotamos os procedimentos de Leon-Gonzalez e Montolio (2004) e Leon-Gonzalez e Vinayagathan (2015), considerando como variável dependente o crescimento médio bianual e ignorando as observações justapostas. Tal análise é dita referir-se ao crescimento econômico de médio prazo. Além disso, como em Moral-Benito (2012) e Caselli, Esquivel e Lefort (1996), tratamos de forma distinta variáveis de fluxo e de estoque. Para as primeiras, computamos médias referentes aos três anos referentes à cada taxa de crescimento bianual, o que resulta em variáveis potencialmente endógenas devido à possível simultaneidade. Para as últimas, consideramos as variáveis explicativas observadas no início do período, ou seja, em $t - 2$, o que nos permite tratá-las como predeterminadas.

Finalmente, com esta configuração, foi possível construir uma amostra balanceada para o período de 1992 a 2016, compreendendo às 27 unidades federativas brasileira. O esquema utilizado acomoda a ausência de dados em alguns anos para variáveis computadas a partir da Pesquisa Nacional por Amostra de Domicílios (PNAD) – cuja coleta de dados não acontece nos anos de Censo – uma vez que tais observações seriam ignoradas pelo critério de seleção de dados discutidos anteriormente. Assim, tem-se um total de 246 observações no painel.

Como mencionado, algumas variáveis usualmente consideradas relevante para o crescimento econômico não estão disponíveis para os estados brasileiros, particularmente a taxa de investimento privado. Por esta razão, construímos uma variável *proxy* que argumentamos captar o comportamento idiossincrático do investimento privado de cada estado. No que se segue, além de descrever como procedemos tal estimação, descrevemos sucintamente as categorias de variáveis selecionadas neste trabalho. Os detalhes sobre a descrição e a origem de cada variável encontram-se no Apêndice A.

5.1 Ambiente Macroeconômico

Muitos autores têm destacado a importância de um ambiente macroeconômico estável para o crescimento econômico, o que envolve taxas de inflação baixas e previsíveis, endividamentos sustentáveis dos governos e taxas de câmbio estáveis relativamente ao seu equilíbrio de longo prazo. De fato, o comprometimento e a credibilidade do governo são fatores que dissipam a incerteza macroeconômica e levam o setor privado a realizar investimentos. Neste sentido, haveria uma relação positiva entre um “bom” ambiente macroeconômico e o investimento.

Há três variáveis comumente utilizadas para capturar as condições do ambiente macroeconômico: inflação ao consumidor, gasto do governo como fração do PIB e grau de abertura

da economia, medido como a soma das exportações e importações como proporção do PIB. No entanto, dado o elevado número de variáveis na base de dados – e, mais importante, o considerável número de possíveis variáveis endógenas –, procedemos como Burnside e Dollar (2000) e estimamos um único Índice de Políticas. Este artifício ainda reduz o problema de multicolinearidade e os custos computacionais associados à incerteza quanto à endogeneidade.

Para construir os índices de políticas estaduais, utilizamos dados do período de 1990 a 2016. O índice é computado como o fator comum das seguintes variáveis: deflator do PIB estadual dividido pelo deflator do PIB nacional – *proxy* para a distorções nos preços relativos estaduais –, a proporção do gasto do governo com relação ao PIB estadual e o grau de abertura estadual.

Os fatores comuns foram estimados utilizando o método bayesiano proposto por Lopes e West (2004), sob um esquema de priors pouco informativas.¹² Por brevidade e dado o principal interesse deste trabalho, não detalhamos questões técnicas, mas indicamos o artigo citado ao leitor interessado.

Reportamos que, de modo similar aos resultados de Burnside e Dollar (2000) e Leon-Gonzalez e Montolio (2015), os índices de política dos estados brasileiros são, em geral, positivamente relacionado com o grau de abertura e negativamente relacionados tanto com o gasto do governo quanto com a taxa de inflação.

5.2 Investimento

A taxa de investimento é outra variável que não está disponível para os estados brasileiros. Como *proxy* para esta variável, empregou-se o fator comum dinâmico não observável de três séries: a taxa de investimento da economia brasileira como um todo – interpretada como uma média ponderada das taxas de investimentos dos estados brasileiros –, o valor adicionado da indústria de construção civil estadual e a taxa de crescimento do consumo elétrico residencial estadual. Assim, assumimos que as características específicas de cada estado são captadas pelas variações no valor adicionado da construção civil e no crescimento do consumo elétrico, uma vez que tais variáveis estão associadas ao investimento e ao consumo de bens duráveis. Como no caso do Índice de Política, os fatores comuns estaduais são estimados utilizando o método proposto por Lopes e West (2004).

Adicionalmente, consideramos também a despesa do governo em capital como medida de investimento público. Assim, nossa análise empírica conta com duas medidas de investimento: a *proxy* de investimento privado e o investimento público. Isso nos permite investigar se ambos os tipos de investimentos são determinantes robustos do crescimento econômico nos estados brasileiros.

