



ECONOMIA APLICADA

BRAZILIAN JOURNAL OF APPLIED ECONOMICS

Vol. 27 – nº 3

Julho – Setembro 2023

ARTIGOS

**Incentivos à reeleição afetam a eficiência dos gestores públicos?
Uma análise para os municípios brasileiros**

Ana Cláudia Annegues, Celina Oliveira, Wallace Souza

**Volume, learning-by-doing, and economies of scale in the Brazilian
hospital sector: the case of hip arthroplasty**

Pedro Henrique Soares Leivas, Paulo de Andrade Jacinto, Anderson
Moreira Aristides dos Santos, Livia Madeira Triaca, Cesar Augusto
Oviedo Tejada

Regional development and convergence clubs in Uruguay

Diego Aboal, Bibiana Lanzilotta, Martín Pereyra, María Paz Queraltó

**Pobreza e desigualdade no Brasil: uma análise da renda para
2004-2008 e 2016-2019**

Otávio Junio Faria Neves, Ana Márcia Rodrigues da Silva, Henrique
Dantas Neder

**Early warning systems via machine learning: a study of currency
crises**

Vitor Ostrensky, Leonardo Matsuno da Frota

ECONOMIA FEA-RP/USP



ECONOMIA APLICADA

Vol. 27 – nº 3 Julho – Setembro 2023

PAPERS

**Do re-election incentives affect the efficiency of public managers?
An analysis for Brazilian municipalities**

Ana Cláudia Annegues, Celina Oliveira, Wallace Souza

**Volume, learning-by-doing, and economies of scale in the Brazilian
hospital sector: the case of hip arthroplasty**

Pedro Henrique Soares Leivas, Paulo de Andrade Jacinto, Anderson
Moreira Aristides dos Santos, Lívia Madeira Triaca, Cesar Augusto
Oviedo Tejada

Regional development and convergence clubs in Uruguay

Diego Aboal, Bibiana Lanzilotta, Martín Pereyra, María Paz Queraltó

**Poverty and inequality in Brazil: an analysis of income for
2004-2008 and 2016-2019**

Otávio Junio Faria Neves, Ana Márcia Rodrigues da Silva, Henrique
Dantas Neder

**Early warning systems via machine learning: a study of currency
crises**

Vitor Ostrensky, Leonardo Matsuno da Frota

ISSN 1980-5330



9 771980 533000

APOIO:



ECONOMIA APLICADA

Vol. 27 – nº 3

Julho – Setembro 2023

Editor

Eliezer Martins Diniz

Editores Associados

Edson Zambon Monte

Fernando Antonio de Barros Júnior

Sergio Naruhiko Sakurai

Eliezer Martins Diniz

Reynaldo Fernandes

Conselho Editorial

Affonso Celso Pastore

USP

Cássio F. Camargo Rolim

UFPR

Clélio Campolina Diniz

CEDEPLAR

Eleuterio F. S. Prado

USP

Francisco Anuatti Neto

USP

Geraldo Sant'Ana de Camargo Barros

USP

José Marcelino da Costa

NAEA/PA

Lízia de Figueirêdo

UFMG

Maria José Willumsen

Florida International University

Mário Luiz Possas

UFRJ

Paulo Nogueira Batista Júnior

FGV/SP

Pedro Luiz Valls Pereira

FGV/SP

Ricardo R. Araújo Lima

UnB

Rodolfo Hoffman

USP

Russell E. Smith

Washburn University

Tomás Málaga

FGV/SP

Wilson Suzigan

Unicamp

Carlos Roberto Azzoni

USP

Cláudio Monteiro Considera

UFF

Denisard C. de Oliveira Alves

USP

Fernando de Holanda Barbosa

FGV/RJ, UFF

Geoffrey J. D. Hewings

University of Illinois

Gustavo Maia Gomes

IPEA

José A. Scheinkman

Princeton University

Marcelo Portugal

UFRGS

Márcio Gomes Pinto Garcia

PUC/ RJ

Paulo César Coutinho

UnB

Pedro Cezar Dutra Fonseca

UFRGS

Pierre Perron

Boston University

Roberto Smith

UFC

Rogério Studart

UFRJ

Sérgio Werlang

FGV/RJ

Victor Bulmer-Thomas

University of London



A Revista ECONOMIA APLICADA é uma publicação trimestral do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo. Os conceitos emitidos são de responsabilidade exclusiva do(s) autor(es). A simples remessa do trabalho implica que o(s) autor(es) concorda(m) que, em caso de aceitação para publicação, a ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS) passa a ter os direitos autorais para a veiculação dos artigos, tanto em formato impresso como eletrônico a eles referentes, os quais se tornarão propriedade exclusiva da ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). É permitida a reprodução total ou parcial dos artigos desta revista, desde que seja explicitamente citada a fonte completa. Esta revista na versão on-line possui vários indexadores, cuja lista atualizada encontra-se no site da revista no Portal Revistas USP.

Secretária: Juliana de Freitas Vitória

Capa: Christof Gunkel

Revisão linguística (Português e Inglês): Iulo Feliciano Afonso (Solution Translations Eireli — Epp)

Diagramação: Beatriz Porfírio Germin

Endereço

Revista Economia Aplicada, Departamento de Economia, FEA-PR/USP

Avenida do Bandeirantes, 3900, Ribeirão Preto, SP.

CEP 14040-905

Telefone: (0xx16) 3315-0407

email: revecap@usp.br

site: <https://www.revistas.usp.br/ecoa>

FICHA CATALOGRÁFICA

Elaborada pela Seção de Publicações e Divulgação do SBD/FEA/USP.

Economia Aplicada / Departamento de Economia, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo – v.1, n.1 (1997–). – Ribeirão Preto, SP: FEA-RP, 1997–

Trimestral

Publicado anteriormente sob responsabilidade de FEA/USP, FIPE e FEA-RP

ISSN 1980-5330

1. Economia I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto. Departamento de Economia.

CDD 22.ed – 330.

ECONOMIA APLICADA

Vol. 27 – nº 3

Julho – Setembro 2023

ARTIGOS

- 291 **Incentivos à reeleição afetam a eficiência dos gestores públicos? Uma análise para os municípios brasileiros**
Ana Cláudia Annegues , Celina Oliveira , Wallace Souza
- 321 **Volume, learning-by-doing, and economies of scale in the Brazilian hospital sector: the case of hip arthroplasty**
Pedro Henrique Soares Leivas , Paulo de Andrade Jacinto , Anderson Moreira Aristides dos Santos , Livia Madeira Triaca , Cesar Augusto Ovi-
edo Tejada
- 351 **Regional development and convergence clubs in Uruguay**
Diego Aboal , Bibiana Lanzilotta , Martín Pereyra , María Paz Queraltó
- 383 **Pobreza e desigualdade no Brasil: uma análise da renda para 2004-2008 e 2016-2019**
Otávio Junio Faria Neves , Ana Márcia Rodrigues da Silva , Henrique Dantas Neder
- 407 **Early warning systems via machine learning: a study of currency crises**
Vitor Ostrensky , Leonardo Matsuno da Frota

INCENTIVOS À REELEIÇÃO AFETAM A EFICIÊNCIA DOS GESTORES PÚBLICOS? UMA ANÁLISE PARA OS MUNICÍPIOS BRASILEIROS

ANA CLÁUDIA ANNEGUES *

CELINA OLIVEIRA †

WALLACE SOUZA ‡

Resumo

O artigo mostra como a reeleição afeta a eficiência na gestão em educação. Selecionou-se como grupo de tratamento municípios com prefeitos candidatos à reeleição, comparando-os com dois grupos de controle: 1) municípios com prefeitos em primeiro mandato não candidatos à reeleição, e 2) municípios com prefeitos já em segundo mandato. No primeiro, utilizou-se um *Propensity Score Weighting* com Diferenças em Diferenças. No segundo, procedeu-se a uma análise de regressão com variável igual a 1 para o tratamento e 0 para o controle. Os resultados indicaram que prefeitos em primeiro mandato candidatos à reeleição são menos ineficientes que aqueles em segundo mandato.

Palavras-chave: reeleição; reputação; DEA; PSW.

Abstract

The article shows how re-election affects the efficiency of education management. The treatment group selected municipalities with mayors running for re-election and compared them with two control groups: 1) municipalities with first-term mayors not running for re-election, and 2) municipalities with mayors already in their second term. In the first, a Propensity Score Weighting was used. In the second, a regression analysis was performed with a variable equal to 1 for the treatment and 0 for the control. The results indicated that first-term mayors running for re-election are less inefficient than those in their second term.

Keywords: re-election; reputation; DEA; PSW.

JEL classification: F1, C1

DOI: <http://dx.doi.org/10.11606/1980-5330/ea184507>

* Departamento de Economia, Universidade Federal da Paraíba. E-mail: anne-gues.ana@gmail.com

† Departamento de Economia, Universidade Federal do Ceará. E-mail: oli.celina@gmail.com

‡ Departamento de Economia, Universidade Federal da Paraíba. E-mail: wpsfarias@gmail.com

1 Introdução

O conceito de eficiência na alocação de recursos públicos diz respeito à capacidade do gestor em oferecer uma maior quantidade ou qualidade dos serviços sem a adição de mais recursos, ou, por outro lado, em prover os mesmos serviços a custos menores. Nesse sentido, uma gestão ineficiente não é desejável ao bem-estar social, tendo em vista que esta implica em desperdício de recursos para uma oferta reduzida de serviços¹. A gestão pública eficiente tem sido pauta recorrente em debates dentro e fora da academia, especialmente com o crescimento da atuação do estado brasileiro na oferta de serviços públicos a partir da Constituição de 1988. A maior presença do estado como resposta aos anseios sociais da população teve como consequência o crescimento acelerado dos gastos públicos, mesmo diante das limitações de financiamento existentes. Não obstante a expansão dos dispêndios, o Brasil apresentou tímidas melhorias em seus indicadores sociais. Na educação, por exemplo, a qualidade do ensino, mensurada por testes padronizados nacionais e internacionais, deixa a desejar. Os resultados obtidos a partir do *Programme for International Student Assessment (PISA)* de 2018 mostram o Brasil entre as piores colocações no ranking das três áreas avaliadas pelo programa.

A literatura econômica fornece várias técnicas paramétricas e não paramétricas para mensuração da eficiência do gasto público, além de apresentar pesquisas que procuram identificar os determinantes das diferenças de capacidade gerencial. As primeiras ajudam a identificar as gestões com maior nível de desempenho e as últimas consistem em importantes contribuições pois identificam quais fatores determinantes podem ser mudados para melhorar a situação de municípios com pior gestão.

Um fator que pode influenciar a maneira como os recursos públicos são geridos, porém pouco explorado nos estudos sobre gestão pública, é o ambiente político-institucional no qual os gestores estão inseridos. Como esses gestores são escolhidos por um processo eleitoral, a assimetria de informação existente entre a população e os agentes políticos suscitam problemas de agência que alteram os incentivos e o comportamento destes quando responsáveis pela administração dos recursos públicos².

Dentro do sistema eleitoral um fator que tende a afetar o comportamento dos gestores públicos é a limitação de mandato (em inglês *term limit*), que é a restrição legal que estabelece o número de vezes que um representante político pode se reeleger consecutivamente a um cargo eletivo. É razoável supor que a possibilidade de se candidatar novamente altera a função objetivo do ocupante do executivo, sendo a intenção de permanecer no cargo uma das preocupações que guiam a sua gestão. A renovação do mandato dependerá da avaliação da população quanto aos resultados da gestão e isso poderá influenciar a eficiência da gestão, seja pelo lado dos gastos públicos, seja pelo lado dos produtos na oferta dos bens e serviços.

Diante disso, o presente artigo tem por objetivo mensurar como a existência do dispositivo da reeleição em um sistema eleitoral pode afetar a eficiência dos gastos públicos em educação. Para tanto, as unidades de observação do

¹Lopes *et al.* (2008) constataam que o aumento da eficiência alocativa tende a elevar o bem-estar social e a reduzir desigualdades.

²Para mais informações sobre a teoria da agência ver por exemplo Spence (1975) e Spence e Zeckhauser (1978).

estudo serão os municípios brasileiros, uma vez que os governos locais são responsáveis por prover boa parte dos serviços públicos à população.

Como variável dependente tem-se os escores de eficiência técnica calculados por meio da Análise por Envoltória de Dados (DEA). Embora exista uma gama de técnicas de avaliação de eficiência para o cálculo desses escores, optamos por utilizar a DEA, a qual se baseia em um método de estimação não paramétrico, que por meio da programação matemática linear procura construir uma fronteira de produção sem a necessidade de estabelecer previamente uma forma funcional.

A variável de insumo contém informações sobre os gastos com educação por aluno dos municípios e para mensurar os produtos foram utilizadas as notas médias entre português e matemática do 5º ano e do 9º ano do ensino fundamental obtidas na Prova Brasil, a qual integra o Sistema de Avaliação da Educação Básica (SAEB), um conjunto de avaliações externas de aprendizagem promovido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), autarquia federal vinculada ao Ministério da Educação do Governo Federal do Brasil. Além das proficiências, também foram incluídas no vetor das variáveis de produto as taxas de aprovação do município nas duas séries analisadas.

Nos modelos estimados tem-se um grupo de tratamento constituído por municípios cujo prefeito em primeiro mandato se candidatou à reeleição. Sendo assim, a variável de interesse será uma *dummy* que assume valor 1 se o município teve o prefeito candidato à reeleição e 0 para os demais casos. São formados dois grupos de controle, um contendo municípios que tinham prefeitos em primeiro mandato, mas que optaram por não concorrerem à reeleição, e outro formado por municípios cujo prefeito já havia sido reeleito anteriormente - já se encontrava em segundo mandato - e que, portanto, não poderia concorrer novamente.

Na estimação com o primeiro grupo de controle, especificamente, recorreu-se a um método de pareamento por reponderação, o *Propensity Score Weighting*. Nessa abordagem, primeiramente são calculados escores de propensão, que medem a probabilidade de um município se submeter ao tratamento de ter um prefeito candidato à reeleição. Essa probabilidade é calculada condicional a algumas variáveis observadas, que incluem alguns atributos do gestor, como gênero, escolaridade e partido ao qual é filiado. Para isso, utilizou-se um modelo logit, com a variável de tratamento como variável dependente e as variáveis observadas como explicativas. Por fim, os escores de propensão estimados são incluídos como pesos na equação de regressão principal, como forma de selecionar os municípios mais semelhantes em termos de probabilidade de sujeição ao tratamento.

Como resultados esperados, pode-se levantar alguns cenários possíveis. Se os municípios geridos por candidatos à reeleição apresentarem um escore de eficiência maior que os municípios cujos gestores em primeiro mandato não se candidataram, esse pode ser um resultado do seu desejo de se reeleger. O prefeito candidato objetiva impulsionar sua reputação diante da população para que esta lhe conceda um segundo mandato e, assim, pode se esforçar mais para gerar o máximo de provisão dos bens públicos. Por outro lado, caso os resultados evidenciem que os prefeitos candidatos sejam menos eficientes, uma explicação possível é que o gestor em época de reeleição aumenta os gastos como forma de sinalizar compromisso com algumas demandas, porém sem o acompanhamento de ações que impactem diretamente os *outputs* das

políticas.

Já quando se compara os candidatos à reeleição com prefeitos em segundo mandato, uma maior eficiência do primeiro grupo corrobora a ideia de que o objetivo de se reeleger funciona como incentivo à melhor gestão, o qual não existe para quem já se reelegeu. Como não concorrerá à reeleição, o político não necessitaria mais se empenhar para melhorar sua reputação junto aos eleitores com melhores resultados, e assim, teria menores índices de eficiência. Por outro lado, se os prefeitos candidatos à reeleição se mostrarem menos eficientes, pode-se apontar a existência de um efeito aprendizagem (*learning-by-doing*), segundo o qual um segundo mandato implica em maior experiência administrativa, fazendo com que esse grupo tenha melhores resultados de eficiência.

As análises serão feitas considerando as eleições municipais de 2012 e 2016. No entanto, os índices de eficiência e demais dados educacionais foram obtidos a partir de informações referentes aos anos de administração imediatamente anteriores aos pleitos, isto é, 2011 e 2015. Dois motivos explicam isso. O primeiro é que as informações do SAEB estão disponíveis apenas para anos ímpares, de maneira que não há dados para 2012 e 2016. O segundo motivo são as limitações de orçamento em anos eleitorais impostas pela legislação eleitoral no Brasil, o que diminui o grau de discricionariedade dos prefeitos quanto à administração dos recursos. Ademais, é razoável supor que a preocupação com a reeleição não se restringe apenas ao último ano do primeiro mandato, sendo um dos norteadores de todos os quatro anos de gestão.

Os dados utilizados têm como fonte diversas bases, a saber: os dados do SAEB de desempenho, conforme já informado, o Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (Siconfi), disponibilizado pela Secretaria do Tesouro Nacional (STN) e estatísticas eleitorais do Tribunal Superior Eleitoral (TSE).

O artigo encontra-se dividido em mais cinco seções, além desta introdução. A segunda seção contém uma revisão teórica e empírica acerca dos impactos dos incentivos eleitorais sobre a gestão pública. A terceira seção contém os procedimentos metodológicos. A quarta seção descreve os dados utilizados. A quinta seção mostra os resultados encontrados. A sexta, e última seção, traz as considerações finais do estudo.

2 Revisão Teórica e Literatura Empírica

Muitos trabalhos têm surgido na literatura nacional com o objetivo de mensurar a eficiência dos gastos públicos, com destaque especial às áreas da educação e da saúde. Dentre estes pode-se citar trabalhos como [Sousa e Ramos \(1999\)](#), [Marinho \(2003\)](#), [Trompieri Neto et al. \(2009\)](#), [Silva e Almeida \(2012\)](#) e [Oliveira, Souza e Annegues \(2018\)](#).

Tão importante quanto estimar a eficiência é saber quais são os seus determinantes, isto é, quais variáveis explicam as diferenças de desempenho dos tomadores de decisão com relação à alocação dos recursos. [Lopes et al. \(2008\)](#) e [Lopes \(2016\)](#) encontram que municípios com maior IDH tendem a apresentar melhores resultados de eficiência alocativa. Já na pesquisa de [Marinho \(2003\)](#) as variáveis PIB per capita municipal e população não tiveram impacto estatisticamente significativo sobre os índices de eficiência.

Conforme já afirmado na seção anterior, um dos fatores que podem afetar a eficiência da gestão dos gastos são as variáveis políticas, dentre elas a limitação de mandato. No Brasil, após a aprovação da Emenda da Reeleição (ER) em 1997, o sistema eleitoral passou a permitir a possibilidade de recondução para um único período subsequente para os cargos executivos de todos os níveis de governo, de tal forma que prefeitos, governadores de estado e o presidente da República podem permanecer por no máximo oito anos consecutivos no poder (ALMEIDA; SAKURAI *et al.*, 2018).

As eleições no Brasil são realizadas em anos pares, com intervalos regulares de dois anos entre cada pleito, alternando-se entre eleições gerais (presidente e vice-presidente da República, os governadores e seus vices, senadores, deputados federais e estaduais) e eleições municipais (prefeitos e vereadores). Até 1994, os ocupantes de cargos executivos eram eleitos para mandatos com duração de cinco anos. Nesse mesmo ano foi aprovada uma emenda ao artigo 82 da Constituição Federal, alterando para quatro anos os mandatos de presidente, governadores, prefeitos e seus respectivos vices.

Mas foi em 1997 que foi promulgada a Emenda 16 ao artigo 82, a qual instituiu a reeleição para mais um único mandato dos chefes do Executivo. Desde então, seis eleições gerais e seis eleições municipais foram realizadas em que os chefes do Executivo puderam concorrer a mais de um mandato do exercício do cargo. Os resultados nacionais e estaduais das eleições apresentam um alto índice de continuísmo, isto é, 100% dos candidatos à presidência tentaram e obtiveram sucesso, assim como 80% dos governadores estaduais, em média, foram reeleitos (MARCINIUK; BUGARIN, 2019). Esse novo cenário político eleitoral que se desenhou há mais de 20 anos, além de outros fatores, possivelmente contribuiu para explicar o comportamento fiscal dos estados e municípios brasileiros, bem como a política pública de uma maneira geral. Para entender os possíveis canais de relação entre o incentivo à reeleição e a forma como os políticos administram os recursos quando responsáveis pela sua execução precisamos mergulhar na teoria econômica e nos testes empíricos a fim de encontrar algumas respostas. Na visão de Nakaguma e Bender (2006), a possibilidade de recondução ao cargo elevou os incentivos à manipulação oportunista das finanças públicas, o que pode gerar ineficiência na administração dos recursos.

A relação entre incentivos eleitorais e políticas públicas é um tema recorrente na literatura econômica. A corrente ligada aos Ciclos Políticos (*political business cycle*) foi uma das primeiras a estudar o assunto, com uma abordagem mais ligada à macroeconomia. Sua premissa é a de que as decisões de política macroeconômica são influenciadas pela agenda do ciclo eleitoral. Em anos de eleição, o ocupante de cargo executivo se comporta de maneira oportunista ao adotar políticas expansionistas de gastos com a intenção de se reeleger, diferentemente do que ocorre em inícios de mandato, onde em geral se adota uma postura mais restritiva no controle das finanças públicas. Segundo Rogoff (1990), os governantes demonstram competência ao seu eleitorado mediante o aumento de gastos perceptíveis ao público, o que pode levar a distorções fiscais.

Algumas correntes teóricas mostram de que forma a reeleição afeta a alocação dos recursos públicos por parte dos gestores. Uma das principais consiste nos Modelos de Construção de Reputação (BESLEY; CASE, 1995, 2003; BESLEY, 2006). Besley e Case (1995) descrevem que em um contexto de informação imperfeita e agentes racionais, os incentivos à reeleição tendem a alterar as

escolhas de política pública por parte dos governantes. Supõe-se inicialmente que a relação entre políticos e eleitores se dá em dois períodos. No primeiro mandato o governante avalia seu nível de esforço tendo consciência de que será avaliado pelos eleitores que se beneficiam ou não da distribuição de bens públicos. Isso altera probabilisticamente a efetividade de certa política pública, e, conseqüentemente, a utilidade dos eleitores, que está diretamente relacionada à provisão de bens públicos (AMARANTE, 2017). Aqueles que desejam se reeleger escolhem aumentar a provisão de bens e serviços públicos para desenvolver sua reputação perante o eleitorado e assim, serem conduzidos a um segundo mandato. Ao realizar testes empíricos com dados de alguns estados americanos, os autores verificaram aumento de dispêndios nas localidades onde os governadores não puderam concorrer à reeleição.

Evidências semelhantes as de Besley e Case (1995) foram encontradas por Johnson e Crain (2004). Por outro lado, Nogare e Ricciuti (2011) não encontraram diferenças significativas na política fiscal de chefes do executivo com limitações de mandato e outros tipos de governantes.

Para o caso brasileiro, a literatura empírica acerca dos efeitos da reeleição e demais variáveis eleitorais sobre a gestão dos recursos ainda é incipiente, visto que o advento da reeleição é recente na política brasileira. Além disso, a maior parte das pesquisas limita-se à análise pelo lado das despesas. Novaes e Mattos (2010) testaram o efeito da intenção de reeleição sobre os gastos em saúde dos municípios brasileiros. Os resultados mostraram um crescimento das despesas em cidades administradas por prefeitos que tentavam um segundo mandato.

Almeida, Sakurai *et al.* (2018) investigam a influência da limitação de mandato e do calendário eleitoral sobre a gestão da conta "restos a pagar". De acordo com os autores, enquanto a aprovação da emenda da reeleição se converteu em um incentivo ao esforço de diferenciação dos gestores candidatos, a aprovação no ano 2000 da Lei de Responsabilidade Fiscal (LRF) buscou restringir seus efeitos adversos às finanças públicas. No entanto, criaram-se artifícios contábeis (contabilidade criativa) para burlar a legislação, como o adiamento, para anos fiscais subsequentes, de pagamentos de despesas reconhecidas, formalmente conhecida como "inscrição de valores em restos a pagar processados". Os resultados obtidos pelos autores apontaram que prefeitos em primeiro mandato inscreveram maiores valores na conta "restos a pagar" em comparação com os prefeitos em segundo mandato.

Conforme já argumentado, o modo como os eleitores julgam as atitudes dos governantes molda o comportamento destes na administração das contas públicas e, portanto, a intenção de reeleição pode gerar efeitos diversos. Um exemplo é o trabalho de Meneguín e Bugarin (2001), que, contrariamente aos trabalhos anteriormente citados, encontrou que prefeitos em primeiro mandato gastaram menos em anos eleitorais do que aqueles que não podiam se candidatar à reeleição.

Levando em conta esse resultado e o desejo do prefeito candidato de construir uma boa reputação, em alguns contextos a austeridade fiscal pode ser vista com bons olhos pela população e considerada um indicador de gestão eficiente. Silva e Braga (2013), em um estudo para municípios do estado da Bahia no período de 2000 a 2008, mostram evidências de que os eleitores punem os prefeitos que aumentam os gastos totais no ano eleitoral. Cavalcante (2016) além de analisar o impacto da reeleição sobre a gestão fiscal, também testa o efeito da política fiscal sobre a probabilidade de reeleição do mandatá-

rio. Os autores verificam que a reeleição não gera efeitos na gestão fiscal e que os candidatos que priorizaram os gastos em investimentos foram premiados pelos eleitores.

Dados os diversos critérios adotados pela população na avaliação da gestão dos seus governantes, a mensuração do efeito da reeleição demanda uma ampliação do escopo da administração pública para além das despesas, incorporando nas análises o lado do resultado, que é a provisão dos bens e serviços públicos. Uma forma de fazer isso é ver o impacto da reeleição não apenas no total de despesas, mas na relação custo-benefício verificada no cálculo da eficiência dos gastos públicos.

Poucos trabalhos são encontrados na literatura testando a relação entre eficiência e reeleições. [Ribeiro et al. \(2020\)](#) investigaram uma série de determinantes da eficiência de governos locais em Portugal, incluindo variáveis políticas, e verificaram que a reeleição não teve impacto estatisticamente significativo. [Motta e Moreira \(2009\)](#) adotam a abordagem da fronteira de produção estocástica, condicionada por variáveis relacionadas ao custo da provisão dos serviços e os incentivos que afetam a eficiência municipal, dentre esses os incentivos eleitorais, como a margem de vitória do prefeito em relação ao concorrente e o fato de o prefeito ter sido reeleito no pleito anterior, ou seja, estar no segundo mandato. O efeito da reeleição do prefeito foi positivo e estatisticamente significativo, indicando uma correlação entre a renovação do mandato e a redução da ineficiência. Esse resultado pode ser uma consequência da despreocupação do governante em se reeleger, restringindo o comportamento oportunista de elevar os gastos sem contrapartida de maiores resultados em termos de serviços públicos, ou a hipótese levantada pelo presente artigo de que maior experiência administrativa gera maior capacidade gerencial, fazendo o administrador ofertar mais bens públicos, sem, no entanto, elevar fortemente os gastos.

Este artigo surge no sentido de oferecer mais contribuições à literatura de eficiência alocativa e seus determinantes políticos, sem se limitar à análise do ponto de vista dos gastos. Além de comparar prefeitos em primeiro e em segundo mandato, como nos trabalhos anteriores, procuramos mensurar também o efeito da candidatura, analisando prefeitos em primeiro mandato concorrentes e não concorrentes à reeleição.

3 Metodologia

Inicialmente, será descrito o cálculo dos escores de eficiência para os municípios por meio do DEA, os quais consistem na variável dependente dos modelos de regressão. Conforme o afirmado na introdução, serão feitas duas estimações considerando um mesmo grupo de tratamento, municípios com prefeitos candidatos à reeleição, e dois diferentes grupos de controles. As subseções seguintes tratarão das estratégias empíricas empregadas nas duas estimações.

3.1 Cálculo dos Escores de Eficiência Utilizando DEA

A análise por Envoltória de Dados (DEA), introduzida por [Charnes, Cooper e Rhodes \(1978\)](#) e [Banker, Charnes e Cooper \(1984\)](#), é um método de programação linear não paramétrico que tem sido aplicado em diversos estudos

relacionados ao setor público como ferramenta para medir o desempenho gerencial, mediante o cálculo da eficiência dos gastos públicos.

O processo de otimização do modelo DEA busca construir uma fronteira de produção ótima em que o desempenho de cada Unidade Tomadora de Decisão (Decision Making Unit - DMU) é calculado tomando como base de comparação as DMU's ditas eficientes, isto é, aquelas que estão sobre a fronteira, e que, portanto, apresentam o melhor desempenho dado os insumos e produtos observados. Dessa forma, desvios em relação à fronteira apresentam-se como medidas de eficiência. (CHARNES; COOPER; RHODES, 1978)

Para cada DMU, a eficiência é medida em termos de uma mudança proporcional nos insumos ou produtos. Para atender a proposta atual, faz-se uso do modelo DEA orientado ao produto, onde uma DMU para atingir a eficiência deve maximizar o produto sem exigir mais de qualquer um dos insumos. Justifica-se o uso desse modelo pelo fato de os recursos utilizados pelos prefeitos se originarem de verbas federais transferidas aos municípios, o que limita o grau de discricionariedade desses gestores sobre o gasto por aluno (variável de insumo). Nesse sentido, a política educacional tende a atuar diretamente na maximização dos resultados educacionais (variável de produto), dados os recursos disponíveis. Para tanto, consideram-se como DMU's os municípios brasileiros, em que um município tecnicamente ineficiente apresenta um nível de produção observado abaixo do nível máximo que poderia alcançar, dada as quantidades de insumos e recursos observados atualmente.

Um outro ponto relevante é a característica das DMU's quanto a sua escala de operação: os municípios apresentam heterogeneidade produtiva. Sendo assim, diferentes municípios podem operar em regime crescente ou decrescente de escala devido às diferenças nas quantidades de insumos utilizados e nos produtos gerados. Dada essa característica, optou-se pela aplicação do modelo com Retornos Variáveis de Escala, também conhecido como BBC, desenvolvido por Banker, Charnes e Cooper (1984). Dessa forma, leva-se em consideração as disparidades de tamanho existentes entre as DMU's. Isto é possível devido às vantagens apresentadas pelo modelo BBC quanto à flexibilidade da forma funcional da fronteira de eficiência, em relação aos modelos mais clássicos do DEA, e ao relaxar a hipótese de proporcionalidade entre inputs e outputs do modelo clássico, assumindo, assim, uma fronteira convexa. (BANKER; CHARNES; COOPER, 1984)

A estimação da eficiência para cada DMU é realizada resolvendo-se o seguinte problema de otimização:

$$\begin{aligned}
 & \max \quad \theta_0 \\
 & \text{Sujeito a:} \\
 & \quad x_{i0} - \sum_{k=1}^n x_{ik} \lambda_k \geq 0, \forall i \\
 & \quad -\theta_0 y_{j0} + \sum_{k=1}^n y_{jk} \lambda_k \geq 0, \forall j \\
 & \quad \sum_{k=1}^n \lambda_k = 1 \\
 & \quad \lambda_k \geq 0, \forall k
 \end{aligned} \tag{1}$$

onde y representa o conjunto de outputs $j = 1, \dots, s$ e x o conjunto de inputs $i = 1, \dots, r$ para $k = 1, \dots, n$ DMU's. O parâmetro λ_k é a contribuição da DMU_k na formação do alvo da DMU_0 , representando assim a restrição de convexidade, classificando o modelo como retornos variáveis à escala. Esse parâmetro é um vetor cujos valores são calculados de forma a obter a solução ótima do problema.

Mantendo-se constantes os recursos, o parâmetro θ_0 representa um fator pelo qual todos os produtos devem ser multiplicados para que a DMU_0 atinja a fronteira eficiente. Sendo assim, θ_0 será maior que 1, pois implicará em aumentos nos valores dos *outputs*, de modo que representará o aumento proporcional no resultado educacional que poderia ser obtido pelo i -ésimo município, mantendo constante a utilização de recursos públicos. Portanto, o cálculo da eficiência será dado por $Ef = \frac{1}{\theta_0}$, variando entre 0 e 1, onde uma DMU com score de eficiência inferior a 1 será classificada como ineficiente (FERREIRA; GOMES, 2009; BERNARDO; RODRIGUES, 2015).

3.2 Efeito do Incentivo à Reeleição

Calculados os escores de eficiência, o próximo passo consiste em estimar a resposta destes ao dispositivo da reeleição e demais variáveis relevantes. Apenas a abordagem não paramétrica de mensuração dos escores de eficiência, como o método DEA, não seria suficiente para fornecer evidências nesse sentido. Por isso, recorre-se ao uso desse procedimento em dois estágios que combina a medição da eficiência pelo DEA com uma análise de regressão que usa a eficiência estimada como variável dependente.

Como colocado por Simar e Wilson (2007), o segundo estágio poderá apresentar falhas e viesar as estimativas. A primeira falha se refere à ausência teórica no processo de geração de dados que justifique o uso de dois estágios e que justifique assumir a censura nos escores de eficiência. O segundo está no fato de tratar os scores de eficiência das DMU's como observações independentes, o que não seria apropriado, uma vez que problemas de correlação serial seriam gerados. Esta última surge pelo fato de a estimação da eficiência ocorrer a partir de uma amostra comum de dados e pela natureza do processo de otimização do modelo DEA, cuja solução de otimização requer restrições de combinação linear entre as DMU's.

Simar e Wilson (2007) elaboraram um processo de estimação de dois estágios que leva em conta as falhas acima mencionadas. O primeiro ponto está no processo de geração de dados do primeiro estágio. As DMUs eficientes encontradas estão sujeitas ao viés de amostra finita. Isto quer dizer que as DMUs eficientes encontradas formam uma fronteira de produção virtual e, portanto, não representam uma característica do verdadeiro processo de geração de dados. Isto implica que o modelo de segundo estágio deve ser um modelo de regressão truncado ao invés de modelos censurados. Para tanto, os autores adotam procedimentos paramétricos de *bootstrap* dos quais geram erros-padrão estimados e intervalos de confiança sem sofrer efeito do viés de correlação serial encontrados entre os escores de eficiência das DMUs. Assim, o modelo econométrico a ser estimado tem a seguinte forma:

$$\theta_k = z_k \beta + \epsilon_k \quad (2)$$

sendo assim, a eficiência θ_k é linearmente depende de z_k , em que este denota um vetor linha de w variáveis ambientais que podem afetar a capacidade da

DMU_k de combinar eficientemente os insumos com o produto a ser gerado. O parâmetro β denota um vetor coluna de coeficientes, ϵ_i são os distúrbios que devem ser considerados estatisticamente independentes entre as DMUs e seguem uma distribuição normal truncada, com parâmetros $\mu = 0$ e σ e truncamento à esquerda em $1 - z_k\beta$.

Dentre as variáveis que compõe o vetor z estão as seguintes: candidatura à reeleição, se o prefeito é filiado ao mesmo partido do governador de estado, uma *dummy* para cada partido (*PT*, *PSB*, *PMDB* e *PSDB*), se o prefeito possui ensino superior, uma variável que mede a educação média do município, o logaritmo do PIB *per capita* municipal e uma *dummy* identificando as capitais dos estados.

A variável de reeleição assume valor 1 para os municípios que possuem prefeitos candidatos à reeleição e 0, caso o município esteja no grupo de controle. Conforme já explanado, serão duas estimações, cada qual com um grupo de controle diferente.

As dummies para os partidos *PT*, *PSB*, *PMDB*, *PSDB* procuram captar o efeito da queda de popularidade do *PT* em decorrência do processo de impeachment da ex-presidente Dilma Rousseff (*PT*), ocorrido em 2016, além de captar preferências dos municípios em relação aos gastos públicos. Alguns partidos podem apresentar maior propensão ou afinidade com políticas de gasto expansionistas.

O grau de escolaridade do prefeito pode indicar o seu nível de habilidades cognitivas e não cognitivas; um prefeito com ensino superior pode ter maior habilidade e capacidade de gerenciar os recursos. O nível de educação médio da população do município é incluído na regressão, uma vez que cidades com população de melhor nível de educação média tendem a apresentar maiores performances nos exames de verificação de aprendizagem e/ou possuem maior capacidade de monitorar e de cobrar as ações dos seus gestores públicos, o que tende a melhorar os níveis de eficiência municipal. O logaritmo do PIB *per capita* municipal e se o município é uma capital procuram capturar o nível de riqueza e desenvolvimento do município e possíveis diferenças regionais quanto ao acesso ao repasse de recursos.

O procedimento de [Simar e Wilson \(2007\)](#) inicia estimando o modelo DEA da forma tradicional. Em seguida, para corrigir a primeira falha, estima-se os erros-padrão e intervalos de confiança para $\hat{\beta}$ por meio do procedimento de bootstrap paramétrico, em que pseudo erros artificiais são extraídos estatisticamente independentes a partir da distribuição normal truncada à esquerda em $1 - z_k\beta$. Já a segunda falha, é tratada no procedimento de [Simar e Wilson \(2007\)](#)³ de duas formas, o presente trabalho faz uso apenas do segundo procedimento de estimativa denominado pelos autores de *algoritmo 2*.

O primeiro passo do *algoritmo 2* é excluir as DMUs ditas eficientes ($\theta_k = 1$) da análise de regressão, pois possuem o viés gerado de amostra finita. Os demais m (com $m < n$) escores de eficiência formam a amostra e entram para um modelo de regressão truncado à esquerda como variável dependente e com o procedimento de bootstrap são produzidos as estimativas $\hat{\beta}$ e o parâmetro de variância $\hat{\sigma}$. A partir das L_1 repetições de bootstrap um conjunto de resíduos ϵ_k são gerados com distribuição $N(0, \hat{\sigma})$ esses resíduos e o $\hat{\beta}$ formam um

³O procedimento proposto pelos autores não leva em conta a existência de *slacks* no trato dos níveis de eficiência.

conjunto de escores de eficiência artificiais que são utilizados para calcular o estimador de eficiência livre do viés de amostra finita.

O estimador livre do viés é utilizado como variável dependente como em 2 e um novo processo de estimação por bootstrap, truncado à esquerda em $1 - z_k \beta$, é realizado e novos $\hat{\beta}$ e seus respectivos $\hat{\sigma}$ são obtidos permitindo o cálculo de intervalos de confiança e de erros-padrão, bem como a aplicação dos testes de hipóteses sobre o conjunto de dados livres das falhas mencionadas no início da seção.

3.3 Propensity Score Weighting

Quando comparamos o grupo de tratados dos municípios com prefeitos candidatos à reeleição com o grupo de controle com prefeitos em primeiro mandato não candidatos, incluir uma *dummy* identificando os primeiros pode não ser suficiente para captar os efeitos sobre os níveis de eficiência. Isso porque, a candidatura à reeleição durante o primeiro mandato não consiste em um evento de atribuição aleatória; o prefeito em primeiro mandato tem a possibilidade de candidatar-se novamente, porém isto não resulta de uma imposição exógena, sendo a escolha pela candidatura definida possivelmente por fatores observáveis e não observáveis, os quais também estão potencialmente relacionados à eficiência na gestão dos recursos. Como forma de contornar esse viés, procurou-se construir um grupo de controle que melhor representasse os municípios com prefeitos candidatos na ausência dessa nova candidatura.

Um método bastante utilizado para essas análises é o método de pareamento. A ideia do método consiste em construir um grupo controle com base na hipótese de uma unidade do grupo de tratamento possuir um conjunto de características observáveis similares a outra unidade no grupo controle. Assim, um vetor X de variáveis observáveis deve conter aquelas relacionadas ao resultado potencial do tratamento e que também podem afetar a decisão de participar ou não do tratamento (no presente artigo, a candidatura à reeleição). Essa hipótese é conhecida como seleção nas observáveis ou ignorabilidade:

$$Y(0) \perp T_i | X_i \quad (3)$$

As características das unidades (municípios) não influenciam a decisão do prefeito de concorrer ou não da reeleição, assim como não afetam os resultados potenciais na ausência da candidatura. A segunda implicação garante que o indivíduo no grupo controle, com o mesmo vetor X , é um bom previsor do que aconteceria com o grupo tratado caso este não tivesse recebido o tratamento, conhecida como hipótese de sobreposição.

Um método de pareamento muito utilizado é o *Propensity Score Matching* (PSM)⁴. No PSM a influência dos fatores observáveis pode ser resumida na probabilidade de candidatura à reeleição, condicionada a esses fatores. Essa probabilidade é chamada de escore de propensão. A estimação do escore de propensão é feita por meio de um modelo de resposta qualitativa (logit ou probit), no qual a variável dependente é a variável de tratamento e as variáveis explicativas correspondem ao conjunto de variáveis observadas do vetor X :

$$Pr(T_i = 1 | X_i) = G(\beta X_i + \epsilon_i) \quad (4)$$

⁴Rosenbaum e Rubin (1983)

onde $G(\cdot)$ representa a função de distribuição acumulada, seguindo por hipótese uma distribuição de probabilidade logística, no caso de um logit. No PSM, os grupos de tratamento e controle são pareados pela comparação entre os escores de propensão, seguida da checagem das condições de balanceamento, isto é, se há similaridades entre os dois grupos quanto às características presentes em X .

Outra maneira de tornar os grupos de tratamento e controle comparáveis é a reponderação da regressão de interesse com base nos escores de propensão, o *Propensity Score Weighting* (PSW). O método consiste fundamentalmente em utilizar o inverso da probabilidade de submeter-se ao tratamento como peso em uma análise de regressão multivariada (GUO; FRASER, 2014).

Uma das dificuldades do pareamento tradicional via PSM, é que este demanda um bom número de observações nos grupos de tratados e não tratados. Uma amostra pequena reduz as chances de se encontrar pares de observações semelhantes. Além disso, a depender da técnica de pareamento utilizada e do número de observações pareadas, é possível que um bom número de observações do grupo de controle, $T_i = 0$, seja descartado da análise. Nesse sentido, uma das vantagens do PSW está na sua capacidade de utilizar todos os participantes de ambos os grupos. Além disso, segundo Guo e Fraser (2014), o PSW apresenta mais duas vantagens frente ao PSM: ele permite a realização de uma gama de análises multivariadas e não exige que a variável de resultado seja contínua ou normalmente distribuída⁵.

O método permite estimar o Efeito Médio do Tratamento sobre os Tratados (*Average Treatment Effect on the Treated* - ATT). Para a estimação do ATT, os autores mostram que os pesos amostrais, representados por ω , assumem a forma:

$$\omega(T_i, X_i) = T_i + (1 - T_i) \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} \quad (5)$$

Quando a observação pertence aos tratados ($T_i = 1$), esta terá peso igual a 1 na regressão. Já para as observações do grupo de controle, o peso será igual a $\frac{p(\hat{X}_i)}{1-p(\hat{X}_i)}$. Sendo assim, quanto mais próxima de um a probabilidade de sujeição ao tratamento, maior o peso dessa observação do grupo de controle na regressão. Os participantes menos prováveis de serem tratados receberão os menores pesos.

O efeito da candidatura à reeleição será obtido estimando-se a equação de regressão (2), considerando como os pesos amostrais calculados com base nos escores de propensão. A utilização dessa técnica assume que a atribuição ao tratamento será independente dos resultados potenciais dos grupos de tratamento e controle, dado o vetor de observáveis, solucionando, portanto, o viés de seleção em observáveis.

Para a estimação do escore de propensão, foi utilizado um modelo logit, sendo o vetor X composto pelos seguintes fatores observáveis: a idade e o sexo do prefeito, se este possuía ensino superior, se era casado, se era filiado ao mesmo partido do governador e/ou do presidente, considerando Dilma Rousseff em 2012 e Michel Temer em 2016. Também foram incluídas *dummies* que identificam se o prefeito era filiado ao PMDB, PT, PSDB e PSB, buscando

⁵Outras referências sobre o PSW podem ser encontradas em Olmos e Govindasamy (2015).

captar a popularidade desses partidos em diferentes municípios e seu efeito sobre a probabilidade de o prefeito concorrer à reeleição.