5.3 Capital Humano

Desde Mankiw, Romer e Weil (1992) várias medidas de capital humano tem sido utilizadas na literatura. Neste trabalho, consideramos algumas das mais comuns, a saber: a fração da população com idade acima de 10 anos sem nenhum ano de estudo (H_1); a fração da população acima de 10 anos com ao menos 8 anos de estudo (H_2); a fração da população acima de 10 anos com ao menos 11 anos de estudo (H_3). Note que H_2 capta a fração da população em idade ativa com educação primária, enquanto H_3 refere-se a educação secundária.

Entretanto, como observa Leon-Gonzalez e Montolio (2004), não há consenso quanto ao uso de tais *proxies* para o capital humano. Por esta razão, segue-se a sugestão dos autores na construção

12. Para ilustração, a estrutura básica utilizada para todos os estados é $h_t = \lambda f_t + \eta_t$, em que f_t é o fator comum não observável estimado pelo método de Lopes e West (2004), λ é a carga do fator e h_t um vetor de variáveis observáveis.

de um indicador adicional *à la* Jones (1997), dado por:

$$H_{i,t} = e^{\gamma_i S_{i,t}} L_{i,t}$$

em que t é o período, i indica o estado, γ é a taxa Minceriana invariante no tempo de retorno de um ano de educação (Mincer 1974), S é a média de anos de estudo na população em idade ativa e L é a força de trabalho. Portanto, H_i é uma medida do estoque de capital humano.

De maneira similar à Leon-Gonzalez e Montolio (2004), a média $S_{i,t}$ é calculada da seguinte forma:

$$S_i = \sum n_e \frac{\omega_{i,e}}{L_i}$$

em que n_e é o número de anos de estudo e $\frac{\omega_{i,e}}{L_i}$ é a fração de trabalhadores com n_e anos de estudo relativamente à população em idade ativa no estado i . Para computar S_i utilizou-se dados da PNAD, que contém informação sobre o número de anos de estudo da população, porém truncada em 15 anos. Assim, $n_e \in \{0, \dots, 15\}$, sendo que o limite superior agrega pessoas com 15 ou mais anos de estudo.¹³ Por sua vez, o parâmetro γ foi calibrado com base nas estimativas referentes a 8 anos de estudo para os estados brasileiros de Dias et al. (2013).

5.4 Fatores Setoriais

Consideram-se, também, medidas da estrutura setorial dos estados, a saber o valor adicionado bruto (VAB) da indústria como proporção do PIB estadual e o VAB da agricultura. O VAB de serviços foi omitido para se evitar multicolinearidade perfeita. A inclusão dessas medidas setoriais é justificada pela possibilidade de que alguns setores crescerem mais rapidamente do que outros devido à diferenças em termos de progresso tecnológico ou políticas governamentais setoriais.

5.5 Gastos do Governo Desagregados

Ao considerar variáveis do governo de modo desagregado, podemos analisar se as políticas tem impactos distintos sobre o crescimento econômico. Assim, além do investimento público e do gasto corrente do governo, considerou-se vários gastos desagregados. Todas as variáveis foram selecionadas tendo em vista o objetivo de construir um painel balanceado, o que limitou a extensão da análise. De todo modo, foram construídas as seguintes categorias de gastos: *i*) assistência e previdência, *ii*) saúde e saneamento, *iii*) segurança nacional e defesa pública, *iv*) transportes e *v*) educação.¹⁴ Finalmente, todas essas medidas de gastos foram calculadas como proporção do gasto total corrente, seguindo a literatura nacional (Rocha e Giuberti 2007).

5.6 Fatores Demográficos

Levamos em conta fatores demográficos motivados tanto pelos modelos neoclássicos quanto pelos modelos de crescimento endógeno, inclusive modelos *à la* Nelson-Phelps. Por exemplo, o modelo neoclássico sugere que o crescimento populacional seja negativamente relacionado com o nível de estado estacionário do produto por trabalhador efetivo.

A respeito da população, considera-se o crescimento populacional, a taxa de urbanização e a taxa de participação. Esta taxa é definida como a proporção da população em idade ativa que se encontra economicamente ativa. A teoria do crescimento endógeno também sugere a inclusão da

13. De fato, para alguns anos, a truncagem ocorreu em 12 anos.

14. Medidas de gastos que se mostraram relevantes em outros estudos, como gastos com habitação, urbanismo e transportes possuíam muitos zeros, possivelmente associados à *missings*, razão pela qual foram excluídas da análise.

taxa de participação, devido a possíveis efeitos de escala associados à economias com grandes populações (ver, por exemplo, Barro e Lee 1994; Sala-i-Martin 1997).

Consideramos uma possível relação empírica entre crescimento econômico e pobreza, captada pelas variáveis proporção de pobres e extremamente pobres na população e coeficiente de Gini. Além disso, levamos em conta a fração de brancos na população em cada estado. Finalmente, consideramos a população economicamente ativa nos estados.

5.7 Fatores Políticos

Para acomodar fatores políticos, incluímos uma *dummy* que assume valor unitário no ano t se na última eleição o partido político do governador compunha a coligação de partidos que declaravam apoio ao presidente em exercício, e zero caso contrário.