4 Dados

4.1 Fontes dos Dados

Foram utilizados dados de fontes distintas para cada estágio da estratégia empírica. No primeiro estágio, que consiste na geração dos escores de eficiência, foi utilizado como insumo as informações das despesas orçamentárias municipais da função Educação coletadas do Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (Siconfi), disponibilizado pela Secretaria do Tesouro Nacional (STN), para os anos de 2011 e 2015.

Para as variáveis de produto foram utilizadas a taxa de rendimento escolar (Taxa de Aprovação) e as notas de Português e Matemática obtidas na Prova Brasil para os anos iniciais e finais do ensino fundamental, por município, para os anos de 2011 e 2015. O desempenho dos alunos nessas provas é avaliado com base nas Escalas de Proficiência, que permitem a interpretação das habilidades e das competências destes associadas a diversos pontos da escala. Essas escalas são construídas com base na Teoria de Resposta ao Item (TRI) e no modelo de Blocos Incompletos Balanceados (BIB)⁶. Esses dados são disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), órgão vinculado ao Ministério da Educação (MEC). O SAEB reflete os níveis de aprendizagem dos estudantes avaliados por meio de testes aplicados a cada dois anos na rede pública e em uma amostra da rede privada. Isso permite que as escolas e as redes municipais e estaduais de ensino avaliem a qualidade da educação oferecida aos estudantes. Para os objetivos do artigo, foram utilizadas apenas informações referentes à rede pública municipal, tanto sobre insumo quanto produtos.

No segundo estágio, as variáveis que podem afetar os níveis de eficiência dos gastos públicos, mas que estão fora do controle do gestor, foram obtidas junto a outras duas fontes. A primeira é a Federação das Indústrias do Estado do Rio de Janeiro (Firjan), que construiu o Índice Firjan de Desenvolvimento Municipal (IFDM), cujo objetivo é acompanhar o desenvolvimento socioeconômico brasileiro monitorando três áreas: Emprego & Renda, Educação e Saúde. Foram coletadas informações referentes apenas à área de Educação, com o propósito de captar o nível de educação médio da população em cada município. A segunda fonte é o Instituto Brasileiro de Geografia e Estatística (IBGE), o qual fornece informações sobre o tamanho da População e o PIB municipal em 2011 e 2015, as quais geraram a variável PIB *per capita*, além de identificar as capitais dos estados.

O vetor de características observáveis utilizado na estimação dos escores de propensão utilizados no cálculo do PSW incluem as variáveis do municí-

⁶A TRI é um conjunto de modelos matemáticos que procura representar a probabilidade de um indivíduo dar uma determinada resposta a um item como função dos seus parâmetros e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. O BIB, por sua vez, é um esquema otimizado para o rodízio de blocos. Isto é especialmente útil nos sistemas de avaliação quando se deseja obter informações amplas sobre o ensino, quando se precisa limitar a quantidade de itens submetida a cada aluno em um valor aceitável e adequado ao tempo de prova. A utilização da TRI e do BIB permite, assim, a construção de uma escala para cada disciplina, englobando as três séries avaliadas e ordenando o desempenho dos alunos do nível mais baixo para o mais alto.

pio citadas no parágrafo anterior mais atributos do prefeito (idade, sexo, se é casado, se possui ensino superior, partido, se é filiado ao mesmo partido do governador). Tais informações coletadas têm como fonte o Repositório de Dados Eleitorais do Tribunal Superior Eleitoral (TSE), que traz uma compilação de informações brutas das eleições tais como: o partido, a coligação dos candidatos e informações de característica pessoal do candidato e se é candidato à reeleição ou não. Essa base possibilitou, assim, identificar os prefeitos em primeiro mandato (candidatos e não candidatos à reeleição) e os gestores em segundo mandato.

4.2 Estatísticas Descritivas

A Tabela 1 apresenta as estatísticas descritivas para o conjunto de variáveis utilizadas no cálculo dos níveis de eficiência, quais sejam: o gasto com educação por aluno, como insumo, e as notas de Português e Matemática e a Taxa de Aprovação como produtos, para os anos iniciais e finais. A tabela apresenta os indicadores de média e desvio-padrão para duas amostras, que se diferenciam pelo grupo de controle utilizado na estimação: o grupo de controle 1 (prefeitos em primeiro mandato não candidatos) e grupo de controle 2 (prefeitos em segundo mandato). Verifica-se no insumo uma média aproximada dos gastos com educação por aluno nas duas amostras. Em relação ao período de tempo, observa-se um aumento nessa média, indicando um aumento dos recursos públicos utilizados na educação.

Com relação às variáveis de produto, a pontuação das avaliações realizadas pela Prova Brasil possui um intervalo de zero a 325 pontos e de 200 a 375 pontos para Língua Portuguesa nos anos iniciais e finais, respectivamente. Para os dados coletados da presente pesquisa, a média de pontos é aproximada entre as amostras para a nota de Língua Portuguesa, tanto para os anos iniciais quanto anos finais. Além disso, observa-se um aumento na média entre os anos 2011 e 2015.

Segundo a escala SAEB, essas pontuações indicam níveis básicos de aprendizagem dos alunos do 5º ano, que compreendem habilidades como identificar o assunto principal e personagens em contos e letras de música; reconhecer relação de causa e consequência e relação entre pronomes e seus referentes em fábulas, poemas, contos e tirinhas. Já para os alunos do 9º ano, assume-se que estes são capazes de reconhecer o sentido estabelecido pelo uso de expressões, de pontuação, de conjunções em poemas, charges e fragmentos de romances e de reconhecer recursos argumentativos em artigos de opinião⁷.

A disciplina de Matemática possui uma pontuação que varia de 0 a 350 pontos para os anos iniciais e de 200 a 450 pontos para os anos finais. A média de pontos para os dois tipos de abordagens foi aproximada. Ao comparar os períodos, verifica-se também um aumento na média das notas de Matemática para os anos finais. Essas médias indicam um aprendizado básico em Matemática cujos alunos dos anos iniciais possuem habilidades de reconhecer grandezas e medidas e de realizar operações de adição, subtração, multiplicação e divisão exata. Já os alunos das séries finais possuem habilidades como determinar a soma, a diferença, o produto ou o quociente de números inteiros em situações-problema e de resolver problemas envolvendo grandezas diretamente proporcionais, representadas por números inteiro.

⁷Para maiores informações, verificar a Escala de Proficiência no portal no INEP: <http://portal.inep.gov.br/web/guest/educacao-basica/saeb/matrizes-e-escalas>.

Tabela 1: Estatística Descritiva das variáveis utilizadas no DEA

Variáveis	Amostra com Grupo de Controle 1				Amostra com Grupo de Controle 2			
	2011		2015		2011		2015	
	Média	DP	Média	DP	Média	DP	Média	DP
Gasto com Educação por Aluno	4681,08	2086,77	5167,85	2298,73	4683,84	2121,66	5130,02	2373,15
Matemática Anos Iniciais	207,43	24,11	215,75	22,6	208,77	24,27	215,45	22,3
Matemática Anos Finais	245,97	19,86	250,35	16,86	246,72	19,81	250,43	16,9
Português Anos Iniciais	186,71	19,54	202,24	21,05	187,78	19,62	202,05	20,87
Português Anos Finais	236,64	16,77	245,77	16,72	237,63	16,9	245,88	16,68
Taxa de Aprovação Anos Iniciais	91,13	6,75	92,86	6,05	91,44	6,56	92,77	6,1
Taxa de Aprovação Anos Finais	83,24	9,71	84,98	9,22	83,4	9,47	84,85	9,13

Fonte: Elaboração dos Autores a partir dos dados coletados.

Legenda: DP = Desvio-Padrão.

Tabela 2: Estatística descritiva - vetor de características observáveis pré-tratamento

Variáveis	Eleições 2012		Eleições 2016	
	Média	DP	Média	DP
Idade	51,2	9,62	51,97	30,36
Homem	90%		87%	
Casado	77%		76%	
Ensino Superior	44%		51%	
Candidatos à Reeleição	19%		65%	
Partido do Governador	17%		21%	
PMDB	20%		17%	
PT	11%		10%	
PSDB	13%		14%	
PSB	5%		8%	

Fonte: Elaboração dos Autores a partir dos dados coletados.

Legenda: DP = Desvio-Padrão.

Com relação às taxas de aprovação, observa-se um aumento na média dessa taxa entre os municípios para o período entre 2011 e 2015 e, entre as abordagens, as médias das taxas de aprovação são também aproximadas. Quanto aos desvios-padrão, ao observar os gastos por aluno verifica-se uma alta dispersão nos insumos entre os municípios brasileiros na provisão de Educação.

Em linhas gerais, entre os anos considerados, observa-se um aumento no valor médio do insumo entre os municípios indicando um aumento nos repasses públicos para a Educação. Junto a isso, há um aumento no desvio-padrão dessa variável indicando uma possível desigualdade no volume de recursos aplicados. Com relação às notas e às taxas de aprovação, verifica-se um aumento no desempenho dos municípios, bem como uma redução no desvio-padrão dessa variável.

No tocante aos fatores observáveis utilizados no cálculo do escores de propensão, a Tabela 2 mostra a média de idade dos prefeitos presentes na amostra e o percentual de representação na amostra de cada categoria de interesse para o período das eleições entre 2012 e 2016. O perfil da amostra utilizada para a estimação por PSW é caracterizado por gestores com uma média de idade de quase 52 anos, com grande parte dos indivíduos homens e casados para o ano de 2012. No entanto, em 2016 a um relativo aumento da participação feminina como ocupante do cargo de Prefeito. Além disso, em 2012 cerca de 44% dos prefeitos possuíam Ensino Superior, aumentando para pouco mais de 50% em 2016. Em 2012, 19% dos prefeitos em primeiro mandato se candidataram à reeleição, percentual que se eleva para 65% em 2016. Quanto às variáveis políticas em 2012, 17% eram filiados ao partido do Governador do seu Estado, 20% ao PMDB, 11% ao PT, 13% ao PSDB e 5% ao PSB. Em 2016 os percentuais se alteram, sendo 21% de filiados ao partido do Governador do seu Estado, e os filiados do PSDB e PSB aumentando suas participações para 14% e 8%, respectivamente. Quanto aos filiados do PT e PMDB tiveram uma queda em suas participações, passando para 10% e 17%, respectivamente.

As estatísticas descritivas das variáveis utilizadas na estimação do modelo

Tabela 3: Estatística Descritiva das Variáveis empregadas nas análises do efeito da reeleição

Variáveis	Amostra com Grupo de Controle 1				Amostra com Grupo de Controle 2			
	Eleições 2012		Eleições 2016		Eleições 2012		Eleições 2016	
	Média	DP	Média	DP	Média	DP	Média	DP
IDFM-Educação	0,71	0,12	0,77	0,11	0,72	0,12	0,76	0,11
PIB per capita	14672,89	17574,84	19626,32	21445,98	15067,29	18724,97	19705,08	21555,1
Capital	0,29%		0,55%		0,58%		0,68%	
Ensino Superior	45,37%		52,57%		47,2%		52,67%	
Reeleição	19,34%		64,8%		21,96%		76,88%	
Partido do Governador	17,99%		21,05%		17,64%		22,24%	
PMDB	19,16%		16,68%		22,08%		17,63%	
PT	10,68%		9,72%		10%		9,82%	
PSDB	13,89%		14,24%		14,57%		13,86%	
PSB	5,59%		8,53%		5,4%		8,38%	

Fonte: Elaboração dos Autores a partir dos dados coletados.

Legenda: DP = Desvio-Padrão.

final, nas duas amostras e considerando as eleições de 2012 e 2016, estão descritas na Tabela 3. A variável *IFDM-Educação* apresentou médias aproximadas para as duas abordagens no valor de cerca de 0,71 para o ano de 2012 e 0,76 para o ano de 2016, apresentando também uma elevação entre o período. Uma vez que o índice *IFDM-Educação* varia entre zero e um, onde quanto mais próximo de 1 maior o desenvolvimento da região em termos educacionais, acredita-se que em média os municípios possuem um desenvolvimento educacional significativo.

Quanto ao *PIB per capita*, as médias e os desvios-padrão dos municípios entre as abordagens se diferenciam minimamente. No entanto, há uma elevação da média e do desvio-padrão entre os anos 2011 e 2015. Em ambos os casos, o desvio-padrão se mostrou elevado, indicando uma dispersão dos valores e revelando, portanto, as disparidades de atividade econômica entre os municípios brasileiros. Com relação ao identificados das capitais, verifica-se uma baixa participação destas, pois o número destas são extremamente menores em relação ao total de municípios. Verificou-se maior participação na segunda abordagem. Ademais, destaque-se que grande parte das capitais está no grupo dos municípios candidatos à reeleição.

Em relação às variáveis de interesse, o percentual de prefeitos candidatos à reeleição se diferenciam um pouco entre as abordagens, mas essa diferença é ainda maior entre os anos de eleição, indicando que na eleição de 2016 havia mais candidatos do que na eleição anterior. Com relação às características políticas dos candidatos, se é filiado ao mesmo partido do Governador do seu Estado e se é filiado ao PT, PMDB, PSB ou PSDB apresentam perfis aproximados em se tratando de abordagens. No entanto, na comparação entre os anos, houve uma elevação no percentual de Prefeitos que possuíam o mesmo partido que o Governador do seu Estado e quanto aos partidos de filiação. Por outro lado, o PMDB e PT tiveram uma redução na participação de Prefeitos ligados a esses partidos, ao passo que aumentou o número de Prefeitos filiados ao PSDB e PSB.

5 Resultados

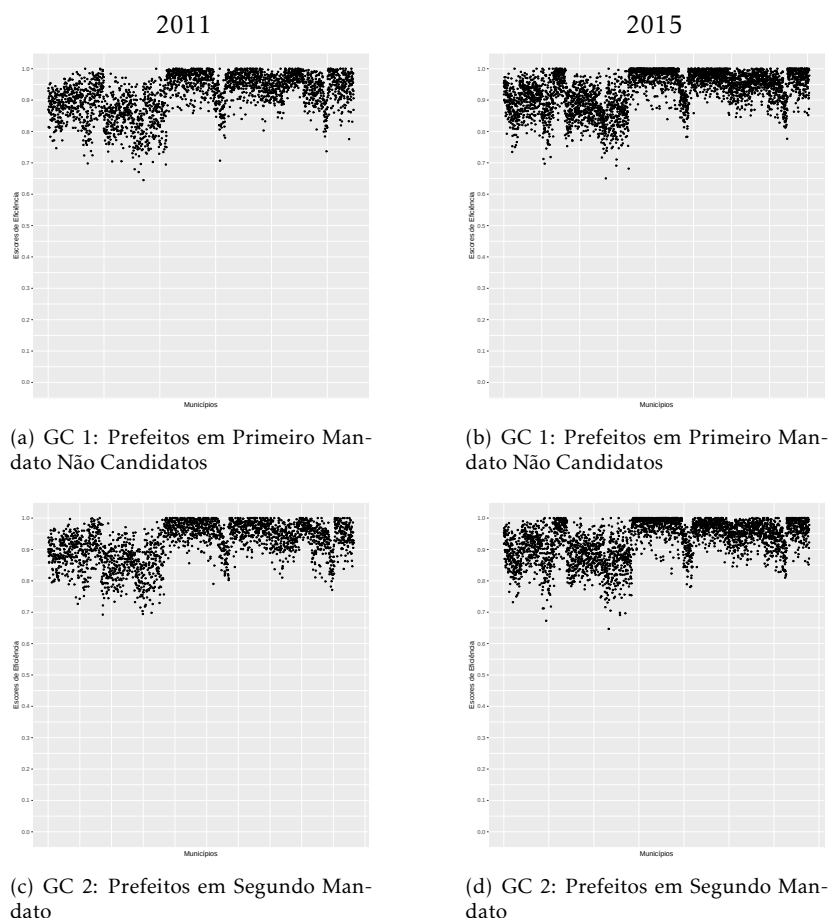
Esta seção contém os resultados estimados da relação entre eficiência dos gastos em educação e candidatura. Primeiramente, é feita uma análise geral da eficiência dos gastos públicos, mostrando o comportamento dos escores de eficiência entre os municípios. Em seguida procede-se à apresentação dos resultados estimados, considerando cada um dos grupos de controle já mencionados.

5.1 Análise da Eficiência dos Gastos Públicos em Educação

Para atender o objetivo de analisar os efeitos da candidatura à reeleição sobre a eficiência dos gastos municipais em Educação, partiu-se da construção dos escores de eficiência pelo DEA com orientação a produto e considerando retornos variáveis de escala (BBC), tendo em vista que os municípios possuem dimensões distintas e podem apresentar rendimentos de escala variados. Como insumo, empregou-se o volume de gastos com Educação por aluno. Como produto considerou-se as notas da Prova Brasil de Matemática e Português e a Taxa de Aprovação para os anos iniciais e finais, tanto para o ano de 2011 quanto para o ano de 2015.

A Figura 1 apresenta a distribuição dos escores de eficiência dos municípios brasileiros. Na primeira coluna estão as eficiências para o ano de 2011 e na segunda coluna estão as eficiências para o ano de 2015, para as amostras com cada grupo de controle. Observa-se que os níveis de eficiência estão em um intervalo entre 0.6 e 1. De todos os municípios considerados, aquele que apresentar escore igual a 1 é considerado eficiente, sendo assim, um município *benchmark*.

Figura 1: Escores de Eficiência dos Gastos Públicos em Educação dos Municípios Brasileiros



Fonte: Elaboração dos Autores a partir dos dados coletados.

Legenda: GC = Grupo de Controle

Pode-se destacar ainda que grande parte dos municípios estão no intervalo entre 0.9 e 1, ou seja, apresentaram escores menores que 1 e, portanto, podem ser considerados fracamente ineficientes na aplicação dos seus recursos em Educação. Por outro lado, o número de municípios *benchmark* chegou a um total de 58 e 213 para a amostra contendo o grupo de controle 1, para os anos de 2011 e 2015, respectivamente, e 72 e 155 municípios eficientes para o mesmo período na amostra que contém o grupo de controle 2. Esses municípios conseguiram obter um maior nível de produto, ou seja, conseguiram um

Tabela 4: Distribuição dos Produtos e do Gasto Público por Nível de Eficiência

Níveis de Eficiência	Matemática Anos Iniciais	Matemática Anos Finais	Português Anos Iniciais	Português Anos Finais	Tx. Aprov. Anos Iniciais	Tx. Aprov. Anos Finais	Despesa Por Aluno
GC 1: Prefeitos em Primeiro Mandato Não Candidatos							
Variável	2011						
Ineficiência Moderada	182,84	229,58	166,6	221,98	75,62	66,28	3161,18
Ineficiência Fraca	208,35	246,62	187,53	237,3	91,98	84,1	4747,91
Eficiente	236,86	263,49	207,76	249,44	97,81	93,63	6056,13
Média Geral	207,43	245,97	186,71	236,64	91,13	83,24	4681,08
Variável	2015						
Ineficiência Moderada	189,78	235,65	176,69	231,61	76,6	68,01	3708,46
Ineficiência Fraca	215,59	250,11	202,14	245,61	92,98	85,04	5099,83
Eficiente	232,29	262,15	217,55	256	99,48	92,96	7122,74
Média Geral	215,75	250,35	202,24	245,77	92,86	84,98	5167,85
GC 2: Prefeitos em Segundo Mandato							
Variável	2011						
Ineficiência Moderada	182,83	230,33	166,13	222,84	76,08	66,88	3154,82
Ineficiência Fraca	209,34	247,09	188,33	237,98	92,04	83,98	4722,44
Eficiente	235	263,2	207,37	251,92	98,88	93,36	6087,47
Média Geral	208,77	246,72	187,78	237,63	91,44	83,4	4683,85
Variável	2015						
Ineficiência Moderada	189,91	235,7	175,89	232,53	76,52	67,43	3615,46
Ineficiência Fraca	215,37	250,28	202,07	245,75	92,97	85,04	5074,78
Eficiente	234,25	263,3	219,1	257,5	99,65	92,72	7262,44
Média Geral	215,45	250,43	202,05	245,88	92,77	84,85	5130,02

Fonte: Elaboração dos Autores a partir dos dados coletados.

melhor desempenho nas notas de Português e Matemática e maiores Taxas de Aprovação na Prova Brasil, dado um nível de recurso público proporcional, em comparação com os demais municípios.

Em termos de níveis de eficiência, considera-se a seguinte classificação: o município k será eficiente nos gastos em educação quando $\theta_k = 1$; municípios com ineficiência fraca devem apresentar $0,8 \leq \theta_k < 1$; quando $0,6 \leq \theta_k < 0,8$ são classificados como sendo de ineficiência moderada e, por fim, municípios que apresentarem $\theta_k < 0,6$ são classificados como tendo ineficiência forte.

Seguindo essa classificação, a Tabela 4 apresenta a distribuição dos produtos (as notas de Matemática e Português e as Taxas de Aprovação nas séries iniciais e finais) e do insumo (gasto público com educação por aluno) para cada nível de eficiência.

Observa-se na Tabela 4 que as médias das notas de Português e Matemática e as taxas de aprovação ampliam-se à medida que aumentam os intervalos de eficiência. Portanto, aqueles municípios que são ditos eficientes possuem de fato melhores desempenhos nas disciplinas, tanto para anos iniciais quanto para anos finais. O mesmo comportamento é observado para a variável de insumo, onde à medida que se aumenta o intervalo de eficiência maior também será a despesa com educação por aluno. No entanto, a média tende a ser aproximada entre os intervalos, sugerindo que o valor médio dos gastos não se amplia muito com o aumento do intervalo de eficiência. Dessa maneira, conforme discutido anteriormente, os municípios eficientes são aqueles que conseguem alinhar um bom resultado, aqui em termos de desempenho educacional, dado o nível de recurso aplicado.

5.2 Relação entre Eficiência e Reeleição

A estratégia empírica para estimar o efeito da candidatura à reeleição considerando o grupo de controle 1 (prefeitos em primeiro mandato não candidatos à reeleição), conforme descrito na seção anterior, foi o pareamento baseado no *Propensity Score Weighting* (PSW), no qual os escores de propensão ao tratamento são utilizados como pesos amostrais na equação de regressão.

A Tabela 5 mostra os coeficientes estimados do modelo logit, utilizado para estimar os escores de propensão:

Os coeficientes da Tabela 5 mostram que com relação à idade os prefeitos mais velhos têm uma menor propensão a tentarem um novo pleito, sendo o coeficiente significativo para 2012 e 2016. É possível que com o passar da idade, esses políticos possuam menos incentivos a enfrentarem as dificuldades de uma nova administração. Prefeitos com ensino superior também aumentam a chance de se candidatarem à reeleição em ambas as eleições, ao passo que ser homem ou mulher não impacta em tal probabilidade. Um maior nível de escolaridade pode estar associado a uma maior capacidade de gestão e de reconhecimento de problemas e gargalos não resolvidos que uma nova gestão poderia resolvê-los. Para prefeitos casados, houve um aumento da propensão a concorrer à reeleição apenas em 2012.

O fato de o prefeito pertencer ao mesmo partido do governador do Estado foi significativo apenas na eleição de 2016. O apoio de um ente superior, manifestado por meio de parcerias em políticas públicas ao município pode incentivar os prefeitos a continuarem nos mandatos. Em um contexto de crise como em 2016, tal apoio pode ser mais importante ainda.

Tabela 5: Estimação do escore de propensão - Modelo logit

Variáveis	Eleições 2012	Eleições 2016
Idade	-0,0238*** (0,0049)	-0,0349*** (0,0031)
Sexo	-0,2261 (0,1413)	0,1163 (0,0966)
Ensino Superior	0,3336*** (0,0910)	0,1595* (0,0645)
Casado	0,2359* (0,1117)	0,0722 (0,0759)
Mesmo Partido do Governador	0,0571 (0,1468)	0,4770*** (0,0919)
PMDB	-0,1346 (0,1212)	0,0759 (0,0899)
PT	0,0447 (0,1507)	-0,4270*** (0,1163)
PSDB	-0,3059 (0,1707)	0,0260 (0,1079)
PSB	-0,4101 (0,2311)	0,0180 (0,1229)
Constante	-0,3009 (0,2800)	2,1084*** (0,1923)
N	3225	4521

Nota: erro-padrão entre parênteses. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

Fonte: elaboração própria a partir das estimações.

Pertencer ao mesmo partido presidente Temer não teve efeito estatisticamente significativo. Já o fato de pertencer ao PT reduziu a possibilidade de um prefeito em primeiro mandato se candidatar novamente para o ano de 2016, possivelmente devido ao efeito do impeachment da presidente Dilma Rousseff e queda de popularidade do partido. As denúncias de corrupção e os equívocos na gestão das finanças públicas durante o governo Dilma culminaram no impeachment da presidente e na ocorrência de uma severa crise econômica, afetando a imagem do partido junto aos eleitores e, portanto, fazendo com que muitos prefeitos reavaliassem suas chances de se submeterem a uma nova eleição. As *dummies* para o PSDB e PSB não foram significativas em nenhum dos anos. Após estimar os coeficientes, é importante analisar o poder preditivo do modelo de resposta binária usado na estimação do *propensity score*. A proporção de casos corretamente previstos foi de 0,30. Com relação às demais medidas de ajustamento, a sensibilidade foi de 97,37% indicando que cerca de 98% dos candidatos à reeleição são corretamente identificados. A medida de especificidade mostra que 7,16% dos não candidatos são corretamente identificados.

Os resultados são apresentados nas duas primeiras colunas da Tabela 6, cada qual com estimativas para as eleições de 2012 e 2016.

Importante lembrar que, como foi dito na introdução e na seção dos dados, as variáveis têm como fonte, informações coletadas nos anos imediatamente anteriores aos pleitos. Além disso, como os escores de eficiência foram calculados por um DEA orientado a produto, estes variam de 1 a ∞ , de maneira que quanto mais próximo de 1, mais eficiente é o município. Ou seja, municípios com maiores valores de escores apresentam maior ineficiência. Se,

Tabela 6: Efeito da Reeleição sobre a Eficiência de Gastos em Educação

Variáveis	Amostra com Grupo de Controle 1		Amostra com Grupo de Controle 2	
	Eleições 2012	Eleições 2016	Eleições 2012	Eleições 2016
Reeleição	0,0027 (0,0055)	-0,0007 (0,0040)	0,0016 (0,0052)	-0,0172*** (0,0047)
Capital	0,0447 (0,0426)	0,0413 (0,0252)	0,0305 (0,0264)	0,0358 (0,0225)
Ln PIB per capita	0,0200*** (0,0048)	0,0167*** (0,0036)	0,0186*** (0,0037)	0,0170*** (0,0036)
IDFM-Educação	-0,8857*** (0,0358)	-0,9733*** (0,0312)	-0,8421*** (0,0278)	-1,0284*** (0,0320)
Ensino Superior	0,0100 (0,0055)	0,0137*** (0,0040)	0,0099* (0,0045)	0,0082* (0,0041)
Partido do Governador	-0,0025 (0,0087)	0,0234*** (0,0057)	-0,0033 (0,0071)	0,0292*** (0,0053)
PMDB	0,0030 (0,0074)	0,0056 (0,0057)	0,0124* (0,0054)	0,0017 (0,0057)
PT	0,0082 (0,0090)	-0,0118 (0,0074)	0,0051 (0,0077)	-0,0107 (0,0075)
PSDB	0,0041 (0,0105)	-0,0255*** (0,0074)	-0,0028 (0,0081)	-0,0244*** (0,0074)
PSB	0,0393** (0,0126)	0,0233*** (0,0069)	0,0472*** (0,0092)	0,0194** (0,0069)
Constante	1,4799*** (0,0374)	1,5836*** (0,0307)	1,4692*** (0,0288)	1,6349*** (0,0313)
N	2735	4023	2409	3391

Nota: erro-padrão entre parênteses. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

Fonte: elaboração própria a partir das estimações.

por exemplo, uma determinada covariada apresenta sinal negativo, isso significa que há uma redução no escore e, portanto, uma tendência de redução da ineficiência.

De acordo com as estimativas, a candidatura à reeleição, tanto em 2012 quanto em 2016, não afetou a eficiência dos gastos públicos do ano anterior à eleição. Ou seja, em termos de eficiência na gestão dos recursos, municípios com prefeitos candidatos à reeleição não apresentam diferenças estatisticamente significantes com relação a municípios cujos prefeitos não concorreram novamente. A evidência encontrada não corrobora a hipótese de que prefeitos candidatos à reeleição tendem a gastar mais ou apresentam resultados de política pública melhores no seu primeiro mandato em comparação com aqueles que mesmo podendo não concorreram novamente.

No tocante às variáveis de controle, considerando as duas eleições, apresentaram efeito estatisticamente significativo o logaritmo do PIB per capita, o nível médio de educação do município e se o partido do prefeito é o PSB. Além disso, em ambas as eleições, essas variáveis apresentaram o mesmo sinal. Já considerando somente o pleito de 2016, apresentaram significância estatística, além dessas mesmas variáveis, o fato de o prefeito ter ensino superior e se este era do mesmo partido do governador do Estado.

O coeficiente do PIB per capita foi positivo indicando que municípios com maior renda média apresentam os maiores escores e, portanto, se mostram mais ineficientes. Esse resultado chama atenção, pois espera-se que cidades mais ricas tenham condições mais propícias para maximização dos seus resultados. Por outro lado, por serem cidades, em geral, mais populosas e com vários problemas sociais, como a violência e desigualdades sociais, demandam políticas públicas mais complexas, as quais são mais difíceis de gerenciar. O acesso mais facilitado aos recursos públicos também pode ser um fator que ajuda a explicar a maior ineficiência dessas localidades, uma vez que estão associadas a uma maior dificuldade em maximizar os resultados das políticas.

Com relação ao nível médio de educação, o coeficiente foi negativo, indicando que municípios cuja população apresenta maior nível de educação média tendem a ser menos ineficientes. O nível de educação médio da população ajuda a explicar parte do melhor desempenho desses municípios nos testes de proficiência.

Quanto às variáveis políticas, ter um prefeito filiado ao PSB tende a aumentar a ineficiência no gerenciamento dos recursos públicos. Para as eleições de 2016, municípios cujos prefeitos estavam filiados ao PSDB apresentaram menor ineficiência comparativamente aos demais municípios. Esse resultado pode ser explicado pelas diferenças ideológicas dos partidos quanto à gestão pública e utilização de um maior quantitativo de gastos, principalmente pelo fato da queda de popularidade do PT e do histórico de antagonismo com o PSDB.

Outro resultado mostrado considerando a candidatura em 2016 é que fazer parte do mesmo partido do governador do estado é um fator que eleva a ineficiência dos gastos em educação. Fazer parte do mesmo grupo político que governa o estado pode criar uma expectativa no prefeito de que as políticas estaduais de educação beneficiem o seu município em maior grau. Assim, é possível que o gestor não apresente tanto empenho em desenvolver políticas locais.

Por fim, o fato de o prefeito ter nível superior está relacionado com maiores níveis de ineficiência. Esse resultado diverge do consenso de que gestores mais

escolarizados apresentam maior capacidade gerencial. Uma das possíveis explicações para esse resultado pode ser pelo fato de o gestor mais qualificado conhecer caminhos alternativos para driblar amarras legais e dessa forma ter um gasto que não se traduz em resultado.

Na segunda amostra, o grupo de controle conta municípios cujos prefeitos já se encontravam em segundo mandato. Aqui, o coeficiente da variável de tratamento apresentou significância estatística para o pleito de 2016. Em comparação com a amostra contendo os municípios do primeiro grupo de controle, houve significância estatística nas mesmas variáveis de controle e verificou-se os mesmos sinais nos coeficientes estimados.

O sinal negativo mostra que municípios com prefeitos em primeiro mandato candidatos à reeleição se mostraram menos ineficientes na gestão dos gastos em educação, comparativamente àqueles cujos prefeitos, em razão da legislação eleitoral, não mais poderiam se reeleger.

Esse resultado chama atenção e é revelador da perda de capacidade gerencial que implica um segundo mandato. Conforme mostrado na introdução, dentre os resultados previstos a partir dessa abordagem, este se justifica pela ausência da possibilidade da candidatura; tomando por base os modelos de reputação, prefeitos em segundo mandato, em tese, não necessitam impulsionar sua reputação junto ao eleitorado, o que pode afetar a sua performance na geração de resultados educacionais. Em suma, para os municípios a expectativa da reeleição é benéfica ao reduzir a ineficiência dos gastos, no entanto, efetivamente reeleger um prefeito tende a reverter esses benefícios.

As evidências encontradas parecem corroborar também as previsões contidas nos modelos de reputação descritos na seção 2, nos quais os prefeitos que desejam se reeleger tendem maximizar os resultados de suas políticas públicas como maneira de melhorar sua reputação junto à população e, com isso, aumentarem as chances de ganharem um segundo mandato.

6 Considerações Finais

O impacto econômico dos incentivos políticos eleitorais é tema recorrente na literatura econômica. A limitação dos mandatos eletivos, particularmente, se mostra como um incentivo importante à tomada de decisão por parte dos agentes públicos, vez que esta afeta a função objetivo do gestor. Com base nesse arcabouço, o objetivo do artigo foi investigar como a existência do dispositivo de reeleição no sistema eleitoral afeta a eficiência na gestão dos gastos. Para isso elegemos como variável de tratamento se o município teve candidato à reeleição. Porém, com uma mesma variável de tratamento, analisamos o efeito da reeleição em duas amostras, cada uma com um grupo de controle diferente.

Na primeira o grupo de controle foi formado por municípios que tinham prefeitos em primeiro mandato, porém que não se candidataram novamente, mesmo a legislação permitindo a renovação. Com essa abordagem tentamos captar se a candidatura faz desses prefeitos mais eficientes em gerir recursos do que aqueles que não quiseram se candidatar novamente, mesmo podendo concorrer. Na segunda o grupo de controle foi formado por municípios com prefeitos já em segundo mandato. Aqui também se investigou como a candidatura afeta a eficiência, porém em comparação com um grupo de prefeitos que já se candidatou à reeleição no passado, ou seja, um grupo semelhante

quanto ao desejo de concorrer, porém que não pode mais se reeleger. Também procurou-se verificar o efeito candidatura, porém foi possível obter respostas sobre a eficiência de quem já se reelegeu. Para a estimação da primeira abordagem foi necessário parear os dois grupos de municípios quanto a probabilidade de ter um prefeito em primeiro mandato candidato à reeleição. Foram calculados escores de propensão, que foram utilizados como ponderadores da regressão principal.

Os resultados mostraram que não há diferença de eficiência de gestão entre prefeitos em primeiro mandato candidatos à reeleição e prefeitos em primeiro mandato não candidatos. Já comparados com prefeitos que se encontravam em segundo mandato, os prefeitos candidatos à reeleição se mostraram menos ineficientes na gestão dos gastos em educação. Em outras palavras, o que esse resultado mostra é que permitir que os prefeitos se candidatem novamente pode reduzir a ineficiência dos gastos durante o primeiro mandato. No entanto, efetivamente reeleger esses prefeitos pode afetar seu empenho em maximizar os resultados das políticas públicas no segundo mandato.

Desse modo, o dispositivo da reeleição gera uma ambiguidade nos incentivos aos administradores públicos quanto à capacidade gerencial. Isto não significa necessariamente que o dispositivo da reeleição deva ser eliminado do sistema eleitoral, pois os seus efeitos sobre a gestão pública envolvem uma série de outras questões complexas, como exposição à corrupção, fatores históricos e culturais intrínsecos à política local, que podem ser melhor explorados em análises empíricas futuras. É necessário trazer ao debate a necessidade de desenvolver outros mecanismos de incentivo que melhorem a qualidade da gestão durante todo o mandato, para além da política partidária. São eles: disponibilizar instrumentos mais eficazes de *accountability* para que a população consiga acompanhar de forma mais transparente a gestão dos recursos, melhorar os indicadores de resultado das políticas públicas, de maneira que estes reflitam de forma mais fidedigna os reais benefícios e custos dessas políticas (avaliação das políticas públicas com base em evidências empíricas robustas) e um trabalho de educação junto à população acerca da importância desses dispositivos. Discussões sobre mecanismos de punição também podem ser importantes; auditorias que evitem manipulações e comportamento oportunista, como por exemplo os sorteios de municípios a serem auditados pelo Tribunal de Contas da União, podem gerar incentivos para uma melhor aplicação dos recursos públicos, independente de se estar no primeiro ou segundo mandato ou ser um candidato a reeleição.

Referências

- ALMEIDA, Rodrigo Borges; SAKURAI, Sergio Naruhiko *et al.* Incentivos eleitorais e regras fiscais (não tão) rígidas: novas evidências para os municípios brasileiros a partir da rubrica restos a pagar. *In: XLIV ENCONTRO NACIONAL DE ECONOMIA. Anais*. Campinas: ANPEC-Associação Nacional dos Centros de Pós-Graduação em Economia, 2018.
- AMARANTE, José Carlos Araújo. *Impactos dos incentivos políticos institucionais sobre a melhoria dos serviços públicos municipais: evidências a partir do processo de reeleição dos prefeitos no Brasil*. Dissertação (Mestrado em Economia) – Universidade Federal da Paraíba, João Pessoa, 2017.

- BANKER, Rajiv D.; CHARNES, Abraham; COOPER, William Wager. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, v. 30, n. 9, p. 1078–1092, 1984.
- BERNARDO, Marcella; RODRIGUES, Lásara Fabrícia. Análise envoltória de dados: aplicação do modelo CCR e do modelo BCC para a avaliação do desempenho de bibliotecas universitárias de uma IFES. *RACEF – Revista de Administração, Contabilidade e Economia da Fundace*, v. 6, n. 2, p. 80–96, 2015.
- BESLEY, Timothy. *Principled agents?: The political economy of good government*. Oxford: Oxford University Press, 2006.
- BESLEY, Timothy; CASE, Anne. Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *Quarterly Journal of Economics*, v. 110, n. 3, p. 769–798, 1995.
- BESLEY, Timothy; CASE, Anne. Political institutions and policy choices: evidence from the United States. *Journal of Economic Literature*, v. 41, n. 1, p. 7–73, 2003.
- CAVALCANTE, Pedro. Desempenho fiscal e eleições no Brasil: uma análise comparada dos governos municipais. *Revista de Administração Pública*, v. 50, n. 2, p. 307–330, 2016.
- CHARNES, Abraham; COOPER, William W.; RHODES, Edwardo. Measuring the efficiency of decision making units. *European Journal of Operational Research*, v. 2, n. 6, p. 429–444, 1978.
- FERREIRA, Carlos Maurício de Carvalho; GOMES, Adriano Provezano. *Introdução à análise envoltória de dados: teoria, modelos e aplicações*. Viçosa, MG: Editora UFV, 2009.
- GUO, Shenyang; FRASER, Mark W. *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: SAGE, 2014.
- JOHNSON, Joseph M.; CRAIN, W. Mark. Effects of term limits on fiscal performance: Evidence from democratic nations. *Public Choice*, v. 119, n. 1-2, p. 73–90, 2004.
- LOPES, Luckas S. *et al.* Eficiência técnica municipal na gestão dos gastos com saúde e educação em Minas Gerais: seus impactos e determinantes. In: XIII SEMINÁRIO SOBRE A ECONOMIA MINEIRA. *Anais*. Diamantina: Universidade Federal de Minas Gerais - Cedeplar, 2008.
- LOPES, Matheus Costa Monteiro. *Análise da eficiência do gasto público em educação para os municípios brasileiros*. Dissertação (Mestrado em Economia) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.
- MARCINIUK, Fernanda L.; BUGARIN, Maurício S. A influência da reeleição nas políticas fiscais subnacionais. *Revista Brasileira de Economia*, v. 73, n. 2, p. 181–210, 2019.

- MARINHO, Alexandre. Avaliação da eficiência técnica nos serviços de saúde nos municípios do Estado do Rio de Janeiro. *Revista Brasileira de Economia*, v. 57, n. 3, p. 515–534, 2003.
- MENEGUIN, Fernando B.; BUGARIN, Mauricio Soares. Reelection e política fiscal: um estudo dos efeitos da reeleição nos gastos públicos. *Economia Aplicada*, v. 5, n. 3, p. 601–622, 2001.
- MOTTA, Ronaldo Serôa; MOREIRA, Ajax. Political factors and the efficiency of municipal expenditure in Brazil. *Economia Aplicada*, v. 13, n. 4, p. 353–370, 2009.
- NAKAGUMA, Marcos Yamada; BENDER, Siegfried. A emenda da reeleição e a Lei de Responsabilidade Fiscal: impactos sobre ciclos políticos e performance fiscal dos Estados (1986-2002). *Economia Aplicada*, v. 10, n. 3, p. 377–397, 2006.
- NOGARE, Chiara Dalle; RICCIUTI, Roberto. Do term limits affect fiscal policy choices? *European Journal of Political Economy*, v. 27, n. 4, p. 681–692, 2011.
- NOVAES, Lucas; MATTOS, Enlinson. O efeito da intenção de reeleição sobre gastos em saúde: uma análise com base no modelo de reputação política. *Brazilian Journal of Political Economy*, v. 30, n. 1, p. 140–158, 2010.
- OLIVEIRA, Victor; SOUZA, Wallace; ANNEGUES, Ana. Eficiência e Autonomia Escolar: Evidências para o Período 2007-2013. *Economia Aplicada*, v. 22, n. 2, p. 5–46, 2018.
- OLMOS, Antonio; GOVINDASAMY, Priyalatha. A practical guide for using propensity score weighting in R. *Practical Assessment, Research, and Evaluation*, v. 20, n. 1, p. 13, 2015.
- RIBEIRO, Nuno *et al.* Local Governments' Efficiency: Study of its Determinants. *IBIMA Business Review*, p. 1–14, 2020.
- ROGOFF, Kenneth. Equilibrium Political Budget Cycles. *American Economic Review*, v. 80, n. 1, p. 21–36, 1990.
- ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41–55, 1983.
- SILVA, Felipe de Figueiredo; BRAGA, Marcelo José. Determinantes da Reelection Municipal da Bahia para o Período de 2000 a 2008. *Planejamento e Políticas Públicas*, n. 40, 2013.
- SILVA, Jorge Luiz Mariano; ALMEIDA, Júlio César Lima. Eficiência no gasto público com educação: uma análise dos municípios do Rio Grande do Norte. *Planejamento e Políticas Públicas*, n. 39, 2012.

- SIMAR, Leopold; WILSON, Paul W. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, v. 136, n. 1, p. 31–64, 2007.
- SOUSA, Maria da Conceição Sampaio de; RAMOS, Francisco S. Eficiência técnica e retornos de escala na produção de serviços públicos municipais: o caso do Nordeste e do Sudeste brasileiros. *Revista Brasileira de Economia*, v. 53, n. 4, p. 433–461, 1999.
- SPENCE, A. Michael. The economics of internal organization: An introduction. *Bell Journal of Economics*, p. 163–172, 1975.
- SPENCE, Michael; ZECKHAUSER, Richard. Insurance, Information, and Individual Action. In: *Uncertainty in Economics*. Organização: Peter Diamond e Michael Rothschild. Amsterdam: Elsevier, 1978. p. 333–343.
- TROMPIERI NETO, Nicolino *et al.* Determinantes da eficiência dos gastos públicos municipais em educação e saúde: o caso do Ceará. In: *Economia do Ceará em Debate*. Organização: Eveline Barbosa Silva Carvalho. Fortaleza: IPECE, 2009. p. 57–72.

VOLUME, LEARNING-BY-DOING, AND ECONOMIES OF SCALE IN THE BRAZILIAN HOSPITAL SECTOR: THE CASE OF HIP ARTHROPLASTY

PEDRO HENRIQUE SOARES LEIVAS ^{*}
PAULO DE ANDRADE JACINTO [†]
ANDERSON MOREIRA ARISTIDES DOS SANTOS [‡]
LÍVIA MADEIRA TRIACA [§]
CESAR AUGUSTO OVIEDO TEJADA [¶]

Resumo

Este estudo visa verificar se o resultado de saúde dos pacientes submetidos a Artroplastia de Quadril são melhores à medida que o hospital realiza mais desse procedimento (efeito volume-resultado). As evidências apontaram a existência desse efeito no setor hospitalar brasileiro. Explorando a distribuição geográfica das residências dos pacientes e dos hospitais como fonte de variação exógena para o volume, foi identificado que o efeito é resultado da hipótese de *practice-makes-perfect*, ocorrendo através de ganhos estáticos de escala. Esse resultado sugere que uma política que visa um volume mínimo por estabelecimento hospitalar pode conduzir a melhores resultados de saúde.

Palavras-chave: efeito volume-resultado; mortalidade hospitalar; economias de escala.

Abstract

This study's objective was to verify if health outcomes of patients undergoing hip replacement are better as a hospital performs more of this procedure – the “volume-outcome effect” – in the Brazilian hospital sector. Evidence pointed to the existence of this effect in Brazil. Exploring the geographical distribution of patients' residences and hospitals' location as a source of exogenous variation for volume, we could identify that the effect is a result of the practice-makes-perfect hypothesis, occurring through static scale economies. This result suggests that a policy of establishing a minimum volume can lead to better health outcomes.

Keywords: volume-outcome effect; in-hospital mortality; practice-makes-perfect; static scale economies.

JEL classification: I11, I18

DOI: <http://dx.doi.org/10.11606/1980-5330/ea192619>

^{*} Universidade Federal do Rio Grande. E-mail: pedroleivas@furg.br

[†] Universidade Federal do Paraná e Doutorado Profissional em Políticas Públicas da Enap. E-mail: paulo.jacinto@ufpr.br

[‡] Universidade Federal de Alagoas. E-mail: anderson.santos@feac.ufal.br

[§] Universidade Federal do Rio Grande. E-mail: livia.triaca@furg.br

[¶] Universidade Federal de Pelotas. E-mail: cesaroviedotejada@gmail.com

1 Introduction

One of the central goals of health-sector policies is to improve the quality and efficiency of health systems. According to data from the World Health Organization (WHO), between 5.7 and 8.4 million deaths per year are attributed to poor quality healthcare in low- and middle-income countries. This inadequate quality represents high costs in terms of the lost productivity of health services and affects the well-being of the population (PAHO, 2020).

Many policy initiatives are based on the idea that higher-volume health-care providers can provide more effective results, that is, better quality (RACHET-JACQUET; GUTACKER; SICILIANI, 2021; SUND, 2010). Empirical evidence suggests a correlation between the volume of a medical procedure performed in hospitals and the health outcomes of treated patients¹. Specifically, it is found that hospitals with a large number of patients with specific diagnoses or procedures have lower mortality rates. This correlation is known in the literature as the volume-outcome effect, which means that the quality of health services improves as the number of patients increases and healthcare staff gets more used to providing a certain treatment (the practice-makes-perfect hypothesis) (LUFT; HUNT; MAERKI, 1987).

The practice-makes-perfect hypothesis rests on the notion that increased experience results in more developed skills, and consequently, better results. Thus, hospitals that care for many patients with a specific condition develop better skills in treating that condition, leading to better outcomes. At the same time, the selective referral hypothesis emerges from the possibility of patients choosing (or being referred by their physicians) hospitals or specialists with better results. Although it is unlikely that patients have information about the quality of hospitals or specialists, it is possible that their physicians have some knowledge in this regard, mainly based on the results of previously referred patients. Furthermore, even if majority patients seek care at the nearest hospital or make their choices regardless of perceived outcomes, a minority seeking the “best physician in town” will result in a pattern of selective referral for specific diagnoses or procedures. In other words, hospitals with better results will have a higher volume than expected (LUFT; HUNT; MAERKI, 1987).

Identifying causality is important for implementing public policies related to improving the quality and efficiency of hospitals. If we find that volume leads to better results, concentration of procedures in a small number of hospitals will be indicated. Hospitals that specialize in performing one or fewer procedures may have the benefit of producing better results. Otherwise, if causality goes from results to volume, these issues become irrelevant (GAYNOR; SEIDER; VOGT, 2005).

The Brazilian hospital system is an interesting case for this analysis. Most hospitals in the country are considered inefficient and of low quality² (BOTEGA;

¹See, for example, Luft, Bunker, and Enthoven (1979), Luft, Hunt, and Maerki (1987), Hannan *et al.* (1992), Ho (2000, 2002), Gaynor, Seider, and Vogt (2005), Kahn, Have, and Iwashyna (2009), Braun (2014), Hentschker and Mennicken (2014, 2015), Birkmeyer *et al.* (2002), and Gowrisankaran, Ho, and Town (2008). However, LaPar *et al.* (2012) found no evidence that volume is a good predictor of mortality and, therefore, should not be used as a proxy for surgical quality. Meanwhile, Reames *et al.* (2014) highlighted that, even with the advances achieved in terms of safety of the surgical process in recent years, volume remains a good proxy for quality, especially for high-risk procedures performed less frequently.

²According to the most recent study, Botega, Andrade, and Guedes (2020), Brazil's hospital efficiency score is 0.59, below the average score of Organisation for Economic Co-operation and

ANDRADE; GUEDES, 2020; LA FORGIA; COUTTOLENC, 2008). This high rate of inefficiency is not a regional phenomenon; it is observed in all states of the country and is mainly associated with the small size of Brazilian hospitals (BOTEAGA; ANDRADE; GUEDES, 2020). According to La Forgia and Couttolenc (2008), most Brazilian hospitals are small, with fewer than 50 beds (about 60%). Despite being important for the decentralization of services, these small hospitals face barriers that impede them from reaching scale and making their management difficult. In contrast, larger hospitals have superior performance resulting from higher volume of care and lower uncertainty regarding the provision of resources (BOTEAGA; ANDRADE; GUEDES, 2020; RAMOS *et al.*, 2015). Some evidence in the international literature, following the logic of economies of scale and scope, suggests that the ideal size of a hospital is 100-450 beds (POSNETT, 1997; SALTMAN; FIGUERAS, 1997). Therefore, the small size of most hospitals in Brazil can have important implications in terms of quality and scale of efficiency.

However, the volume-outcome effect has not received much attention in Brazil, given the relatively small number of studies that address this topic in the country³. In a systematic review of the literature on the effectiveness of care in the Brazilian hospital sector between 1990 and 2011, only six studies considered the number of procedures in their analyses⁴ (MACHADO; MARTINS; MARTINS, 2013).

Although these studies make important contributions to the literature that investigates the volume-outcome effect, all of them have an essential limitation: they do not look at the relationship from a causal point of view. These studies do not differentiate between the hypotheses of “practice-makes-perfect” and “selective referral,” which is crucial for the formulation of public policies. The concentration of procedures in a smaller number of hospitals with the objective of enhancing the results will make sense only if it is possible to isolate the hypothesis of “practice-makes-perfect”. In view of these methodological limitations, combined with a hospital sector characterized by small establishments, it seems opportune to conduct studies that seek to investigate possible gains of scale in this sector.

This study aims to investigate the volume-outcome effect in the Brazilian hospital sector for the hip arthroplasty procedure in the period 2008-2014. Hip arthroplasty is a surgical procedure that involves replacing an injured natural hip with another prosthetic hip. Despite being an elective surgery that is reasonably safe and performed in relatively healthy individuals, there is a risk that the patient may develop venous thrombosis, pulmonary embolism, or even surgical site infection (CASSONE *et al.*, 2002; ERCOLE; CHIANCA, 2002; LENZA *et al.*, 2013; PEREIRA *et al.*, 2014). Given these possible surgical complications, perioperative care is of paramount importance and is directly related to the quality of hospital care, which makes hip arthroplasty, an outcome widely used in the international literature, an appropriate measure for the study.

It is worth mentioning that this study advances the national literature in at least three aspects. First, it investigates the volume-outcome effect for a pro-

Development (OECD) countries, which ranges from 0.77 to 0.82.

³See, for example, Noronha *et al.* (2003), Martins, Blais, and Leite (2004), Ribeiro *et al.* (2006), Godoy *et al.* (2007), Piegas, Bittar, and Haddad (2009), Gomes *et al.* (2010), and Piegas and Haddad (2011).

⁴Noronha *et al.* (2003), Martins, Blais, and Leite (2004), Godoy *et al.* (2007), Piegas, Bittar, and Haddad (2009), Gomes *et al.* (2010), and Piegas and Haddad (2011).

cedure that, to the best of our knowledge, has not yet been studied in Brazil. With the demographic transition process in Brazil, which began in 1970, the number of elderly people is expected to exceed that of children and teenagers in 2050 by more than 38 million (BRASIL – MINISTÉRIO DA SAÚDE, 2010). With advancing age, individuals become more susceptible to illnesses and functional limitations (MESQUITA *et al.*, 2009). Fractures of the femoral neck are among the diseases that most affect the elderly. These fractures are one of the most common causes of hospitalization among the elderly and deserve attention not only for their frequency but also for their severity because they have a comparatively high mortality rate (MESQUITA *et al.*, 2009). Second, it seeks to analyze the relationship from a causal perspective, differentiating the hypotheses of practice-makes-perfect from selective referral, which is indispensable for the formulation of public policies that aim to improve the quality of hospital care in Brazil. Finally, we seek to verify whether the volume-outcome effect occurs due to static scale gains or through learning-by-doing. As highlighted by Gaynor, Seider, and Vogt (2005), if scale gains are observed, it does not matter much in which hospitals the procedures are concentrated because an increase in the volume of any hospital implies better results. In the case of learning-by-doing, reallocation of volume from one hospital to another can imply the loss of experience accumulated in the hospital that lost volume. In this case, concentrating volume in a hospital can reduce the net benefit.

To achieve this objective, we used an instrumental variable approach to explore the geographic distribution of patients' homes and hospitals as sources of exogenous variation in volume. Through this analysis, we were able to identify the existence of a volume-outcome effect for hip arthroplasty in the Brazilian hospital sector, and that this effect is a result of the practice-makes-perfect hypothesis and occurs through static gains of scale.

In addition to this introduction, this paper has four sections. Section 2 presents a review of the literature relevant to the topic, both international and national. Section 3 presents the data and methodology used to achieve the proposed objectives. In Section 4, the results are presented and discussed, and Section 5 concludes the paper.

2 Literature review

Halm, Lee, and Chassin (2002) analyzed 135 studies in a systematic review of the literature and concluded that the volume-outcome effect is found for a wide range of procedures and conditions; however, the magnitude of the association varies considerably. Many studies, due to their methodological limitations, emphasize that the implications of results, in clinical terms and in terms of the implementation of public policies aimed at the hospital sector, become complicated. Scholars also pointed out that differences in the case mix and service processes between providers with high and low volumes may explain part of the observed relationship.

In Table A.1 in the Appendix A, we present some studies that seek to analyze the volume-outcome effect, both internationally and nationally. Regarding international literature, despite the fact that it is heterogeneous in terms of methods and procedures and/or conditions investigated, it is clear that most studies focus on the hospital sector in the United States of America (USA). Exceptions are the studies by Braun (2014), Hentschker and Mennicken (2014,

2015), Kaier *et al.* (2018), Avdic, Lundborg, and Vikström (2019), and Rachet-Jacquet, Gutacker, and Siciliani (2021). Another important characteristic of this literature is that most studies seek to deal with both the hypotheses: practice-makes-perfect and selective referral.

Gaynor, Seider, and Vogt (2005) sought to verify whether the effect of volume on results is mainly contemporary – the result of static economies of scale – or occurs via lags under the hypothesis of learning-by-doing. Their study focused on a specific procedure, myocardial revascularization surgery (MRS). Using data from California for the period 1983-1999, this study showed that the direction of causality goes mainly from volume to result and that the effect is a function mostly of economies of scale.

Braun (2014) sought to verify this relationship in the context of the National Health Service in Portugal. Considering hospital discharges from 2001 to 2008, this study's results showed evidence favorable to the volume-outcome effect for seven of the 21 homogeneous diagnosis groups (HDG) investigated at the hospital level and for 10 GDH at the patient level. Their evidence, as well as that of Gaynor, Seider, and Vogt (2005), show the impact of static economies of scale, with few results suggesting the existence of learning-by-doing. Avdic, Lundborg, and Vikström (2019) also found a positive relationship between volume and survival in more than 100,000 patients, who underwent surgery for advanced cancer in Sweden in the period 1998-2007; however the study found evidence that the relationship occurs through the learning-by-doing hypothesis.

Unlike previous studies, Rachet-Jacquet, Gutacker, and Siciliani (2021) did not find causal evidence for the existence of the volume-outcome effect. This study analyzed the data of 105,229 patients undergoing hip replacement surgery in England and found evidence of only a positive association between volume and better health outcomes; however, when analyzing the relationship from a causal perspective, the effect became statistically non-significant.

In Brazil, the topic has received little attention given the small number of studies that relate the quality of care to the volume of procedures (Table A.1 in the Appendix A). Machado, Martins, and Martins (2013) showed 32 studies that analyzed the effectiveness of care in Brazilian hospitals, and of these studies, only six seek to consider the volume of procedures in their analyses.

Similar to Gaynor, Seider, and Vogt (2005), Noronha *et al.* (2003) sought to highlight the relationship considering MRS. These studies considered MRS financed by the Unified Health System (SUS) from 1996 to 1999. Further, these studies emphasized that they considered only a certain type of MRS carried out in the country, as they did not include procedures financed by private health plans or those paid with resources of patients themselves. The results showed that the risk of death decreases considerably as the number of procedures increases.

Other studies, despite not having the main objective of analysis, ended up relating the volume of procedures to health results achieved by patients. Martins, Blais, and Leite (2004) sought to assess hospital performance by focusing on two dimensions associated with hospital production: effectiveness of care (measured by mortality) and technical efficiency (measured by the length of stay). Specifically, these scholars sought to verify whether there are variations in such dimensions when considering different legal natures, namely public and private. They employed a multilevel logistics model in a two-level hierarchical structure, in which the first corresponded to hospitalizations and

the second to hospitals. They emphasized that the model was chosen due to the lack of independence between the observations because patients treated in the same hospital share common characteristics, using, in addition to other hospital-related variables, the volume of cases as a control. Their evidence suggested that the volume of procedures does not have a statistically significant effect on either mortality or the length of stay.

[Godoy et al. \(2007\)](#) analyzed the lethality of SUS-funded coronary angioplasty in the state of Rio de Janeiro and showed that the lowest gross lethality rates were concentrated in hospitals with a higher volume of procedures. However, when adjusting for sex, age, and diagnostic groups, the magnitude of the inverse case-fatality-volume effect reduced considerably.

[Gomes et al. \(2010\)](#) sought to verify the factors associated with hospital mortality in the SUS network in Rio Grande do Sul, also applying a multilevel logistics model. These scholars found no evidence that the volume of procedures has an effect on mortality, stressing that this result may reflect the fact that their study assessed hospital mortality globally, while the volume is more relevant for the analysis of mortality for specific causes.

Another study that addressed the relationship between the volume of procedures and patient outcomes is [Piegas and Haddad \(2011\)](#), who studied percutaneous coronary interventions performed under SUS funding. These scholars found no evidence of an association between volume and mortality. However, [Piegas, Bittar, and Haddad \(2009\)](#) showed that mortality after MRS is higher in low-volume hospitals.

Although studies carried out in Brazil have made important contributions to the literature on the volume-outcome effect, they have some methodological limitations. For example, only [Martins, Blais, and Leite \(2004\)](#) and [Godoy et al. \(2007\)](#) addressed the problem of dependence between observations. [Piegas, Bittar, and Haddad \(2009\)](#) and [Piegas and Haddad \(2011\)](#) related mortality and volume without controlling for characteristics, such as sex, age, and character of care. However, the biggest limitation of these studies is not that they did not differentiate between the hypotheses of “practice-makes-perfect” and “selective referral,” which is crucial for the formulation of public policies. The indication of concentration of procedures with the objective of enhancing the results will only be supported if it is possible to isolate the hypothesis of “practice-makes-perfect.”

3 Data and Methodology

3.1 Data

The procedure studied in this study (hip arthroplasty) is the surgical intervention that consists of “...replacement of the biological hip joint with inorganic metallic or polyethylene joint components...” ([BRASIL – MINISTÉRIO DA SAÚDE, 2016](#)). The procedure is mainly used for coxarthrosis, which involves wear and tear of the hip cartilage. To analyze the effect of volume on the results of this procedure, the data from the Hospital Information System (SIH), supplied by the Hospital Admission Authorizations (AIH), as well as by the National Registry of Health Establishments (CNES) and the Unified Health System (SUS), for the period 2008-2014 was used⁵.

⁵In some cases, data from 2004 to 2007 were used to calculate lagged volumes. The period chosen is justified by the availability of georeferencing data, which were used to construct the

SIH and SUS do not have universal coverage, which means data from these institutions are available for only those procedures financed by them, as they do not have data for all procedures performed in hospitals. However, [Piegas and Haddad \(2011\)](#) pointed out that, for some procedures, these institutions cover approximately 80% of hospital admissions performed in Brazil.

AIH is issued for the reimbursement of expenses incurred during hospitalization. Each hospitalization is classified according to the SUS procedure table. The SUS table underwent changes in the coding of procedures in 2008. Therefore, the codes of the two versions of the table were considered. Table 1 presents the hip arthroplasty codes for the two versions of the SUS table.

Some restrictions were imposed on the sample: 1) only hospitals that performed at least 10 procedures in the year were considered⁶; 2) only patients whose health results at the time of getting discharged could be obtained were considered, that is, only patients who were discharged, cured/improved, or who died⁷; and 3) only individuals aged 50 years and above were considered, as the procedure tends to occur in older adults ([RACHET-JACQUET; GUTACKER; SICILIANI, 2021](#)).

AIHs have information about patient characteristics, such as age, sex, main diagnosis, length of stay, date of admission and discharge from the hospital, hospital discharge status, character of care, days of intensive care unit (ICU) stay, and zip code. Based on the AIHs, it is possible to calculate the total number of patients (SUS) treated by a hospital and the average length of stay of these patients for each year. CNES allows obtaining information about the characteristics of hospitals, including the number of beds, ownership (public, private, or philanthropic), whether it conducts teaching activities, and the zip code where it is located.

The quality indicator used was in-hospital mortality during hospitalization. The Agency for Healthcare Research and Quality ([AHRQ, 2007](#)) highlights mortality as a valid indicator for assessing the difference in quality between hospitals. In-hospital mortality is the outcome most commonly used in the literature to investigate the volume-outcome effect ([HALM; LEE; CHASSIN, 2002](#)). Despite its wide use, it is important to highlight that this outcome has limitations, the greatest of which is that it may reflect the health status prior to a patient's admission into a hospital.

The key variable in the estimated models is the volume of procedures performed by a hospital in a given year. [AHRQ \(2007\)](#) highlighted that volume indicators are proxies or valid indirect measures of the quality of hospital care, as evidence suggests that hospitals that perform a greater number of intensive, high-technology, or complex procedures have better results. The volume of procedures is defined as the number of times a hospital performs a certain procedure in a given period. Therefore, for calculating volume, all patients who underwent hip arthroplasty were considered, as all procedures performed, regardless of the outcome, contributed to gains in scale and/or

instruments. However, it is believed that the period used does not affect the relationship studied over time.

⁶Gaynor, Seider, and Vogt (2005) and Hentschker and Mennicken (2015) employed a similar strategy. Only those hospitals that performed a minimum number of procedures in a year were considered, while those hospitals that performed procedures without minimum infrastructure were excluded.

⁷It was not possible to infer the health outcome of approximately 3% of patients at the time of getting discharged from hospitals.

Table 1: Procedure analyzed

Procedure	SUS procedures table		
	Name	Code	
		<2008	≥2008
Hip Arthroplasty (HA)	Hip arthroplasty (unconventional)	39029123	408040041
	Partial hip arthroplasty	39003124	408040050
	Conversion total hip arthroplasty	39026124	408040068
	Revision arthroplasty or hip reconstruction	39025128	408040076
		39027120	
	Cemented primary total hip arthroplasty	39018121	408040084
	Cementless/hybrid primary total hip arthroplasty	39016129	408040092

Source: Author's elaboration based on [Brasil – Ministério da Saúde \(2017\)](#)

learning-by-doing. Furthermore, the proportion of patients that were not discharged or died was only 3.63% of the total⁸.

The choice of controls followed the literature, covering individual characteristics such as age, sex, urgency, the Charlson comorbidity index, which categorically measures the severity of the comorbidity⁹, and whether the hospitalization was in an ICU. Hospital controls were employed, for example, based on their size, such as the number of beds and patients served. Furthermore, considering possible regional inequalities in the patient's place of residence that may affect the outcome, controls such as state dummies, municipal gross domestic product (GDP) per capita, and some health infrastructure variables were considered. The instruments¹⁰ were measured by two variables, potential patient (the number of patients residing within 10 km of a hospital) and the number of other hospitals (number of hospitals within a radius of up to 30 km of a hospital)¹¹. Table 2 provides a more detailed description of the variables. The data source is (SIH), except for the municipal GDP per capita variable, for which the data was obtained from IBGE (primary source) and DATASUS.

Table 3 presents the descriptive statistics for patients and hospitals. Approximately 3% patients died during their hospital stay. The mean age of patients was 72 years and the majority (61.19%) were women. Almost 60% patients were hospitalized on an emergency basis. Moreover, a vast majority of patients had a CCI equal to zero, and just over 20% of them had to be admitted to an ICU.

Regarding the characteristics of the hospitals, it was observed that the average volume of procedures increased over time, rising from 34 in $t - 4$ (four years before the year in question, e.g., for the year 2008, $t - 4$ refers to the year 2004) to 44 in t (refers to the year in question, which can be from 2008 to 2014). Approximately 35% hospitals conducted teaching activities, while 13.34% were run by private organizations for profit. For parameters such as annual number of beds in a hospital, patients treated by a hospital, patients within a radius of 10 km from a hospital, hospitals within a radius of 30 kms from a hospital, average annual length of stay in a hospital (in days), the average stood at 221, 8089, 70, 5, and 5, respectively¹².

3.2 Empirical Strategy

In the empirical analysis, a probit model is used, in which y_{ihmt} is a binary variable that indicates whether patient i , admitted to hospital h in municipality m in year t , died ($y_{ihmt} = 1$) or not ($y_{ihmt} = 0$) after hip arthroplasty. Thus, the following equation is estimated:

⁸In any case, alternative estimates were performed, considering only discharge and death when calculating the volume, and the results were qualitatively identical. They are available upon request.

⁹The Charlson Comorbidity Index (CCI) is composed of 20 clinical conditions, which are weighted from 1 to 6 according to the risk of death and the severity of the disease, and then added together to compose the total ICC score. For more details see [Charlson et al. \(1987\)](#).

¹⁰The starting point for the creation of the instruments was to georeference the zip codes of residence of patients and hospitals, and later calculate the distance between them with the command of Stata "geonear," and finally, add the number of patients and hospitals within a certain radius of distance from each hospital.

¹¹The number of patients and other hospitals are computed considering the database of this study, all SUS procedures related to hip arthroplasty.

¹²These numbers correspond to the overall mean at the hospital level.

Table 2: Description variable

Variables	Description
Outcome	
Death	Binary variable that is equal to 1 if a patient dies after performing the procedure.
Individual characteristics	
Age	Patient age in years.
Sex	Binary variable that is equal to 1 when a patient is female.
Urgency	Binary variable that is equal to 1 when a patient is admitted on an emergency basis.
CCI	Categorical variable ranging from 0 to 6 according to the severity of a patient's comorbidity.
Use of ICU	Binary variable that is equal to 1 if a patient is admitted to an ICU.
Variable of interest	
Volume	Continuous variable that indicates the number of patients who underwent hip arthroplasty procedure in a hospital in a year.
Hospital controls	
Carry out teaching activity	Binary variable that is equal to 1 when a hospital carries out teaching activities.
Private (for profit)	Binary variable that is equal to 1 when a hospital is run by a private organization for profit.
Number of beds	Continuous variable that indicates the number of beds a hospital has in a year.
Number of patients treated	Continuous variable that indicates the total number of patients seen by a hospital in a year.
Average length of stay	Continuous variable that indicates the average length of stay of patients in a hospital in a year.
Instruments	
Number of patients (0 to 10 km)	Continuous variable that indicates the number of patients residing within a radius of 0 to 10 kilometers from a hospital.
Number of hospitals (0 to 30 km)	Continuous variable that indicates the number of hospitals within a radius of 0 to 30 kilometers from a hospital in the year.
Controls of the individual's region of residence	
FU of location	Binary variables that indicate the Federation Unit (FU) where a hospital is located.
GDP per capita	Continuous variable that indicates the real GDP per capita of the municipality where a patient resides in a year.
Number of orthopedists/traumatologists	Continuous variable that indicates the number of orthopedists/traumatologists per thousand inhabitants in the municipality where a patient resides in a year.
Number of doctors	Continuous variable that indicates the number of doctors per thousand inhabitants in the municipality where a patient resides in a year.
Number of diagnostic imaging devices	Continuous variable that indicates the number of diagnostic imaging devices in the municipality where a patient resides in a year.
Year dummies	
Time FE	Binary variables that indicate the year in which a procedure was performed.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#)

Table 3: Descriptive statistics (2008-2014)

	Mean	S.E.	Min.	Max.
Patient (N = 108,376)				
Death	0.0350	0.1839	0	1
Age	72	11	50	113
Women	0.6119	0.4873	0	1
Urgency	0.5950	0.4909	0	1
CCI	0.9984	0.0396	0	1
0			0	1
1	0.0012	0.0346	0	1
2	0.0003	0.0166	0	1
6	0.0001	0.0096	0	1
ICU use	0.2207	0.4147	0	1
Hospital (Hospitals = 630; N = 2,504)				
Volume	44	54	10	938
Volume ($t - 1$)	41	52	0	608
Volume ($t - 2$)	38	49	0	546
Volume ($t - 3$)	36	48	0	565
Volume ($t - 4$)	34	48	0	546
Carry out teaching activity	0.3498	0.4770	0	1
Private (for profit)	0.1334	0.3401	0	1
Number of beds	221	184	18	1,980
Number of patients treated	8.089	6.224	191	51,631
Average length of stay (in days)	5	2	1	19
Number of patients (0 to 10 km)	70	109	0	651
Number of hospitals (0 to 30 km)	5	10	0	47

Source: Author's elaboration based on [Brasil – Ministério da Saúde \(2017\)](#)

$$y_{ihmt}^* = \beta_0 + \beta_1 \sqrt{vol_{ihmt}} + x'_{ihmt} \beta_2 + f'_{hmt} \beta_3 + m'_{mt} \beta_4 + year'_t \beta_5 + \epsilon_{ihmt} \quad (1)$$

$$y_{ihmt} = 1[y_{ihmt}^* > 0],$$

where y_{ihmt}^* is an unobserved latent variable that indicates whether the patient died in hospital; $\sqrt{vol}$ is the square root of the volume¹³; x is a vector of patient characteristics (age, sex, and character of care [urgency = 1], ICU utilization, and CCI); f are characteristics of the hospital (number of beds, teaching activity, number of patients treated, average length of stay, and FU of location); m are characteristics of the municipality in which the patient resides (GDP per capita, number of orthopedists/traumatologists per thousand inhabitants, number of doctors per thousand inhabitants, and number of diagnostic imaging devices per thousand inhabitants); and $year$ is dummy of the year. β_0 and β_1 are coefficients, while β_2 , β_3 , β_4 and β_5 are vectors of the parameters to be estimated. Finally, ϵ is the normally distributed error term. It is necessary to note that the standard errors are clustered at the hospital level, as the estimated model presents aggregated variables at the hospital level. In this case, the hypothesis of independence between the observations is replaced by the hypothesis of independence between the clusters, that is, hospitals ([BRAUN, 2014](#)).

¹³The model specification with the square root of volume is used due to the expectation that the impact of volume will decrease as the number of procedures performed in a hospital increases ([GAYNOR; SEIDER; VOGT, 2005](#); [GOWRISANKARAN; HO; TOWN, 2008](#))

Estimates based on equation (1) do not allow differentiation between practice-makes-perfect and selective referral hypotheses. As the objective of this study is to focus on the first hypothesis, it is necessary to discount the potential bias due to the selective referral problem from the total effect.

Studies have considered the geographic distribution of patients treated as a source of exogenous variation to deal with this problem (GAYNOR; SEIDER; VOGT, 2005; HENTSCHKER; MENNICKEN, 2014, 2018; SEIDER; GAYNOR; VOGT, 2004; GOWRISANKARAN; HO; TOWN, 2008). To constitute a valid source of exogenous variation, this instrumental variable (Z) must be highly correlated with volume ($cov(\sqrt{vol_{hmt}}, Z_{hmt}) \neq 0$) and uncorrelated with the error term ($cov(\sqrt{vol_{hmt}}, \epsilon_{ihmt}) = 0$). Generally, patients choose hospitals close to their homes (SEIDER; GAYNOR; VOGT, 2004). In this sense, the greater the distance between a hospital and patient's residence, the lower the probability that the hospital will be chosen. Likewise, the greater the number of other hospitals close to a patient's home that offer the same procedure, the less likely they will choose a particular hospital¹⁴. Given this observation, we can assume that the volume of a hospital depends on the number of potential patients and that of nearby hospitals offering the same service.

Hentschker and Mennicken (2014) emphasized that the number of potential patients and distance from each patient's home to a hospital should not have a direct influence on the quality of treatment. Patient homes can be considered exogenous to the quality of a hospital, as patients are unlikely to choose where to live based on the quality of care at nearby hospitals. However, other factors, such as income, may be correlated with patients' place of residence and hospital outcomes. In this sense, we believe that the controls of the municipality where a patient resides can capture these differences.

Thus, the observed volume of a hospital can be written as a function of the number of potential patients (p) within a given radius of hospital h as well as the number of other hospitals close (h) to the same hospital h that offer the same procedure within a certain radius, as highlighted by Seider, Gaynor, and Vogt (2004).

In this way, the equation for the observed volume can be written as:

$$\sqrt{vol_{ihmt}} = \alpha_0 + \alpha_1 p_{hmt} + \alpha_2 h_{hmt} + x'_{ihmt} \alpha_3 + f'_{hmt} \alpha_4 + m'_{mt} \alpha_5 + year'_t \alpha_6 + \epsilon_{hmt} \quad (2)$$

Equations (1) and (2) will be jointly estimated using the maximum likelihood through an ivprobit¹⁵. Once the exogeneity of the volume is evidenced, that is, once it is verified that the correlation observed between the volume and the outcome of patients is due to the practice-makes-perfect hypothesis, the model will be specified to verify whether this is the result of static economies of scale or learning-by-doing.

To verify the learning-by-doing hypothesis, the econometric model specification includes the volume of previous years, following Gaynor, Seider, and

¹⁴It would be ideal to build the instrument taking into account the transport infrastructure and ease of access to a hospital. The presence of a subway or bus station close to an individual's residence, for example, can facilitate access to a hospital, even if the geographic distance is considerable. However, a patient who is geographically close to a hospital may have greater difficulty in accessing it, if the location where he or she lives does not have adequate transport infrastructure. Unfortunately, it is not possible to consider these factors in the construction of our instrument.

¹⁵For more details, see www.stata.com/manuals13/rivprobit.pdf

Vogt (2005). Although the use of volume from previous years in a disaggregated way may imply problems of multicollinearity, such a strategy allows us to consider the impact of the volume of each previous year differently, which is not possible by considering the accumulated volume of previous years, as was done by Braun (2014) and Ho (2002). Thus, the equation to be estimated is

$$y_{ihmt}^* = \gamma_0 + \gamma_1 \sqrt{vol_{hmt}} + \sum_{\tau=1}^4 \theta_{\tau} \sqrt{vol_{hmt-\tau}} + x'_{ihmt} \gamma_2 + f'_{hmt} \gamma_3 + year'_t \gamma_4 + \mu_{ihmt}$$

$$y_{ihmt} = 1[y_{ihmt}^* > 0], \quad (3)$$

The hypotheses to be tested are as follows: (1) the existence of static economies of scale, that is, the coefficient of the volume variable is negative and statistically significant, and the coefficients of the lagged volume variables are equal to zero; and (2) the existence of learning-by-doing, in which case the coefficient of the volume variable is less than or equal to zero and that of the lagged volume variable is less than zero and statistically significant for the previous year.

4 Results

Initially, we analyzed the association between volume and probability of death without considering the endogeneity of the variable of interest. These results are presented in Table 4 based on the average marginal effect of the square root of volume. Models (1) to (4) show the results of the probit models with different specifications. In Model (1), in addition to the square root of volume, a year-fixed effect (FE) is included. Model (2) includes patient characteristics (age, sex, urgency, ICU use, and CCI). Model (3) also includes hospital characteristics (whether it conducts teaching activities, whether it is private, the number of beds, the number of patients treated, the average length of stay, and the FU of location). Finally, Model (4) also includes the characteristics of the municipality in which patients reside (GDP per capita, number of orthopedists/traumatologists per thousand inhabitants, number of doctors per thousand inhabitants, and number of diagnostic imaging equipment per thousand inhabitants).

The results of the four estimated models showed a significant negative correlation between procedure volume and patient outcome ($p < 0.01$). In other words, patients treated in hospitals with a high volume of procedures are less likely to die, a result that remains significant, even after controlling for patient and hospital characteristics. However, it should be noted that the inclusion of patient controls substantially reduced the effect of volume on the probability of death. The fact that the volume square root coefficient reduced in magnitude when we controlled for patient characteristics indicates that these characteristics are correlated with the volume of procedures and unevenly distributed among hospitals with different volumes.

Other studies that investigated the volume-outcome effect for hip arthroplasty (hip replacement) or hip fracture found results that corroborated the existence of this relationship. Hentschker and Mennicken (2014, 2015, 2018),

Table 4: Average marginal effects of the probit model - Hip Arthroplasty

Variables	(1)	(2)	(3)	(4)
Volume (square root)	-0.0016*** (0.0002)	-0.0008*** (0.0003)	-0.0009*** (0.0002)	-0.0009*** (0.0002)
Observations	108,376	108,376	108,376	108,376
Number of hospitals	630	630	630	630
Time FE	YES	YES	YES	YES
Patient controls	NO	YES	YES	YES
Hospital controls	NO	NO	YES	YES
Region controls	NO	NO	NO	YES

Notes: Cluster-robust hospital-level standard errors in parentheses. For details on variables, see Table 2. Model (1) includes, in addition to volume, a year FE. In Model (2), the characteristics of patients are included. In Model (3), hospital controls are included. Model (4) includes controls for a patient's home region. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

for example, showed the existence of a volume-outcome effect in Germany for hip fractures using different methodologies. [Luft, Bunker, and Enthoven \(1979\)](#) showed the volume outcome effect for total hip arthroplasty in the USA.

However, the results in Table 4 do not differentiate between the hypotheses of practice-makes-perfect and selective referral. Such a differentiation is essential for public policy. If the results presented in Table 4 are the result of the practice-makes-perfect hypothesis, the establishment of a minimum volume will most likely improve the results achieved by patients undergoing hip arthroplasty.

As with other studies ([GAYNOR; SEIDER; VOGT, 2005](#); [HENTSCHKER; MENNICKEN, 2014](#); [GOWRISANKARAN; HO; TOWN, 2008](#)), we also explored the geographic distribution of patients and hospitals to differentiate between practice-makes-perfect and selective referral hypotheses. We used the number of patients within a 0–10 km radius and the number of hospitals that performed hip arthroplasty within a 0–30 km radius of the hospital as a source of exogenous variation in volume.

Table 5 presents the results of the first and second stages of the probit model. Model 1 [columns (1) and (2)] does not include patient and hospital controls, whereas Model 2 [columns (3) and (4)] includes controls for patient characteristics. Model 3 [columns (5) and (6)] includes both patient and hospital controls. In addition to these, Model 4 [columns (7) and (8)] includes municipal controls. The first-stage results of the four estimated models show that the number of patients within a 0–10 km radius and the number of hospitals within a 0–30 km radius of the hospital are highly correlated with the volume of procedures. The greater the number of patients close to the hospital, the greater the volume. Likewise, the greater the number of nearby hospitals, the lower the volume of procedures. In other words, the first-stage results were consistent with expectations.

The results of the second stage of the probit model showed a causal effect of volume on mortality, considering that the instruments used satisfied the hypotheses of relevance ($cov(\sqrt{vol_{ht}}, Z_{ht}) \neq 0$) and exogeneity ($cov(\sqrt{vol_{ht}}, \epsilon_{iht}) = 0$). This showed that the volume-outcome effect observed for hip arthroplasty

Table 5: Exogenous volume - *ivprobit*

Variables	Model 1		Model 2		Model 3		Model 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First Stage								
No. of patients (0 to 10 km)		0.0363*** (0.0049)		0.0349*** (0.0046)		0.0358*** (0.0050)		0.0360*** (0.0051)
No. of hospitals (0 to 30 km)		-0.2129*** (0.0441)		-0.2142*** (0.0445)		-0.2292*** (0.0474)		-0.2305*** (0.0483)
Second Stage								
Volume (root square)	-0.0018*** (0.0004)		-0.0011** (0.0006)		-0.0010** (0.0004)		-0.0010*** (0.0004)	
Observations	108,376	108,376	108,376	108,376	108,376	108,376	108,376	108,376
Number of hospitals	630	630	630	630	630	630	630	630
Wald Exogeneity Test (p-value)		0.4808		0.6006		0.0022		0.0040
Time FE	YES	YES	YES	YES	YES	YES	YES	YES
Patient controls	NO	NO	YES	YES	YES	YES	YES	YES
Hospital controls	NO	NO	NO	NO	YES	YES	YES	YES
Region controls	NO	NO	NO	NO	NO	NO	YES	YES

Notes: Cluster-robust hospital-level standard errors in parentheses. For details on variables, see Table 2. Model (1) includes, in addition to volume, a year FE. In Model (2), the characteristics of patients are included. In Model (3), hospital controls are included. Model (4) includes controls for a patient's home region. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

Table 6: Marginal effects of the probit model: scale gains or learning-by-doing

Variables	Marginal Effect
Volume (root square)	-0.0011*** (0.0004)
Volume in $t - 1$ (root square)	0.0003 (0.0004)
Volume in $t - 2$ (root square)	0.0002 (0.0004)
Volume in $t - 3$ (root square)	-0.0001 (0.0004)
Volume in $t - 4$ (root square)	-0.0003 (0.0003)
Observations	108,763
Number of hospitals	630
Time FE	YES
Patient controls	YES
Hospital controls	YES
Region controls	YES

Notes: Cluster-robust hospital-level standard errors in parentheses. In addition to volume level and lagged volumes, the model includes the following controls: year fixed effect (time FE), patient characteristics, hospital controls, and controls for the municipality in which a patient resides. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

in Brazil is the result of the practice-makes-perfect hypothesis and not the selective referral hypothesis. Comparing the complete models in Tables 5 and 6 [columns (4) and (7), respectively], we can observe a bias caused by the endogeneity of volume, as its coefficient increases in absolute terms when we instrumentalize it by the number of patients and from nearby hospitals.

Finally, to verify whether the volume-outcome effect is due to scale gains or the learning-by-doing process, a model was specified with the volume of procedures at level and lagged. The results in Table 6 show that the volume-outcome effect is due to scale gains and not the learning-by-doing process, as the marginal effect of volume in level is significant, while the marginal effect of volume lags are not. In this sense, the establishment of a minimum volume with the objective of concentrating procedures in hospitals with greater volume to improve patient health outcomes will not result in loss of experience in hospitals that no longer perform the procedure.

4.1 Robustness Analysis

To verify the robustness of the ivprobit results, in which the volume is considered endogenous and instrumentalized by the number of patients residing in a radius of 0-10 kilometers and the number of hospitals in a radius of 0-30 kilometers from the hospital, a “false y ” type falsification strategy was used. Specifically, two variables were used that were not expected to have any relationship with the quality of a hospital, that is, with the volume variable. The variables used as false y , as can be seen in Table 7, were a patient's sex and the

Table 7: Falsification test

Variables	Gender		Urgency	
	(1) 2° Stage	(2) 1° Stage	(3) 2° Stage	(4) 1° Stage
No. of patients (0 to 10 km)		0.0360*** (0.0051)		0.0364*** (0.0051)
No. of hospitals (0 to 30 km)		-0.2305*** (0.0483)		-0.2329*** (0.0485)
Volume (root square)	0.0016 (0.0026)		-0.0220 (0.0153)	
Observations	108,376	108,376	108,369	108,369
Number of hospitals	630	630	629	629
Wald Exogeneity Test (p-value)	0.0411		0.1162	
Time FE	YES	YES	YES	YES
Patient controls	YES	YES	YES	YES
Hospital controls	YES	YES	YES	YES
Region controls	YES	YES	YES	YES

Notes: Cluster-robust hospital-level standard errors in parentheses. The models use controls and instrumental variables as described in Table 2, however, by changing the mortality outcome for sex and urgency. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

fact that he/she was hospitalized as an emergency. This last variable can even measure, to a certain extent, the health status of patients prior to the procedure. As can be seen from the results of the second stage [columns (1) and (3)], there is no relationship between the volume of procedures and gender and/or the nature of care (urgency), indicating that, most likely, the results obtained in Table 5 are not the result of chance or any previous false outcome.

In Table 8, we tested the influence of the metropolitan region variable, either as an additional control or as a subsample, in a similar way to Table 5. Looking at the results in columns (1) and (2), it is observed that the results remain similar with regard to the effect of volume on results. The metropolitan region variable showed a positive effect in the first stage, indicating that hospitals located in metropolitan regions serve more patients than those located in other regions. However, there was no direct effect of this variable on the mortality indicator, indicating the robustness of our results. When the estimation is performed for the subsamples, the volume effect is significant only in the case of metropolitan regions. Some possible explanations for this result may be related to the lower variability of the volume variable and its instruments, especially related to the number of surrounding hospitals. Therefore, there are clear effects of the outcome volume for metropolitan regions.

Finally, the results in Table 9 indicate that even when a metropolitan region is considered either via additional control or as a subsample, the volume-result effect occurs via static gains of scale, similar to the previous analysis, showing the robustness of our results.