5.8 Geografia

A literatura de crescimento empírico sugere uma lista de variáveis que relacionam aspectos geográficos ao crescimento econômico. Parte dessas variáveis não apresentam variação temporal, o que inviabiliza a inclusão das mesmas em nosso modelo, uma vez que empregamos métodos em painel com efeito fixo. Consequentemente, isso limita ainda mais à reduzida disponibilidade de dados com a qual nos defrontamos.

Para levar em conta as diferenças geográficas nas regiões brasileiras, constrói-se uma variável que indica a fração da área total do estado (em Km^2) destinada à agricultura. Isso é motivado pela possibilidade de efeitos de escala, que podem ser relevantes para o crescimento econômico. Esta medida foi construída somando-se a fração das terras destinadas à produção agrícola permanente e temporária. Assim, esta é uma variável que não é constante ao longo do tempo, o que permite sua inclusão no nosso modelo em painel.

5.9 Saúde e Padrão de Vida

Por fim, consideramos a possibilidade de que o crescimento econômico seja correlacionado com às condições de saúde da população. Os argumentos para a inclusão de variáveis relacionadas à saúde devem-se à chamada abordagem “Nelson-Phelps”, na qual uma elevação no nível de saúde acarreta um aumento no nível de produto.

Aghion, Howitt e Murtin (2011) enfatizam que indivíduos mais saudáveis tendem a ser mais produtivos, mais criativos e mais capazes de se adaptarem à novas tecnologias e à mudanças nas características associadas a um ambiente de grande crescimento econômico. Por isso, consideramos como *proxies* da saúde da população em cada estado a taxa de mortalidade por diarreia aguda e a mortalidade por aids. Consideramos, também, as taxas de mortalidade e fecundidade, já que tais variáveis estão associadas ao nível e ao crescimento populacional, e são comumente relacionadas ao crescimento econômico. Ainda, incluímos indicadores relacionados à qualidade de vida nos estados como acesso à água potável e à rede de esgoto e a fração das residências com material resistente.

Finalmente, consideramos o logaritmo da expectativa de vida em cada estado. Conforme discutido em Aghion, Howitt e Murtin (2011), indivíduos com maior expectativa de vida tendem a poupar mais, implicando maior acumulação de capital e, em última instância, maior taxa de crescimento econômico. Além do mais, tais indivíduos comumente investem mais em educação, o que tende a impactar positivamente o crescimento econômico.

5.10 Comprimindo o Espaço de Modelos

Há duas importantes razões para reduzirmos o espaço de modelos possíveis deste trabalho. A primeira é o elevado custo computacional. Como acomodamos incertezas sobre o modelo, os parâmetros e os instrumentos, o espaço de modelos é demasiadamente grande e altamente custoso para ser considerado. A segunda é que muitas das variáveis selecionadas podem conter informações semelhantes, o que acarretaria multicolinearidade severa.

De fato, recentemente, alguns trabalhos na literatura que utilizam ponderação de modelos tem sugerido que as estimativas da probabilidade posterior de inclusão são sensíveis ao uso de mais de uma *proxy* para captar um mesmo efeito ou por erro de medida (Doppelhofer e Weeks 2011; Ciccone e Jarocinski 2010; Moral-Benito 2012). Em outras palavras, variáveis adicionais em um certo subgrupo podem conter pouca informação adicional, podendo induzir à multicolinearidade e instabilidade.

Para evitar esses problemas, a literatura tem escolhido arbitrariamente uma espécie de “*proxy* representativa”, excluindo as variáveis similares do procedimento econométrico. Entretanto, esta estratégia se defronta com o seguinte problema: o econometrista não sabe, a priori, qual variável representa a contra-partida amostral de certa teoria. Por exemplo, não é claro qual seria, em nossa amostra, a variável que melhor captaria o capital humano no agregado. O mesmo raciocínio vale para a abordagem “Nelson-Phelps” para o fator saúde. Assim, ao invés de escolher arbitrariamente certas *proxies* dentro um subgrupo de potenciais determinantes do crescimento que deverão ser considerados no BMA, optamos por combinar indicadores similares de modo a maximizar a informação disponível na amostra, ao mesmo tempo em que evitamos a multicolinearidade. Para tanto, utilizamos componentes comuns.¹⁵

Para que os componentes comuns possam ser estimados adequadamente, procedemos do seguinte modo. Primeiramente, para cada estado, utilizamos todas as observações entre 1991 e 2016, o que inclui os anos que são excluídos devido a sobreposição ao estimarmos o modelo (2).¹⁶ Todas as variáveis são padronizadas para terem média zero e variância unitária. Em seguida, estimamos: (i) um componente principal para capital humano, CP_H , com base em $-H_1$, H_2 , H_3 e S_i ; (ii) dois componentes principais ortogonais para demografia, CP_D^1 e CP_D^2 , a partir das variáveis número de brancos, gini, PEA, PIA, fração de população rural, crescimento populacional, taxa de participação; (iii) dois componentes principais ortogonais para saúde e padrão de vida, CP_L^1 e CP_L^2 , baseados nas variáveis domicílio com água potável, domicílio com material durável, expectativa de vida, fecundidade, mortalidade e mortalidade por diarreia. Em geral, os componentes estaduais explicam aproximadamente 90% da variação das variáveis consideradas, indicando que tais variáveis componentes comuns muito expressivos. Assim, a compressão das variáveis em um conjunto menor de dados é importante em nosso procedimento econométrico devido ao potencial problema de multicolinearidade, uma vez que tal problema inflacionaria a variância dos parâmetros estimados.

Evidentemente, há um custo em se utilizar fatores comuns em nosso contexto, quer seja, a dificuldade de interpretação e o escopo para política pública. Entretanto, reportamos que os componentes comuns possuem claro padrão que, se não nos permite fazer análises marginais sobre certas variáveis selecionadas, ainda assim nos permitem inferir sobre impactos causais de certa *categoria* de variáveis sobre o crescimento econômico.

Por fim, convém discutir o critério que utilizamos para decidir se certa variável é classificada como determinante robusta do crescimento. A regra comumente utilizada na literatura de ponderação de modelos é observar se a probabilidade posterior de inclusão de certa variável é maior do que o tamanho esperado do modelo dividido pelo número de potenciais regressores

15. Diferentemente dos indicadores de política e investimento, no presente caso utilizamos componentes principais simples, uma vez que as variáveis aqui são, grosso modo, estoque e variam pouco de ano para ano.

16. Variáveis com *missing data* são interpoladas utilizando o método Pchip (*Piecewise Cubic Hermite Interpolating Polynomial*).

(Sala-i-Martin, Doppelhofer e Miller 2004; Moral-Benito 2012; Doppelhofer e Weeks 2011). Na prática, as escolhas comuns consideram que variáveis com probabilidade posterior de inclusão abaixo de 0.10 não são tratadas como determinantes robustas. Seguimos tal valor como “regra-de-bolo”, complementando-o com a exigência de que variáveis com probabilidade posterior de inclusão maior do que 0.10 não tenham parâmetros estimados distribuídos ao redor do zero.

6 Resultados

Pelo fato de nosso procedimento econométrico ser flexível, possibilitando que algumas variáveis sejam consideradas a priori endógenas, o primeiro é definir justamente quais seriam essas variáveis. Assim, tratamos como potencialmente endógenas o PIB inicial, os investimentos público e privado, e o valor adicionada da indústria e da agropecuária. Demais variáveis são assumidas fracamente exógenas – lembrando que as variáveis estoque são tratadas como no início de período, implicando na exogeneidade fraca. As variáveis de despesa pública são potencialmente endógenas. Contudo, testes preliminares nos levaram a concluir que tais variáveis tem baixa probabilidade de serem endógenas, indicando que a política fiscal estadual tende a responder lentamente aos ciclos de negócio.

Procedemos 800.000 iterações baseadas no algoritmo RJMCMC, tendo descartado as 200.000 iniciais como *burn-in*. Para atestar a convergência, seguindo Leon-Gonzalez e Montolio (2015) e Leon-Gonzalez e Vinayagathan (2015), computamos a TVP. O resultado obtido, 99%, sugere que todo o espaço de modelos é visitado com frequência, o que indica boa convergência das cadeiras markovianas. Ademais, repetimos a análise com diferentes escolhas para os valores iniciais e priors, e obtivemos resultados semelhantes.

Tabela 1: Resultados do BMA

VARIÁVEL	Prob. Inc.	2.5%	50%	97.5%	Prob. Endóg.
DETERMINANTES ROBUSTOS					
PIB Inicial	1.00000	-0.71171	-0.48302	-0.32170	1.00000
CP_L^1	0.99919	0.00012	0.00710	0.02345	-
CP_H	0.97479	0.00001	0.02020	0.03847	-
Índice de Investimento	0.71138	0.00001	0.00948	0.02189	0.64221
PIB Agropecuário	0.65696	-0.01941	-0.01016	0.00000	0.65502
Despesa com Saúde	0.59723	0.00000	0.25879	0.57494	-
PIB Indústria	0.39720	0.00000	0.00573	0.00840	0.31971
NÃO ROBUSTOS					
Investimento Público	0.16825	-0.06786	0.00000	0.06307	0.00079
Despesa com Educação	0.06657	-0.17600	-0.05200	0.00012	-
Índice de Políticas	0.04038	0.00000	0.00000	0.00061	0.88032
Despesa Assistencial	0.01194	0.00000	0.00000	0.00000	-
CP_D^1	0.00059	0.00000	0.00000	0.00000	-
Despesa com Segurança	0.00000	0.00000	0.00000	0.00000	-
Área Plantada	0.00000	0.00000	0.00000	0.00000	-
CP_D^2	0.00000	0.00000	0.00000	0.00000	-
CP_L^2	0.00000	0.00000	0.00000	0.00000	-
Coligações	0.00000	0.00000	0.00000	0.00000	-

Notas: Prob. Incl.: Probabilidade de Inclusão. Prob. Endóg.: Probabilidade de ser Endógena.
Número de obs.: 189

A tabela 1 apresenta os resultados por meio da estrutura em painel e da aplicação da transformação FOD – conforme equação (3)–, sendo reportados a probabilidade posterior de

inclusão e o intervalo de credibilidade de cada variável explicativa, além da probabilidade das variáveis potencialmente endógenas serem endógenas. Os resultados estão ordenados da maior para a menor probabilidade de inclusão da variável.