5 Conclusion

In this study, we examine the relationship between the volume of procedures and health outcomes achieved by patients undergoing hip arthroplasty in Brazil, thus seeking to contribute to the literature that discusses the imple-

Table 8: Metropolitan region (MR)

Variables	MR Dummy		Metropolitan Region		Non-Metropolitan Region	
	(1)	(2)	(3)	(4)	(5)	(6)
Metropolitan Region	-0.0038 (0.0026)	1.4438** (0.5985)				
No. of patients (0 to 10 km)		0.0347*** (0.0048)		0.0341*** (0.0050)		0.0368*** (0.0055)
No. of hospitals (0 to 30 km)		-0.2495*** (0.0488)		-0.2457*** (0.0533)		-0.1630 (0.1024)
Volume (root square)	-0.0009** (0.0004)		-0.0012*** (0.0004)		-0.0002 (0.0013)	
Observations	108,376	108,376	59,203	59,203	49,163	49,163
Number of hospitals	630	630	265	265	364	364
Wald Exogeneity Test (p-value)	0.0166		0.0338		0.2730	
Time FE	YES	YES	YES	YES	YES	YES
Patient controls	YES	YES	YES	YES	YES	YES
Hospital controls	YES	YES	YES	YES	YES	YES
Region controls	YES	YES	YES	YES	YES	YES
MR controls	YES	YES	NO	NO	NO	NO

Notes: MR is Metropolitan region. Cluster-robust hospital-level standard errors in parentheses. Results of the first and second stages of three models, considering the inclusion of all the variables described in Table 2: in the first case, a variable is added indicating whether the hospital is in a metropolitan region; in the other two cases, estimations by subsample are performed, divided following the criterion of this metropolis dummy variable. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

Table 9: Learning-by-doing considering metropolitan region

Variables	MR dummy	MR sample
	(1)	(2)
Volume (root square)	-0.0010** (0.0004)	-0.0011*** (0.0004)
Volume in $t - 1$ (root square)	0.0003 (0.0004)	0.0005 (0.0005)
Volume in $t - 2$ (root square)	0.0002 (0.0004)	-0.0001 (0.0004)
Volume in $t - 3$ (root square)	-0.0000 (0.0004)	0.0000 (0.0004)
Volume in $t - 4$ (root square)	-0.0003 (0.0003)	-0.0002 (0.0003)
Metropolitan region	-0.0032** (0.0016)	
Observations	108,763	59,203
Number of hospitals	630	265
Time FE	YES	YES
Patient controls	YES	YES
Hospital controls	YES	YES
Region controls	YES	YES
MR controls	YES	NO

Notes: MR is Metropolitan region. Cluster-robust hospital-level standard errors in parentheses. The dependent variable presented in the table is a dummy that indicates whether a patient died (= 1) or not (= 0) after undergoing hip arthroplasty. The model includes, in addition to volume level and lagged volumes, these controls: year FE, patient characteristics, hospital controls, and controls for the municipality in which a patient resides. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

mentation of a minimum volume of hospital procedures. Using data from 2008 to 2014 with more than 120,000 observations, we contribute to the literature by analyzing the relationship from a causal perspective, seeking to differentiate the hypotheses of practice-makes-perfect from selective referral, an indispensable condition for the formulation of policies to improve the quality of hospital care. Furthermore, we verify whether the volume-outcome effect is due to static scale gains or the learning-by-doing process.

Our results support the existence of a volume-outcome effect in Brazil for hip arthroplasty and, by exploring the geographical distribution of patients and hospitals, the study finds evidence that the volume-outcome effect is the result of the practice-makes-perfect hypothesis, and not the selective referral hypothesis. To the best of our knowledge, this is the first study that provides such evidence using Brazilian data. Furthermore, we have an indication that it occurs through static gains of scale and is not due to the learning-by-doing process. However, it should be noted that such results were not observed for the subsample of hospitals belonging to non-metropolitan regions, which may be related to the lower variability of the variable of interest and its instruments.

Given these results, the establishment of a minimum hospital volume must have the potential to improve patient health outcomes and reduce the differences in the quality of care between hospitals. Evidence of static gains in scale

also supports the argument of concentrating supply.

However, it should be noted that the study has some limitations, and that before implementing any policy, further analysis should be carried out. First, this study uses administrative data instead of clinical data, with which it will be possible to obtain a much more detailed picture of patients' health condition. Second, our observation unit is hospitalization and not the patient. The database does not allow identifying a patient if he/she was hospitalized more than once; as a result, he/she may be considered more than once. Third, this study considers only the hospitalizations performed within the scope of SUS; that is, we did not have data about all hospitalizations and hip arthroplasty procedures. Fourth limitation is related to the creation of the instrument, as the study considers only the geographical distance between patients' homes and hospitals, whereas the most appropriate would have been to consider transport infrastructure access. Fifth, we did not observe that the volume-outcome effect occurs because of the learning-by-doing process; however, it is possible that we did not find significant marginal effects of volume lags as a function of the time period analyzed. Thus, additional analyses considering other periods must be carried out.

Future research should seek to determine the minimum threshold and investigate the limits of the number of procedures that will result in significant differences in quality. Another factor that deserves attention is the effect that concentration of procedures can have in terms of access to care. It is possible that there will be deterioration in access to care, as many hospitals that do not reach the minimum volume will be unable to provide care. Such a restriction could further worsen the financial sustainability of these hospitals, leading to greater concentration in the Brazilian hospital sector. It will be interesting to investigate the extent to which a minimum volume policy can affect the well-being of the population in terms of increasing the distance between patients' homes and hospitals that will remain in operation, in addition to the possible formation of queues for procedures to be performed. Thus, it will be possible to measure the net benefit of concentrating procedures in a smaller number of hospitals.

References

- AGENCY FOR HEALTHCARE RESEARCH AND QUALITY — AHRQ. *AHRQ Quality Indicators - Guide to Inpatient Quality Indicators: Quality of Care in Hospitals - Volume, Mortality, and Utilization*. Rockville, MD: Agency for Healthcare Research and Quality, 2007.
- AVDIC, Daniel; LUNDBORG, Petter; VIKSTRÖM, Johan. Estimating returns to hospital volume: Evidence from advanced cancer surgery. *Journal of Health Economics*, v. 63, p. 81–99, 2019.
- BIRKMEYER, John D. *et al.* Hospital volume and surgical mortality in the United States. *New England Journal of Medicine*, v. 346, n. 15, p. 1128–1137, 2002.
- BOTEGA, Laura de Almeida; ANDRADE, Mônica Viegas; GUEDES, Gilvan Ramalho. Brazilian hospitals' performance: an assessment of the unified health system (SUS). *Health Care Management Science*, v. 23, p. 443–452, 2020.

- BRASIL – MINISTÉRIO DA SAÚDE. DATASUS–Departamento de Informática do SUS. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=0901>. Acesso em: 25 ago. 2017.
- BRASIL – MINISTÉRIO DA SAÚDE. *Sistema de Planejamento do SUS: Uma construção coletiva: Plano Nacional de Saúde (PNS) 2008/2009-2011*. Brasília: Ministério da Saúde, 2010. (Série Cadernos de Planejamento. Série B: Textos Básicos de Saúde, nº 9). Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/sistema_planejamento_sus_v9.pdf. Acesso em: 27 ago. 2017.
- BRASIL – MINISTÉRIO DA SAÚDE. *Tabela de Procedimentos, Medicamentos e OPM do SUS*. 2016. Disponível em: <http://sigtap.datasus.gov.br/tabela-unificada/app/sec/procedimento/publicados/consultar>. Acesso em: 25 ago. 2017.
- BRAUN, Gisele. *Essays in Health Economics: Volume-Outcome Effect, Upcoding Behaviour and the Evaluation of a Short-Term Intervention Program*. Dissertation (Doctor of Philosophy in Economics) – Universidade NOVA de Lisboa, Lisbon, 2014.
- CASSONE, Alejandro *et al.* Trombose venosa profunda em artroplastia total de quadril. *Revista Brasileira de Ortopedia*, v. 37, n. 5, p. 153–61, 2002.
- CHARLSON, Mary E. *et al.* A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*, v. 40, n. 5, p. 373–383, 1987.
- ERCOLE, Flávia Falci; CHIANCA, Tânia Couto Machado. Infecção de sítio cirúrgico em pacientes submetidos a artroplastias de quadril. *Revista Latino-Americana de Enfermagem*, v. 10, p. 157–165, 2002.
- GAYNOR, Martin; SEIDER, Harald; VOGT, William B. The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, v. 95, n. 2, p. 243–247, 2005.
- GODOY, Paulo Henrique *et al.* Hospital case fatality associated with coronary angioplasty in Rio de Janeiro State, Brazil, 1999-2003. *Cadernos de Saúde Pública*, v. 23, n. 4, p. 845–851, 2007.
- GOMES, Andréa Silveira *et al.* Factors associated with hospital mortality in Rio Grande do Sul SUS network in 2005: application of a Multilevel Model. *Revista Brasileira de Epidemiologia*, v. 13, n. 3, p. 533–542, 2010.
- GOWRISANKARAN, Gautam; HO, Vivian; TOWN, Robert J. *Causality, learning and forgetting in surgery*. Federal Trade Commission. 2008. Disponível em: <http://www.bristol.ac.uk/media-library/sites/comp/migrated/documents/gowrisankaran.pdf>. Acesso em: 7 jul. 2015.
- HALM, Ethan A.; LEE, Clara; CHASSIN, Mark R. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine*, v. 137, n. 6, p. 511–520, 2002.

- HANNAN, Edward L. *et al.* A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Services Research*, v. 27, n. 4, p. 517–542, 1992.
- HENTSCHKER, Corinna; MENNICKEN, Roman. *Selective-Referral and Unobserved Patient Heterogeneity–Bias in the Volume-Outcome Relationship*. Rochester, NY: Social Science Research Network, 2014.
- HENTSCHKER, Corinna; MENNICKEN, Roman. The volume–outcome relationship revisited: practice indeed makes perfect. *Health Services Research*, v. 53, n. 1, p. 15–34, 2018.
- HENTSCHKER, Corinna; MENNICKEN, Roman. The volume–outcome relationship and minimum volume standards–empirical evidence for Germany. *Health Economics*, v. 24, n. 6, p. 644–658, 2015.
- HO, Vivian. Evolution of the volume–outcome relation for hospitals performing coronary angioplasty. *Circulation*, v. 101, n. 15, p. 1806–1811, 2000.
- HO, Vivian. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Health Economics*, v. 21, n. 5, p. 873–885, 2002.
- KAHN, Jeremy M.; HAVE, Thomas R. T.; IWASHYNA, Theodore J. The relationship between hospital volume and mortality in mechanical ventilation: an instrumental variable analysis. *Health Services Research*, v. 44, n. 3, p. 862–879, 2009.
- KAIER, Klaus *et al.* Volume–outcome relationship in transcatheter aortic valve implantations in Germany 2008–2014: a secondary data analysis of electronic health records. *BMJ Open*, v. 8, n. 7, 2018.
- LA FORGIA, Gerard Martin; COUTTOLENC, Bernard. *Hospital performance in Brazil: the search for excellence*. Washington, D.C.: World Bank Publications, 2008. Disponível em: <http://documents.worldbank.org/curated/en/815061468015870054/Hospital-performance-in-Brazil-the-search-for-excellence>. Acesso em: 25 ago. 2017.
- LAPAR, Damien J. *et al.* Hospital procedure volume should not be used as a measure of surgical quality. *Annals of Surgery*, v. 256, n. 4, p. 606–615, 2012.
- LENZA, Mario *et al.* Epidemiology of total hip and knee replacement: a cross-sectional study. *Einstein (Sao Paulo)*, v. 11, n. 2, p. 197–202, 2013.
- LUFT, Harold S.; BUNKER, John P.; ENTHOVEN, Alain C. Should operations be regionalized? The empirical relation between surgical volume and mortality. *New England Journal of Medicine*, v. 301, n. 25, p. 1364–1369, 1979.

- LUFT, Harold S.; HUNT, Sandra S.; MAERKI, Susan C. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Services Research*, v. 22, n. 2, p. 157–182, 1987.
- MACHADO, Juliana Pires; MARTINS, Ana Cristina Marques; MARTINS, Mônica Silva. Quality assessment of hospital care in Brazil: a systematic review. *Cadernos de Saúde Pública*, v. 29, p. 1063–1082, 2013.
- MARTINS, Mônica; BLAIS, Régis; LEITE, Iúri da Costa. Hospital mortality and length of stay: comparison between public and private hospitals in Ribeirão Preto, São Paulo State, Brazil. *Cadernos de Saúde Pública*, v. 20, s268–s282, 2004.
- MESQUITA, Gerardo Vasconcelos *et al.* Morbid-mortality in elderly due to proximal fractures of the femur. *Texto & Contexto-Enfermagem*, v. 18, n. 1, p. 67–73, 2009.
- NORONHA, José Carvalho de *et al.* Volume and quality of care in coronary artery bypass grafting in Brazil. *Cadernos de Saúde Pública*, v. 19, n. 6, p. 1781–1789, 2003.
- PAN AMERICAN HEALTH ORGANIZATION — PAHO. *OMS revela principais causas de morte e incapacidade em todo o mundo entre 2000 e 2019*. [S. l.: s. n.], 2020. Disponível em: <https://www.paho.org/pt/noticias/9-12-2020-oms-revela-principais-causas-morte-e-incapacidade-em-todo-mundo-entre-2000-e>. Acesso em: 11 ago. 2021.
- PEREIRA, Bruna Rogeliane Rodrigues *et al.* Artroplastia do quadril: prevenção de infecção do sítio cirúrgico. *Revista SOBECC*, v. 19, n. 4, p. 181–187, 2014.
- PIEGAS, Leopoldo Soares; BITTAR, Olímpio J. Nogueira V.; HADDAD, Nagib. Myocardial revascularization surgery (MRS): results from national health system (SUS). *Arquivos Brasileiros de Cardiologia*, v. 93, n. 5, p. 555–560, 2009.
- PIEGAS, Leopoldo Soares; HADDAD, Nagib. Percutaneous coronary intervention in Brazil: results from the Brazilian Public Health System. *Arquivos Brasileiros de Cardiologia*, v. 96, n. 4, p. 317–324, 2011.
- POSNETT, John. Concentration and choice in healthcare. *Financial Times Healthcare*, v. 32, 1997.
- RACHET-JACQUET, Laurie; GUTACKER, Nils; SICILIANI, Luigi. Scale economies in the health sector: The effect of hospital volume on health gains from hip replacement surgery. *Journal of Economic Behavior & Organization*, v. 190, p. 704–729, 2021.
- RAMOS, Marcelo Cristiano de Azevedo *et al.* Performance evaluation of hospitals that provide care in the public health system, Brazil. *Revista de Saúde Pública*, v. 49, p. 1–9, 2015.

- REAMES, Bradley N. *et al.* Hospital volume and operative mortality in the modern era. *Annals of Surgery*, v. 260, n. 2, p. 244–251, 2014.
- RIBEIRO, Antonio Luiz P. *et al.* Mortality related to cardiac surgery in Brazil, 2000–2003. *The Journal of Thoracic and Cardiovascular Surgery*, v. 131, n. 4, p. 907–909, 2006.
- SALTMAN, Richard B.; FIGUERAS, Josep. *European health care reform: analysis of current strategies*. Copenhagen: World Health Organization. Regional Office for Europe, 1997.
- SEIDER, Harald; GAYNOR, Martin; VOGT, William B. *Volume-outcome and antitrust in US health care markets*. Carnegie Mellon University: Unpublished Manuscript, 2004.
- SUND, Reijo. Modeling the volume-effectiveness relationship in the case of hip fracture treatment in Finland. *BMC Health Services Research*, v. 10, n. 1, p. 1–12, 2010.

Appendix A

Table A.1: Systematization of Literature

International Literature					
Author(s) (year)	Sample	Method	Dependent variable	Controls	Results
Luft, Bunker, and Enthoven (1979)	842,622 patients who underwent 12 surgical procedures between 1974 and 1975 (USA)	Ordinary Least Square	Risk-adjusted hospital mortality	Patient and hospital characteristics	Inverse relationship between mortality and volume.
Luft, Hunt, and Maerki (1987)*	1,008,502 patients who underwent 17 procedures in 1972 (USA)	Simultaneous Equations Model	Volume and hospital mortality	Patient and hospital characteristics	Practice-makes-perfect and selective referral hypotheses are plausible.
Hannan <i>et al.</i> (1992)*	4,524 patients who underwent AAA ^a repair between 1985 and 1987 (NY-USA)	Logistic Regression	Hospital mortality	Patient and hospital characteristics	Inverse relationship between mortality and volume; little evidence of selective referral.
Ho (2000)	353,488 patients who underwent PCI ^b between 1984 and 1996 (CA-USA)	Logistic Regression	Hospital mortality and CABS ^c	Patient characteristics	Inverse relationship between mortality (and CABS ^c) and volume.
Ho (2002)	343,737 patients who underwent PCI ^b between 1984 and 1996 (CA-USA)	Probit Model	Hospital mortality and CABS ^c	Patient and hospital characteristics	Inverse relationship between mortality (and CABS ^c) and volume, but the effect is small; no evidence of learning-by-doing.
Birkmeyer <i>et al.</i> (2002)	2.5 million patients who underwent 14 procedures between 1994 and 1999 (USA)	Logistic Regression	Hospital mortality	Patient characteristics	Inverse relationship between mortality and volume for 14 procedures, but with great variation in magnitude.
Gaynor, Seider, and Vogt (2005)*	363,994 patients who underwent CABS ^c between 1983 and 1999 (CA-USA)	Probit Model	Hospital mortality	Patient characteristics, time (year), and hospital dummies	Inverse relationship between mortality and volume; static scale effect.
Gowrisankaran, Ho, and Town (2008)* [¶]	459,540 patients who underwent PD ^d , AAA ^a repair, or CABS ^c between 1988 to 1999 (CA, FL-USA)	Conditional Logit Model and Probit Model	Hospital mortality	Patient and hospital characteristics	Evidence of the practice-makes-perfect (learning-by-doing) hypothesis for AAA ^a and CABS ^c .
Kahn, Have, and Iwashyna (2009)*	78,427 non-surgical patients who underwent mechanical ventilation between 2004 and 2006 (PA-USA)	Linear Probability Model with Instrumental Variable (IV)	Hospital mortality	Patient and hospital characteristics	Inverse relationship between mortality and volume using clinical data, which is higher with the use of IV; non-significant result for administrative data.

Notes: * Studies that seek, in some ways, to test the hypotheses of practice-makes-perfect and selective referral; [¶] Working papers; ^a: Abdominal Aortic Aneurysm; ^b: Percutaneous Coronary Intervention; ^c: Coronary-Artery Bypass Surgery; ^d: Pancreato-Duodenectomy; ^e: Homogeneous Diagnostics Group; ^f: AIH corresponds to a Hospitalization Authorization of Health Public System in Brasil (SUS).

Table A.1: Systematization of Literature (*continued*)

International Literature					
Author(s) (year)	Sample	Method	Dependent variable	Controls	Results
Braun (2014)* [¶]	880,449 patients in 21 GDHs ^e who were hospitalized between 2001 and 2008 (Portugal)	Fixed Effects Model (FE) and Probit Model (MP)	Hospital mortality	Patient and hospital characteristics	FE: inverse relationship between volume and mortality in seven GDHs ^e ; MP: inverse relationship between volume and mortality in 10 GDHs ^e .
Hentschker and Mennicken (2014)* [¶]	89,541 hip fracture patients in 2007 (Germany)	IV-GMM	Hospital mortality	Patient, hospital, and municipality characteristics	Inverse relationship between mortality and volume.
Hentschker and Mennicken (2015)	97,183 patients with hip fracture or AAA ^a in 2007 (Germany)	Logistic Regression	Hospital mortality	Patient and hospital characteristics	The probability of dying is lower in a higher volume hospital for both conditions.
Hentschker and Mennicken (2018)*	89,541 hip fracture patients in 2007 (Germany)	Probit Model (MP) and VI-Probit	Hospital mortality	Patient and hospital characteristics	Inverse relationship between mortality and volume.
Kaier <i>et al.</i> (2018)	43,996 patients, who underwent transcatheter aortic valve implantation (Germany)	Linear and logistic regressions	Hospital mortality, post-procedure complications, length of stay, and proportion of patients with ventilation	Patient characteristics	Inverse relationship between mortality and volume, but this effect decreases over the years analyzed.
Avdic, Lundborg, and Vikström (2019)*	105,021 patients, who underwent surgery for advanced cancer (Sweden)	Linear Probability Model with Instrumented Variable (IV)	Patient survival four years after surgery, likelihood of readmission, subsequent cancer surgery, and length of hospital stay	Patient, hospital, and municipality characteristics	Positive relationship between survival and volume and evidence supporting the learning-by-doing hypothesis.
Rachet-Jacquet, Gutacker, and Siciliani (2021)*	105,229 patients, who underwent hip replacement surgery (England)	Linear Probability Model and Conditional Logit	Oxford hip score (OHS)	Patient and hospital characteristics	No evidence of causal relationship between volume and the health outcome.

Notes: * Studies that seek, in some ways, to test the hypotheses of practice-makes-perfect and selective referral; [¶] Working papers; *a*: Abdominal Aortic Aneurysm; *b*: Percutaneous Coronary Intervention; *c*: Coronary-Artery Bypass Surgery; *d*: Pancreato-Duodenectomy; *e*: Homogeneous Diagnostics Group; *f*: AIH corresponds to a Hospitalization Authorization of Health Public System in Brasil (SUS).

Table A.1: Systematization of Literature (*continued*)

National Literature					
Author(s) (year)	Sample	Method	Dependent variable	Controls	Results
Noronha <i>et al.</i> (2003)	41,989 patients submitted to CABS ^c between 1996 and 1999 paid for by SUS	Cox Regression Model	Hospital mortality	Patient and hospital characteristics	Relative risk of death after CABS ^c is higher in low-volume hospitals.
Martins, Blais, and Leite (2004)	32,906 patients from 27 hospitals in Ribeirão Preto/SP with cardiovascular and respiratory diseases between 1996 and 1998	Multilevel Logistic Model	Hospital mortality and length of stay	Patient and hospital characteristics	Non-significant relationship for mortality and length of stay.
Ribeiro <i>et al.</i> (2006)	115,021 patients, who underwent cardiovascular surgery between 2000 and 2003 paid for by SUS	Logistic Regression	Hospital mortality	Patient characteristics	Odds ratio for mortality is higher in low-volume hospitals.
Godoy <i>et al.</i> (2007)	8,735 patients, who underwent PCI ^b in Rio de Janeiro between 1999 and 2003	Poisson Regression	Hospital mortality	Patient characteristics	After adjustment, the volume-lethal ratio is considerably reduced.
Piegas, Bittar, and Haddad (2009)	63,272 patients submitted to CABS ^c between 2005 and 2007 paid by SUS	χ^2 test, t de Student test, Mann-Whitney test, ANOVA, Scheffé test, Kruskal-Wallis test	Hospital mortality and length of stay	Without controls	Mortality after CABS ^c is higher in low-volume hospitals.
Gomes <i>et al.</i> (2010)	10,000 AIHs ^f for RS (state of Brazil) in 2005	Multilevel Logistic Model	Hospital mortality	Patient and hospital characteristics	Volume has no significant effect on mortality.
Piegas and Haddad (2011)	166.514 pacientes submetidos a ICPb entre 2005 e 2008 pagas pelo SUS	χ^2 test, t de Student test, ANOVA	Hospital mortality	Without controls	Non-significant relationship

Notes: * Studies that seek, in some ways, to test the hypotheses of practice-makes-perfect and selective referral; [¶] Working papers; *a*: Abdominal Aortic Aneurysm; *b*: Percutaneous Coronary Intervention; *c*: Coronary-Artery Bypass Surgery; *d*: Pancreato-Duodenectomy; *e*: Homogeneous Diagnostics Group; *f*: AIH corresponds to a Hospitalization Authorization of Health Public System in Brasil (SUS).

Table A.2: Marginal effects of probit model - Hip Arthroplasty - Full Model

Variables	Model 1	Model 2	Model 3	Model 4
Volume (root square)	-0.0016*** (0.0002)	-0.0008*** (0.0003)	-0.0009*** (0.0002)	-0.0009*** (0.0002)
Age (years)		0.0014*** (0.0001)	0.0014*** (0.0001)	0.0014*** (0.0001)
Woman		-0.0039*** (0.0008)	-0.0037*** (0.0008)	-0.0037*** (0.0008)
Urgency		0.0145*** (0.0016)	0.0144*** (0.0014)	0.0143*** (0.0014)
Admitted to an ICU		0.0454*** (0.0040)	0.0482*** (0.0036)	0.0485*** (0.0036)
CCI				
1		0.0084 (0.0156)	0.0030 (0.0127)	0.0030 (0.0125)
2		0.1264** (0.0554)	0.0964** (0.0487)	0.0951** (0.0478)
6		0.2998* (0.1566)	0.2417* (0.1395)	0.2423* (0.1375)
Conducts teaching activity			0.0012 (0.0018)	0.0014 (0.0017)
Public			-0.0014 (0.0021)	-0.0011 (0.0021)
Number of beds			-0.0000 (0.0000)	-0.0000 (0.0000)
Patients served			0.0000 (0.0000)	0.0000* (0.0000)
Average length of stay			0.0015*** (0.0004)	0.0016*** (0.0004)
GDP per capita				-0.0000 (0.0000)
Orthopedists/traumatologists per thousand inhabitants				-0.0305** (0.0135)
Doctors per thousand inhabitants				-0.0000 (0.0002)
Diagnostic imaging equipment per thousand inhabitants				0.0024 (0.0017)
Observations	108,376	108,376	108,376	108,376
Number of hospitals	630	630	630	630
Time FE	YES	YES	YES	YES
Patient controls	NO	YES	YES	YES
Hospital controls	NO	NO	YES	YES
Region controls	NO	NO	NO	YES

Notes: MR is Metropolitan region. Cluster-robust hospital-level standard errors in parentheses. The dependent variable presented in the table is a dummy that indicates whether a patient died (= 1) or not (= 0) after undergoing hip arthroplasty. The model includes, in addition to volume level and lagged volumes, these controls: year FE, patient characteristics, hospital controls, and controls for the municipality in which a patient resides. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Source: Author's elaboration with information from [Brasil – Ministério da Saúde \(2017\)](#).

REGIONAL DEVELOPMENT AND CONVERGENCE CLUBS IN URUGUAY

DIEGO ABOAL *

BIBIANA LANZILOTTA †

MARTÍN PEREYRA ‡

MARÍA PAZ QUERALTÓ §

Resumo

Construímos um indicador multidimensional de desenvolvimento regional e realizamos uma análise de convergência usando dados uruguaios para o período 2006–2015. O indicador multidimensional departamental colapsa quatro dimensões de desenvolvimento, permitindo rastrear o desempenho de regiões em diferentes fatores de desenvolvimento. Nossos achados descartam a hipótese de convergência global para o período analisado. Identificamos, no entanto, convergências nos clubes, distinguindo três grupos de departamentos, com correspondência geográfica, que apresentam dinâmicas de desenvolvimento específicas. Esses clubes seguem trajetória própria e não convergem no período analisado, ainda que suas curvas de transição se aproximem um pouco.

Palavras-chave: desenvolvimento econômico regional; indicador multidimensional; convergência; clubes.

Abstract

We build a multidimensional indicator of regional development and perform a convergence analysis using Uruguayan data for the period 2006 – 2015. The multidimensional departmental indicator collapses four dimensions of development, allowing us to track the performance of regions on different factors of development. Our findings rule out the hypothesis of global convergence for the period analyzed. We nevertheless identify convergence in clubs, distinguishing three groups of departments, with a geographical correspondence, that exhibit specific development dynamics. These clubs follow their own path and do not converge in the period analyzed, even though their transition curves get slightly closer.

Keywords: regional economic development; multidimensional indicator; convergence; clubs.

JEL classification: O18, C33

DOI: <http://dx.doi.org/10.11606/1980-5330/ea188185>

* Centro de Investigaciones Económicas (CINVE). Universidad ORT Uruguay. Universidad de la República del Uruguay. E-mail: diego.aboal@gmail.com

† Centro de Investigaciones Económicas (CINVE). Universidad de la República del Uruguay. E-mail: bibiana.lanzilotta@gmail.com

‡ Centro de Investigaciones Económicas (CINVE). Universidad ORT Uruguay. E-mail: mpe-reyra@cinve.org.uy

§ Universidad de la República del Uruguay. E-mail: mpazqueralto@gmail.com

1 Introduction

The application of regional development policies is based on rich theoretical grounds, with a large toolkit of instruments at hand. Despite theoretical and empirical achievements, regional disparities persist across countries' regions or administrative divisions. This could be the consequence of ill-designed regional development policies. Too much emphasis on top-down, supply-side, "one-size-fits-all" policies eventually resulted in unbalanced policies, only relevant to the formal sector, and ultimately incapable of delivering sustainable development (PIKE; RODRÍGUEZ-POSE; TOMANEY, 2017). On the other hand, the emphasis could be placed on one instrument only, rather than on a comprehensive battery of programs. A large body of research highlights that, even if the aggregate impact of infrastructure policies has sometimes been positive, they have often led to greater economic agglomeration, regional polarization, and to an increasing economic marginalization of many peripheral regions where significant infrastructure investments have taken place, both in Europe (VANHOUDT; MATHÄ; SCHMID, 2000; PUGA, 2002; DALL'ERBA; LE GALLO, 2008), and in the emerging world (e.g., Roberts *et al.* (2010) for China). Similarly, state aid and industrial intervention has wasted resources on declining industries, lame ducks, and big projects (ULLTVEIT-MOE, 2008). In general, these policies have struggled to cope with the more heterogeneous economic reality emerging from globalization (ROBERTS, 1993), often ending as "strategies of waste" (RODRÍGUEZ-POSE; ARBIX, 2001).

We intend to shed light on the understanding of the imbalanced growth paths of regions in Uruguay by constructing a multidimensional indicator of development at the regional level. The provision of rigorous evidence on regional development disparities is a critical ingredient for both the guidance and the design of public policy. Multidimensional regional development indicators constitute an effort to provide policy makers with relevant information on aspects of regional development processes that need to be improved or fostered. As such, they constitute a measure to monitor the results of regional development policy. Our multidimensional indicator follows the premises of previous social indicators in that it considers relevant aspects of development. In the same lines as the Human Development Index (HDI) evaluates income, education and health, we consider various dimensions of human life that intent to capture the development attainment for the inhabitants of a region and to guide policy making. The multidimensional characteristic of our measure provides regional decision-making authorities with an opportunity to better comprehend the shortcomings of their areas. At the same time, the convergence exercise provides evidence of the dynamics of regional growth paths during a period in which the country grew at positive rates uninterruptedly. We also thereby contribute to the debate about the design of regional development policies in Uruguay.

To test one aspect of the regional development process in Uruguay, namely the tendency for differences across regions to fade away with time, we conduct a convergence analysis. Most of the existing literature analyses regional development convergence using GDP per capita (e.g., Barro (1991), Bartkowska and Riedl (2012), and Borsi and Metiu (2015)). Here we study regional development convergence using a multidimensional indicator of development.

We use the method of Phillips and Sul (2007) to study regional development convergence. Even though this method has been widely applied to ana-

lyze convergence at the country level, only a few recent studies in the regional development literature have employed it (e.g., [Rodríguez Benavides, Herrera, and González \(2016\)](#) and [Tian *et al.* \(2016\)](#)). We do not follow an exhaustive spatial exploratory analysis of regions in the spirit of [Rey and Janikas \(2005\)](#). Instead, we pose that as an alternative methodological approach to capture regional aspects of development in Uruguay, we can resort to the [Phillips and Sul \(2007\)](#) method.

As far as we know, there are no studies in the literature performing an analysis of regional convergence with a multidimensional development indicator using the methodology of [Phillips and Sul \(2007\)](#). We contribute to the literature by building a synthetic indicator based on four dimensions of development that provides relevant information for policy making and conduct a convergence analysis of the regional development process in Uruguay.

In this paper we analyze the process of convergence of the 19 departments of Uruguay for the period of 2007-2015. Uruguay is a small South American country which has no significant geographical accidents that could make regions particularly difficult to either access (i.e., connect with the rest of the country) or to produce at (extreme weather or adverse climate conditions). Examples of such conditions are abundant in other South American countries: the entire Brazilian Northeast and the Chaco, in Paraguay, are arid regions that are also relatively poorer. The Southern part of Chile and the jungle regions in Peru and Bolivia also constitute examples of relatively difficult to access areas of countries that are also laggards in the development process.

Uruguay is a unitary republic, where each of the 19 departments enjoy limited autonomy. Major investments in infrastructure (highways, communication, etc.), education policy (location of educational centers), and healthcare policy (public hospitals and number of doctors) are mostly determined by the central government. The institutional framework is also essentially the same for all departments: income and corporate tax structure, labor market conditions (i.e., same regulations and minimum wage), etc.

In addition, Uruguay has a strong institutional framework: it ranks second in the Americas according to the corruption perception index 2020 (and 21st in the world), and it is considered a full democracy by the democracy index of the Economist Intelligence Unit, ranking 15th in the world. Strong institutions coexist with long lasting political agreements on areas like energy (Uruguay currently produces electricity entirely from renewable sources). Uruguay grew at positive rates between 2003 and 2019, substantially decreasing poverty and income-based inequality measures. The international financial crisis that took place in 2008 barely affected the Uruguayan economy. It is interesting to study regional economic development in such circumstances.

Altogether, these characteristics suggest that the hypothesis of convergence across all departments seems plausible for Uruguay. With no significant geographical barriers and a set of regulations that is essentially the same for all regions, one could expect factors of production to move across internal borders at nearly zero cost, thereby contributing to a relatively homogeneous growth and development process.

Uruguay is an emerging economy that shares many characteristics with other Latin American countries. Relatively backward regions, most notable the northeast region, coexist with more developed areas, in the southern region of the country. This regionalization pattern is essentially the same since the emancipation of Uruguay from the rule of the Spanish and Portuguese em-

pires in the early stages of the XIX century. By then, Uruguay had no industries other than extractive primary activities related to the huge stock of cattle introduced in the region in the XVI century. The country pursued the elusive quest for growth followed by many other less developed countries in the 1900, alternating periods of trade openness with import substitution industrialization and free market reforms. After an economic crisis originated in the financial sector in 2002, Uruguay experienced its largest recorded expansion from 2003 to 2019 but could not eradicate the relative backwardness of some regions, in which some “poverty traps” persist. Uruguay therefore constitutes an example of an emerging economy that has worked hard to build strong institutions and create conditions for long term growth. However, it still exhibits a weak opportunity net for the relatively laggard regions, as many Latin American countries (RODRÍGUEZ MIRANDA; COSSANI; CENTURIÓN, 2022). In this paper we document this process of no convergence and intend to learn from the experience of an emerging economy and its persistent regional disparity.

Our strategy is based on the analysis of multidimensional regional development indicators, which are instruments designed to assess public policy aimed at bridging regional inequality gaps. We test the implications of the neoclassical growth theory in terms of convergence, and interpret the result of no global convergence as evidence of a richer regional development process, consistent with the literature on two persistent equilibria (QUAH, 1997). We therefore argue that the experience of Uruguay should emphasize the need for a debate centered on current regional disparities.

The rest of the paper is structured as follows. The literature review is presented in the second section, including a brief introduction to the methodological aspects of regional development indicators. The empirical methodology is discussed in the third section. The fourth section shows the results. Finally, section five concludes.

2 Literature Review

The literature on convergence originates from the predictions of the neoclassical growth model (SOLOW, 1957; SWAN, 1956). One of the testable propositions of this model is that countries or regions tend towards the same stationary state (i.e., the same GDP per capita) if the economies differ only in the initial GDP per capita (that is, if all the parameters of the models are the same across countries). This proposition was named beta-convergence. In accordance with this proposition, relatively poorer countries or regions should experience higher growth rates than those of the relatively richer ones. Therefore, they will converge in terms of GDP per capita.

At the same time, convergence analyses incorporate the concept of sigma convergence, which states that the dispersion of GDP per capita across countries or regions declines over time.

The empirical evidence using data from countries and regions generally rejected the validity of these propositions (BARRO; SALA-I-MARTIN, 1991; DE LONG, 1988; REBELO, 1990). However, the traditional neoclassical model yields the testable hypotheses of beta and sigma convergence only for countries or regions with equal structures. Therefore, if their structures differ, convergence (if any) would be conditional on these differences. Since convergence and dispersion is conditional on the stationary level of each country or region,

convergence under these conditions are labelled beta and sigma conditional convergence.

The concepts of sigma and beta conditional convergence, were applied to numerous sets of data, including countries with similar characteristics (e.g., members of the European Union) or administrative divisions of countries (e.g., States of the United States of America, Prefectures of Japan, etc.). Different studies present evidence supporting the existence both of beta and sigma conditional convergence for groups of countries or regions (BARRO; SALA-I-MARTIN, 1991, 1992; MANKIW; ROMER; WEIL, 1992; SALA-I-MARTIN, 1996).

The best-known convergence tests assume that, for example, the process of technological growth is homogenous for all countries and/or regions (i.e., all countries and regions experience the same technological change at the same rate during the period of analysis). However, it is possible that due to heterogeneities in the technological growth process, countries or regions with similar economic structures end up converging to different stationary states.

This last proposition gives rise to the hypothesis of convergence clubs: economies that are similar in their structural and dynamic characteristics can converge with each other. Therefore, it can be stated that it is possible for economies to experience a process of convergence within a given group, but that each group converges to different long-term equilibrium (AZARIADIS; DRAZEN, 1990; GALOR, 1996). The econometric techniques to test the convergence in clubs should therefore incorporate heterogeneities in the technological growth processes.

Based on regional and national data for European countries and regions, Canova (2004) documents evidence of convergence in clubs in the product per capita of European countries and regions. The technique proposed by Canova (2004) allows for countries or regions to exhibit different dynamic of growth and posterior steady states to where they converge, also different. The differences found in the processes rest on the existent technological heterogeneity between countries and between regions. The author finds that the European regions converge into four different clubs, while countries converge into two clubs, clearly distinguishable by their structural characteristics.

In order to account for spatial heterogeneity, alternative techniques to the analysis of convergence were proposed (see, among others, Postiglione, Benedetti, and Lafratta (2010), Postiglione, Andreano, and Benedetti (2013), and Panzera and Postiglione (2014)). Postiglione, Benedetti, and Lafratta (2010) analyze 191 European regions between 1980 and 2002 by applying an algorithm based on a modified version of the regression trees procedure, which attend to identify the local stationarity in the economic growth of the different regions. According to their procedure, they initially identify the local stationary states of the different regions, and subsequently divide the regions into groups, if the estimated parameters are significantly different. Moreover, Postiglione, Andreano, and Benedetti (2013) propose a method for identifying clubs of convergence for the European regions based on a spatial augmented version of the conditional β -convergence model. Two different optimization algorithms for the identification of convergence clubs are proposed and compared: Simulated Annealing and Iterated Conditional Modes. Panzera and Postiglione (2014) analyze the conditional β -convergence hypothesis for NUTS 3 Italian provinces considering spatial dependence (by assuming a Spatial Durbin Model specification) and spatial heterogeneity, identifying convergence clubs by applying the modified simulated annealing algorithm

introduced by Postiglione, Andreano, and Benedetti (2013).

In this same research line but with a different approach, Phillips and Sul (2007) assume that technological change depends on each country or region characteristics and that it also evolves in a different way over time (the technological growth rate is not constant throughout the period). Phillips and Sul (2007) apply this criterion to study the convergence patterns in clubs of the cost of living indexes for 19 metropolitan regions of the United States.

Applications of the Phillips and Sul (2007) approach at a country level can be found Monfort, Cuestas, and Ordonez (2013) and Borsi and Metiu (2015). Monfort, Cuestas, and Ordonez (2013) analyze the convergence in clubs of the product per worker in 14 countries of the European Union and find evidence that supports the hypothesis of convergence into two clubs: Central Europe and the Eastern countries plus Greece. Borsi and Metiu (2015) study patterns of convergence for countries of the European Union between 1970 and 2010. According to their results, no convergence exists between the countries, although they find convergence in clubs. At a regional level, Tian *et al.* (2016) and Rodríguez Benavides, Herrera, and González (2016) study convergence from the perspective proposed by Phillips and Sul (2007). Tian *et al.* (2016) analyze regional income inequality convergence for Chinese provinces and find that provincial incomes are converging into two clubs. Furthermore, Rodríguez Benavides, Herrera, and González (2016) study club convergence for the states of the Mexican Republic between 1970 and 2012 and find relative convergence in six groups.

Finally, an extension of Phillips and Sul (2007) was proposed by Bartkowska and Riedl (2012). This research investigates the convergence of clubs for the income per capita of 206 European regions between 1990 and 2002 using a two-stage process. In the first stage they identify groups of regions which converge to the same level of stationary state, while in the second stage authors investigate the role of the initial conditions as determinants of membership of that club. Following this methodology, they find evidence of five convergence clubs, each one of which converges following its own growth path.

As noted in most of the empirical background reviewed, club-convergence analysis provides useful inputs for economic policy and to spatially orientate actions focused on promoting regional equality (Tian *et al.*, 2016; Bartkowska; Riedl, 2012, among others).

Most of the applied studies that follow this research line analyze convergence on the base of uni-dimensional indicators. Some of them, use a set of variables to condition the model, in order to account for factors that could affect the process of convergence. See, among others, Tian *et al.* (2016) for Chinese regions.

Nevertheless, as far as we know, empirical studies that consider multidimensional indicators to analyze club convergence, either at country or at the regional level, are scarce. The present paper, which performs a convergence analysis of clubs for the 19 political divisions of Uruguay¹ on the base of a multi-dimensional developing indicator, contributes to the empirical literature in that sense. Considering a synthetic and multidimensional indicator (which includes institutional, social and economic factors) to analyze regional convergence may provide richer and precise orientations to the design of pol-

¹ See [here](#) for a political map of Uruguay showing its 19 departments.

icy actions to promote equality in the regional development steady states and paths.

This paper contributes to the literature on regional development indicators. These measures usually synthesize many dimensions of development into a single indicator and are instruments for the improvement of public policies. In a couple of aspects, the *Indice de Competitividad Estatal* (State Competitiveness Index), built by the Mexican Institute for Competitiveness constitutes a precedent of our indicator. The latter index evaluates the competitiveness of Mexican states based on ten dimensions related to the ability of each region to generate, attract and retain investment and human capital (IMCO, 2016, 2020). Each dimension intends to capture different aspects of competitiveness (i.e., property rights, inclusion, sustainability, poverty, health, etc.) and is evaluated according to 97 variables. Criteria for selecting variables include their periodicity, transparency, availability, and non-redundancy. In previous versions of this index², dimensions (also referred to as subindexes) were constructed by means of a principal component analysis (PCA) applied to each set of variables that comprised each dimension. Dimensions were then aggregated into the index using weights that were determined in part by econometric methods (50% of the weight derived from a regression of the average of investment and human capital³ on all dimensions), and in part (the other 50%) the result of surveys to experts at the IMCO.

Argentina and Chile follow somewhat related methodologies to construct regional development indexes. In the former, the *Dirección Nacional de Relaciones Económicas con las Provincias* is in charge of the *Indicador de Desarrollo Relativo Provincial* (Relative Development Provincial Index), composed of two dimensions and 16 variables in total. In this case, however, no specific weights are assigned to variables (each dimension is a simple average of variables, and the index is a simple average of dimensions). Chile, through the *Instituto Chileno de Estudios Municipales* measures imbalances in regional development using the *Índice de Desarrollo Regional* (Regional Development Index). This measure is composed of seven dimensions of development that comprise a total of 32 variables. Variables are aggregated through simple averages in each dimension and dimensions are given weights assigned by a panel of 64 experts (four dimensions have a weight equal to 17.5% and the other three have a 10% weight). All indexes reviewed consider more than one dimension of development, whereas they differ in how to aggregate the information contained in the variables used. We take from them many aspects, but choose not to incorporate the opinion of experts, to construct an objective measure.

More recently, efforts have been made to construct a multidimensional regional development indicator for some Latin American countries. This index is composed of eight selected dimensions of development aggregated using weights determined by the opinion of experts (RODRÍGUEZ MIRANDA; COSSANI; PARRAO, 2021).

These indicators are built to improve the design and implementation of re-

²In the current edition (2020), the PCA analysis is dropped and 70% of the weight is assigned to reduce the variance of dimensions (the weight is the inverse of each dimension variance), and 30% of the weight comes from surveys to experts at the IMCO.

³State Investment was calculated as the national share of investment applied to estimates of Gross State Product. Human Capital was estimated as the share of population aged 25 or more with secondary education complete. The dependent variable used in regressions was a simple average of these two.

gional development policies. By measuring the performance of many aspects of the development process, these indexes contribute to a better understanding of the dynamics of regional processes, and identify which aspects need to be strengthened.