Após acomodar a incerteza sobre o modelo econométrico, encontramos que fatores distintos estão relacionados ao crescimento econômico nos estados brasileiros. A variável com maior probabilidade de inclusão ($=1$) é o PIB Inicial, cujo quinquagésimo percentil da posterior foi calculado ao redor de -0.48 , sugerindo fortemente a existência de convergência condicional dos estados após acomodar tanto o efeito fixo quanto a incerteza sobre os modelos.¹⁷ O segundo determinante robusto do crescimento é o componente comum CP_L^1 – relacionado às condições de saúde e ao padrão de vida –, cuja probabilidade de inclusão foi estimada em 0.999 . Tal fator é positivamente correlacionado com o número de domicílios com água potável e esgoto, expectativa de vida e fecundidade e negativamente correlacionado à mortalidade e mortalidade por diarreia. Assim, os resultados indicam que melhora nestes indicadores tendem a elevar o crescimento no médio prazo. O seguinte determinante robusto é o componente comum do capital humano, CP_H , cuja a probabilidade de inclusão é de 97% . Como este componente comum foi construído de modo a representar melhoras no capital humano e o seu coeficiente está concentrado em valores positivos, os resultados indicam que aumentos no capital humano tendem a resultar em maior crescimento econômico no médio prazo.

O quarto determinante robusto é o índice de investimento com probabilidade de inclusão de aproximadamente 71% . O coeficiente desta variável está concentrado em valores positivos, os resultados indicam que melhorias no investimento promovem o crescimento econômico. O quinto determinante robusto do crescimento econômico é o PIB agropecuário, com probabilidade de inclusão de 65% . Encontramos que aumentos relativos do setor agropecuário em detrimento dos outros setores tendem a reduzir o PIB. Finalmente, o último indicador que pode ser considerado robusto é a despesa com saúde, com 59% de probabilidade de inclusão no modelo. O resultado indica que aumentos no orçamento dos serviços públicos de saúde tendem a elevar o crescimento econômico no médio prazo.

Um resultado relevante diz respeito ao papel da indústria no crescimento brasileiro. Embora o coeficiente associado a tal variável seja sempre positivo, sua magnitude é baixa e a probabilidade de inclusão desta variável é menor do que 40% . Isso indica que a indústria cumpre papel marginal em termos do crescimento de médio prazo nos estados brasileiros.

Demais variáveis ou apresentam coeficientes indistinguíveis de zero e/ou tem baixa probabilidade de seleção, não sendo portanto classificadas como determinantes robustos do crescimento. Conseqüentemente, os resultados indicam que o crescimento de médio prazo no Brasil está associado a pelo menos uma variável relacionada a cada um dos seguintes fatores: condições iniciais, padrão e qualidade de vida, educação, investimento privado, composição setorial e despesas em saúde pública.

É importante destacar alguns pontos. Primeiro, em se tratando de composição dos gastos nos estados, apenas a despesa com saúde parece ser relevante para o crescimento econômico. As demais variáveis desta natureza raramente são selecionadas pelo algoritmo bayesiano. Segundo, a probabilidade do PIB industrial e do investimento público serem variáveis endógenas é inferior a 50% . Por outro lado, há evidência de endogeneidade para o PIB inicial, o índice de investimento, o valor adicionado da agropecuária e o índice de políticas. Terceiro, a despeito do crescimento populacional ser visto na literatura como um fator relevante, tal variável não se mostrou relevante em nossa análise. Quarto, merece discussão a não seleção do índice de políticas. Pelo fato de tal fator ter sido construído dentro de um mesmo país – o que seria equivalente a uma espécie de união monetária –, é possível que ele não adequadamente a qualidade das políticas estaduais, mas sim fatores nacionais.

17. A taxa de convergência condicional implicada por tal resultado é $\gamma \approx 0.008$, similar àquelas calculadas tanto em painel como em *cross-sections* para países em Mankiw, Romer e Weil (1992) e Moral-Benito (2012).

Em virtude desses resultados, como exercício de robustez, reestimamos o modelo com as seguintes modificações: (i) excluimos o crescimento populacional dos componentes comuns demográficos e o adicionamos como variável independente e, (ii) retiramos as variáveis de gastos desagregados e o índice de política e incluímos a razão gasto do governo/PIB, exportações líquidas/PIB e inflação relativa, esta última sendo considerada fracamente exógena.

Os resultados deste novo modelo são apresentados na tabela 2 no apêndice. Nota-se, inicialmente, que o desmembramento do índice de políticas parece ter afetado os intervalos de credibilidade das estimativas, o que pode ser devido ao problema de multicolinearidade. De todo modo, qualitativamente, os resultados estão em linha com os reportados na tabela 1. alguns destaques são o grau de abertura e o valor adicionado da indústria. O grau de abertura parece cumprir um papel relevante no crescimento econômico nos estados. Embora sua probabilidade de inclusão não seja superior a 50%, seu coeficiente está concentrado em valores positivos, com magnitude relativamente elevada. Quanto ao valor adicionado da indústria, sua probabilidade de inclusão tornou-se superior a 50%. Ademais, não há evidências à favor do crescimento populacional ser um determinante robusto do crescimento econômico nos estados brasileiros.

Considerando em conjunto os resultados das tabelas 1 e 2 há forte evidência de endogeneidade com respeito ao PIB inicial e ao índice de investimento. Há evidência de endogeneidade para o PIB agropecuário e industrial para o índice de políticas e o gasto como proporção do PIB. Portanto, estudos prévios que não consideram tais variáveis como sendo endógenas devem ser vistos com cautela.

7 Conclusões

Nos anos recentes a literatura tem empenhado grandes esforços na construção de arcabouços empíricos capazes de acomodar às incertezas presentes na determinação dos determinantes robustos do crescimento econômico, já que as bases teóricas para tais análises ainda carecem de consenso. Para lidar com a incerteza sobre o modelo empírico, este trabalho utilizou métodos bayesianos de ponderação de modelos em painel para identificar os determinantes robustos do crescimento econômico nos estados brasileiros de 1992 a 2016.

Após levar em conta as incertezas sobre o modelo e sobre a endogeneidade de certos regressores, além de características idiossincráticas dos estados, os resultados sugerem que vários fatores estão correlacionados com o crescimento econômico nos estados brasileiros, sendo eles o PIB inicial, o padrão e a qualidade de vida, a educação, o investimento privado, a composição setorial e as despesas em saúde pública.

Analisou-se ainda a capacidade do governo de afetar o crescimento econômico considerando-se várias categorias de gastos desagregados, incluindo as despesas de capital aqui interpretadas como o investimento público. Entre todas, a única variável que pode ser considerada determinante robusto do crescimento é o gasto com saúde. Finalmente, como o PIB inicial mostrou-se relevante e com coeficiente negativo, concluímos que há evidência de convergência condicional nos estados brasileiros.

8 Apêndice A

Tabela 2: Resultados do BMA

VARIÁVEL	Prob. Inc.	2.5%	50%	97.5%	Prob. Endóg.
DETERMINANTES ROBUSTOS					
PIB Inicial	0.98668	-0.49065	-0.35075	-0.00001	1.00000
CP_H	0.97709	0.00001	0.01985	0.03954	-
Índice de Investimento	0.92119	0.00001	0.00878	0.02985	0.74221
CP_L^1	0.84538	0.00001	0.00650	0.03905	-
PIB Agropecuário	0.60786	-0.02065	-0.00051	-0.00000	0.20101
PIB Indústria	0.56085	0.00187	0.00555	0.00879	0.92119
Grau de Abertura	0.43011	0.00000	0.08398	0.19282	0.19282
NÃO ROBUSTOS					
Gasto/PIB	0.09455	-0.03167	0.0000	0.00399	0.66261
CP_D^1	0.00000	0.00000	0.00000	0.00000	-
Área Plantada	0.00000	0.00000	0.00000	0.00000	-
CP_D^2	0.00000	0.00000	0.00000	0.00000	-
CP_L^2	0.00000	0.00000	0.00000	0.00000	-
Cresc. Pop.	0.00000	0.00000	0.00000	0.00000	-
Coligações	0.00000	0.00000	0.00000	0.00000	-
Inflação Relativa	0.00000	0.00000	0.00000	0.00000	-

Notas: Prob. Incl.: Probabilidade de Inclusão. Prob. Endóg.: Probabilidade de ser Endógena.
Número de obs.: 189

Referências

- Aghion, Philippe, Peter Howitt e Fabrice Murtin. 2011. “The Relationship Between Health and Growth: When Lucas Meets Nelson-Phelps”. *Review of Economics and Institutions* 2 (1): 1–24.
- Barro, R. J. 1991. “Economic growth in a cross section of countries”. *The quarterly journal of economics* 106 (2): 407–443.
- Barro, Robert J., e Jong-Wha Lee. 1994. “Sources of economic growth”. *Carnegie-Rochester Conference Series on Public Policy* 40 (): 1–46. ISSN: 01672231.
- Bertussi, Geovana, Pedro Capeluppi e Roberto Goes Ellery. 2011. “Gastos públicos com infraestrutura de transporte e crescimento econômico: uma análise para os estados brasileiros”. *Working papers - Textos para Discussão do Departamento de Economia da Universidade de Brasília* 357.
- Brock, William, e Steven N. Durlauf. 2001. “Growth Empirics and Reality”. *World Bank Economic Review* 15 (2): 229–272.
- Burnside, Craig, e David Dollar. 2000. “Aid, policies and growth”. *American Economic Review* 90 (4): 847–868.
- Caselli, Francesco, Gerardo Esquivel e Fernando Lefort. 1996. “Reopening the convergence debate: a new look at cross-country growth empirics”. *Journal of economic growth* 02138.