3 Methodology

3.1 Multidimensional Indicator of Departmental Development (ID)

Development consists of a process which spans multiple dimensions of human life, that transcend income measures. It is understood as comprising not only the improvement of the economic structure and the satisfaction of material needs, but also encompassing dimensions such as security, institutions, inclusion and education.

The regional development indicator is constructed in two stages. The first step is the selection of variables (i.e., Indicators) that capture aspects of economic development. Variable selection is based on the literature on regional development indicators, and have to come from official sources, be periodically published, and provide non-redundant information. Also based on the literature, we group variables (i.e., Indicators: I) into factors (F), according to the following expression:

$$F_j = \alpha_1 I_{j1} + \dots + \alpha_m I_{jm} \quad (1)$$

where F_j denotes the factor of development j , which is assumed to consist of m indicators I , appropriately weighted using weights α . The contribution of each of the indicators to the factor (weights α) are estimated using Principal Component Analysis and reported on Table 1. We extract the weights α 's associated with the first principal component for each set of variables for each Factor. The PCA allows to optimally represent in a space of small dimension, observations of an n -dimensional general space, while transforming the original variables, generally correlated, into new uncorrelated variables (PEÑA *et al.*, 2002). The PCA analysis implies an indicator structure temporarily unchanged within each factor.

The second step is to obtain the weight corresponding to each Factor. To this end, we run the following regression:

$$y_{it} = \beta_0 + \beta_1 F_{1it} + \beta_2 F_{2it} + \dots + \beta_n F_{nit} + \epsilon_{it} \quad (2)$$

where y_{it} is the GDP per capita of department i at the year t . The estimated β_1 to β_n will be the weights of each dimension in the multidimensional indicator of development, assuming there are n Factors.

The multidimensional development indicator, the Index of Departmental development (ID), is the weighted sum of the Factors. Here we end up considering four Factors of regional development, to be discussed in detail in section 4.1. By following this procedure, we maintain the ability to measure the performance of departments on each factor, thereby generating useful information for the design of public policy. At the same time, we construct an objective indicator, relying heavily on statistical and econometric criteria.

3.2 Analysis of Convergence

In order to perform the analysis of convergence, we follow the methodology proposed by Phillips and Sul (2007), and the Stata package proposed by Du

(2017). The methodology of Phillips and Sul (2007) allows for the endogenous identification of clubs of regions following the same growth path where global convergence does not exist. This represents an advantage over other methodologies in which the determination of the clubs is performed *ex ante* (GONZÁLEZ; VARO; NAVARRO, 2017).

The starting point for the model is to break down the data of the panel X_{it} as:

$$X_{it} = g_{it} + a_{it} \quad (3)$$

where g_{it} represents systematic components such as permanent common components and a_{it} represents transitory components.

This specification may contain a mixture of common and idiosyncratic components in the elements g_{it} and a_{it} . To be able to separate the common elements from the idiosyncratic in the panel, we transform equation (3) in the following one:

$$X_{it} = \left(\frac{g_{it} + a_{it}}{u_t} \right) u_t = \delta_{it} u_t \quad (4)$$

where δ_{it} is an idiosyncratic component that changes over time and u_t is a component common to all regions which changes over time. If u_t represents a component of a common trend in the panel, then δ_{it} measures the relative participation in u_t of the individual i at moment t . Equation (4) is a dynamic factorial model, where u_t captures some deterministic or stochastic behavior of the trend, and δ_{it} measures the idiosyncratic distance between the component of common trend u_t and X_{it} .

In equation (4), the number of observations in the panel is less than the number of unknowns in the model. Therefore, to be able to estimate the load coefficients δ_{it} some structure for δ_{it} and u_t must be given. Phillips and Sul (2007), propose removing the common factor in the following way:

$$h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}} \quad (5)$$

where h_{it} is the parameter of relative transition which measures the trajectory of each region or country i from the relative position of departure towards the common growth path. That is, it shows the transition trajectory of the region i in relation to the average of the panel. As can be seen in equation (5), the regional average of h_{it} is 1 by definition, and the variance meets the following condition:

$$H_t = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 \rightarrow 0 \text{ if } \lim_{t \rightarrow \infty} \delta_{it} = \delta, \text{ for all } i \quad (6)$$

Therefore, the hypothesis of relative convergence of Phillips and Sul (2007) requires that equation (6) is met. To be able to specify the null hypothesis of convergence, Phillips and Sul (2007) model δ_{it} in the following way:

$$\delta_{it} = \delta_i + \sigma_{it} \epsilon_{it}, \sigma_{it} = \frac{\sigma_i}{L(t)t^\alpha}, t \geq 1, \sigma_i > 0 \text{ for all } i, t \quad (7)$$

where α represents the speed of convergence.

The following hypothesis test is proposed to test the existence of global convergence:

$$H_0 : \delta_i = \delta \text{ and } \alpha \geq 0,$$

$$H_1 : \delta_i \neq \delta \text{ and } \alpha < 0$$

To decide whether or not to reject the null hypothesis, a model of regression $\log t$ is applied:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log(\log(t)) = a + b\log(t) + \mu_t, t = 1, 2, \dots, T \quad (8)$$

If there is convergence the variance across regions tends to zero, then $h_{it} \rightarrow 0$ and $H_t \rightarrow 0$, which implies that $\log(\frac{H_1}{H_t}) \rightarrow \infty$. For this to happen, $b \geq 0$ must happen. If $b < 0$, the hypothesis of global convergence is rejected, and we therefore proceed with the identification of possible convergence clubs. The identification of convergence clubs is performed through the application of an iterative algorithm developed by [Phillips and Sul \(2007\)](#). The iterative procedure to identify the convergence clubs is summarized in five steps:

1. **Ordering of the data panel by cross-section:** sort the data panel from highest to lowest based on observations in the last period (in this case 2015).
2. **Formation of convergence clubs:** groups of regions start to form (in this case departments) from the highest value of each variable in 2015, in such a way that the groups will be formed by a number of regions $2 \leq k < N$. The regression $\log t$ test is applied for the first group and the statistical convergence statistic t_k is calculated, choosing the value of k^* that maximizes t_k (at a significance level of 5%) in accordance with the following criterion:

$$k^* = \arg \max k \{t_k\} \text{ conditioned to } \min \{t_k\} > -1.65$$

This is done for the first two regions, and if it happens that the criterion is not met, then it is repeated with the second and the third and so on successively until a pair of regions meet the criterion. If it should be the case that there are no pairs of regions/departments that meet the criterion, we can conclude that no convergence clubs exist in the data panel.

3. **Screen the data to form convergence clubs:** if in the previous step there is a pair of departments which meet the established criterion, then further departments are added in the order that they appear in the data panel (considering that they are already ordered) until the criterion is no longer met. At this point we set the critical value c^* to the conservative value of zero, in accordance to [Phillips and Sul \(2007\)](#) and [Sichera and Pizzuto \(2019\)](#), since $T = 9$ in our sample.
4. **Repetition and detention rule:** we begin with the department that broke the rule in the previous step. Department after department is added while the established criterion is met. Once the criterion is broken, we stop and begin again. In the case where k does not exist in step 2 whose $\{t_k\} > -1.65$, it is concluded that the departments are divergent.
5. **Club merging:** [Schnurbus, Haupt, and Meier \(2017\)](#) propose a fifth step, which we incorporate into the analysis following [Du \(2017\)](#). The procedure, known as “club merging”, implies the union of those clubs that meet the convergence hypothesis together. The procedure consists

of running the regression $\log t$ test for the initial clubs 1 and 2, and if they meet the convergence hypothesis together, joining them to form a new convergence club, named now club 1. Later, run the $\log t$ test for the new club 1 and the initial convergence club 3 and continue identifying whether the convergence hypothesis is met by the two clubs together, and so on, successively, forming all the possible combinations of clubs. This way, one would reach the lowest possible number of convergence clubs.

4 Results

4.1 Multidimensional development indicator

At the first stage of the construction of the indicator we considered 6 dimensions or Factors and 44 variables or Indicators. Variables with scarce variation, low quality and intermittent availability were discarded from the analysis at this stage. Dimensions reflect factors of regional development, following previous studies (IMCO, 2020; MINISTERIO DE HACIENDA, 2018; UNIVERSIDAD AUTÓNOMA DE CHILE, 2019) that seek to provide useful inputs to public policy. After discarding low quality variables, we proceeded to conduct the PCA analysis with 32 variables, grouped in six dimensions or Factors. Weights α 's are obtained from conducting a PCA on each set of variables (i.e., each Factor) separately. We construct each Factor by applying the α weights emerged from each PCA to the standardized values of each variable.

At the second stage, equation (2) is estimated through a Panel Corrected Standard Errors (PCSE) model, which corrects for heteroscedasticity in the panel and contemporary autocorrelation to ensure reliable standard errors (BECK; KATZ, 1995). When estimating equation (2), two Factors (containing a total of 14 variables) did not prove to be significant, and were consequently discarded⁴. Therefore, our multidimensional indicator of departmental development (ID) is constructed as the weighted sum of four factors⁵ that aggregate a total of 18 variables (a brief description of final Factors and variable sources are included in Table A.1 in the Appendix A).

Table 1 shows the results from the application of the principal component analysis for final significant dimensions. Table 1 also includes the total explained variation (EV) by the principal component, estimated as the percentage of the total variation, and the overall Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, for each PCA. Variables included within each dimension have the expected sign and explain, in all cases, more than 50% of the variation of each dimension⁶. Thereby, we consider heterogeneous variation in the values of indicators for different departments.

⁴Factors discarded represented environmental and citizen participation dimensions. Full set of variables, dimensions, and initial estimations are available upon request

⁵Factor 1 includes variables related to some characteristics of the judicial system and the efficiency of the police. Although there should be no huge differences in institutional aspects within a unitary republic like Uruguay (we thank an anonymous referee for pointing this to us), this Factor illustrates significant differences across Departments, and it is therefore included in the indicator

⁶The variable number of crimes for every 1,000 inhabitants is calculated as $1/\text{crimes}$, and therefore the expected sign is positive.

Table 1: First principal component for final Factors (F)

1. Citizen security and legal system (F1)					
Component 1	Coef.	Std.Err	Z	P > z	EV
Number of judges per thousand inhabitants (log_judges)	0.35	0.08	4.12	0	64%
Number of crimes per thousand inhabitants (log_oneovercrimes)	0.67	0.03	25.31	0	
Efficiency indicator of the police (efficiency_justice)	0.66	0.03	20.62	0	
KMO statistic	0.5378				
2. Inclusive, educated and healthy society (F2)					
Component 1	Coef.	Std.Err	Z	P > z	EV
Life expectancy (log_expectancy)	0.23	0.05	4.98	0	51%
Percentage of households in poverty (log_poverty)	-0.42	0.03	-15.83	0	
Health sector employees (log_emp_health)	0.37	0.03	10.97	0	
Access to drinking water at home (log_drinking_water)	0.47	0.02	23.31	0	
Rate of female activity (log_fem_act)	0.41	0.03	14.81	0	
Proportion of EAP who have secondary education or higher (log_eap_sec&ter)	0.47	0.02	25.2	0	
Students enrolled in secondary education. In percentage of 11 to 18 years old (log_grad_sec)	0.14	0.05	2.75	0.01	
KMO statistic	0.8255				
3. Efficient and dynamic factors markets (F3)					
Component 1	Coef.	Std.Err	Z	P > z	EV
Average income of employees who have secondary education or higher (log_income_sec&ter)	0.56	0.04	15.58	0	57.8%
Average income (without locative value) per capita per hour worked (log_income_perhour)	0.60	0.02	25.97	0	
Industry productivity (log_prod_industry)	0.44	0.05	8.56	0	
Productivity of commerce, repairs, restaurants and hotels (log_prod_crrrh)	0.37	0.06	6.02	0	
KMO statistic	0.6139				
4. Physical and technological infrastructure (F4)					
Component 1	Coef.	Std.Err	Z	P > z	EV
Internet access at home (log_internet)	0.47	0.06	7.51	0	50.7%
Ownership of landline at home (log_tel)	0.57	0.04	12.98	0	
Passenger movement in ports (log_passengers)	0.29	0.09	3.27	0	
Energy distributed per capita (log_energydistpc)	0.60	0.04	14.68	0	
KMO statistic	0.5050				

Source: own calculations.

Table 2: Estimation for the period 2007-2015 (Eq.2)

Dependent variable: $\log(y)$				
	Coef.	Std. Err.	z	$P > z$
F1	0.05	0.01	4.37	0.00
F2	0.05	0.02	2.9	0.00
F3	0.15	0.04	3.72	0.00
F4(-1)	0.05	0.02	2.62	0.01
_cons	11.92	0.07	171.32	0.00
Rho	0.2921278			

Note: regressions include time dummies.

Source: own computations.

Table 2 presents the estimation of equation (2) for the whole period: from 2007 to 2015⁷. Factor four is included with a one period lag since this improves the estimation. Lagging Factor four improves both the estimation of the coefficient in its corresponding Factor and the overall fit of the model. We follow this procedure because the overall objective of estimating equation 2 is to obtain reasonable weights to aggregate Factors into one indicator, not to conduct any forecasting or inference analysis. Furthermore, we aim to provide an alternative (more rigorous) method for aggregating Factors than currently used in the literature (i.e., development experts suggestions on the relative importance of Factors, as in [Rodríguez Miranda, Cossani, and Parrao \(2021\)](#), and others). Coefficients evidence the weight of each dimension in the variability of the dependent variable explanation, the log of the departmental per capita GDP. Estimation results show that dimension 3 (F3: Efficient and dynamic market of factors) is the one with the higher weight. Nevertheless, dimensions 1, 2 and 4 (F1: Citizen security and legal system, F2: Inclusive, educated, and healthy society, and F4: Physical and technological infrastructure) also provide information about the variability of GDP per capita of departments, since all their coefficients are significant.

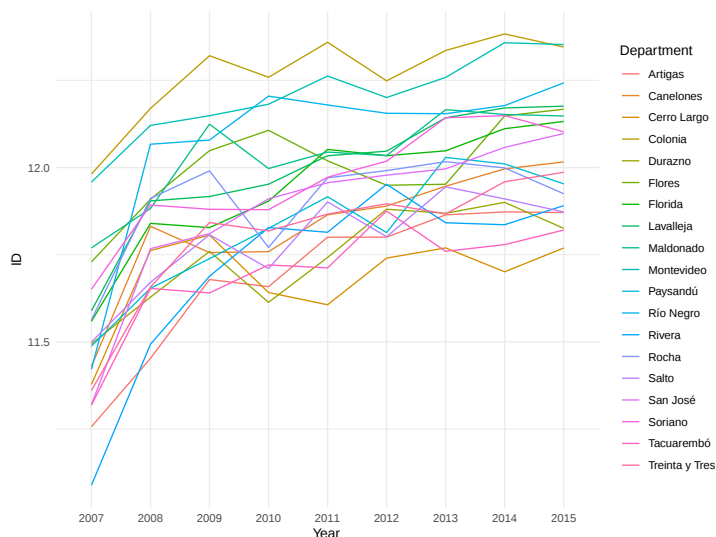
We re-estimate equation (2) using alternative time windows to check the robustness of our results. These exercises show that neither the coefficients, nor their significance change significantly in the alternative estimations (see Tables A.2 and A.3 in the Appendix A).

Using the estimated weights of each factor presented in Table 2, the 19 regional IDs are constructed as the weighted sum of the 4 dimensions (or factors). Figure 1 shows the index for the period 2007-2015 for the 19 Uruguayan departments.

Figure 1 indicates that in general the ID of all departments increases over the period considered, apparently showing signs of sigma convergence, given that the dispersion of the ID of departments tends to decline over time. To test the latter claim, we plot the densities of the 19 departments for three years (see Figure 2): the beginning of the sample (2007), four years later (2011), and the end year of the sample (2015). Casual observation suggests that the variance of the distribution declines from 2007 until 2011 (the distribution is thinner for 2011 than for 2007), but the change from 2011 to 2015 is not

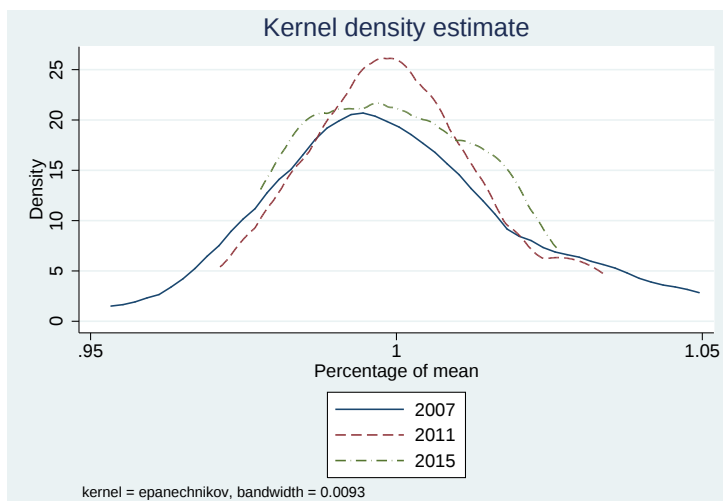
⁷Given that dimension 4 is included with a lag, we lose 1 year and it is only possible to estimate the multidimensional development indicator for 9 years (i.e. from 2007 to 2015).

Figure 1: Multidimensional index of departmental development of Uruguay (2007-2015)



Source: own computations.

Figure 2: ID Densities for 2007, 2011 and 2015



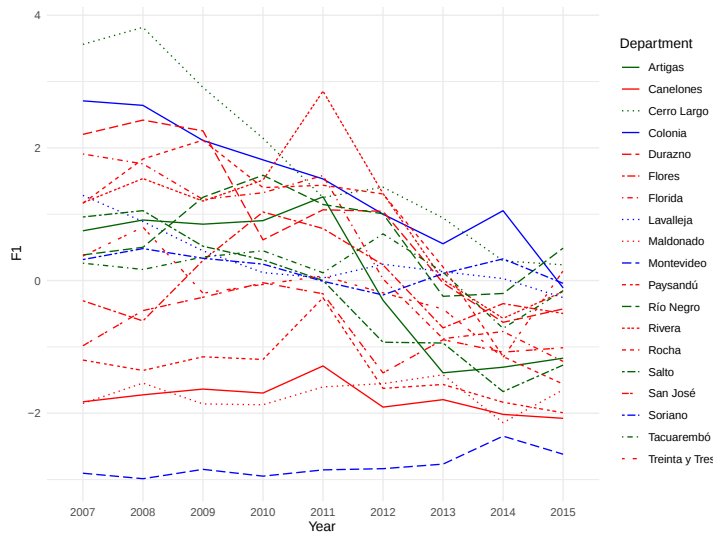
Source: own computations.

that clear. Estimated standard deviations for the ID are 0.29, 0.19, and 0.18 for 2007, 2011, and 2015, respectively⁸. We can therefore not conclude on the existence of sigma convergence in the ID during the period.

Our methodology permits to disentangle the evolution of the ID by considering each one of the Factors (development factors or dimensions) and analyze its evolution over time. Figures 3 to 6 display each one of the four Factors over the whole period.

Figure 3 shows a negative trend in the dimension “Citizen Security and

⁸Summary statistics are available at the Appendix A (see Table A.4).

Figure 3: Factor 1: Citizen Security and Legal System

Source: own computations.

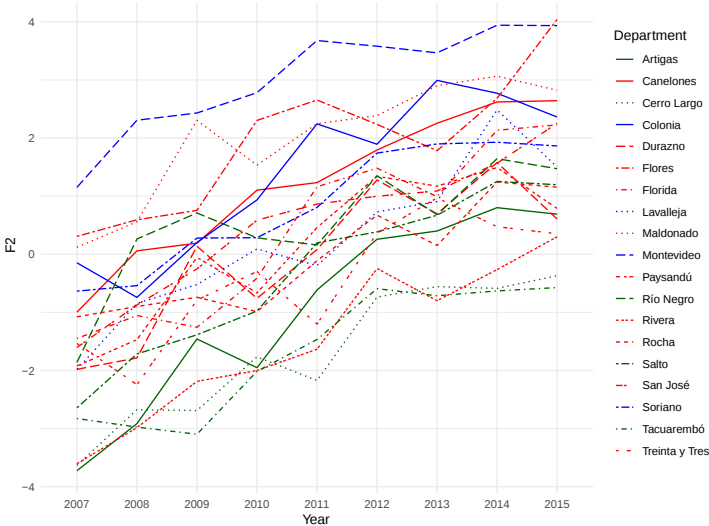
Legal System” for most departments. Although some departments exhibit an improvement in its relative position, overall, a convergence towards lower absolute values in this dimension can be observed. The overall downward trend exhibited by most departments in this Factor indicates that the benefits of the growth process experienced by the country during the period did not affect all dimensions of human development in the same way. The decrease in the values of this dimension of development indicates that policymakers should consider improvements in this area. The evolution of the Montevideo department is interesting to note, it is consistently the lowest ranked region throughout the analysis, with only a modest improvement at the end of the period.

As opposed to Figure 3, Figures 4 and 5 show an upward trend in their dimensions. Figure 4 plots the departmental values for the second factor of the ID: “Inclusive, Prepared and Healthy Society”. In this case we observe values that are somewhat more dispersed than in Figure 3, indicating a higher variance in the second dimension of development than in the former one. Also, departments that belong to club 3 (in green color) are consistently at the bottom of the graphic representation, indicating that improvements in this factor are critical for these Departments.

Figure 5 plots the third factor: “Efficient and Dynamic Factors Markets”. The most important characteristic exhibited by this factor is the relatively (when considering the three other factors) low variability of the values across Departments and years. This indicates that in the most relevant Factor (in terms of the weight associated to it), departments with relatively low/high values are not that far apart. Income and productivity measures for most Departments increased consistently during the period, at a higher pace during the period 2007 – 2010.

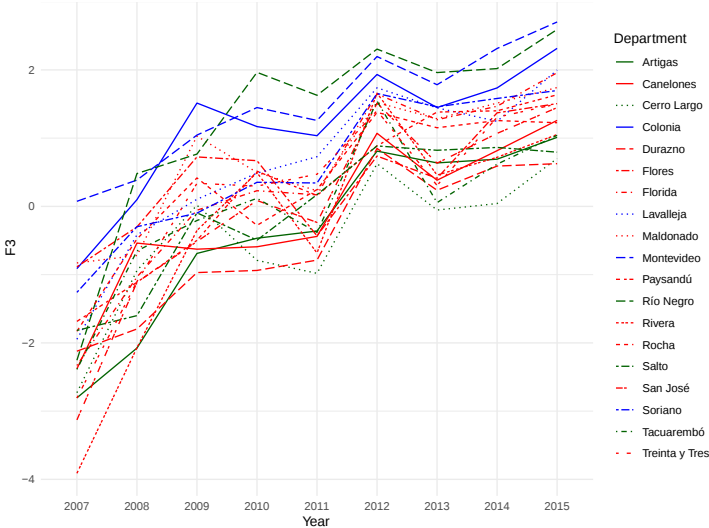
In Figure 6 a positive trend in the dimension “Physical and Technological Infrastructure” is noted, although not as marked as in the two previous cases. Values for this factor show a relatively higher dispersion than for the previous

Figure 4: Factor 2: Inclusive, Educated and Healthy Society

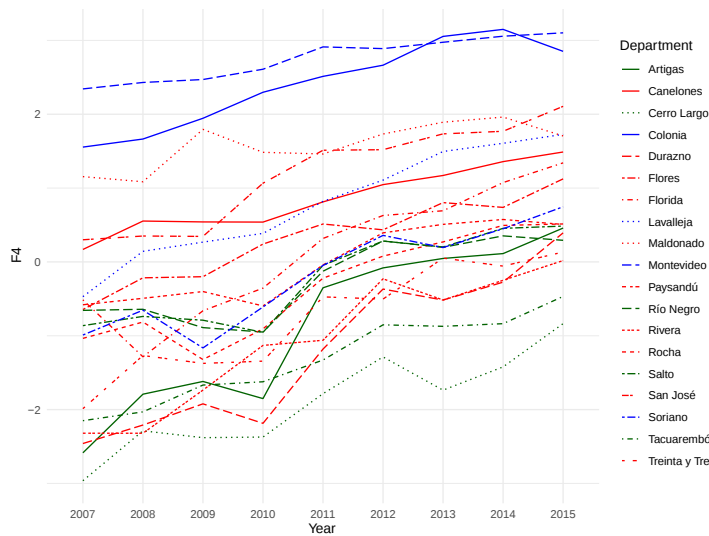


Source: own computations.

Figure 5: Factor 3: Efficient and Dynamic Factors Markets



Source: own computations.

Figure 6: Factor 4: Physical and Technological Infrastructure

Source: own computations.

factor. In this factor, both Montevideo and Colonia (two Departments that are in club 1) are consistently at the first and second place of the distribution.

4.2 Convergence

Once the indicator of departmental development (ID) has been constructed, the convergence analysis is performed following the methodology proposed by Phillips and Sul (2007). The data panel contains information of the ID for the 19 departments of Uruguay for the period of 2007 to 2015. Thus, the panel data is composed of 171 observations.

To perform this analysis we follow the steps proposed by Du (2017). Firstly, the existence of global convergence in the 19 departments is analyzed, and the convergence hypothesis is estimated at a significance level of 5%. The log t test is applied to the variable ID. The coefficient estimated is $b = -0.89$ and the value of the statistic is $t = -2.73$. In accordance with the decision rule ($t < -1.65$), the null hypothesis of joint convergence is rejected, and therefore, the possible existence of “clubs” of convergence between departments must be analyzed.

An iterative process is then performed as proposed by Phillips and Sul (2007) to determine the existence of convergence clubs. Table 3 shows the results of the club analysis applied to the data panel. As can be seen in the table, three convergence clubs are found, and no divergent department is identified. The values of the t statistic in the three cases show that convergence cannot be rejected (we cannot reject that b is equal or greater than 0).

Club 1 is formed by four departments: Colonia, Lavalleya, Montevideo and Soriano. This is the club which on average, considering the ID, shows the best performance throughout the period and includes the capital of Uruguay where more than 40% of the national population resides. The value of b is linked to the speed of convergence, and hence this first club experiences the fastest convergence of the three clubs.

Table 3: Convergence Clubs in the Multidimensional Index of Departmental Development

Club	Number of Members	Departments	b	t-Stat.	ID Averages by year		
					2013	2014	2015
1	4	Colonia, Lavalleja, Montevideo, Soriano	0.19	0.34	12.22	12.27	12.24
2	10	Canelones, Durazno, Flores, Florida, Maldonado, Paysandú, Rocha, Río Negro, San José, Treinta y Tres	-0.14	-0.27	12.00	12.05	12.05
3	5	Artigas, Cerro Largo, Rivera, Salto, Tacuarembó	-0.19	-0.30	11.84	11.82	11.85

Source: own computations.

Club 2 is the largest of the three by size and comprises 10 departments: Canelones, Durazno, Flores, Florida, Maldonado, Paysandú, Rocha, Río Negro, San José and Treinta y Tres. In this case, the coefficient estimated for b is less than 0 (even though the t -statistic shows that we cannot reject the hypothesis that $b \geq 0$), which indicates weak evidence of club convergence. This club 2 shows a medium performance of its ID index.

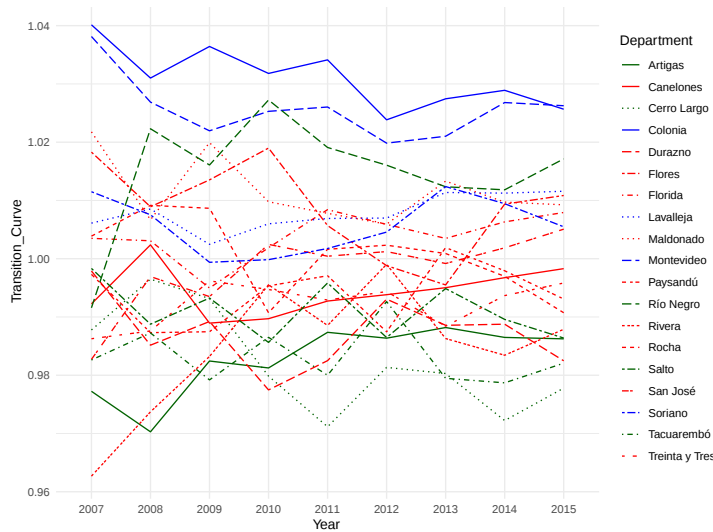
Club 3 is formed by 5 departments: Artigas, Cerro Largo, Rivera, Salto and Tacuarembó and represents those departments which have had the worst average performance in the value of the ID. As with Club 2, the coefficient estimated for b is less than 0 (even though we cannot reject the hypothesis that it is 0 or greater than 0 in statistical terms), which represents weak evidence of convergence within the club. Club 3 shows the worst ID performance throughout the period.

Results from the convergence analysis are consistent with the findings of the local spatial exploratory analysis: for the year 2015, club 3 members are identified only as cold spots, whereas club 1 members are identified only as hot spots. Hot spots departments are members of Clubs 1 and 2 only (see Table A.7).

For clubs 2 and 3, the number of departments that maximize the t statistic for each subgroup determines that b point estimates are negative, although in neither case sufficiently large to reject the null hypothesis of convergence. In the spirit of Phillips and Sul (2007, pp. 1811) this constitutes rather weak evidence of convergence.

Finally, the fifth step (*club merging*) is applied as suggested by Schnurbus, Haupt, and Meier (2017). That is, we analyze the possibility of merging those clubs which together satisfy the convergence hypothesis. This exercise did not result in the union of any of the clubs. As a result, the final classification is the 3 clubs which are shown in Table 3.

As noted above, h_{it} measures the path of each region i from its relative position of departure towards the common growth path. This parameter can be interpreted as showing the extent to which a region shares at each point in time in the common growth component. Additionally, since h_{it} is time dependent and provides a description of how this share evolves over time, and therefore traces out a transition curve for region i .

Figure 7: Departments Relative Transition Curves

Source: own computations.

Note: Club 1: blue, Club 2: red, Club 3: green.

According to Figure 7, transition curves⁹ (h_{it}) for each department show that clubs converge to different stationary states. Between 2008 and 2011 there was a slight convergence between the three clubs, while since 2012 the transition behavior differs across clubs.

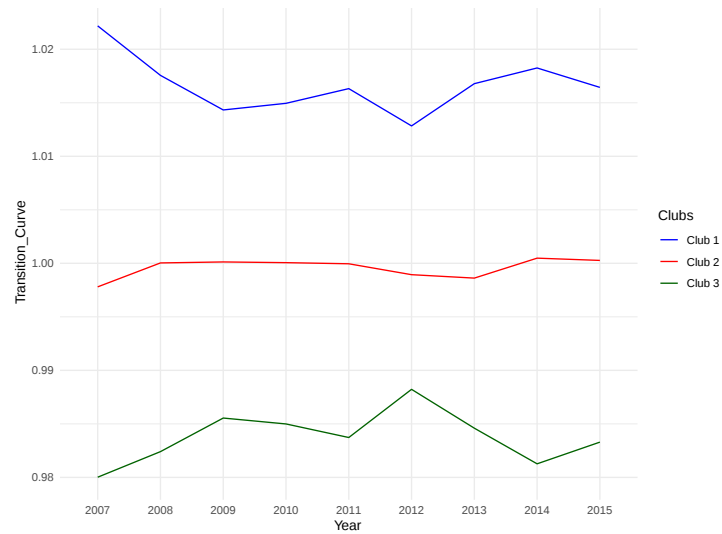
Alternatively, in Figure 8, transition curves are represented by clubs and in terms of the global average of the panel. Values of h_{it} above one indicate regions (or a club in this case) whose share in the common growth path exceeds the panel average at time t . As previously mentioned, Club 1 is the one that shows the best performance with values always better than the average, Club 2 shows values around the average (oscillating around 1), and Club 3 always presents below average values. According to theory, under the assumption of convergence for the full panel of departments, the relative transition path tends to unity for all departments. On the other hand, under the assumption of club convergence (i.e., when groups of departments converge to different equilibria) the relative transition paths of the members of each club converge to different constants.

Observing the geographical location of the convergence clubs, it can be noted that there is an association between belonging to a club and its location on the map of Uruguay. In other words, clubs are mostly defined by its geographical localization. Therefore, it could be said that the club convergence method captured the geographical dimension of the development index.

As Figure 9 indicates, Club 1 is comprised of departments which are in the southern half of the country, including the capital. On the other hand, among the 10 members of Club 2, only 2 departments are in the northern half. Finally, Club 3 is comprised of departments from the north and northwest of the country, presenting a clear territorial distribution. Results are similar but not equal to those found by national studies [Rodríguez Miranda and Menén-](#)

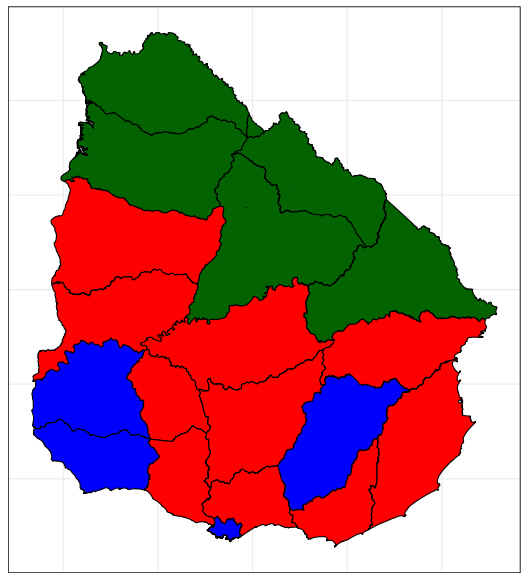
⁹To estimate transition curves, we use equation (5) of section 3.2

Figure 8: Clubs Relative Transition Curves



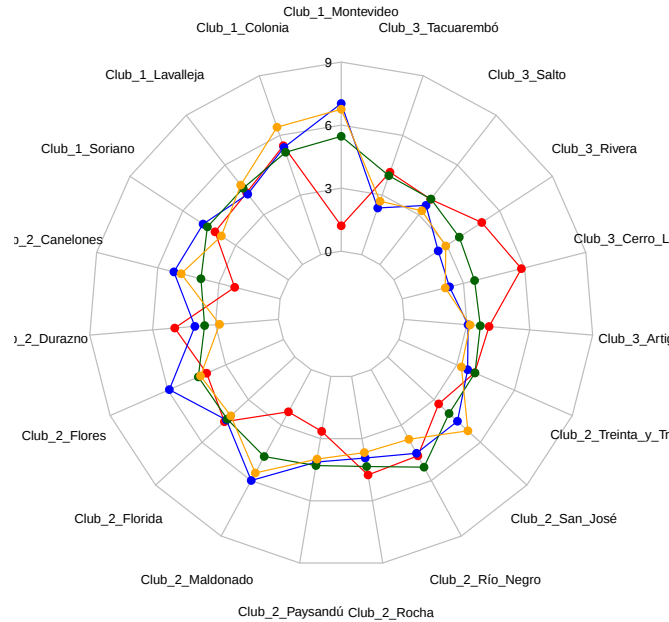
Source: own computations.

Figure 9: Convergence clubs in Uruguay



Source: own computations.
Note: Club 1: blue, Club 2: red, Club 3: green.

Figure 10: Average of factor values by department (2007-2015). Averages +4 for ease of presentation.



Source: own computations.

dez (2020) that use different methodologies for the regionalization as well as a one-dimensional indicator (GDP per capita).

The results could be interpreted as a reflection of a relative lag in development that exists in the departments in the north and northwest region of the country. Therefore, regional development grade seems to fade as departments or regions distance from the country's capital, which boasts one of the highest values in the development index. Additionally, the multidimensionality of our indicator identifies the factors or dimensions that may explain club-convergence results. Figure 10 could give a first approach to this issue, but an extensive analysis has to be conducted in future research in order to identify the causality from development factors to regionalization results in the convergence analysis.

Figure 10 is constructed by averaging values in each factor across departments. This graphic representation aims to identify areas in which a department needs to significantly improve its performance. As can be noted, departments that are grouped in Club 1 have the highest average values in almost all the dimensions (except for Montevideo in the first factor), mainly in factor 3 (which has the highest weight in the index). Departments that belong to Club 3 shows small average values in almost all dimensions, meanwhile Club 2 shows more heterogeneity. It is clear from the graph that Montevideo, Maldonado and Canelones need to design and implement policies that can modify aspects of development referred to citizen security and measures of efficiency of the police. Cerro Largo has the lowest average values on factors 2 and 3, indicating a critical situation of that department in these factors.

5 Conclusions

We build a multidimensional indicator of regional development (ID) based purely on statistical and econometric techniques. Then, we investigate the dynamics of regions by conducting a convergence analysis that follows the methodology of [Phillips and Sul \(2007\)](#). As far as we know, no previous studies have applied this analysis to a multidimensional indicator of regional development as the one constructed here. The periodical generation of development indicators similar to ours, together with the convergence analysis, could become an important instrument for territorial development policy aimed at closing regional development gaps.

The convergence analysis performed in this paper indicates that, even though Uruguay is a small size country that exhibited a strong dynamism in the period analyzed, there is no evidence of global convergence between its 19 departments. This means that, although all departments show an improvement in their level of development, disparities between them do not seem to be reduced, except for specific regions. These results are in line with those found by [Rodríguez Miranda and Menéndez \(2020\)](#), who suggest that the positive impact of the economic expansion was not homogeneous in all regions and point out the persistence of inequities linked to structural features. In addition, [Rodríguez Miranda, Cossani, and Centurión \(2022\)](#) also find that most Latin American countries exhibit persistent inequalities across regions.

We found a regionalization in terms of convergence, in which 3 distinct convergence clubs can be distinguished which seems to have a geographical correspondence. Clubs are characterized by different levels of development and speed of convergence and have a territorial correspondence.

The development indicator and the convergence analysis performed in this paper could be useful for regional and national public policy. During the period analyzed Uruguay exhibited characteristics instrumental to closing inequality gaps. Although income-based measures of inequality were certainly reduced at the country level, departmental convergence was not found. Our paper constitutes evidence in favor of the persistence of regional inequalities, common to most countries in the world, that do not disappear even in the presence of periods of positive growth at the country level. Our methodology can be readily applied to other contexts, to monitor the performance of different regions in terms of development.

The findings of this paper contribute to the debate of the optimal design of development policies. Here we document the persistence of geographical inequalities in the development process of Uruguay. These inequalities have been present for a long time in this country. Starting at the country's birth as an independent republic (in 1828), the departments located in the North East region have always been laggards in the growth process. This process is consistent with the no overall convergence and the convergence in three clubs findings. The former means that inequalities persist (as all departments have not or are not converging to the same ID level), while the latter documents the existence of a set of departments that converge to a relatively low ID level (club 3).

When the different dimensions of development are analyzed separately, factors 2 and 4 (Inclusive, Educated and Healthy Society; and Physical and Technological Infrastructure) stand out as those that reflect better the persistent heterogeneity in development of the three clubs.

Factor 4 is linked to the productive structure of the country, where most of the GDP is generated close to the capital of the country, Montevideo. This also implies an imbalance in infrastructure investment and the incapacity for generating policies that attract more private investment to less developed regions, with its associated endogenous capacity of generating the attraction of public infrastructure. This is partially changing with the construction of mega paper mills in the less developed areas of the country in the last years. However, this is an ongoing process that probably is not yet captured by the data that we are working with.

Factor 2 includes variables associated with poverty, health and education. Some of these variables are endogenously associated with the productive structure of the different regions. However, variables related to the education policy can also be an attraction factor for private investment to the different regions. Recently, a policy of installing tertiary education institutions in less developed areas of the country has been put in place. This again is a very recent development in the country, and it could probably have an impact in the convergence of regions in the future. In our data, this is not yet reflected.

As we cannot attribute the absence in general convergence, or the inability of some regions to take off, to autonomous departmental economic policies or significant geographical accidents, we can conjecture that economic development policies would not have been suitable enough for lagging regions. According to our findings, one should further delve into which are the ultimate reasons for these differences in development paths to be so persistent. Our paper, thus, emphasizes relevant regional development debates, like the place-based versus place-neutral [Barca, McCann, and Rodríguez-Pose \(2012\)](#) approaches to regional development.

Our analysis is not exempt from limitations that should be addressed in future research. First, data availability restricts the sample of the convergence analysis to 9 years. This constitutes a relatively short horizon to analyze development processes. Second, we do not perform an exhaustive spatial exploratory analysis. Finally, our methodological approach with respect to principal component analysis does not incorporate spatial and temporal dimensions. All these issues constitute avenues for further research.

6 Acknowledgments

We thank the financial support of the National Agency for Innovation and Research of Uruguay (ANII). We are grateful for the comments to earlier versions of this paper by participants at the 2018 Annual Meetings of the Central Bank of Uruguay; Seminar Series at the Economics Institute of the University of the Republic (Uruguay); and presentations of the conclusions of this paper at the Office of Budget and Planning (Presidency of the Republic – Uruguay) and the Development National Agency of Uruguay. Any remaining errors and limitations are our responsibility.

References

- AZARIADIS, Costas; DRAZEN, Allan. Threshold externalities in economic development. *Quarterly Journal of Economics*, v. 105, n. 2, p. 501–526, 1990.

- BARCA, Fabrizio; MCCANN, Philip; RODRÍGUEZ-POSE, Andrés. The case for regional development intervention: place-based versus place-neutral approaches. *Journal of Regional Science*, v. 52, n. 1, p. 134–152, 2012.
- BARRO, Robert J. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, v. 106, n. 2, p. 407–443, 1991.
- BARRO, Robert J.; SALA-I-MARTIN, Xavier. Convergence. *Journal of Political Economy*, v. 100, n. 2, p. 223–251, 1992.
- BARRO, Robert J.; SALA-I-MARTIN, Xavier. Convergence across states and regions. *Brookings Papers on Economic Activity*, v. 22, n. 1, p. 107–182, 1991.
- BARTKOWSKA, Monika; RIEDL, Aleksandra. Regional convergence clubs in Europe: Identification and conditioning factors. *Economic Modelling*, v. 29, n. 1, p. 22–31, 2012.
- BECK, Nathaniel; KATZ, Jonathan N. What to do (and not to do) with Time-Series Cross-Section Data. *American Political Science Review*, v. 89, n. 3, p. 634–647, 1995.
- BORSI, Mihály Tamás; METIU, Norbert. The evolution of economic convergence in the European Union. *Empirical Economics*, v. 48, p. 657–681, 2015.
- CANOVA, Fabio. Testing for convergence clubs in income per capita: a predictive density approach. *International Economic Review*, v. 45, n. 1, p. 49–77, 2004.
- DALL'ERBA, Sandy; LE GALLO, Julie. Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis. *Papers in Regional Science*, v. 87, n. 2, p. 219–244, 2008.
- DE LONG, J. Bradford. Productivity growth, convergence, and welfare: comment. *American Economic Review*, v. 78, n. 5, p. 1138–1154, 1988.
- DU, Kerui. Econometric convergence test and club clustering using Stata. *Stata Journal*, v. 17, n. 4, p. 882–900, 2017.
- GALOR, Oded. Convergence? Inferences from theoretical models. *Economic Journal*, v. 106, n. 437, p. 1056–1069, 1996.
- GONZÁLEZ, M^a Candelaria Barrios; VARO, Esther Flores; NAVARRO, M^a Ángeles Martínez. Patrones de convergencia en las regiones españolas: Una aplicación de la metodología de Phillips-Sul. *Revista de Estudios Regionales*, n. 109, p. 165–190, 2017.
- INSTITUTO MEXICANO PARA LA COMPETITIVIDAD — IMCO. *Que no vuelva a pasar: Estados prevenidos valen por dos. Índice de Competitividad Estatal 2020*. 2020.

- INSTITUTO MEXICANO PARA LA COMPETITIVIDAD — IMCO. *Un Puente entre Dos Mexicos. Índice de Competitividad Estatal 2016*. 2016. Disponible em: <https://imco.org.mx/indice-de-competitividad-estatal-2016/>.
- MANKIW, N. Gregory; ROMER, David; WEIL, David N. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, v. 107, n. 2, p. 407–437, 1992.
- MINISTERIO DE HACIENDA. *Indicador de Desarrollo Relativo Provincial. Resultados Regionales. Evolución en el período 2004-2017*. Argentina: Dirección Nacional de Asuntos Provinciales, 2018.
- MONFORT, Mercedes; CUESTAS, Juan Carlos; ORDONEZ, Javier. Real convergence in Europe: A cluster analysis. *Economic Modelling*, v. 33, p. 689–694, 2013.
- PANZERA, Domenica; POSTIGLIONE, Paolo. Economic growth in Italian NUTS 3 provinces. *Annals of Regional Science*, v. 53, n. 1, p. 273–293, 2014.
- PEÑA, Daniel *et al.* *Análisis de datos multivariantes*. Madrid: McGraw-Hill, 2002.
- PHILLIPS, Peter C. B.; SUL, Donggyu. Transition modeling and econometric convergence tests. *Econometrica*, v. 75, n. 6, p. 1771–1855, 2007.
- PIKE, Andy; RODRÍGUEZ-POSE, Andrés; TOMANEY, John. *Local and Regional Development*. 2. ed. London: Routledge, 2017.
- POSTIGLIONE, Paolo; ANDREANO, M. Simona; BENEDETTI, Roberto. Using Constrained Optimization for the Identification of Convergence Clubs. *Computational Economics*, v. 42, p. 151–74, 2013.
- POSTIGLIONE, Paolo; BENEDETTI, Roberto; LAFRATTA, Giovanni. A regression tree algorithm for the identification of convergence clubs. *Computational Statistics & Data Analysis*, v. 54, n. 11, p. 2776–2785, 2010.
- PUGA, Diego. European regional policies in light of recent location theories. *Journal of Economic Geography*, v. 2, n. 4, p. 373–406, 2002.
- QUAH, Danny T. Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of Economic Growth*, v. 2, p. 27–59, 1997.
- REBELO, Sergio Tavares. *Long Run Policy Analysis and Long Run Growth*. Cambridge, MA: NBER, 1990.
- REY, Sergio J.; JANIKAS, Mark V. Regional convergence, inequality, and space. *Journal of Economic Geography*, v. 5, n. 2, p. 155–176, 2005.
- ROBERTS, Mark *et al.* *On the road to prosperity? The economic geography of China's national expressway network*. Washington, DC: World Bank, 2010. (Policy Research Working Paper, n. 5479).

- ROBERTS, Peter. Managing the strategic planning and development of regions: lessons from a European perspective. *Regional Studies*, v. 27, n. 8, p. 759–768, 1993.
- RODRÍGUEZ BENAVIDES, Domingo; HERRERA, Francisco López; GONZÁLEZ, Miguel Ángel Mendoza. Clubs de convergencia regional en México: un análisis a través de un modelo no lineal de un solo factor. *Investigaciones Regionales – Journal of Regional Research*, n. 34, p. 7–22, 2016.
- RODRÍGUEZ MIRANDA, Adrián; COSSANI, Camilo Vial; CENTURIÓN, Irene. El desarrollo como fenómeno multidimensional y territorial en América latina: Un aporte para la construcción de agendas regionales de impacto. In: MAZA ÁVILA, Francisco Javier; SALAS, Antonio Martínez; PÉREZ, María Del Carmen González. *Problemas y Retos del Desarrollo en América Latina*. Colombia: Universidad de Cartagena, 2022. p. 61–92.
- RODRÍGUEZ MIRANDA, Adrián; COSSANI, Camilo Vial; PARRAO, Alejandra. Índice Compuesto y Multidimensional de Desarrollo Regional: Una propuesta para América Latina. *Revista Iberoamericana de Estudios Municipales*, n. 23, p. 1–30, 2021.
- RODRÍGUEZ MIRANDA, Adrián; MENÉNDEZ, María de las Mercedes. *Desigualdades regionales, crecimiento económico y cambio estructural en Uruguay: 1983–2017*. Montevideo: Universidad de la República Uruguay, 2020. (Serie Documentos de Trabajo 12/20).
- RODRÍGUEZ-POSE, André; ARBIX, Glauco. Strategies of waste: bidding wars in the Brazilian automobile sector. *International Journal of Urban and Regional Research*, v. 25, n. 1, p. 134–154, 2001.
- SALA-I-MARTIN, Xavier. Regional cohesion: evidence and theories of regional growth and convergence. *European Economic Review*, v. 40, n. 6, p. 1325–1352, 1996.
- SCHNURBUS, Joachim; HAUPT, Harry; MEIER, Verena. Economic transition and growth: a replication. *Journal of Applied Econometrics*, v. 32, n. 5, p. 1039–1042, 2017.
- SICHERA, Roberto; PIZZUTO, Pietro. ConvergenceClubs: A Package for Performing the Phillips and Sul's Club Convergence Clustering Procedure. *R Journal*, v. 11, n. 2, p. 142–151, 2019.
- SOLOW, Robert M. Technical change and the aggregate production function. *Review of Economics and Statistics*, v. 39, n. 3, p. 312–320, 1957.
- SWAN, Trevor W. Economic growth and capital accumulation. *Economic Record*, v. 32, n. 2, p. 334–361, 1956.
- TIAN, Xu *et al.* Regional income inequality in China revisited: A perspective from club convergence. *Economic Modelling*, v. 56, p. 50–58, 2016.

ULLTVEIT-MOE, Karen Helene. *Live and Let Die: Industrial Policy in a Globalised World*. Expert Report No 22 to Sweden's Globalization Council. 2008.

UNIVERSIDAD AUTÓNOMA DE CHILE. *Índice de Desarrollo Regional IDERE 2019*. M. Camilo Vial Cossani, editor. 2019.

VANHOUDT, Patrick; MATHÄ, Thomas; SCHMID, Bert. How productive are capital investments in Europe? *EIB Papers*, v. 5, n. 2, p. 81–106, 2000.

Appendix A

Table A.1: Dimensions and variables (a list of acronyms is available below)

Variable (variable name)	Description	Unit	Source
1. Citizen security and legal system			
Number of judges per thousand inhabitants (log_judges)	Number of judges/1,000 inhabitants	N	Judicial Branch and INE
Number of crimes per thousand inhabitants (log_oneovercrimes)	Number of crimes/1,000 inhabitants	N	Interior Ministry and INE
Efficiency indicator of the police (efficiency_justice)	New cases/crimes	N	Interior Ministry and Judicial Branch
2. Inclusive, prepared and healthy society			
Life expectancy (log_expectancy)	Life expectancy at birth, both sexes	Years	INE
Percentage of households in poverty (log_poverty)	Number of households with per capita income below the poverty line/total pop.	Percentage	MIDES
Health sector employees (log_emp_health)	Health sector employees/total population	Percentage	ECH-INE
Access to drinking water at home (log_drinking_water)	Households with access to drinking water/total population	Percentage	MIDES
Rate of female activity (log_fem_act)	Female labor force participation rate	Percentage	MIDES
Proportion of EAP who have secondary education or higher (log_eap_sec&ter)	Number of individuals who have secondary education or higher/EAP	Percentage	ECH-INE
Students enrolled in secondary education. In percentage of 11 to 18 years old (log_grad_sec)	Number of students enrolled in secondary education/pop. between 11 and 18 years	Percentage	ANEP-ECH
3. Market of efficient and dynamic factors			
Average income of employees who have secondary education or higher (log_income_sec&ter)	Income measured in constant local currency units (LCU) of 2015	LCU in thousands	ECH-INE
Average income (without locative value) per capita per hour worked (log_income_perhour)	Income measured in constant local currency units (LCU) of 2015	LCU in thousands	ECH-INE
Industry productivity (log_prod_industry)	Industry value added/hours worked	N/A	OPP-BCU-ECH-INE
Productivity of commerce, repairs, restaurants and hotels (log_prod_crrrh)	Commerce, repairs, restaurants and hotels value added/hours worked	N/A	OPP-BCU-ECH-INE

Source: Author's compiled database.

Table A.1: Dimensions and variables (a list of acronyms is available below) (*continued*)

Variable (variable name)	Description	Unit	Source
4. Physical and technological infrastructure			
Internet access at home (log_internet)	Number of households with internet access/total number of households	Percentage	MIDES
Ownership of landline at home (log_tel)	Number of households with phone landline/total number of households	Percentage	MIDES
Passenger movement in ports (log_passengers)	Number of passengers	N	INE
Energy distributed per capita (log_energydistpc)	Total electric energy distributed/population	MWh	UTE- OPP- INE

Source: Author's compiled database.

Table A.2: Estimation window 2007-2014. Dependent variable: log(y)

	Coef.	Std. Err.	z	P > z
F1	0.05	0.01	4.35	0.00
F2	0.05	0.02	2.79	0.01
F3	0.15	0.04	3.46	0.00
F4(-1)	0.04	0.02	2.31	0.02
_cons	11.93	0.07	160.03	0.00
Rho	0.2637349			

Note: regressions include time dummies.

Table A.3: Estimation window 2007-2013. Dependent variable: log(y)

	Coef.	Std. Err.	z	P > z
F1	0.05	0.01	3.91	0.00
F2	0.05	0.02	2.68	0.01
F3	0.15	0.05	3.18	0.00
F4(-1)	0.04	0.02	2.10	0.04
_cons	11.92	0.08	150.49	0.00
Rho	0.24635			

Note: regressions include time dummies.

Table A.4: Development Indicator (ID) Summary Statistics

Variable	Obs.	Mean			Std. Dev.		
		2007	2011	2015	2007	2011	2015
ID	19	11.51882	11.95122	12.03652	.2282308	.1864784	.1770549
pctofmID	19	1	1	1	.0198137	.0156033	.0147098

Note: "pctofmID" is the variable ID as a percentage of its mean for each year (it is the variable whose density is plotted in Figure 2).

Table A.5: Moran's I Statistic (2015). Distance: 200 kms.

	Moran's I	Mean	SE	z-stat	p-value
pcgdp	0.357244	-0.05556	0.113975	3.621835	0.000293
d1	0.35658	-0.05556	0.113462	3.632366	0.000281
d2	0.357234	-0.05556	0.113747	3.629019	0.000285
d3	0.388686	-0.05556	0.113935	3.899093	9.66E-05
d4	0.437373	-0.05556	0.114513	4.304571	1.67E-05

Source: Author's own calculations.

Table A.6: Number of Hot and Cold Spots According to Getis-Ord Statistic (2015). Distance 200 kms.

	Cold Spots		Hot Spots		
	Significance Level		Significance Level		
	1%	5%	5%	1%	
dpcgdp	0	2	14	3	0
d1	0	0	19	0	0
d2	1	3	8	1	6
d3	1	0	17	1	0
d4	0	2	10	3	4

Source: Author's own calculations.

Table A.7: Departments Identified as Hot and Cold Spots According to Getis-Ord Statistic (2015). Distance: 200 kms.

Cold Spots		
	Significance Level	
	5%	1%
dpcgdp	N/A	Tacuarembó, Cerro Largo
d1	N/A	N/A
d2	Rivera	Tacuarembó, Cerro Largo, Artigas
d3	Tacuarembó	N/A
d4	N/A	Tacuarembó, Cerro Largo
Hot Spots		
	Significance Level	
	5%	1%
dpcgdp	Colonia, Soriano, Flores	N/A
d1	N/A	N/A
d2	Lavalleja	Flores, Florida, Colonia, Canelones, Montevideo, San José
d3	Flores	N/A
d4	Maldonado, Colonia, Flores	Florida, Canelones, Montevideo, San José

Source: Author's own calculations.

Table A.8: List of Acronyms

ANEP	National Administration of Public Education (Uruguay)
BCU	Central Bank of Uruguay
ECH	Households Continual Survey
IMCO	Mexican Institute for Competitiveness (Mexico)
INE	National Statistics Institute (Uruguay)
MIDES	Social Development Ministry (Uruguay)
OPP	Office of Planning and Budget (Presidency of the Republic – Uruguay)
UTE	National Administration of Electricity Transmission and Power Plants

POBREZA E DESIGUALDADE NO BRASIL: UMA ANÁLISE DA RENDA PARA 2004-2008 E 2016-2019

OTÁVIO JUNIO FARIA NEVES ^{*}
ANA MÁRCIA RODRIGUES DA SILVA [†]
HENRIQUE DANTAS NEDER [‡]

Resumo

O presente artigo analisa a pobreza e desigualdade pela ótica da renda e multidimensional no Brasil em 2004-2008 e 2016-2019. Desse modo, empregaram-se os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) dos anos de 2004 e 2008 e da PNAD Contínua para 2016 e 2019. Utilizamram-se as metodologias de Foster, Greer e Thorbecke (1984) e Alkire e Foster (2011) e o índice de Gini. Os resultados demonstram que a pobreza e desigualdade pelas duas óticas reduziu em 2004-2008. Todavia, em 2016-2019 houve crescimento da pobreza multidimensional. Tais resultados podem servir para a elaboração de políticas para a redução dos índices.

Palavras-chave: pobreza unidimensional; pobreza multidimensional; desigualdade.

Abstract

This paper analyzes poverty and inequality from an income and multidimensional perspective in Brazil between 2004-2008 and 2016-2019. Thus, data from the Pesquisa Nacional por Amostra de Domicílios (PNAD) for the years 2004 and 2008 and the PNAD Contínua for 2016 and 2019 were used. The methodologies of Foster, Greer, and Thorbecke (1984) and Alkire and Foster (2011) and the Gini index were used. The results show that poverty and inequality from both perspectives reduced in 2004-2008. However, in 2016-2019 there was an increase in multidimensional poverty. Such results can be used to design policies to reduce these indexes.

Keywords: one-dimensional poverty; multidimensional poverty; inequality.

JEL classification: I31, I32, D31

DOI: <http://dx.doi.org/10.11606/1980-5330/ea193248>

^{*} Universidade Estadual de Campinas. E-mail: o203580@dac.unicamp.br

[†] Universidade Federal de Alfenas. E-mail: ana.rodrigues@unifal-mg.edu.br

[‡] Universidade Federal de Uberlândia. E-mail: hdneder@gmail.com

1 Introdução

A importância de combater a pobreza tem sido pauta relevante, principalmente, para os países em desenvolvimento. Isso porque a contração da pobreza e a melhora das condições de vida da população são importantes para estimular o desenvolvimento e crescimento econômico. Mesmo com toda essa preocupação, não há um consenso sobre o conceito e mensuração da pobreza. No decorrer da história, foram introduzidas diversas definições sobre o tema, havendo alguns pontos em comum entre os autores que abordam o assunto.

A mensuração da pobreza é usualmente estudada em termos da renda, sendo a insuficiência de renda a maneira utilizada para a identificação das pessoas pobres. Devido às constantes críticas atribuídas a essa abordagem, como a identificação das necessidades humanas com base unicamente em requerimentos fisiológicos e a permanência da pobreza e das condições de vida inadequadas mesmo diante de cenários com crescimento econômico, surgiram outras concepções para a mensuração da pobreza. A análise dessa “nova” conceituação é feita por meio de evidências que a mensuração do bem-estar deve agregar uma multiplicidade de dimensões, que vão muito além da questão monetária, incluindo dimensões como educação, habitação, saneamento, entre outras, chegando-se um conceito multidimensional. A mensuração baseada meramente na insuficiência de renda, não leva em conta as dimensões do bem-estar que podem não ser atendidas por meio da renda. Esse mesmo raciocínio se aplica à desigualdade, que pode ser estudada com base estritamente na renda ou em um conjunto de dimensões.

A política econômica e social no Brasil adotada na primeira década do século XXI contribuiu para a redução da pobreza por insuficiência de renda e da concentração da renda no país. Mas, especialmente a partir de 2015, houve mudanças nos rumos da política econômica, e esses aspectos podem afetar os indicadores sociais. Assim, é importante analisar a pobreza e a desigualdade no contexto brasileiro tanto pela ótica da insuficiência de renda, quanto pela ótica multidimensional, nesse período.

Perante a esse cenário, neste trabalho, objetiva-se analisar a pobreza e a desigualdade pela ótica da renda e multidimensional no Brasil nos períodos de 2004-2008 e 2016-2019. Para alcançar esse objetivo utilizam-se microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD) dos anos de 2004 e 2008, e da PNAD Contínua (PNADC) dos anos de 2016 e 2019, ambas realizadas pelo Instituto Brasileiro de Pesquisa e Geografia (IBGE). A opção por essas bases de dados se justifica pela disponibilidade de informações nos anos analisados nesta pesquisa.

A escolha do período de investigação se deve às peculiaridades e especificidades do desempenho econômico do país nesses recortes temporais. O primeiro período representa um momento de crescimento econômico, enquanto, o segundo momento de análise, corresponde a um período de crise econômica. Para verificar a pobreza por insuficiência de renda e multidimensional utilizam-se respectivamente os índices de [Foster, Greer e Thorbecke \(1984\)](#) e de [Alkire e Foster \(2011\)](#). Além disso, para a mensuração da desigualdade utiliza-se o índice de Gini.

Além desta seção introdutória, este trabalho está estruturado em quatro seções. A primeira parte evidencia a pobreza e a desigualdade, apontando as principais características e abordagens sobre o tema na literatura econômica. Em seguida, é demonstrado os materiais e métodos utilizados para atingir o

objetivo central do trabalho. Na terceira parte, são apresentados os resultados da pesquisa, dando ênfase para as principais considerações sobre a pobreza e desigualdade. Por fim, são expostas as considerações finais, destacando-se as principais conclusões como forma de direcionar o enfrentamento da pobreza e desigualdade no território brasileiro.

2 Pobreza e desigualdade: muito além da renda

No decorrer da história foram debatidas diferentes conceituações de pobreza, sendo que ainda não há uma consonância na literatura sobre a sua definição e mensuração. A representação do tema atravessou importantes transformações, interpretada por distintas escolas de pensamento. As diferentes ideias de abordagem demonstram as suas formas de tratá-la e mensurá-la. Ou seja, além de buscar conceituá-la, também se propõe medidas para dimensioná-la.

As pesquisas científicas sobre a pobreza se intensificaram no final do século XIX e no início do século XX. A definição sobre o assunto foi se moldando no decorrer dos séculos, sendo que no fim do século XIX, a pobreza era interpretada como uma ideia de subsistência e sua mensuração era baseada em termos da renda. Os pesquisadores que iniciaram esses estudos sobre a pobreza fundamentada em apenas uma dimensão – com base na insuficiência de renda ou consumo – foram Charles Booth e Seebohm Rowntree ([SERRA, 2017](#)). Nessa abordagem, para a identificação das pessoas pobres, normalmente, define-se uma linha de pobreza ou nível de corte abaixo do qual as pessoas são classificadas como pobres.

A pobreza pode ser conceituada de forma absoluta¹, que sob o enfoque da renda ou consumo pode ser entendida como o mínimo de rendimento suficiente para que a pessoa possa adquirir calorias mínimas fundamentais para a reprodução fisiológica. A esse valor mínimo, agregam-se despesas com moradia, transporte, dentre outros ([SALAMA; DESTREMAU, 1999](#)). Essa definição está relacionada com as necessidades mínimas necessárias para à sobrevivência do indivíduo ([SILVA, 2009](#)).

Outra forma de se conceituar a pobreza se dá de maneira relativa, que se refere às satisfações das necessidades dos indivíduos segundo o padrão de vida prevalente na sociedade ([ROCHA, 2006](#)). As famílias pobres são aquelas que não podem usufruir dos recursos e necessidades essenciais da sociedade, as quais lhes permitem concretizarem funções enquanto integrantes do corpo social. De uma maneira ou de outra, o esforço das nações em atingir uma sociedade mais justa e, consequentemente, o combate à pobreza, requer, por exemplo, da defesa da cidadania² como direito de todos ([CODES, 2005](#)).

A abordagem da pobreza baseada na renda ou consumo sofreu questionamentos devido à permanência da pobreza e das condições de vida inapropriadas dos indivíduos, mesmo diante de contextos de crescimento da renda. Esses aspectos sucederam-se tendo em vista que certas necessidades básicas

¹ Os indivíduos que estão em situação de extrema pobreza são chamados de indigentes, isto é, são aqueles que estão localizados abaixo da linha de indigência ([SILVA, 2009](#)).

² As ponderações sobre a questão relativa da pobreza concebem uma discussão em torno da identificação da pobreza sobre o âmbito da cidadania, tendo em pauta a negação desses direitos para a sociedade. Desta maneira, instiga um grande desafio para criarem populações mais igualitárias, sendo esse o modelo a ser empregado para averiguar o desempenho do desenvolvimento que se deve atingir. Desse modo, o argumento da cidadania alicerça no cerne do debate sobre a pobreza ([CODES, 2005](#)).

humanas não foram alcançadas mediante à renda. Assim, surgiram as abordagens das necessidades básicas e das capacitações, que são bases teóricas para o entendimento da pobreza multidimensional. A ideia da multidimensionalidade da pobreza diz respeito à agregação de dimensões, além da renda, que afetam o bem-estar.

Na definição das abordagens das necessidades básicas é fundamental gozar de condições mínimas para satisfazer as necessidades de subsistência, moradia, vestuário e equipamentos. Isso demonstra uma expansão da definição de subsistência, uma vez que acentua as aquisições mínimas requeridas, e não apenas condições de sobrevivência e competência de cada família (CODES, 2005).

Na abordagem das capacitações, apregoada pelo economista Amartya Sen, a liberdade é um fator relevante para a vida das pessoas. Desse modo, para eliminar a pobreza é necessário garantir e expandir as liberdades individuais que ampliam as capacitações dos indivíduos. Nesse caso, a pobreza é vista como privação de capacitações³ básicas que impede os indivíduos de atingir funcionamentos⁴ (SEN, 2001).

As abordagens das capacitações e das necessidades humanas básicas podem ser agregadas para constituir uma conceituação mais completa da pobreza multidimensional. Portanto, a pobreza pode ser classificada como um conjunto de privações: materiais, relativas, atribuições políticas e sociais que estão em torno dos índices (CODES, 2005). Além disso, são incorporadas questões voltadas ao trabalho, saúde, ocupação e outras dimensões que colaboram para a identificação das pessoas pobres (CODES, 2005; SERRA, 2017). Nesse sentido, para o enfrentamento da pobreza é necessário considerá-la em um sentido amplo com a inclusão de fatores que viabilizem o acesso à uma vida digna e plena.

A desigualdade é constantemente relacionada à pobreza, seja por insuficiência de renda, seja em sua perspectiva multidimensional. Desse modo, a forma de definir e mensurar a pobreza também interfere na forma de examinar a desigualdade. Assim como a pobreza, a desigualdade é complexa que se revela, sobretudo, por meio de duas dimensões: econômica e social. A primeira dimensão é aquela que diferencia pessoas ou grupos sociais por intermédio de diversos fatores, como: salários, meios de produção, patrimônio, consumo, dentre outros. Por outro lado, a social aponta as condições de acesso a bens e serviços públicos dentro de uma sociedade, que não estão garantidas pela renda (TROVÃO, 2015).

O capitalismo não conseguiu amenizar o problema da pobreza, tornando-se uma preocupação central dos governos. A desigualdade acompanhou esse processo e no decorrer da trajetória do desenvolvimento capitalista se agravou diante das consequências desse sistema. Diante disso, a relevância de combater a pobreza e desigualdade tem sido pauta importante para as nações. A elevação desses índices tem se apresentado como importantes agendas para as políticas públicas do território brasileiro (MONTALI; LESSA, 2016). Logo, a

³Na definição de capacitações, a pobreza não é entendida somente pela restrita condição de vida em que a pessoa se encontra, mas também pela ausência de oportunidades reais para alcançar uma vida plena (SEN, 1997).

⁴O funcionamento é compreendido como as diferentes ações que um ser humano possa fazer ou ser, como viver muitos anos, ter boa alimentação, ter um ótimo convívio com as pessoas da sociedade e muitas outras coisas valiosas que envolvem o cotidiano (SEN, 1997).

redução desses índices, juntamente com a melhora das condições de vida da sociedade, é justificativa para estimular o desenvolvimento socioeconômico.

3 Contexto brasileiro nos anos analisados

No Brasil, a primeira década do século XXI foi um período em que houve a retomada na dinâmica econômica. A distribuição de renda e a redução da pobreza foram aspectos fundamentais no contexto de desenvolvimento do país. Nos anos 2000, a economia foi estimulada por meio da ampliação do mercado interno, impulsionado pelas políticas de renda e sociais propostas pelo Governo Lula. De acordo com [Paula e Pires \(2017\)](#), desde o início dos anos de 1980 a economia do país oscilou em pequenos ciclos de crescimento, tornando-se padrão no decorrer dos anos 2000.

No governo Lula, reuniu-se o crescimento econômico com a geração de empregos e melhora na distribuição de renda, para que assim, fosse possível reduzir a pobreza por insuficiência de renda no país ([OLIVEIRA, 2015](#)). O Estado passou a empregar uma parcela do Produto Interno Bruto (PIB) em políticas sociais com enfoque na pobreza e desigualdade. Segundo [Pochmann \(2011\)](#), o gasto social agregado alcançou cerca de 23% do PIB no governo Lula, sendo que no ano de 1985 era de 13,3%. Dessa maneira, foi possível expandir os níveis de renda e consumo da população brasileira ([LOPES, 2018](#)).

Esses resultados também surtiram efeitos positivos na concentração de renda, rompendo com a tendência histórica do país. Isso porque, a política de valorização do salário-mínimo, associada aos programas de transferência de renda emergiram efeitos positivos para a população de menor renda. Esses fatores colaboraram para a diminuição da concentração de renda ([CARVALHO, 2018](#); [KERSTENETZKY, 2016](#); [SALATA, 2018](#)).

Então, a recuperação da economia foi importante para a distribuição de renda e para a diminuição da pobreza ([CURADO, 2018](#)). Na evolução da taxa de crescimento da economia brasileira é perceptível que o ano de 2004 representa a retomada do crescimento do PIB no Brasil, atingindo conforme as Contas Nacionais Trimestrais do IBGE uma taxa acumulada de 5,8%. No ano de 2008, nota-se que o crescimento econômico ainda estava em constante evolução. Mesmo com uma redução, a taxa de crescimento ainda representava 5,1%.

Após esse período, houve desaceleração do crescimento econômico, pois, após uma breve recuperação econômica em 2013, a economia perdeu ritmo, marcando uma recessão econômica. Além dos choques enfrentados em 2015-2016 (deterioração nos termos de troca, inflação em ascensão, forte desvalorização da moeda, crise hídrica etc.), o aperto nas políticas fiscal e monetária contribuiu para o agravamento da crise, resultando em salários e lucros decrescentes. Nesse contexto, a oferta de crédito sofreu forte retração, o que retardou ainda mais a recuperação da economia ([PRATES; FRITZ; PAULA, 2017](#), p.197).

A distribuição de renda e a redução da pobreza por insuficiência de renda foram pautas importantes para o desenvolvimento do Brasil na primeira década do século XXI. A dinâmica da economia nesse primeiro decênio foi acompanhada pela queda da desigualdade e pela melhora nas condições de vida da população brasileira. Contudo, a partir de 2015 com a crise econômica e política, o curso dos índices de pobreza e desigualdade tendem a ser alterados.

Esse contexto está relacionado com a crise política e com protestos contra o governo de Dilma Rousseff. Fatores que originaram um clima de instabilidade e descontentamento social, que acarretaram o desejo da população por mudanças (CARVALHO, 2018). Tanto é que o período de 2016-2018, já representado pelo governo Temer, apresentou altas taxas de desemprego, correspondendo a uma taxa média de desemprego de 12,3% em 2018, conforme divulgado pelo IBGE.

Por consequências de todos os acontecimentos vivenciados no país, os anos de 2016 e 2019 representam momentos de baixas taxas de crescimento, principalmente em 2016 (-3,3%). Houve uma lenta recuperação no ano de 2019, mas insuficiente para amenizar os problemas da pobreza e desigualdade no país. Nesse momento, um novo governo liderado por Bolsonaro se inseriu no Brasil, com o prosseguimento das políticas liberalizantes e com foco nas reformas, sobretudo a Reforma da Previdência. Em meio a isso, as medidas impostas no território brasileiro seguiam rumo à adoção de políticas com cunho ortodoxo, tendo como objetivo a mínima participação do Estado na economia, e consequentemente, a redução da atenção às políticas sociais.

Por meio da exposição dos cenários dos anos analisados neste artigo, verifica-se que um período de análise neste estudo corresponde a um momento de crescimento econômico (2004-2008) e o outro se caracteriza por uma crise econômica (2016-2019) do país. Esses dois cenários serão importantes para a compreensão do comportamento da pobreza e da desigualdade nesse período, expandindo a análise para a ótica multidimensional.

4 Material e Métodos

Este trabalho é um estudo descritivo-analítico, de natureza quantitativa, com o propósito de analisar a pobreza e a desigualdade pela ótica da renda e multidimensional no Brasil nos períodos de 2004-2008 e 2016-2019. Para o alcance do objetivo traçado foram utilizados os índices Foster, Greer e Thorbecke (1984) para mensuração da pobreza por renda; a metodologia de Alkire e Foster (2011) para calcular a pobreza multidimensional; e o índice de Gini para mensuração da desigualdade.

Para analisar a pobreza e desigualdade por insuficiência de renda e multidimensional no Brasil, recorreu-se às bases de dados da PNAD e da PNADC, realizadas pelo IBGE. A opção por adotar essas duas bases de dados se justifica pela disponibilidade de informações nos períodos analisados. A PNAD era uma pesquisa realizada anualmente, porém, foi suspensa a partir de 2016 e substituída pela PNADC. Desse modo, para organizar os dados do período 2004 e 2008 é empregada a PNAD, e no outro período de análise (2016-2019), a PNADC é a base de dados utilizada. O recorte geográfico utilizado neste trabalho foi a população do território brasileiro nos anos mencionados.

4.1 Pobreza por insuficiência de renda: índice de Foster, Greer e Thorbecke (FGT)

Os índices formulados por Foster, Greer e Thorbecke (1984) podem ser considerados os mais empregados na literatura econômica, sendo eles: o *Poverty Headcount* (Proporção de pobres - P_0), o *Poverty Gap* (Gap de Pobreza - P_1) e o *Squared Poverty Gap* (Gap de Pobreza elevado ao quadrado - P_2). O conjunto de índices FGT são calculados por meio das seguintes expressões:

$$P_0 = \frac{q}{n} \quad (1)$$

$$P_1 = \frac{1}{n} \sum_{i=1}^q \frac{z - y_i}{z} \quad (2)$$

$$P_2 = \frac{1}{n} \sum_{i=1}^q \left(\frac{z - y_i}{z} \right)^2 \quad (3)$$

em que:

q é o número de pobres (pessoas cuja renda *per capita* domiciliar é menor que a linha de pobreza);

n é o tamanho da população;

z é a linha de pobreza;

y_i é a renda *per capita* domiciliar da i -ésima pessoa.

O índice P_0 mensura a proporção de indivíduos pobres. Em outros termos, representa a parcela de pessoas que aufera uma renda *per capita* domiciliar abaixo da linha de pobreza. Esse indicador é fundamental para as pesquisas relacionadas à pobreza, porém, pode-se analisar o índice juntamente com os outros dois indicadores. Conforme [Neder \(2020\)](#), o índice *Poverty Gap* mensura a intensidade da pobreza para o conjunto da população pobre por meio do cálculo do desvio médio entre o valor da linha de pobreza e a renda dos indivíduos pobres; e pode ser compreendido com um indicador do déficit de pobreza. Por fim, ainda em consonância com [Neder \(2020\)](#), o índice P_2 normalmente é interpretado como um indicador de severidade da pobreza e na elaboração do índice emprega-se um maior peso para os indivíduos mais pobres, considerando a desigualdade de renda entre as pessoas pobres.

Quanto à linha de pobreza utilizada, de acordo com [Soares \(2009\)](#), não existe uma consonância sobre uma linha de pobreza oficial no território brasileiro. No presente estudo foram utilizadas as linhas de pobreza propostas pelo Banco Mundial, conhecida como *dollar a day*. A definição é realizada por meio das médias das linhas de pobreza nacionais de 115 países em desenvolvimento, dispondo da Paridade do poder de compra (PPC) mesmo diante de moedas distintas ([TRONCO; RAMOS, 2017](#)). A aplicação da PPC é importante, pois essa metodologia possibilita a comparação da pobreza entre os países tendo como premissa um parâmetro em comum, mediante os distintos poderes de compra entre países ([SOUZA; JANNUZZI, 2014](#)). O valor da linha de pobreza para países como o Brasil é de US\$ 5,50 *per capita* por dia. Esse valor convertido em reais por meio da PPC foi respectivamente de R\$ 229,06; R\$ 245,78; R\$ 376,20; e R\$ 401,28 mensais em 2004, 2008, 2016 e 2019.

4.2 Pobreza multidimensional: Índice de Alkire e Foster

A mensuração da medida de pobreza multidimensional proposta por [Alkire e Foster \(2011\)](#) denominada como M_0 , é dividida em duas etapas essenciais, a saber: método de identificação e método de agregação. A etapa do método de identificação, apresentada na subseção 4.2 se fundamenta em dois níveis de corte: i) linha de pobreza na dimensão, que identifica se o indivíduo está ou não privado dentro da dimensão; ii) quantidade mínima de dimensões em que os indivíduos precisam estar privados para serem considerados multidimensionalmente pobres. A etapa do método de agregação, apresentada na

subseção 4.2 consiste na obtenção das medidas de pobreza multidimensional a partir da identificação dos pobres realizada na etapa anterior.

Método de identificação

Os indicadores considerados neste trabalho estão baseados nas seguintes dimensões⁵: Vulnerabilidade Econômica; Condições de Moradia; Condições Sanitárias; Ausência de Bens; Características Educacionais e Condições Ocupacionais. Na Tabela 1, pode ser observada a caracterização das variáveis em concordância com as dimensões impostas.

Os níveis de corte z caracterizam as condições de privação em cada indicador empregado, conforme pode ser visto na Tabela 1. As dimensões, os indicadores para representá-las e o nível de privações foram estabelecidos de acordo com o critério de que, se não atendidos, impossibilitam que os indivíduos levem uma vida autônoma e plena, assim como é definido os indicadores utilizados pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) no cálculo do Índice de Pobreza Multidimensional (IPM) global. Eles foram selecionados para representarem necessidades que se não forem satisfeitas limitam a capacitação dos indivíduos para atingirem um tipo de vida valorizado por eles. Sendo assim, a quantidade de dimensões leva em consideração as variáveis que, em conjunto, são capazes de identificar os indivíduos pobres multidimensionalmente.

Levando em conta os dados em termos de privações, para qualquer y , dado $g^0 = [g_{ij}^0]$, indica uma matriz de privação de 0-1 relacionada a y , em que g_{ij}^0 é dado por $g_{ij}^0 = 1$ quando $y_{ij} < z_j$, e $y_{ij} > z_j$ quando $g_{ij}^0 = 0$. Dessa forma, g^0 é uma matriz $n \times d$ cuja entrada g_{ij}^0 é equivalente a 1 quando o indivíduo i é privado na dimensão j -ésima, e quando o valor corresponde a 0, o indivíduo não é privado na dimensão j -ésima. Por meio da matriz g^0 , pode-se construir um vetor de coluna c de contagem de privação, em que a i -ésima entrada $c_i = |g_i^0|$ caracteriza a quantidade de privações enfrentadas pelo indivíduo i (ALKIRE; FOSTER, 2011).

Apenas os cortes nas variáveis que identificam privações não são eficientes para a identificação dos pobres. Desse modo, é necessário considerar outro nível de corte para a especificação do método, e esse corte é denominado como k . Para tanto, é necessário considerar um nível de corte intermediário para c_i que se situa nos extremos entre 1 e d , em que d é o número de dimensões. Então, para $k = 1, \dots, d$, p_k obtido pelo método de identificação é determinado

⁵Considerou-se como *Abastecimento de água inadequado* quando a água não era proveniente de rede geral de distribuição nas áreas urbanas ou não proveniente de rede geral de distribuição ou poço ou nascente nas áreas rurais. Como *Material inadequado das paredes* foi considerado outros materiais que não fossem alvenaria e como *Material inadequado do telhado* considerou-se os materiais diferentes de telha e laje de concreto. A situação de privação de *Ausência de rede de esgoto ou fossa séptica* foi observada para áreas urbanas. Nas áreas rurais considerou-se rede de esgoto e também fossa séptica pela particularidade de não haver rede de esgoto na maioria das localidades rurais. Da mesma maneira, por não haver coleta direta de lixo em grande parte das áreas rurais, considerou-se também a possibilidade de o lixo ser queimado ou enterrado na propriedade quando o indivíduo se situava em localidades rurais, na variável *Destino inadequado do lixo*. O objetivo do indicador *Ausência de trabalhador com carteira assinada* se deu em relação aos benefícios que a formalidade no trabalho permite. Dessa forma, foram considerados juntamente trabalhadores no RJFP, Militares e Empregadores. Essas delimitações desses níveis de corte foram empregadas conforme a literatura sobre o tema, como nos trabalhos de Silva (2009) e nos relatórios do PNUD.

Tabela 1: Dimensões selecionadas, níveis de corte e pesos dos indicadores

Dimensão	Indicadores com o nível de corte z	Peso
Vulnerabilidade econômica	Renda inferior à linha de pobreza	1/3
	Número de dependentes superior ao de não dependentes	1/3
	Ausência de ocupação remunerada	1/3
Condições de moradia	Domicílio não próprio	1/5
	Existência de mais de três pessoas por cômodo	1/5
	Ausência de energia elétrica	1/5
	Material inadequado das paredes	1/5
	Material inadequado do telhado	1/5
Condições sanitárias	Ausência de sanitário	1/4
	Ausência de rede esgoto ou fossa séptica	1/4
	Destino inadequado do lixo	1/4
	Abastecimento de água inadequado	1/4
Ausência de bens	Ausência de telefone fixo ou celular	1/5
	Ausência de televisor	1/5
	Ausência de geladeira	1/5
	Ausência de máquina de lavar	1/5
	Ausência de microcomputador	1/5
Características Educacionais	Existência de adulto analfabeto	1/3
	Existência de adulto sem Ensino Fundamental completo	1/3
	Existência de criança entre 6 e 15 anos fora da escola	1/3
Condições Ocupacionais	Ausência de participação da mulher no mercado de trabalho	1/4
	Ocupação inadequada	1/4
	Ausência de trabalhador com carteira assinada	1/4
	Ausência de ocupado com rendimento superior a um salário mínimo	1/4

Fonte: Elaboração própria com base nos dados da PNAD e PNADC dos anos de 2004, 2008, 2016 e 2019.

por $p_k(y_i; z) = 1$ sempre que $c_i \geq k$, e será igual a 0, caso contrário (ALKIRE; FOSTER, 2011). Em outros termos, p_k assume valor 1 quando a quantidade de dimensões privadas é maior ou igual a k , apontando que o indivíduo i é multidimensionalmente pobre. Por outro lado, quando o valor de p_k for equivalente a 0, a quantidade de dimensões privadas é menor que k , indicando que o indivíduo i não é multidimensionalmente pobre.

Não existe um nível de k predeterminado na literatura econômica. Diferentes níveis de k podem ser condizentes com a realidade de cada objeto de estudo, sendo assim, trata-se de uma medida arbitrária. De acordo com Alkire (2011), maiores valores de k indicam maiores possibilidades de privações simultaneamente. Nesta pesquisa, foi estabelecido o nível de $k = 2$, após uma análise comparativa dos valores de k . Julgou-se mais coerente o valor igual a 2, por representar uma medida satisfatória em relação aos outros valores.

Método de agregação

Supondo o estabelecimento de uma função de identificação específica de p_k para a construção da medida multidimensional de pobreza M_0 , inicia-se com o cálculo da proporção de pobres, cujo índice faz parte da classe de índices FGT, já apresentado. Dessa forma, pode-se calcular a incidência da população classificada multidimensionalmente pobre (H), similar ao da pobreza por

insuficiência de renda, delimitado como:

$$H = \frac{q}{n} \quad (4)$$

Em que, q é a quantidade de indivíduos pobres identificados por meio de duplo corte e n é o número total de indivíduos estudados. O índice H pode ser multiplicado por 100 para a investigação em percentual.

O próximo passo constitui-se no estabelecimento da privação média entre os pobres (A). Para melhor percepção de A , é preciso conhecer a fração de indicadores ponderados em que os indivíduos estão privados, caracterizada por $\frac{c_i(k)}{d}$, em que c_i é o vetor de contagem de privações (explicado na subseção anterior), k é o nível de corte entre as dimensões, e d é o número de dimensões consideradas na análise. Então, conforme [Alkire e Foster \(2011\)](#), a parcela média de privação em todos os indivíduos multidimensionalmente pobres é dada por:

$$A = \frac{|c(k)|}{qd} \quad (5)$$

Em que, q é a quantidade de indivíduos pobres identificados por meio de duplo corte e A aponta a intensidade da pobreza multidimensional. Quanto maior o nível de corte k maior será o índice A . Assim como o índice H , esse índice também pode ser multiplicado por 100 para a análise em percentual.

Por sua vez, a medida de pobreza multidimensional (M_0) combina informações sobre a prevalência da pobreza e a extensão média da privação dos indivíduos pobres. De maneira simplória, a medida M_0 é o produto dos dois índices parciais, H e A :

$$M_0 = H \cdot A \quad (4)$$

O índice de pobreza multidimensional se move no mesmo sentido do índice H , isto é, quanto maior o nível de k , maior a pobreza multidimensional entre os indivíduos. Esse índice pode ser verificado de modo que quanto mais próximo de 0 menor a pobreza.

4.3 Desigualdade de renda e multidimensional: Índice de Gini

A exposição do Índice de Gini será apresentada por meio da análise de [Medeiros \(2012\)](#) e [Hoffmann \(1998\)](#). Esse índice está relacionado com a Curva de Lorenz, que é uma análise gráfica que possibilita verificar a distribuição de renda em determinado país, além de viabilizar a ordenação dessas distribuições sobre o aspecto do bem-estar. Conforme [Medeiros \(2012\)](#), de todos os instrumentos gráficos empregados para analisar a desigualdade em uma distribuição, a Curva de Lorenz é a mais conhecida.

Para o desenvolvimento da Curva de Lorenz, considera-se n valores para uma variável x_i , os quais são ordenados da seguinte maneira: $x_1 \leq x_2 \leq \dots \leq x_n$. A proporção acumulada da população até a i -ésima pessoa corresponde a $p_i = \frac{i}{n}$ e a proporção acumulada de x_i é dada por $\Phi_i = \frac{1}{n\mu} \sum_{j=1}^i x_j$. Feito isso, os pares de valores dados por (p_i, Φ_i) equivalem a pontos que, se unidos, estabelecem a Curva de Lorenz.

O índice de Gini pode ser calculado por meio da Curva de Lorenz, haja vista que corresponde ao dobro do valor da área entre a curva e a linha da perfeita igualdade. Dessa forma, é uma medida de afastamento de certa distribuição de renda em associação a um cenário de perfeita igualdade. É necessário o dobro do valor, para se obter um índice que varia entre 0 e 1 e não 0 e 1/2. Então, quando o Gini for equivalente a 0 significa que não existe desigualdade na sociedade e um valor igual a 1, tem-se um cenário de desigualdade máxima (MEDEIROS, 2012). Como esse índice é o dobro da área entre a Curva de Lorenz e a linha da perfeita igualdade, tal medida pode ser representada por:

$$G = 1 - 2 \int_0^1 L_x(y) dy \quad (5)$$

Em que, G é o coeficiente de Gini e $L_x(y)$ a Curva de Lorenz da distribuição y entre as pessoas x .