Tabela 3: Descrição das variáveis e fontes

VARIÁVEL	DESCRIÇÃO	FONTE	
Variável Dependente	Crescimento médio do PIB per capita no período de 2 anos	IBGE	
PIB Inicial	Logaritmo do PIB inicial	IPEADATA e IBGE	
Cresc. da População	Taxa de crescimento populacional	IBGE	*
Despesa Assistencial	Despesas com assistência e previdência	IPEADATA e EOE	*, ***
Despesa com Educação	Despesas educação e cultura	IPEADATA e EOE	*, ***
Despesa com Saúde	Despesas saúde e saneamento	IPEADATA e EOE	*, ***
Despesa com Segurança	Despesas segurança nacional e defesa pública	IPEADATA e EOE	*, ***
Dom. com Água Potável	Domicílios com instalação adequada de água potável (fração)	IPEADATA	
Dom. com Esgoto	Domicílios com instalação adequada de esgoto (fração)	IPEADATA	
Expectativa de Vida	Expectativa de vida ao nascer	DATASUS	
Fecundidade	Taxa de fecundidade	IPEADATA e DATASUS	
Gini	Coefficiente de Gini	IPEADATA	
H1	Proporção de analfabetos (15 anos ou mais)	PNAD	
H2	Fração da pop. com ao menos primeiro grau	PNAD	
H3	Fração da pop. com ao menos segundo grau	PNAD	
Si	Capital humano. Ver seção 5.3	PNAD	
Homicídios	Taxa de homicídios	IBGE	
Índice de Investimento	Ver seção 5.2	Ver tabela 4	*
Índice de Política	Ver seção 5.1	Ver tabela 4	*
Investimento Público	Despesas com capital - investimento	IPEADATA e EOE	*, **
Lavoura	Fração da área estadual destinada à lavoura	IBGE	
Material do Domicílio	Domicílios com material de construção durável (fração)	IPEADATA	
Mortalidade	Taxa de mortalidade	DATASUS	
Óbitos por Diarreia	Incidência de mortes por diarreia aguda (fração)	DATASUS	
Partidos	Dummy: 1 se partido do governo coligado ao da presidência, 0 c.c.	Elaboração dos autores	
PIA	População em idade ativa	IPEADATA	
PEA	População economicamente ativa	IPEADATA	
PIB Agropecuário	Valor adicionado da agropecuária	IBGE	*, **
PIB Indústria	Valor adicionado da indústria	IBGE	*, **
Pobreza extrema	Fração de pessoas extremamente pobres	IPEADATA	
População Branca	Fração de pessoas brancas	IPEADATA	
População Rural	Fração da população vivendo em zona rural	PNAD e Censo	***
Tx. Participação	Relação entre PEA e PIA	IPEADATA	

Notas: Variáveis monetárias denominadas em Reais de 1995; EOE: Execução Orçamentária de Estados.

*: Média no período; **: Proporção do PIB; ***: Proporção do Gasto Corrente;

**** Regiões exceto a Norte: PNAD. Região Norte: PNAD e Censo combinadas para

calcular a taxa de crescimento e retopolar as observações ausentes na PNAD.

Critérios para construção das variáveis das a partir das contas regionais: 1990-1994: referência 1985;

1995-2001: referência 2002; 2002-2016: referência 2010

Tabela 4: Construção dos índices

ÍNDICE	VARIÁVEIS	FONTE	PERÍODO
Investimento	Taxa de investimento brasileira	IBGE	1990-2016
	Valor adicionado da ind. de construção civil/PIB	IBGE	1990-2016
	Crescimento do consumo elétrico residencial	AEEE	1990-2016
Policy	Gasto corrente/PIB	IPEADATA e EOE	1990-2016
	Grau de abertura/PIB = (Exp + Imp)/PIB	BACEN	1990-2016
	Deflator relativo do PIB Estadual	IPEADATA e IBGE*	1990-2016

Notas: Variáveis denominadas em Reais de 1995; EOE: Execução Orçamentária de Estados;

AEEE: Anuário Estatístico de Energia Elétrica 2018; *: Construção dos autores com base em dados

do IPEADATA e IBGE. Dados nominais anteriores a 1995: IPEADATA. Demais dados, ver nota da tabela 3.