Para a mensuração da desigualdade de renda no Brasil para os anos de 2004, 2008, 2016 e 2019 foi empregada à renda domiciliar *per capita*. Para a desigualdade multidimensional foi necessário, determinar um indicador do bem-estar multidimensional, sendo realizados alguns procedimentos.

Conforme apresentado na metodologia de Alkire e Foster (2011), foi construído um vetor de contagem de privações (c_i) para verificar a quantidade de privações enfrentadas pelo indivíduo. Dessa maneira, tem-se um vetor de privações. Entretanto, para avaliação da desigualdade multidimensional é necessário um indicador de bem-estar que se comporte de maneira similar à renda. Para transformar essa representação de privações em bem-estar, foi empregado o inverso, isto é, dividiu-se 1 pelo vetor de contagem de privações $\frac{1}{c_i}$. Assim, tem-se um indicador de bem-estar multidimensional, que foi aplicado ao índice de Gini em substituição à tradicional medida de renda para obtenção da desigualdade multidimensional.

5 Análise da pobreza e desigualdade

Com propósito de verificar o contexto da pobreza e desigualdade no Brasil nos anos de 2004, 2008, 2016 e 2019 foram realizadas as estatísticas apresentadas nos gráficos e tabelas abaixo. Inicialmente, na Tabela 2 demonstram-se os resultados dos indicadores sociais analisados para o país como um todo.

Com relação à proporção de pobres por insuficiência de renda (P_0), observa-se que no Brasil houve uma redução do índice, sobretudo, no período de 2004-2008. Nesse período ocorreu uma melhora nos indicadores de pobreza, muito em decorrência das políticas econômicas e sociais introduzidas pelo Governo Lula. Ao analisar o período de 2016-2019 nota-se que também houve uma redução da proporção de pobres. Porém, essa diminuição foi em menor magnitude se comparada aos anos de 2004 e 2008. Vale ressaltar que nos anos recentes da pesquisa o país se encontrava em um momento de baixo dinamismo da economia e uma crise política que culminaram em novas ideias⁶ sobre políticas econômicas e sociais.

A intensidade da pobreza por insuficiência de renda (P_1) reduziu entre os anos de 2004 e 2008. Em média, os pobres no ano de 2004 tinham uma renda

⁶Foi introduzido na economia um pacote de medidas liberais como a Reforma Trabalhista, Reforma da Previdência, congelamento de gastos, dentre outras.

Tabela 2: Indicadores Sociais para o território brasileiro nos anos de 2004, 2008, 2016 e 2019

Indicadores	2004	2008	2016	2019
P_0	53,18	35,99	23,98	23,05
P_1	26,46	15,68	10,69	10,61
P_2	16,63	9,24	6,73	7,00
H	49,13	38,34	30,33	34,80
A	32,52	31,38	30,41	30,97
M_0	0,159	0,120	0,092	0,108
Gini da renda	0,568	0,542	0,539	0,554
Gini Multidimensional	0,249	0,258	0,278	0,273

Fonte: Elaboração própria com base nos dados da PNAD e PNADC dos anos de 2004, 2008, 2016 e 2019.

26,46% inferior à renda correspondente à linha de pobreza (R\$ 229,06). Por outro lado, em 2008 a renda foi 15,68% inferior à linha de pobreza (R\$ 245,78). Esse desempenho também pode ser observado nos anos de 2016 e 2018, pois também houve a diminuição do indicador.

Em oposição, no que tange à severidade da pobreza (P_2) percebe-se que houve uma redução apenas no primeiro período de análise. Esse indicador coloca ênfase nas pessoas que estão muito abaixo da linha de pobreza, isto é, os mais pobres dos pobres (NEDER, 2020). Portanto, aponta-se para uma piora da situação dessas pessoas entre 2016 e 2019.

Em referência à proporção de pobres analisada pela ótica multidimensional (H), com base na metodologia de Alkire e Foster (2011), verifica-se que houve uma redução do índice de 2004 para 2008. Por outro lado, de 2016 para 2019 ocorreu um aumento do índice H . Ademais, os valores que mensuram a proporção de indivíduos multidimensionalmente pobres no país são mais elevados que na análise por meio da renda, exceto para 2004. No período recente, mais especificamente no ano de 2019, no Brasil, havia 34,80% de indivíduos privados, considerando o nível de corte $k = 2$. Dessa maneira, se questiona a análise da pobreza apenas por renda, uma vez que no período de 2016 e 2019 a proporção de pobres pela análise multidimensional foi maior em relação à ótica da renda. Sendo assim, torna-se necessário pautar a atenção para outras dimensões importantes no cotidiano dos indivíduos brasileiros, como: moradia, saneamento básico, escolaridade, entre outros.

No que se refere à intensidade da pobreza multidimensional (A), o índice diminuiu 1,14 p.p de 2004 para 2008, passando de 32,52% para 31,38%. Diferentemente desse período, de 2016 para 2018 esse indicador aumentou 0,56 p.p, uma vez que o índice passou de 30,41% para 30,97%.

Quanto ao índice de pobreza multidimensional (M_0), ele pode ser analisado em uma escala que varia entre 0 e 1, de maneira que, quanto mais próximo de 0, menor a pobreza multidimensional e quanto mais próximo de 1, maior a condição de pobreza. Essa medida pode ser considerada a mais importante da pobreza multidimensional, uma vez que agrega as duas medidas antecedentes.

Por meio da Tabela 2, percebe-se que o índice M_0 do ano de 2004 relatou uma pobreza multidimensional de 0,1597 no Brasil para o nível de corte $k = 2$. Em 2008 houve uma queda para 0,1203, manifestando uma redução em ter-

mos absolutos de 0,0394 no índice. Com referência aos anos correspondentes a PNADC, percebe-se que em 2016 o índice M_0 registrou uma pobreza multidimensional de 0,0922 no país. Em contrapartida, em 2019 o índice aumentou para 0,1080, evidenciando uma elevação absoluta de 0,0158.

Quanto à desigualdade constata-se que houve uma redução de 4,57% da desigualdade de renda no Brasil do ano de 2004 para o ano de 2008. Porém, o decréscimo em 2016 foi ainda maior no país, representando um índice no valor de 0,5389. Apesar disso, como já mencionado, a crise que o território brasileiro se deparou após 2015 afetou os indicadores sociais. Por consequência desse contexto, a desigualdade de renda de 2016 para 2019 aumentou em 2,78%, simbolizando um índice de 0,5544 em 2019.

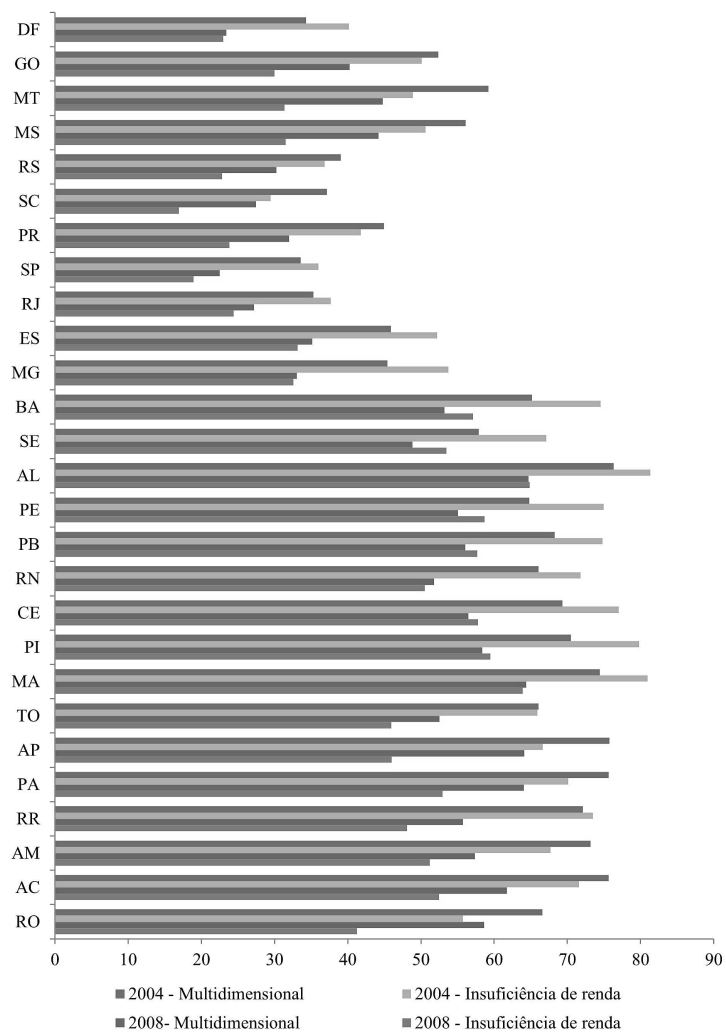
A trajetória da desigualdade multidimensional é diferente da análise da renda. Os índices são mais baixos porque a concentração de privação da população brasileira quanto às variáveis aqui estudadas é menos dispersa que em relação à renda. Na Tabela 2, conclui-se que a desigualdade multidimensional aumentou em 3,61% no período de 2004-2008. Por outro lado, na análise de 2016-2019 a desigualdade multidimensional reduziu em 1,79%. Contudo, destaca-se que os índices de desigualdade multidimensional de 2016 (0,278) e 2019 (0,273) são maiores que nos anos de 2004 (0,249) e 2008 (0,258). Conforme [Trovão \(2015\)](#), a desigualdade multidimensional ainda é um aspecto de destaque na sociedade brasileira, o que induz a concluir que o planejamento para modificar esse cenário necessitará do aumento dos investimentos em infraestrutura social.

O Brasil, por ser um país de dimensões continentais, apresenta estados muito diversos entre si, não apenas cultural e social, mas com situações e condições econômicas que destoam um dos outros. Conforme [Guimarães Neto \(1997\)](#), estruturas produtivas, relações de trabalho, condições de vida e outros fatores se apresentam de forma desigual não apenas entre as macrorregiões do país, como também no interior de cada uma destas. Em um território marcado historicamente por desigualdades regionais que geraram impactos na modernização e no desenvolvimento econômico e social do país, o trabalho também se propõe analisar os estados brasileiros, por meio da proporção de pobres e da desigualdade.

A escolha de analisar apenas as proporções de pobres para as unidades federativas do país se justifica por ser um indicador simples e intuitivo para ser analisado. Ademais, é um índice comum na análise da renda e na multidimensional, o que permite a comparação. Assim sendo, na Figura 1, serão apresentados os resultados descritivos da proporção de pobres por insuficiência de renda (P_0) e multidimensional (H) para todas as unidades federativas brasileiras nos anos de 2004 e 2008.

Destaca-se que houve uma diminuição da proporção de pobres por insuficiência de renda em todas as unidades federativas do Brasil de 2004 para 2008. Observa-se também que a porcentagem mais elevada da população pobre por meio da renda estava localizada nos estados da região Nordeste, com destaque para Alagoas (2004 – 81,33%; 2008 – 64,85%), Maranhão (2004 – 80,97%; 2008 – 63,91%), Piauí (2004 – 79,82%; 2008 – 59,5%), Ceará (2004 – 77,00%; 2008 – 57,81%) e Pernambuco (2004 – 74,98%; 2008 – 58,71%). Com referência aos estados da região Norte, ressalta-se a importante queda desse índice no ano de 2008. As unidades federativas do Sudeste e Sul apresentaram os menores números de proporções de pobres no período observado, com destaque para Santa Catarina (2004 – 29,44%; 2008 – 16,92%), São Paulo (2004 –

Figura 1: Proporção de pobres por insuficiência de renda (P_0) e multidimensional (H) por Unidade de Federação – Brasil (2004 e 2008) (em %)



Fonte: Elaboração própria com base nos dados da PNAD dos anos de 2004 e 2008.

36,01%; 2008 – 18,95%), Rio Grande do Sul (2004 – 36,81%; 2008 – 22,83%) e Rio de Janeiro (2004 – 37,70%; 2008 – 24,41%). Trajetória semelhante pode ser verificada nos estados do Centro-Oeste, uma vez que evidenciaram um baixo índice em 2008, com ênfase para o Distrito Federal (2004 – 40,16%; 2008 – 23,00%).

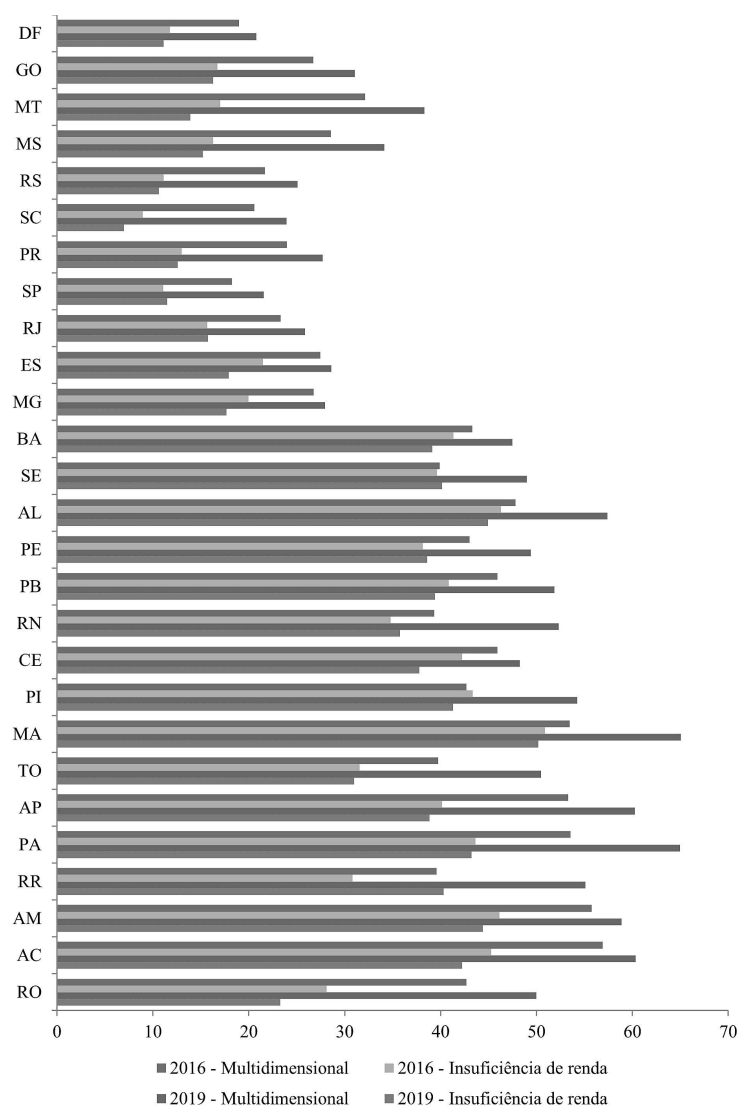
Com relação à proporção de pobres pela análise multidimensional, verifica-se na Figura 1 que os resultados demonstraram uma relevante redução do indicador em todos os estados do país. Comparando-se as unidades federativas, conclui-se que os estados detentores dos maiores índices nos respectivos anos eram Alagoas (2004 – 76,32%; 2008 – 64,71%), Amapá (2004 – 75,73%; 2008 – 64,13%), Pará (2004 – 75,66%; 2008 – 64,06%), Acre (2004 – 75,65%; 2008 – 61,75%) e Maranhão (2004 – 74,45%; 2008 – 64,39%). Em contrapartida, Distrito Federal (2004 – 34,32%; 2008 – 23,42%), São Paulo (2004 – 33,55%; 2008 – 22,53%), Rio de Janeiro (2004 – 35,29%; 2008 – 27,21%), Santa Catarina (2004 – 37,17%; 2008 – 27,45%) e Rio Grande do Sul (2004 – 39,04%; 2008 – 30,25%) foram os estados com as menores proporções de indivíduos multidimensionalmente pobres. Os estados da região Norte e Nordeste possuíam uma quantidade elevada de indivíduos privados, apesar da redução do índice.

Compilando os números das proporções de pobres unidimensional e multidimensional, constata-se que estados como de Alagoas e Maranhão nas duas análises apresentaram os maiores resultados desse índice. Ademais, na maioria dos estados a quantidade de indivíduos multidimensionalmente pobres foi maior que pela insuficiência de renda, concluindo-se que os indivíduos estão privados de condições de moradia, educação, saúde, dentre outros. Esses resultados demonstram a importância de estudar a pobreza por meio de uma agregação de dimensões que fazem parte do cotidiano das pessoas (ASSELIN, 2002; CODES, 2005; KAGEYAMA; HOFFMANN, 2006). Dessa maneira, considera-se importante observar essas localidades com maior atenção, partindo-se do propósito que se deve formular políticas públicas que reduzam esses problemas, como exposto por Sen (2000). Ou seja, para o autor é importante expandir as liberdades das pessoas para a realização das capacitações, isto é, para que esse indivíduo possa ter acesso à saúde, prevenção de doenças, boa educação e ótimas condições de moradia. As liberdades não devem deixar de ser o assunto principal das discussões sobre políticas públicas (SEN, 2000).

A redução da pobreza se deve a diversos aspectos. Contudo, é notório que a trajetória da economia brasileira nesse período e a expansão do mercado formal de trabalho apresentaram uma função importante para a diminuição dos índices de pobreza. Além disso, é importante mencionar a relevância das políticas de transferência de renda e da política de valorização do salário-mínimo nos resultados dos indicadores sociais como a pobreza (KERSTENETZKY, 2016). Na Figura 2, apresentam-se as proporções de pobres para os anos de 2016 e 2019.

Diferentemente do que se demonstrou na comparação entre 2004 e 2008, na análise de 2016 e 2019, não houve uma redução das proporções de pobres por insuficiência de renda em todas as unidades federativas do país. Estados como Roraima (2016 – 30,81%; 2019 – 40,33%), Rio Grande do Norte (2016 – 34,77%; 2019 – 35,77%), Pernambuco (2016 – 38,16%; 2019 – 38,58%), Rio de Janeiro (2016 – 15,64%; 2019 – 15,74%) e São Paulo (2016 – 11,06%; 2019 – 11,48%) acentuaram as suas populações pobres pela ótica da renda. Além disso, os estados com as maiores proporções de pobres foram Maranhão (2016

Figura 2: Proporção de pobres por insuficiência de renda (P_0) e multidimensional (H) por Unidade de Federação – Brasil (2016 e 2019) (em %)



Fonte: Elaboração própria com base nos dados da PNADC dos anos de 2016 e 2019.

– 50,87%; 2019 – 50,17%), Alagoas (2016 – 46,31%; 2019 – 44,95%), Amazonas (2016 – 46,15%; 2019 – 44,41%), Acre (2016 – 45,29%; 2019 – 42,23%) e Pará (2016 – 43,65%; 2019 – 43,22%).

Nota-se que Alagoas e Maranhão, que possuíam acentuados índices de proporção de pobres por renda em 2004 e 2008, no período mais recente dessa análise, ainda continuam apresentando resultados que se destacam. Contudo, ressalta-se a melhora no estado alagoano e uma preocupação com a proporção de pobres no Maranhão. Unidades federativas como Santa Catarina (2016 – 8,95%; 2019 – 6,98%), Rio Grande do Sul (2016 – 11,13%; 2019 – 10,64%) e Distrito Federal (2016 – 11,78%; 2019 – 11,10%) apresentaram as menores proporções de pobres, quando comparados aos outros estados brasileiros.

A proporção de indivíduos multidimensionalmente pobres no período observado cresceu em todos os estados brasileiros, ao contrário da análise da renda (Figura 2). Entre as unidades federativas salienta-se a preocupação com os elevados índices nos estados do Maranhão, Pará, Acre e Amapá, sendo respectivamente 65,07%, 64,98%, 60,33% e 60,27% no ano de 2019. Em contrapartida, os menores índices estavam localizados nos estados de São Paulo (2016 – 18,26%; 2019 – 21,55%), Distrito Federal (2016 – 18,97%; 2019 – 21,55%), Santa Catarina (2016 – 20,58%; 2019 – 23,93%) e Rio Grande do Sul (2016 – 21,68%; 2019 – 25,1%).

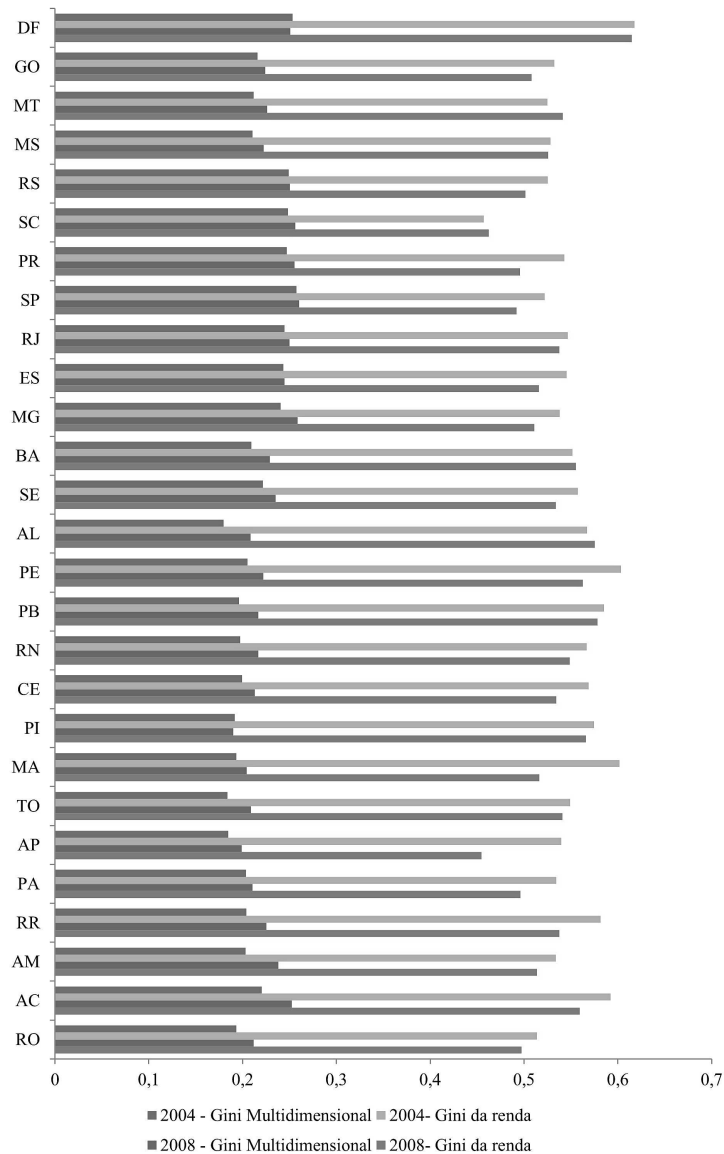
Observa-se que, na análise da proporção de pobres por renda e multidimensional, os estados nordestinos são os que mais se destacam pelos seus acentuados índices. As características históricas do Nordeste indicam um atraso em relação às outras regiões brasileiras (LOIOLA, 2003). Ao longo do tempo essa região tem passado por grandes transformações estruturais, indicando uma atenção e relevância nas pautas de políticas econômicas do Governo Federal. Apesar de todas essas mudanças na localidade, essa região ainda enfrenta elevados níveis de pobreza e desigualdade.

A expansão do mercado consumidor brasileiro teve papel importante na melhoria da distribuição de rendimentos da sociedade na primeira década do século XXI. O motor de elevação dos gastos do consumo perpassa pelo crescimento do consumo das classes mais inferiores. Esse resultado foi fundamental após o período da crise de 2008, já que a estabilidade dos gastos do consumo dessa classe impediu uma queda mais intensa da demanda agregada (CURADO, 2018). A expansão do consumo, juntamente com o progresso da distribuição da renda, foi preponderante para a trajetória da atividade econômica brasileira no período. Desde modo, na Figura 3 será verificada a trajetória da desigualdade de renda e multidimensional nos estados brasileiros nos anos de 2004 e 2008, demonstrando a sua evolução nesse período da economia brasileira.

Por meio do exame da Figura 3 percebe-se que maioria dos estados reduziram sua desigualdade de renda na comparação entre os anos de 2004 e 2008, exceto Alagoas (2004 – 0,5673; 2008 – 0,5753), Bahia (2004 – 0,5518; 2008 – 0,5553), Santa Catarina (2004 – 0,4573; 2008 – 0,4623) e Mato Grosso (2004 – 0,5247; 2008 – 0,5414). Verificou-se (Gráfico 1) que o estado de Santa Catarina possuía os menores índices de proporção de pobres entre as unidades federativas. No entanto, ao observar os dados da desigualdade de renda houve um aumento da concentração dos rendimentos no estado catarinense.

Os estados com maior desigualdade de renda no período analisado eram Paraíba (2004 – 0,5853; 2008 – 0,5784), Alagoas, Acre (2004 – 0,5922; 2008 – 0,5596), Mato Grosso e Distrito Federal (2004 – 0,6178; 2008 – 0,6147). A

Figura 3: Gini da renda e multidimensional por Unidade da Federação - Brasil (2004 e 2008)



Fonte: Elaboração própria com base nos dados da PNAD dos anos de 2004 e 2008.

capital brasileira apresentou o maior índice de desigualdade em comparação com os outros estados, expressando em 2008 um índice equivalente a 0,6147. Em suma, é fundamental destacar que, nesse período, as desigualdades de renda nas unidades federativas reduziram, muito devido às políticas econômicas e sociais impulsionadas pelo governo da época. Esses resultados interromperam uma trajetória de crescente desigualdade de renda no Brasil, que perdurou ao longo de sua história. Quanto à desigualdade multidimensional verificou-se uma elevação em quase todas as unidades federativas, com exceção para os estados do Piauí (2004 – 0,1915; 2008 – 0,1901) e Distrito Federal (2004 – 0,2532; 2008 – 0,251). Mesmo com a alta pobreza multidimensional no Piauí, essa localidade obteve uma redução na desigualdade multidimensional.

Nos mandatos do Governo Lula notou-se uma estabilidade econômica em consonância com as contínuas crises do capitalismo, uma consolidação e expansão de programas de transferências de renda em prol da população mais pobre, com destaque para o Programa Bolsa Família, Programa Minha Casa, Minha Vida e o acesso às universidades públicas. As consequências desses fatores foram totalmente positivas, sendo significativo para a queda da pobreza e desigualdade no período observado (GUIMARÃES, 2018).

Na Figura 4, apresenta-se o panorama recente desses indicadores sociais nos estados brasileiros para os anos de 2016 e 2019. Na primeira década do século XXI, o país apresentou uma relevante melhora na concentração de renda. Entretanto, os resultados se reverteram, voltando a impactar os brasileiros, como verificado na Figura 3.

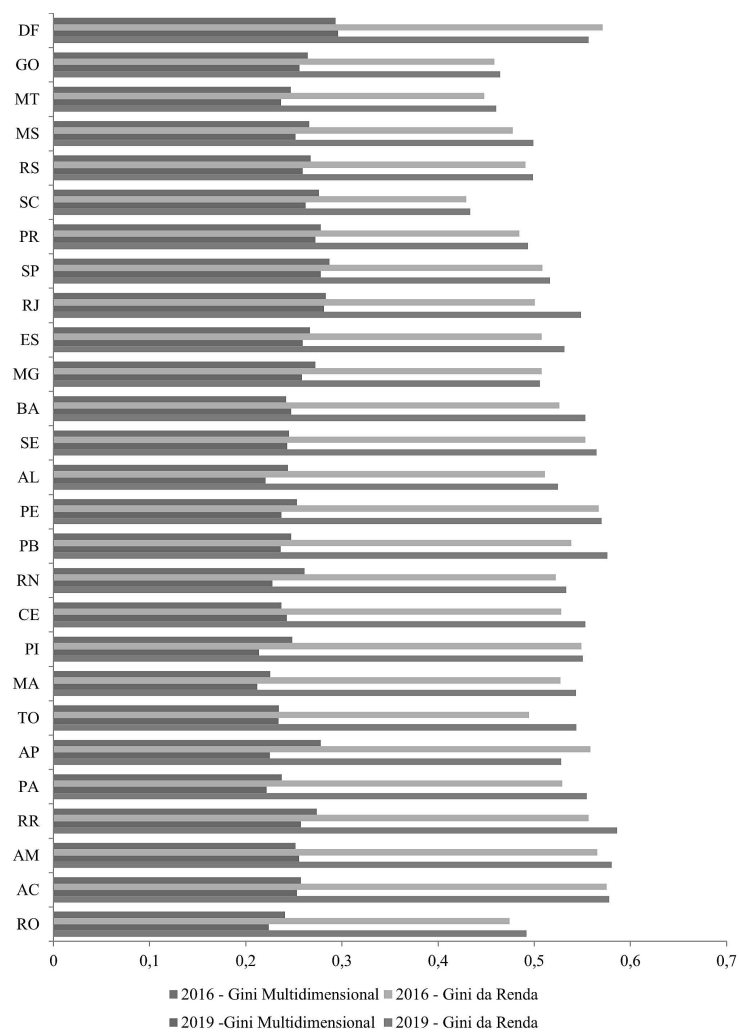
No período 2016-2019, o nível da desigualdade de renda voltou a se elevar nos estados brasileiros. Verifica-se que em todas as unidades federativas houve um aumento, exceto em Minas Gerais (2016 – 0,5077; 2019 – 0,5057) e no Distrito Federal (2016 – 0,5711; 2019 – 0,5567). Ademais, percebe-se que os índices de desigualdade de renda das unidades federativas do país estão semelhantes, evidenciando ainda mais a preocupação em torno da concentração de renda do Brasil.

Os resultados dos anos de 2016 e 2019 evidenciam a grande riqueza acumulada nas mãos de poucas pessoas, colocando-se em questionamento a forma de distribuição de renda no território brasileiro. Mesmo com todas as transformações positivas dos índices de pobreza e desigualdade na primeira década do século XXI, o Brasil, após 2015 passou a se deparar com a trajetória ascendente desses indicadores, conforme apontado por Neri (2019), CEPAL (2021).

A desigualdade multidimensional reduziu nas unidades federativas do país, exceto no Amapá (2016 – 0,2518; 2019 – 0,2555), Ceará (2016 – 0,2372; 2019 – 0,2425), Bahia (2016 – 0,2421; 2019 – 0,2470) e Distrito Federal (2016 – 0,2934; 2019 – 0,2958). Diferentemente da análise da pobreza, os estados da região Sudeste e Sul apresentam elevados níveis de desigualdade multidimensional. Destaca-se o alto índice do Distrito Federal, representando um valor de 0,2958 em 2019. Semelhante aos resultados de 2004 e 2008, no período mais recente abordado neste estudo, nota-se que a desigualdade de renda no país é mais alta se comparada à desigualdade multidimensional. Isso não necessariamente aponta para uma boa condição da distribuição. Ao contrário, isso acontece porque as privações multidimensionais estão menos dispersas entre a população brasileira.

A análise dos dados da pobreza e desigualdade por renda e multidimensional para os períodos 2004-2008 e 2016-2018 demonstraram que, no pri-

Figura 4: Gini da renda e multidimensional por Unidade da Federação - Brasil (2016 e 2019)



Fonte: Elaboração própria com base nos dados da PNADC dos anos de 2016 e 2019.

meio momento da análise, houve melhora dos índices de pobreza e desigualdade. Por outro lado, no segundo período se evidenciou um crescimento da proporção de indivíduos multidimensionalmente pobres. Em termos da desigualdade multidimensional houve aumento em alguns estados brasileiros. Diante de toda a magnitude da desigualdade e pobreza no caso brasileiro, é necessária e importante a abrangência de estudos e pesquisas sobre o tema. Esses conteúdos podem colaborar para a eficácia e execução de políticas sociais no Brasil. A desigualdade se transfigura como um aspecto do problema da distribuição de recursos e das dimensões de bem-estar social excessivamente injusta no território brasileiro.

É importante ressaltar que a proporção de indivíduos multidimensionalmente pobres foi superior em relação à análise da renda nos dois períodos, retratando que as pessoas estão privadas em dimensões que fazem parte das suas necessidades básicas como condições de moradia, sanitárias, educação, ocupacionais, dentre outras. Conhecer esses resultados torna-se importante para elaboração de políticas públicas.

6 Considerações Finais

Na análise da pobreza e da desigualdade é importante que se extrapole a dimensão da insuficiência de renda, para que se possa evidenciar o tamanho desses problemas quando se considera um conjunto de dimensões além da renda. No Brasil, os anos de 2004 e 2008 correspondem a um período de crescimento econômico, enquanto os anos de 2016 e 2019 correspondem a um período de recessão. O contexto econômico desses anos torna relevante as análises sobre o tema nesses períodos.

Sendo assim, neste trabalho, objetivou-se analisar a pobreza e a desigualdade pela ótica da renda e multidimensional no Brasil nos períodos de 2004-2008 e 2016-2019. Nesse intuito, foram utilizados dados da PNAD e da PNADC para o cálculo de índices de pobreza e desigualdade baseando-se nos índices FGT, na metodologia de [Alkire e Foster \(2011\)](#) e no índice de Gini.

Os estudos dos indicadores de pobreza e desigualdade pela ótica da renda e multidimensional para os anos analisados no território brasileiro evidenciaram que no período de 2004-2008, ocorreu uma melhora na pobreza e desigualdade. Em contrapartida, no período de 2016-2018 verificou-se uma elevação da proporção da população multidimensionalmente pobre e aumento da desigualdade multidimensional em alguns dos estados do Brasil.

Mediante a esse cenário de desigualdade e pobreza no Brasil, é preciso e relevante a expansão de trabalhos que visem analisar esses indicadores. As conclusões das diferentes pesquisas sobre o tema podem auxiliar na tomada de decisão e elaboração de políticas públicas que visem amenizar esses índices. A desigualdade enraizada no território brasileiro se evidencia como um problema da injusta distribuição de renda e das dimensões de bem-estar social. Além disso, é relevante mencionar que a população multidimensionalmente pobre foi maior em comparação com a renda nos dois momentos de análise, evidenciando que os indivíduos estão privados em dimensões que fazem parte das suas necessidades básicas. Entender e refletir esses resultados julga-se relevante, pois, tem a finalidade de colaborar para a formulação de medidas que tencionem a redução da pobreza e desigualdade.

7 Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), código de financiamento 001.

Referências

- ALKIRE, Sabina. *Multidimensional poverty and its discontents*. Oxford: University of Oxford, 2011. (OPHI Working Paper No. 46).
- ALKIRE, Sabina; FOSTER, James. Counting and multidimensional poverty measurement. *Journal of Public Economics*, v. 95, n. 7-8, p. 476–487, 2011.
- ASSELIN, Louis-Marie. *Composite indicator of multidimensional poverty*. Québec: Institut de Mathématique Gauss, 2002.
- CARVALHO, Augusto Xavier de. A reforma trabalhista e o mito da geração de empregos. *Carta Social e do Trabalho*, v. 38, p. 16–35, 2018.
- CODES, Ana Luiza Machado de. *Modelagem de equações estruturais: uma contribuição metodológica para o estudo da pobreza*. Tese (Doutorado em Economia) – Universidade Federal da Bahia, Salvador, 2005.
- COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA E O CARIBE — CEPAL. *Panorama Social de América Latina 2020*. Santiago: United Nations, 2021. Disponível em: <https://www.cepal.org/es/publicaciones/46687-panorama-social-america-latina-2020>. Acesso em: 13 ago. 2021.
- CURADO, Marcelo. Do esgotamento do projeto neoliberal ao governo Lula: em busca de um projeto nacional de desenvolvimento. *Cadernos do Desenvolvimento*, v. 8, n. 12, p. 67–81, 2018.
- FOSTER, James; GREER, Joel; THORBECKE, Erik. A class of decomposable poverty measures. *Econometrica*, v. 52, n. 3, p. 761–766, 1984.
- GUIMARÃES, Simone de Jesus. Desigualdades sociais, questão social e políticas públicas. *Revista de Políticas Públicas*, v. 22, p. 607–624, 2018.
- GUIMARÃES NETO, Leonardo. Desigualdades e políticas regionais no Brasil: caminhos e descaminhos. *Planejamento e Políticas Públicas*, v. 15, p. 41–99, 1997.
- HOFFMANN, Rodolfo. *Distribuição de renda: medidas de desigualdade e pobreza*. São Paulo: EDUSP, 1998.
- KAGEYAMA, Angela; HOFFMANN, Rodolfo. Pobreza no Brasil: uma perspectiva multidimensional. *Economia e Sociedade*, v. 15, n. 1, p. 79–112, 2006.
- KERSTENETZKY, Celia Lessa. Consumo social e crescimento redistributivo: Notas para se pensar um modelo de crescimento para o Brasil. *Revista de Economia Política*, v. 36, n. 1, p. 29–45, 2016.

- LOIOLA, Elizabeth. Desenvolvimento regional e as políticas públicas: o caso do nordeste brasileiro. *Revista Econômica do Nordeste*, v. 34, n. 2, p. 225–241, 2003.
- LOPES, Patrícia Ribeiro. A “gestão da pobreza” nos governos petistas: plano Brasil sem miséria em debate. In: ENCONTRO NACIONAL DE PESQUISADORES EM SERVIÇO SOCIAL, 16., 2018, Vitória. *Anais*. Vitória: Universidade Federal do Espírito Santo, 2018.
- MEDEIROS, Marcelo. *Medidas de desigualdade e pobreza*. Brasília: Editora Universidade de Brasília, 2012.
- MONTALI, Lilia; LESSA, Luiz Henrique. Pobreza e mobilidade de renda nas regiões metropolitanas brasileiras. *Cadernos Metrópole*, v. 18, n. 36, p. 503–533, 2016.
- NEDER, Henrique Dantas. *Análise de indicadores sociais utilizando o Stata e o R*. [S. l.]: Mimeo, 2020.
- NERI, Marcelo. *A escalada da desigualdade: Qual foi o Impacto da crise sobre distribuição de renda e pobreza?* 2019. Disponível em: <https://cps.fgv.br/desigualdade>. Acesso em: 15 ago. 2021.
- OLIVEIRA, Thais Diniz. Determinantes da retomada do crescimento no Governo Lula: interpretação do modelo de crescimento com equidade. *Revista Debate Econômico*, v. 3, n. 2, p. 25–51, 2015.
- PAULA, Luiz Fernando de; PIRES, Manoel. Crise e perspectivas para a economia brasileira. *Estudos Avançados*, v. 31, n. 89, p. 125–144, 2017.
- POCHMANN, Marcio. Políticas sociais e padrão de mudanças no Brasil durante o governo Lula. *SER Social*, v. 13, n. 28, p. 12–40, 2011.
- PRATES, Daniela M.; FRITZ, Barbara; PAULA, Luiz Fernando de. Uma avaliação das políticas desenvolvimentistas nos governos do PT. *Cadernos do Desenvolvimento*, v. 12, n. 21, p. 187–215, 2017.
- ROCHA, Sonia. *Pobreza no Brasil: afinal, de que se trata?* 3. ed. Rio de Janeiro: FGV Editora, 2006.
- SALAMA, Pierre; DESTREMAU, Blandine. *O tamanho da pobreza*. Rio de Janeiro: Editora Garamond, 1999.
- SALATA, Andre. Distribuição de renda no Brasil entre 2002 e 2013: redução das desigualdades entre classes? *Latin American Research Review*, v. 53, n. 1, p. 76–95, 2018.
- SEN, Amartya. *Desenvolvimento como liberdade*. São Paulo: Companhia das Letras, 2000.
- SEN, Amartya. *Desigualdade reexaminada*. Rio de Janeiro: Record, 2001.

- SEN, Amartya. Poverty in the human development perspective: concept and measurement. *United Nations Development Programme (UNDP). Human Development Report*, v. 1, p. 15–23, 1997.
- SERRA, Adriana Stankiewics. *Pobreza Multidimensional no Brasil rural e urbano*. Tese (Doutorado em Economia) – Universidade Estadual de Campinas, Campinas, 2017.
- SILVA, Ana Márcia Rodrigues da. *Um estudo sobre a pobreza multidimensional na Região Nordeste do Brasil*. Dissertação (Mestrado em Economia) – Universidade Federal de Uberlândia, Uberlândia, 2009.
- SOARES, Sergei Suarez Dillon. *Metodologias para estabelecer a linha de pobreza: objetivas, subjetivas, relativas e multidimensionais*. Rio de Janeiro: IPEA, 2009.
- SOUZA, M.; JANNUZZI, P. M. *Questões metodológicas acerca do dimensionamento da extrema pobreza no Brasil nos anos 2000*. Brasília: Ministério do Desenvolvimento Social e Combate à Fome, 2014.
- TRONCO, Giordano Benites; RAMOS, Marília Patta. Linhas de pobreza no Plano Brasil Sem Miséria: análise crítica e proposta de alternativas para a medição da pobreza conforme metodologia de Sonia Rocha. *Revista de Administração Pública*, v. 51, n. 2, p. 294–311, 2017.
- TROVÃO, Cassiano José Bezerra Marques. *Desigualdade Multidimensional: uma abordagem keynesiana para o seu enfrentamento*. Tese (Doutorado em Economia) – Universidade Estadual de Campinas, Campinas, 2015.

EARLY WARNING SYSTEMS VIA MACHINE LEARNING: A STUDY OF CURRENCY CRISES

VITOR OSTRENSKY
LEONARDO MATSUNO DA FROTA *

Resumo

Sistemas de Alerta Antecipado (EWS) para crises cambiais é um tópico essencial em macroeconomia. Recentemente, houve uma renovação dessa literatura com a introdução de métodos de aprendizado de máquina. No entanto, argumentamos que a maioria dos trabalhos publicados tem métricas de precisão excessivamente otimistas causadas pela desconsideração da autocorrelação ou atrasos na publicação de dados. Nossa contribuição é construir um Sistema de Alerta Antecipado baseado em um conjunto de modelos de aprendizado de máquina apropriados para dados de séries temporais. Usando dados de 25 países entre 1995 e 2020, nossas descobertas são mais modestas do que trabalhos recentes, mas destacam a utilidade e as limitações dos sistemas de alerta antecipado na prática.

Palavras-chave: crises cambiais; previsão; machine learning.

Abstract

Early Warning Systems (EWS) for currency crises is an essential topic in macroeconomics. Recently, there has been a renewal of this literature with the introduction of machine learning methods. However, we argue that most published works have overly optimistic accuracy metrics caused by disregarding autocorrelation or data publication lags. Our contribution is to build an Early Warning System based on an ensemble of machine learning models appropriate for time series data. Using data from 25 countries between 1995 to 2020, our findings are more modest than recent works but highlight the usefulness and limitations of Early Warning Systems in practice.

Keywords: currency crises; forecasting; machine learning.

JEL classification: C53, G01

DOI: <http://dx.doi.org/10.11606/1980-5330/ea189768>

* Universidade Federal do Paraná

1 Introduction

Currency crises take a heavy toll on any country's asset markets and the real economy, with heavy losses in output and productivity, especially in emerging markets (NAKATANI, 2018). Therefore, predicting or anticipating such crises is relevant to policymakers and market participants. However, predicting the timing of significant currency depreciation is very challenging. Nevertheless, there is no shortage of attempts. In particular, there is much discussion on Early Warning Systems (EWS): tools that give advanced warning signals about an impending currency crisis (Eichengreen, Rose, and Wyplosz (1995) and Kaminsky, Lizondo, and Reinhart (1998).