- Cicchone, Antonio, e Marek Jarocinski. 2010. “Determinants of Economic Growth: Will Data Tell?” *American Economic Journal: Macroeconomics* 2 (4): 222–246. ISSN: 1945-7707.
- Dias, Joilson, Waleska de Fátima Monteiro, Maria Helena Ambrosio Dias e Letícia Xander Russo. 2013. “Função de capital humano dos estados brasileiros : retornos crescentes ou decrescentes da educação?” *Pesquisa e Planejamento Econômico* 43 (2).
- Divino, José Angelo, e Rogério Lúcio Soares da Silva Junior. 2012. “Composição dos Gastos Públicos e Crescimento Econômico dos Municípios Brasileiros”. *Revista Economia* 13 (3): 507–528.
- Doppelhofer, Gernot, e Melvyn Weeks. 2011. *Robust Growth Determinants*. Working Paper, CESifo Working Paper Series 3354. CESifo Economic Studies.
- Durlauf, S. N., P. A. Johnson e J. R. W. Temple. 2005. “Growth Econometrics”. Capítulo 8 em *Handbook of economic growth*, editado por Philippe Aghion e Steven N. Durlauf, volume 1A, 555–675. 05. Elsevier B.V.
- Durlauf, Steven N. 2001. “Manifesto for a growth econometrics”. *Journal of econometrics* 100 (1): 65–69.
- Durlauf, Steven N., Andros Kourtellis e Chih Ming Tan. 2008. “Are Any Growth Theories Robust?” *The Economic Journal* 118 (527): 329–346. ISSN: 0013-0133.
- Fernández, Carmen, Eduardo Ley e Mark F. J. Steel. 2001. “Model uncertainty in cross-country growth regressions”. *Journal of Applied Econometrics* 16 (5): 563–576. ISSN: 0883-7252.
- Hjort, Nils Lid, e Gerda Claeskens. 2003. “Frequentist Model Average Estimators”. *Journal of the American Statistical Association* 98 (464): 879–899.
- Islam, N. 1995. “Growth empirics: a panel data approach”. *The Quarterly Journal of Economics*.
- Jones, Charles I. 1997. “Convergence revisited”. *Journal of Economic Growth* 2:131–153.
- Koop, Gary, Roberto Leon-Gonzalez e Rodney Strachan. 2012. “Bayesian model averaging in the instrumental variable regression model”. *Journal of Econometrics* 171 (2): 237–250. ISSN: 03044076.
- Kormendi, RC, e PG Meguire. 1985. “Macroeconomic determinants of growth: cross-country evidence”. *Journal of Monetary economics* 16:141–163.
- Lenkoski, A, TS Eicher e AE Raftery. 2014. “Two-stage Bayesian model averaging in endogenous variable models”. *Econometric reviews*: 1–32.
- Leon-Gonzalez, Roberto, e Daniel Montolio. 2004. “Growth, convergence and public investment. A Bayesian model averaging approach”. *Applied Economics* 36 (17): 1925–1936. ISSN: 0003-6846.
- . 2015. “Endogeneity and Panel Data in Growth Regressions: A Bayesian Model Averaging”. *Journal of Macroeconomics* 46:23–39.
- Leon-Gonzalez, Roberto, e Thanabalasingam Vinayagathan. 2015. “Robust determinants of growth in Asian developing economies: A Bayesian panel data model averaging approach”. *Journal of Asian Economics*, número 36: 34–46.
- Levine, Ross, e David Renelt. 1992. “A sensitivity analysis of cross-country growth regressions”. *The American economic review*: 942–963.
- Lopes, Hedibert, e Mike West. 2004. “Bayesian model assessment in factor analysis”. *Statistica Sinica* 14:41–67.

- Mankiw, NG, David Romer e DN Weil. 1992. “A contribution to the empirics of economic growth”. *Quarterly Journal of Economics* 107 (2): 407–437.
- Mincer, Jacob A. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- Moral-Benito, Enrique. 2012. “Determinants of Economic Growth: A Bayesian Panel Data Approach”. *Review of Economics and Statistics* 94 (2): 566–579. ISSN: 0034-6535.
- Rocha, F, e AC Giuberti. 2007. “Composição do gasto público e crescimento econômico: uma avaliação macroeconômica da qualidade dos gastos dos Estados brasileiros”. *Economia Aplicada* 11 (4): 463–485.
- Rodrigues, R. V., e E. C. Teixeira. 2010. “Gasto público e crescimento econômico no Brasil: uma análise comparativa dos gastos das esferas de governo”. *Revista Brasileira de Economia* 64 (4): 423–438.
- Roodman, D. 2009. “How to do xtabond2: An introduction to difference and system GMM in Stata”. *Stata Journal* 9 (1): 86–136(51).
- Sala-i-Martin, Xavier, Gernot Doppelhofer e Ronald I Miller. 2004. “Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach”. *American Economic Review* 94 (4): 813–835.
- Sala-i-Martin, XX. 1997. “I just ran two million regressions”. *The American Economic Review* 87 (2): 178–183.
- Silva, G. J. C. da, e R. S Scatolin. 2012. “Gastos públicos e crescimento econômico recente dos estados brasileiros”. *Revista Economia & Tecnologia* 8 (3): 19–38.
- Solow, RM. 1956. “A contribution to the theory of economic growth”. *The quarterly journal of economics* 70 (1): 65–94.
- Steel, Mark F. J. 2017. “Model Averaging and its Use in Economics”. *MPRA Paper*, número 81568.