To build an EWS to predict a currency crisis, we need to know where to look for signals. The theoretical literature on currency crisis is typically divided into three main phases - or *generations* -, with each one having different explanations for what may cause a crisis. The first-generation begins with Krugman (1979) and Flood and Garber (1984). A crisis arises from the inconsistency of the conducted macroeconomic policy, specifically, the combination of fixed exchange rates with a public deficit financed by a monetary expansion in rising inflation. In the second generation models, a crisis arises when the social cost of maintaining the currency regime is higher than the benefit of keeping it (OBSTFELD, 1996). Last, the third generation is concentrated on the financial aspects of the economy. Variables like debt, capital inflows, private credit, and the behavior of asset markets come into play (KRUGMAN, 2003). Another concern is the contagion between countries. In this perspective, a crisis in one place can trigger crises in similar countries, especially those with geographical proximity and/or a strong commercial relationship (MASSON, 1998). The currency crises of the late 1990s are an example of this possibility.

Cuaresma *et al.* (2008) summarizes the relevant economic variables for each generation of currency crisis models: 1st generation - variables that show the fiscal deficits financed by domestic credit; in particular, high inflation (or high money growth), current account deficits, and real interest rate raises; 2nd generation - variables that try to infer the market sentiment and expectations that can lead to herding behavior or self-fulfilling crisis; finally, third-generation - variables that show liquidity problems or more broad balance sheet vulnerabilities in the financial sector.

Early Warning Systems are designed in a way that allows the monitoring of key indicators that are supposed to show out-of-the-ordinary behavior in the months or years preceding a crisis. The job then is to calculate threshold values for these indicators; when they go above it, a good Early Warning System should point out that there is trouble ahead (KAMINSKY; LIZONDO; REINHART, 1998). These models are constructed so that they could be able to set off an alarm before the occurrence of a currency crisis, giving policymakers, investors, and other stakeholders, sufficient time to implement adequate policies to avoid - or at least attenuate - the adverse effects of the crisis (CANDELON; DUMITRESCU; HURLIN, 2014).

Typically, these models were made with standard statistical or econometric tools, such as logit models (BUSSIERE; FRATZSCHER, 2006; CANDELON; DUMITRESCU; HURLIN, 2014; EMIN; AYTAÇ, 2016; BOONMAN *et al.*, 2019), signal approaches (EICHENGREEN; ROSE; WYPLOSZ, 1995) and Markov Switching (DU; YU; LAI, 2020). Recently, with the rise of machine learning tools and applications in finance and macroeconomics, there has been an opportunity to re-

assess the Early Warning System framework with novel techniques such as random forests and neural networks, which are remarkably efficient in making predictions in several areas¹. In the last few years, some papers have used models from the machine learning framework to predict currency crises. The methodological approach is quite different among them, given that several types of models have been used, such as decision trees (KINKYO, 2020), support vector machine (RAMLI; ISMAIL; WOOL, 2015) and neural networks (SEVIM *et al.*, 2014; ALAMINOS *et al.*, 2019).

However, there is still room for advances in terms of the methodological approach to Early Warning Systems. One of the most recurring problems is the frequency of the data used. The foreign exchange market is quite volatile and sensitive to shocks, which makes forecasts on an annual, half-yearly, or even quarterly basis very ineffective. In this respect, many papers do not consider the data publication lag in the analysis. Since data over a period (excluding financial market data) will only be available many weeks later, or even a couple of months later, the forecast cannot be made considering that the data was available on its reference date. Another issue in the literature is the indiscriminate use of the K-fold validation technique in macroeconomic and financial data, without accounting for the time structure of the series. A procedure that can lead to inaccurate estimates when the data is non-stationary Cerqueira, Torgo, and Mozetič (2020). Also, macroeconomic data usually shows some considerable percentage of missing observations for some countries and variables and there is little clarity on the major part of the literature on how they deal with this problem. Finally, many papers use accuracy as the model's performance metric. The problem is that accuracy is not easy to interpret with unbalanced classes, which is the case for currency crises.

Despite these methodological issues, most authors claim to have excellent forecasting results, so a question arises. Is it possible to predict exchange rate crises with such accuracy, or is this performance caused by methodological flaws? To answer this question, besides looking for alternatives to the mentioned problems, we combine standard econometric tools such as a logit model with novel machine learning algorithms such as random forests and neural networks to produce an Early Warning System for currency crises. Our results indicate that much of the literature may present models with overestimated forecast performance, resulting in Early Warning Systems that would probably not be workable in practice. Nevertheless, we show that it is possible to produce forecasts with practical applicability and predictive power, when the temporal structure of the data is taken into account.

The rest of this paper is organized as follows. Section II explains the methodological structure of the model as well as the data used in the models. Section III gives model output examples and how to interpret them. Section IV concludes and Section V is devoted to the data appendix.

¹ It is worth mentioning that the theoretical development of these techniques is not recent. However, because of an increase in computational capacity in recent years, there has been a great popularization of them.

2 Data

Several different forecasting frequencies have been used in the literature: Annual [Emin and Aytaç \(2016\)](#) and [Alaminos *et al.* \(2019\)](#), semiannual [Kinkyo \(2020\)](#), quarterly [Ramli, Ismail, and Wooi \(2015\)](#), [Chong and Yan \(2018\)](#) and [Boonman *et al.* \(2019\)](#) and monthly [Sevim *et al.* \(2014\)](#) and [Du, Yu, and Lai \(2020\)](#). However, we did not find any published work with data frequency higher than monthly. In a forecasting model, the frequency of the data is directly linked to the objective and purpose of the forecast. A model with annual data, for example, most likely will have an annual update of forecasts when used in practice. This can be especially damaging to the model's forecasting ability in the event of currency crises, as, unlike other macroeconomic variables, substantial changes can occur suddenly. Countries can start a year with stability and, in a few months, grow into a situation of greater risk of strong exchange rate depreciation. An example of this is the exchange rates depreciation that occurred in several countries in April 2020, after the coronavirus outbreak. Therefore, as it is important for the usefulness of the models that they have higher frequencies, we choose to estimate our model with a monthly time frequency. However, this also requires that data be available more frequently for countries that are the focus of the forecast, which is not always true.

We based the choice of countries for the study on two criteria. The first is to be a country more vulnerable to this type of event. Therefore, with this criterion, countries in the Eurozone, the United Kingdom, the United States, Japan, and others without a history of major exchange rate depreciation were excluded. The second criterion was to select countries that had sufficient data for the forecast. In international macroeconomic databases, low-income and small countries typically have a higher proportion of missing data, so we chose to remove them from the analysis. The final sample has 25 countries:

America	Asia	Europe	Africa
• Argentina	• China	• Czechia	• Egypt
• Brazil	• Korea	• Hungary	• Nigeria
• Chile	• India	• Poland	• South Africa
• Colombia	• Myanmar	• Romania	
• Mexico	• Indonesia	• Russia	
• Peru	• Israel		
• Uruguay	• Philippines		
	• Pakistan		
	• Thailand		
	• Turkey		

Our sample corresponds to the period of 01/1995-12/2020, being divided between in-sample (training set), from the start to 12/2015, and out-of-sample (test set), from 01/2016 to 12/2020. As our time frequency is monthly, we

mostly use variables with this frequency. However, variables with lower frequencies, such as GDP and debt data, were included. For this, we interpolate the quarterly and annual data using the Kalman filter method [Gómez and Maravall \(1994\)](#). Also, we use financial variables, such as the VIX and the oil price, that have higher frequencies. In these cases, we use the closing value of the previous month. Most of the variables underwent transformations to make them stationary, such as year-over-year percentage change. The indicators, the transformations, and the source of all variables are presented in Table A.2 in the appendix.

Most papers do not use real-time data, that is, considering the typical lag in making macroeconomic data available. This problem can seriously affect the forecast, since the model may end up being trained by observing data it would not observe in practice [Bańbura and Rünstler \(2011\)](#). However, it is not simple to build a vintage database for several countries with so many variables, since for each of these there will be a different lag for data dissemination, not to mention the issue of data revisions. Therefore, we consider an average difference of 2 months between the month of correspondence of the macroeconomic data and its release. Financial variables do not have a disclosure delay so they were not affected.

3 Method

3.1 Crisis definition

The first step in designing an Early Warning System is the precise definition of a currency crisis ([BOONMAN *et al.*, 2019](#)). We tested two different metrics that are used in the literature. First, the most straightforward option is to define it as a substantial devaluation of the nominal exchange rate in a relatively short period (i.e., 20 percent in a month). The nominal depreciation is a value that can be easily understandable and comparable between different countries or periods. However, it does not capture downward currency pressures that are dealt by the central banks' policies. For that reason, we also included the Exchange Market Pressure index [Kaminsky, Lizondo, and Reinhart \(1998\)](#).

From a macroeconomic point of view, it is more important to know whether there will be a crisis within a certain horizon than in a certain month because this period allows the authorities to prevent the crisis ([CANDELON; DUMITRESCU; HURLIN, 2014](#)). Therefore, our goal is to predict whether the month in question precedes a crisis, using a predetermined time interval, which in this case is 12 months. Thus, our first definition of the dependent variable is constructed as follows:

$$Y_{i,t} = \begin{cases} 1 & \text{if there is any month in next 12 with a depreciation above the defined threshold,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where $Y_{i,t}$ is our dependent variable for country i in period t .

It is also necessary to define the thresholds. As a way of measuring the effect of different crisis specifications, we used three different values: 10%, 15%, and 20%. There is no exact definition of an exchange rate crisis using the percentage devaluation of a currency in one month². Thus, we are not

²[Kaminsky, Lizondo, and Reinhart \(1998\)](#) uses 25% in a year, for instance. We have tested larger

only concerned with capturing heavy crises but also with extraordinary depreciations.

However, as we mentioned above, sometimes countries spend a vast amount of resources in terms of international reserves to defend their currencies in some given range. This way, [Kaminsky, Lizondo, and Reinhart \(1998\)](#) incorporated this variable to create the concept of Exchange Market Pressure (EMP)³ to gauge currency crises. So, for any country the EMP index is:

$$EMP_t = \frac{\Delta e_t}{e_t} - \frac{\sigma_{\Delta e,t}}{\sigma_{e,t}} \quad (2)$$

where $\frac{\Delta e_t}{e_t}$ is the rate of change in the exchange rate in period t in relation to US dollars; $\sigma_{e,t}$ is the standard deviation of the rate of change calculated until period t ; $\sigma_{r,t}$ is the standard deviation of rate of change in reserves calculated until period t ; $\frac{\sigma_{e,t}}{\sigma_{r,t}}$ is the rate of change in reserves.

To obtain a dependent variable, we need a specific threshold. As the value of the EMP varies significantly between countries, the limit must be individual, even though the rule is the same for all. Following [Kinkyō \(2020\)](#), our dummy variable returns a value of 1 when the value of the EMP index exceeds the value of 95 percentile, calculated by country to date:

$$Y_{i,t} = \begin{cases} 1 & \text{if there is any month in next 12 with a EMP value above the 95th percentile,} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Using these definitions, we show the incidence of pre-crisis periods in our sample of countries in [Figure 1](#). We can see that there is a correlation between countries, that is, some periods are characterized by several simultaneous crises. Two simple examples of this are the subprime crisis in 2008 and the beginning of the coronavirus pandemic in 2020. It is also interesting to note that some periods are characterized by crises only when we use the 10% threshold, as in the period from 2010 to 2012. While in the late 1990s, the share of crises with depreciation above 20% was higher.

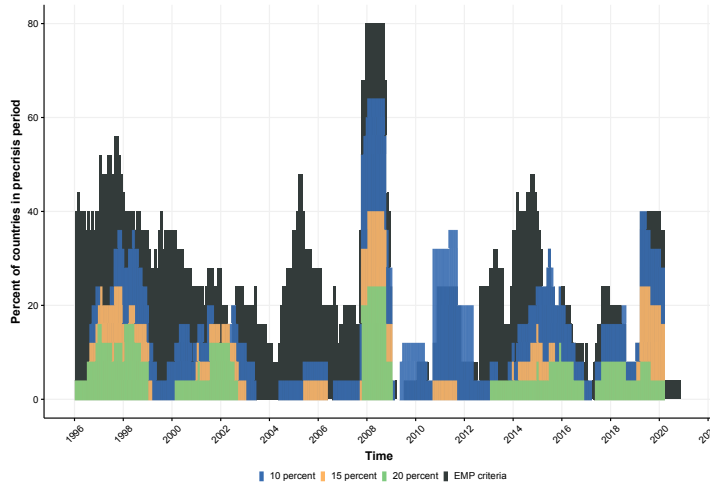
3.2 Models

Our strategy for this forecasting model is to build an ensemble of different models. This type of approach is widely used in machine learning because it is proven to be quite effective in increasing the efficiency of the models and balancing their weaknesses ([ZHANG; MA, 2012](#); [AMPOMAH; QIN; NYAME, 2020](#)). There are different ensemble methods, but we will use one of the simplest, soft voting. In this technique, we estimate the different models individually, which will return, for each month and country, a crisis probability value. Then, we average these multiple outputs to get a single probability value. There is also the possibility to include weights in each model, but here we used fixed weights.

To control the complexity of the model and to obtain a good generalization in out-of-sample settings, the hyperparameters for the models were obtained

thresholds, but the count and intensity of currency crisis have decreased since the nineties, so a 25% month depreciation has become rarer.

³There are different ways to build this index. For example, in [Eichengreen, Rose, and Wyplosz \(1995\)](#), there is the inclusion of the interest rate differential relative to the United States rate. However, we prefer to keep the simpler and more widespread version of the indicator.

Figure 1: Crises over time

through a grid-search algorithm - a standard approach in the early warning literature [Holopainen and Sarlin \(2017\)](#) and [Beutel, List, and Von Schweinitz \(2018\)](#). The optimized hyperparameters for the different models are shown in Table A.1.

We estimated four different models to get our Early Warning System. First, them logistic regression, which is the simplest and the most used in this type of forecast. So we can consider it our baseline estimation. The parameters are not country-specific, because we have unbalanced classes: there are many more “tranquil” periods than crisis periods. Estimating an individual model for each country could generate a loss of relevant information, and could lead to a problem of too many variables (features) for few observations. This approach is followed by the other models, as we estimate all the data together

The Random Forest is the second model ([Ho, 1998](#)). This is an ensemble method that averages several decision trees using a bagging algorithm. Its primary goal is to reduce the issue of over-fitting, which is extremely prevalent when using deep decision trees. The random forest has recently attracted more interest in time series applications because it has shown to be quite successful in solving problems of this nature, such as in [Lohrmann and Luukka \(2019\)](#) and [Kinkyo \(2020\)](#).

Our third model is the XGBoost, which is also an ensemble method with decision trees. The principle is similar to that of the random forest but uses gradient boosting - a procedure that optimizes a loss function to aggregate the trees. This method has gained considerable relevance in recent years, as it is used to win several machine learning competitions. In terms of time series forecasting, it demonstrated excellent forecasting ability compared to other statistical and ML models ([Abolghasemi et al., 2019](#)), in addition to being used in the FFORMA model, it achieved second place in the M4 competition ([Montero-Manso et al., 2020](#)).

Finally, we will use a deep learning model, Long Short Term Memory (LSTM). This is a recurrent artificial neural network. This method makes it possible to process data sequentially. For this reason, it is used for time series problems, identifying relationships with lags between variables. Like the

other two previous models, it was also shown that it outperforms traditional models in time series problems ([SIAMI-NAMINI; NAMIN, 2018](#)).

For the Random Forest, XGBoost, and LSTM models, we use the log loss criterion, also known as the cross-entropy loss, as a loss function to be minimized in the validation step. Most articles in the literature primarily use other metrics, such as AUC or accuracy. In contrast to these, which evaluate only the final classification of the model (in our case, crisis or non-crisis), the log loss penalizes the model based on the estimated probability. For example, an erroneous prediction that there is a 90% probability of a crisis has more penalties than a prediction that is also wrong, but which pointed out only a 60% probability of a crisis. Using the log loss is justified by the intrinsic difficulty of predicting a currency crisis. In this context, an increase in the probability of a crisis from 10% to 40% is relevant and informative regarding the heightened risk of a currency crisis, even though the crisis probability for the model is still less than 50% - below the threshold that would make the model indicate a currency crisis in the next 12 months.

3.3 Cross-Validation and Time Series

The standard approach to assess the generalizability of the results of machine learning applications is to use the K-fold cross-validation (CV) algorithm ([HASTIE; TIBSHIRANI; FRIEDMAN, 2009](#)). However, time series data can present patterns - such as serial correlation and non-stationarity - that violate key hypotheses that are needed to guarantee the desirable statistical properties of the K-fold cross-validation method. Also, there is the basic intuition that we should not use data from the future to predict the past [Bergmeir, Hyndman, and Koo \(2018\)](#).

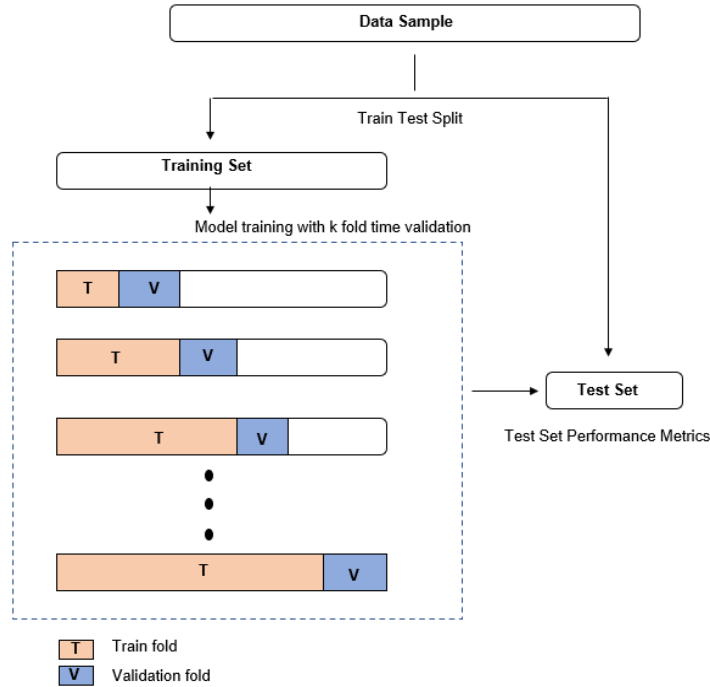
We are dealing with a set of macroeconomic and financial time series that in general are serially correlated and non-stationary. Therefore, we use a slightly different validation technique, called time series validation or forward validation, which respects the temporal structure of the data and has desirable statistical properties in the presence of non-stationary data [Schnaubelt \(2019\)](#).

Instead of a random division, as done by K-fold, the data is split in an orderly fashion by time. Also, the data and validated parameters are only trained using previous data. We split the data using 5 sequential folds. The model metrics are then assessed in the test set. Figure 2 describes the validation and testing strategy.

3.4 Imputing missing data

The issue of missing observations is an important theme which is often overlooked in the Early Warning literature. As far as our knowledge goes, only [Kinkyo \(2020\)](#) details explicitly a method to correct this problem. Of course, most papers used lower time frequencies, which can mitigate the missing observations problem. Using fewer predictive variables or countries can also alleviate the problem, however, it is unlikely that the issue will be completely eliminated, and at limit, it makes the estimation of machine learning models impossible.

The incidence of missing observations in the sample is nearly 10%. This number varies considerably concerning each country and each variable. Some

Figure 2: Time validation method

series start after the first month (January 1995), in others there is no data for some countries. Attributing the missing observations is a challenge, especially in this case. Removing variables or countries with a lot of missing data can compromise the model by removing important information. The opposite case is equally damaging, since imputing a large chunk of data can cause a lot of noise to be included. We, thus, removed many countries from the sample as they did not have enough data. Following, [Kinkyo \(2020\)](#), we use the K—Nearest Neighbors method to impute the missing data. It is a multivariate impute method, which models each feature with missing values as a function of the other ones.

4 Results

4.1 Metrics evaluation

Forecasting a currency crisis is an imbalanced classification problem, that is, the distribution of the dependent variable between 0 (tranquil periods) and 1 (crisis periods) is far from balanced, which would be 50% of the observations for each. Naturally, the more unbalanced the problem is, the easier it is to achieve greater accuracy, even if the model is not necessarily better. For example, if we have a sample with only 1% of pre-crisis periods, a model that does not predict any of them may have 99% of accuracy by predicting only “tranquil periods”. In summary, this is why in this case it is much more important to analyze the value of other metrics, such as the AUC. However, many studies, such as ([SEVIM et al., 2014](#); [ALAMINOS et al., 2019](#)) use accuracy with-

Table 1: Metrics results

		10% drop	15% drop	20% drop	EMP index criteria
Accuracy	Logit	0.843	0.832	0.957	0.829
	Random Forest	0.786	0.932	0.957	0.720
	XGBoost	0.850	0.929	0.964	0.776
	LSTM	0.825	0.671	0.894	0.596
	Ensemble	0.839	0.879	0.952	0.596
AUC	Logit	0.537	0.546	0.500	0.500
	Random Forest	0.684	0.611	0.652	0.476
	XGBoost	0.629	0.627	0.677	0.501
	LSTM	0.497	0.528	0.619	0.499
	Ensemble	0.637	0.636	0.663	0.484
Log loss	Logit	0.387	0.241	0.178	0.484
	Random Forest	0.463	0.374	0.207	0.645
	XGBoost	0.347	0.255	0.140	0.481
	LSTM	1.144	0.495	0.248	2.681
	Ensemble	0.350	0.242	0.149	0.516

out any further discussion, which can make the reported results misleading in relation to the quality of the model.

It is necessary to define a probability value that separates the forecasts between tranquil periods and pre-crisis periods. As we are using log loss to validate the parameters, the default value of 50% may not be the most appropriate depending on the criteria you want to optimize. Similar to [Candelon, Dumitrescu, and Hurlin \(2012\)](#), we use an optimal cut-off value based on maximizing AUC on the validation set.

To evaluate the models, we only return the results of the out-of-sample metrics, as shown in Table 1. The log loss, which was used as a loss function in the validation set, has a rather simple interpretation. Nonetheless, for completeness, we also include the results of the AUC and the accuracy of all models. For each metric, the table shows the values for each of the models individually and the value of the ensemble by soft voting.

The estimated model using the dependent variable based on the EMP index has a much worse performance than models based on the simple nominal depreciation rate. This may be due to monetary policy, given that the central bank can avoid a strong devaluation using monetary reserves. Thus, this variable appears to have more noise than the variable with simple depreciation, which makes it less efficient to be used in the forecast.

An important point of these results is to show the effectiveness of using the model ensemble. This model returned the best AUC for the estimation with a 15% threshold, the second-best for 20% and 10%. It is worth mentioning that the adopted ensemble method was simple, in addition to being affected by the poor individual performance of Logit and LSTM. Some more robust ensemble methods, such as boosting, may bring even better results. Regarding the individual models, the consistent performance of the random forest algorithm is noteworthy, a not-so-complex model that provided results similar to the more complex XGBoost. On the other hand, the LSTM has failed in this regard. One possible explanation is the low number of observations in relation to what is necessary for deep learning models.

At first, the results of the performance metrics may seem discouraging. There are studies with accuracy close to 0.99 ([ALAMINOS *et al.*, 2019](#)), oth-

ers with AUC above 0.95 (RAMLI; ISMAIL; WOOL, 2015), despite the differences in the dependent variable definition. However, we seek to use parameters that would allow our forecast to be performed with this metric in real policy-making and investment settings. If it were possible to predict a currency crisis with such precision, months before it occurred, using public databases and widely used methods, we would be facing a clear violation of the efficient markets hypothesis (MALKIEL; FAMA, 1970). Therefore, our forecast, with an AUC of 0.636 to 0.663, depending on the threshold used, can be considered adequate and realistic.

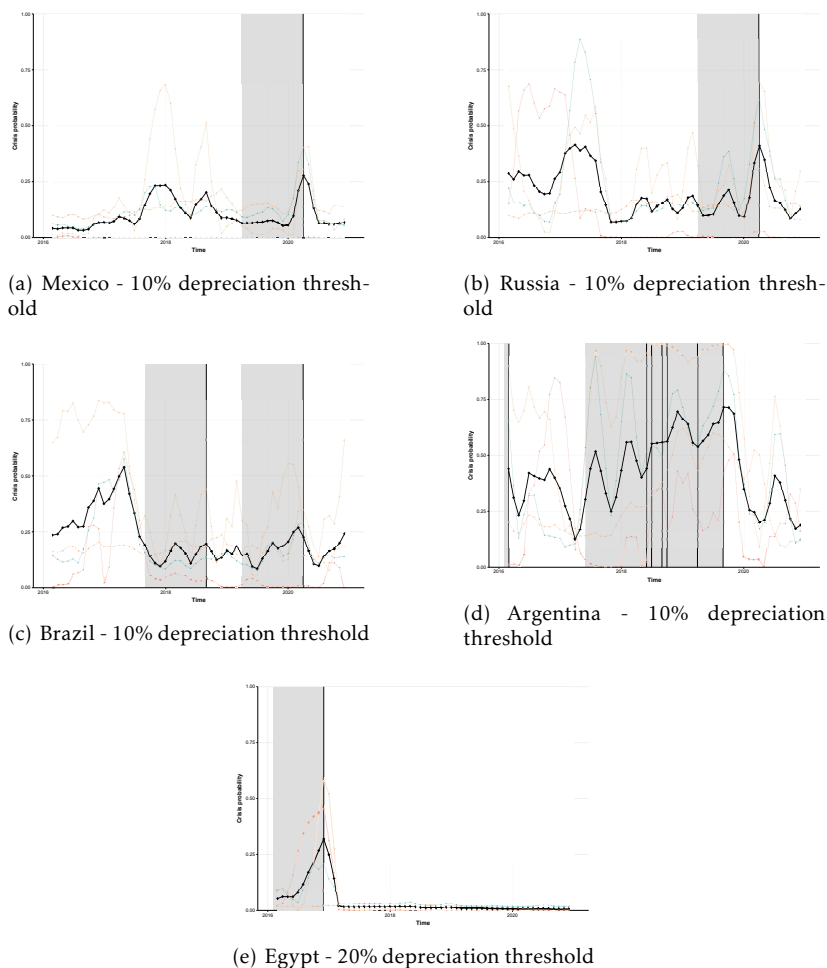
4.2 Some practical examples

The values of the performance metrics are very important, but to evaluate an Early Warning System it is also interesting to understand how it behaves in showing us the probability of a crisis. For this, we selected some countries to monitor the performance of the model in the out-of-sample period. In all the graphs that are shown in this section, the black line is the probability estimated by the ensemble of the four models (logit, random forest, XGBoost and LSTM). The colored lines represent the individual probabilities estimated by these models. Furthermore, the vertical line represents that there was a depreciation above the specific threshold in that month and the shaded area represents the period 12 months before the crisis.

It is interesting to note that in April 2020 we had several major currency devaluations in developing countries caused by the worsening of the coronavirus pandemic. One of the countries most affected was Mexico, which has the probability shown in Figure 3. It is possible to see that the model detects an increase in instability in February and starts to rise. The same is true for Russia, which also suffered a sharp depreciation that month. However, in both cases, there is an increase in probability in previous periods that is not followed by a crisis. For the same period, the model for Brazil has an increase in probability, but one that is much less marked and evident than in the previous two cases. Also, there is a period with a high probability of depreciation, but without an event. The case of Argentina is more emblematic, given that it is a country that has been going through several exchange rate crises in the last 20 years. In this country, the model behaves very clearly and correctly, reflecting the political and economic conditions that the country has been through. In the last individual case, Egypt, unlike the other four, we chose to show the model with a 20% threshold. This case behaves in the ideal way of an Early Warning System, with very low probabilities in tranquil periods, but with a substantial increase in the probability right before a currency crisis.

5 Conclusion

Our goal in this article was to build an Early Warning System for currency crises using machine learning models. Our contribution is to build a model that is applicable in policy making and investment settings and deals with recurrent problems in the literature such as: data with low frequency, that does not consider publication lag, misuse, and/or misreporting of validation techniques in time series data and inappropriate use of model evaluation metrics. Our model showed that although it is not possible to predict exchange rate

Figure 3: Probability of a crisis - Output of ensemble model

crises with perfect accuracy, it is possible to build a model that returns useful probabilities for decision-making.

Still, there are limitations to our work. For example, due to the availability of data, we have chosen to work with a restricted sample of countries, almost all of which are middle-income. A wider sample - with the inclusion of low-income and frontier economies - would enrich the model predictions and robustness. Also due to data quality considerations, we started our dataset in 1995. A longer sample - reaching the 80s and 70s - would improve our understanding of how currency crises change as the global economy changes.

Finally, currency crises are notably very challenging to predict and the blind application of machine learning techniques and metrics to forecast them can provide misleading results and give potentially wrong signals to policy-makers and other stakeholders. We do not consider that the objective of modeling a currency crisis should be to aim for unrealistic accuracy metrics rather it should be to understand the signals and thresholds that typically have preceded one. We argue that the usage of machine learning techniques - which are very powerful for prediction - should be coupled with an understanding

of macroeconomic theory and policymaking objectives. While this approach, as we have shown in this paper, produces more sober results in terms of prediction accuracy it has the potential to provide more value in terms of guiding macro-prudential measures for policymakers or asset allocation strategies for investors.

References

- ABOLGHASEMI, Mahdi *et al.* Machine learning applications in time series hierarchical forecasting. *arXiv preprint arXiv:1912.00370*, 2019.
- ALAMINOS, David *et al.* Currency crises prediction using deep neural decision trees. *Applied Sciences*, v. 9, n. 23, p. 5227, 2019.
- AMPOMAH, Ernest Kwame; QIN, Zhiguang; NYAME, Gabriel. Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement. *Information*, v. 11, n. 6, p. 332, 2020.
- BAÑBURA, Marta; RÜNSTLER, Gerhard. A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, v. 27, n. 2, p. 333–346, 2011.
- BERGMEIR, Christoph; HYNDMAN, Rob J.; KOO, Bonsoo. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, v. 120, p. 70–83, 2018.
- BEUTEL, Johannes; LIST, Sophia; VON SCHWEINITZ, Gregor. An evaluation of early warning models for systemic banking crises: Does machine learning improve predictions? *Deutsche Bundesbank Discussion Paper No. 48/2018*, 2018.
- BOONMAN, Tjeerd M. *et al.* Early warning systems for currency crises with real-time data. *Open Economies Review*, v. 30, n. 4, p. 813–835, 2019.
- BUSSIERE, Matthieu; FRATZSCHER, Marcel. Towards a new early warning system of financial crises. *Journal of International Money and Finance*, v. 25, n. 6, p. 953–973, 2006.
- CANDELON, Bertrand; DUMITRESCU, Elena-Ivona; HURLIN, Christophe. Currency crisis early warning systems: Why they should be dynamic. *International Journal of Forecasting*, v. 30, n. 4, p. 1016–1029, 2014.
- CANDELON, Bertrand; DUMITRESCU, Elena-Ivona; HURLIN, Christophe. How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods. *IMF Economic Review*, v. 60, n. 1, p. 75–113, 2012.
- CERQUEIRA, Vitor; TORGO, Luis; MOZETIČ, Igor. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, v. 109, n. 11, p. 1997–2028, 2020.

- CHONG, Terence T. L.; YAN, Isabel K. Forecasting currency crises with threshold models. *International Economics*, v. 156, p. 156–174, 2018.
- CUARESMA, Jesús Crespo *et al.* Determinants of currency crises: A conflict of generations? *Focus on European Economic Integration*, n. 1, p. 126–141, 2008.
- DU, Jiangze; YU, Runfang; LAI, Kin Keung. Identification and prediction of currency crisis: Markov Switching-Based approach. *The Singapore Economic Review*, v. 65, n. 06, p. 1667–1698, 2020.
- EICHENGREEN, Barry; ROSE, Andrew K.; WYPLOSZ, Charles. Exchange market mayhem: the antecedents and aftermath of speculative attacks. *Economic Policy*, v. 10, n. 21, p. 249–312, 1995.
- EMIN, Dogus; AYTAÇ, Ayşegül. Predicting currency crises: how do indicators differ according to crisis definition? *Çankırı Karatekin University Journal of The Faculty of Economics and Administrative Sciences Y*, v. 6, n. 2, p. 85–102, 2016.
- FLOOD, Robert P.; GARBER, Peter M. Collapsing exchange-rate regimes: Some linear examples. *Journal of International Economics*, v. 17, n. 1-2, p. 1–13, 1984.
- GÓMEZ, Víctor; MARAVALL, Agustín. Estimation, prediction, and interpolation for nonstationary series with the Kalman filter. *Journal of the American Statistical Association*, v. 89, n. 426, p. 611–624, 1994.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The elements of statistical learning*. New York: Springer, 2009.
- HO, Tin Kam. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, v. 20, n. 8, p. 832–844, 1998.
- HOLOPAINEN, Markus; SARLIN, Peter. Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, v. 17, n. 12, p. 1933–1963, 2017.
- KAMINSKY, Graciela; LIZONDO, Saul; REINHART, Carmen M. Leading indicators of currency crises. *IMF Staff Papers*, v. 45, n. 1, p. 1–48, 1998.
- KINKYO, Takuji. A bi-annual forecasting model of currency crises. *Applied Economics Letters*, v. 27, n. 4, p. 255–261, 2020.
- KRUGMAN, Paul. A model of balance-of-payments crises. *Journal of Money, Credit and Banking*, v. 11, n. 3, p. 311–325, 1979.
- KRUGMAN, Paul. Crises: The Next Generation? In: *Economic Policy in the International Economy: Essays in Honor of Assaf Razin*. Organized by: E. Helpman and E. Sadka. Cambridge: Cambridge University Press, 2003. p. 15–32.

- LOHRMANN, Christoph; LUUKKA, Pasi. Classification of intraday S&P500 returns with a Random Forest. *International Journal of Forecasting*, v. 35, n. 1, p. 390–407, 2019.
- MALKIEL, Burton G.; FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, v. 25, n. 2, p. 383–417, 1970.
- MASSON, Paul R. *Contagion: monsoonal effects, spillovers, and jumps between multiple equilibria*. Washington, DC: International Monetary Fund, 1998.
- MONTERO-MANSO, Pablo *et al.* FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, v. 36, n. 1, p. 86–92, 2020.
- NAKATANI, Ryota. *Output costs of currency crises: shocks, policies and cycles*. Munich: Munich Personal RePEc Archive, 2018. (MPRA Paper No. 83549).
- OBSTFELD, Maurice. Models of currency crises with self-fulfilling features. *European Economic Review*, v. 40, n. 3-5, p. 1037–1047, 1996.
- RAMLI, Nor Azuana; ISMAIL, Mohd Tahir; WOOL, Hooy Chee. Measuring the accuracy of currency crisis prediction with combined classifiers in designing early warning system. *Machine Learning*, v. 101, n. 1, p. 85–103, 2015.
- SCHNAUBELT, Matthias. *A comparison of machine learning model validation schemes for non-stationary time series data*. Nürnberg: Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019. (FAU Discussion Papers in Economics No. 11/2019).
- SEVIM, Cuneyt *et al.* Developing an early warning system to predict currency crises. *European Journal of Operational Research*, v. 237, n. 3, p. 1095–1104, 2014.
- SIAMI-NAMINI, Sima; NAMIN, Akbar Siami. Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*, 2018.
- ZHANG, Cha; MA, Yunqian. *Ensemble machine learning: methods and applications*. New York: Springer, 2012.

Appendix A**Table A.1:** Hyperparameter Tuning via Grid-Search Algorithm

		10% drop	15% drop	20% drop
XGBOOST	Max Depth	3	2	3
	Number of Estimators	100	50	50
	Learning Rate	0.01	0.1	0.05
	Minimum Child Weight	10	5	7
Random Forest	Max Depth	20	50	60
	Min Samples Leaf	2	2	2
	Min Samples Split	5	2	5
	Number of Estimators	600	1000	200
LSTM	Hidden Layer	3	3	2
	Neurons per layer	64	32	8
	Dropout rate	0.1	0.01	0.1
	Learning Rate	0.001	0.001	0.001
	Batch Size	5	10	5
	Epoch Size	100	300	300

Table A.2: Metrics results

Data (Source or Distributor)	Transformations applied
Nominal Exchange Rate - Local Currency relative to USD (Fixer.io)	Month-over-month percentage change
Real Effective Exchange Rate - Based on consumer price index (CPI) (World Bank)	Hodrick–Prescott filter
Total Reserves excluding Gold, Foreign Exchange, US Dollars (World Bank)	Ratio in relation to GDP and monthly percentage change
Value of Exports, Free on board (FOB), US Dollars (World Bank)	year-over-year percentage change
Value of Imports, Free on board (FOB), US Dollars (World Bank)	year-over-year percentage change
Claims on private sector (IMF)	year-over-year percentage change
Money Market Rate (IMF)	year-over-year percentage change
Lending Rate (IMF)	year-over-year percentage change
Net Foreign Assets (IMF)	Ratio in relation to total reserves and Monetary Base, then year-over-year percentage change
Broad Money Liabilities (IMF)	Ratio in relation to total reserves and Monetary Base, then year-over-year percentage change
Consumer Prices (World Bank)	year-over-year percentage change
Monetary Base (IMF)	Used in other variable calculations
Balance of payments on goods & services (IMF)	year-over-year percentage change
Balance of Payments (current account) (IMF)	Ratio in relation to GDP and year-over-year percentage change
Commodity net export price index (IMF)	year-over-year percentage change
Coupcast (probability of political coup) (OEF)	No transformation applied
Election Month (Binary) (OEF)	Transformed to indicate a future election in the next 6 months (1 if positive, 0 if negative)
General government gross debt (IMF)	year-over-year percentage change
Stock market index (US Dollars) (World Bank)	year-over-year percentage change
Import Coverage (World Bank)	Level
Gross Domestic Product (World Bank)	year-over-year percentage change
Unemployment (World Bank)	year-over-year percentage change
Industrial Production Index (World Bank)	year-over-year percentage change
General government net lending/borrowing (IMF)	year-over-year percentage change
West Texas Intermediate 40 API Midland Texas; US\$ per barrel (IMF)	year-over-year percentage change
NASDAQ Emerging Markets Index (NQEM) (NASDAQ)	No transformation applied
Gold Spot Prices - USD - Daily (Perth Mint)	No transformation applied
MSCI Emerging Markets Index Futures (Wiki Continuous Futures)	No transformation applied
US Dollar Index Futures (Wiki Continuous Futures)	No transformation applied
VIX (Yahoo Finance)	year-over-year percentage change

Informações aos autores

Objetivo e escopo

A ECONOMIA APLICADA é uma publicação trimestral do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, com uma linha editorial voltada exclusivamente para artigos científicos em economia aplicada. Tem por foco a publicação de trabalhos de análise econômica aplicada a problemas específicos, tanto de interesse do setor público como privado, e que contenham, particularmente, estudos quantitativos, cujos resultados possam aproximar a teoria da realidade. A interação entre trabalho empírico e teoria, análise e política econômica é, portanto, um importante diferencial da revista. A revista ECONOMIA APLICADA coloca-se como um espaço aberto para toda a comunidade acadêmica, nacional e internacional.

Política editorial

Os artigos submetidos são avaliados inicialmente pelos Editores quanto a sua adequação à linha editorial da Revista. Uma vez aceita a submissão, o artigo é encaminhado para avaliação por pareceristas designados pelos Editores, no sistema double-blind.

Apresentação de Artigos

Os artigos submetidos para avaliação devem obedecer às seguintes normas:

1. Apresentar ineditismos no país e no exterior
2. Na primeira página devem constar o título do artigo, nome dos autores com endereço para contato e afiliação institucional, resumo com até 100 palavras em português e inglês, palavras chave, em português e inglês e classificação JEL.
3. Na versão para “avaliação”, basta um arquivo no formato pdf com todas as equações, tabelas e figuras no corpo do texto, e referências no final.
4. Caso o artigo venha a ser aceito para publicação, a equipe da revista contatará os editores para as atividades de editoração. Nessa ocasião serão requeridos os seguintes arquivos:
 - Arquivo fonte no formato L^AT_EX.
 - Arquivo de referências bibliográficas no padrão B_BT_EX.
 - Arquivo com as figuras em formato TIFF ou EPS.
 - Planilha eletrônica com todas as tabelas.

Responsabilidade e Direitos Autorais

Os autores que submeterem ou tiverem o artigo aceito para publicação na Revista ECONOMIA APLICADA declaram:

- Ter participado da concepção do artigo para tornarem públicas suas responsabilidades pelos seus conteúdos, que não omitiram quaisquer ligações ou acordos de financiamento entre autores e companhias que possam ter interesse na publicação deste artigo.
- Que o manuscrito é inédito e que o artigo, em parte ou na íntegra, ou qualquer outro artigo com conteúdo substancialmente similar, de sua autoria, não foi enviado à outra revista e não o será enquanto sua publicação estiver sendo considerada pela revista ECONOMIA APLICADA, que seja no formato impresso ou no eletrônico, exceto o descrito em anexo.
- Que a revista ECONOMIA APLICADA passa a ter os direitos autorais para a veiculação dos artigos, tanto em formato impresso como eletrônico a eles referentes, os quais se tornarão propriedade exclusiva da ECONOMIA APLICADA. É permitida a reprodução total ou parcial dos artigos desta revista, desde que seja explicitamente citada a fonte completa.

ISSN 1980-5330



The Brazilian Journal of Applied Economics is a quarterly publication of the Departments of Economics of the FEA-RP of the University of São Paulo. The published papers are sole responsibility of the author(s). As the author(s) send a manuscript for submission to the Journal they tacitly agree that, if accepted for publication, the copyright of the material will be transferred to the ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). The Journal may then publish the material in paper and electronic formats that will be exclusive property of the ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). The publication of the whole papers or parts of it is permitted as long as the source is explicitly mentioned. This journal, in its online version, has several indexers, whose updated list can be found on the journal's website at Portal Revistas USP.

Secretary: Juliana de Freitas Vitória.

Cover: Christof Gunkel.

Linguistic revision (portuguese and english): Iulo Feliciano Afonso (Solution Translations Eireli — Epp)

Diagramming: Beatriz Porfirio Germin.

Mailing Address

Revista Economia Aplicada, Departamento de Economia, FEA-PR/USP

Avenida do Bandeirantes, 3900, Ribeirão Preto, SP.

CEP 14040-905

Telefone: (0xx16) 3315-0407

email: revecap@usp.br

Website: <https://www.revistas.usp.br/ecoa>

Editor

Eliezer Martins Diniz

Associate Editors

Edson Zambon Monte	Eliezer Martins Diniz
Fernando Antonio de Barros Júnior	Reynaldo Fernandes
Sergio Naruhiko Sakurai	

Editorial Board

Affonso Celso Pastore USP	Carlos Roberto Azzoni USP
Cássio F. Camargo Rolim UFPR	Cláudio Monteiro Considera UFF
Clélio Campolina Diniz CEDEPLAR	Denisard C. de Oliveira Alves USP
Eleuterio F. S. Prado USP	Fernando de Holanda Barbosa FGV/RJ, UFF
Francisco Anuatti Neto USP	Geoffrey J. D. Hewings University of Illinois
Geraldo Sant'Ana de Camargo Barros USP	Gustavo Maia Gomes IPEA
José Marcelino da Costa NAEA/PA	José A. Scheinkman Princeton University
Lízia de Figueirêdo UFMG	Marcelo Portugal UFRGS
Maria José Willumsen Florida International University	Márcio Gomes Pinto Garcia PUC/ RJ
Mário Luiz Possas UFRJ	Paulo César Coutinho UnB
Paulo Nogueira Batista Júnior FGV/SP	Pedro Cezar Dutra Fonseca UFRGS
Pedro Luiz Valls Pereira FGV/SP	Pierre Perron Boston University
Ricardo R. Araújo Lima UnB	Roberto Smith UFC
Rodolfo Hoffman USP	Rogério Studart UFRJ
Russell E. Smith Washburn University	Sérgio Werlang FGV/RJ
Tomás Málaga FGV/SP	Victor Bulmer-Thomas University of London
Wilson Suzigan Unicamp	