

Francisco Paulo Alves de Oliveira Filho

Uma breve análise de

MÉTODOS
QUANTITATIVOS
DE AVALIAÇÃO DE
POLÍTICAS PÚBLICAS:

aplicações ao Programa Bolsa Família,
desigualdade e transferências de renda
no Brasil



2025 - Thesis Editora Científica

Copyright © Thesis Editora Científica.

Copyright do texto © 2025 O autor.

Copyright da edição © 2025 Thesis Editora Científica

Direitos para esta edição cedidos à Thesis Editora Científica pelo autor.

Open access publication by Thesis Editora Científica.

Editor Chefe: Felipe Cardoso Rodrigues Vieira.

Diagramação, Projeto Gráfico e Design da Capa: Thesis Ed. Cien.

Revisão: O autor.



Licença Creative Commons

Uma breve análise de MÉTODOS QUANTITATIVOS DE AVALIAÇÃO DE POLÍTICAS PÚBLICAS: Aplicações ao Programa Bolsa Família, desigualdade e transferências de renda no Brasil da Thesis Editora Científica está licenciada com uma Licença Creative Commons - Atribuição-NãoComercial-SemDerivações 4.0 Internacional. (CC BY-NC-ND 4.0).

O conteúdo da obra e seus dados em sua forma, correção e confiabilidade é de responsabilidade exclusiva do autor, não representando a posição oficial da Thesis Editora Científica. É permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos ao autor, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares (*blind peer review*), membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

ISBN: 978-65-83199-82-9

Thesis Editora Científica
Teresina – PI – Brasil
contato@thesiseditora.com.br
www.thesiseditora.com.br

**Uma breve análise de MÉTODOS QUANTITATIVOS
DE AVALIAÇÃO DE POLÍTICAS PÚBLICAS:
Aplicações ao Programa Bolsa Família, desigualdade e
transferências de renda no Brasil**

Francisco Paulo Alves de Oliveira Filho

CONSELHO EDITORIAL

Felipe Cardoso Rodrigues Vieira – lattes.cnpq.br/9585477678289843
Adilson Tadeu Basquerote Silva – lattes.cnpq.br/8318350738705473
Andréia Barcellos Teixeira Macedo – lattes.cnpq.br/1637177044438320
Eliana Napoleão Cozendey da Silva – lattes.cnpq.br/2784584976313535
Rodolfo Ritchelle Lima dos Santos – lattes.cnpq.br/8295495634814963
Luís Carlos Ribeiro Alves – lattes.cnpq.br/9634019972654177
João Vitor Andrade – lattes.cnpq.br/1079560019523176
Bruna Aparecida Lisboa – lattes.cnpq.br/1321523568431354
Júlio César Coelho do Nascimento – lattes.cnpq.br/7514376995749628
Ana Paula Cordeiro Chaves – lattes.cnpq.br/4006977507638703
Stanley Keynes Duarte dos Santos – lattes.cnpq.br/3992636884325637
Brena Silva dos Santos – lattes.cnpq.br/8427724475551636
Jessica da Silva Campos – lattes.cnpq.br/7849599391816074
Milena Cordeiro de Freitas – lattes.cnpq.br/5913862860839738
Thiago Alves Xavier dos Santos – lattes.cnpq.br/4830258002967482
Clarice Bezerra – lattes.cnpq.br/8568045874935183
Bianca Thaís Silva do Nascimento – lattes.cnpq.br/4437575769985694
Ana Claudia Rodrigues da Silva – lattes.cnpq.br/6594386344012975
Francisco Ronner Andrade da Silva – lattes.cnpq.br/5014107373013731
Maria Isabel de Vasconcelos Mavignier Neta – lattes.cnpq.br/8440258181190366
Anita de Souza Silva – lattes.cnpq.br/9954744050650291
Sara Milena Gois Santos – lattes.cnpq.br/6669488863792604
Leônidas Luiz Rubiano de Assunção – lattes.cnpq.br/4636315219294766
Jose Henrique de Lacerda Furtado – lattes.cnpq.br/8839359674024233
Noeme Madeira Moura Fé Soares – lattes.cnpq.br/7107491370408847
Luciene Rodrigues Barbosa – lattes.cnpq.br/2146096901386355
Mário César de Oliveira – lattes.cnpq.br/8924508898024445
Antonio da Costa Cardoso Neto – lattes.cnpq.br/9036328153320126
Elisabete Soares de Santana – lattes.cnpq.br/1149505575311414
Thiago Santos Borges – lattes.cnpq.br/4495115911159968
Alessandra Souza dos Santos – lattes.cnpq.br/5313671194713784
Mateus Henrique Dias Guimarães – lattes.cnpq.br/7137001589681910
Emelyn Glorys Herasme Henríquez – lattes.cnpq.br/7314295633439757
George Luiz Nérís Caetano – lattes.cnpq.br/0598052051026256



2025 - Thesis Editora Científica

Copyright © Thesis Editora Científica.

Copyright do texto © 2025 O autor.

Copyright da edição © 2025 Thesis Editora Científica

Direitos para esta edição cedidos à Thesis Editora Científica pelo autor.

Open access publication by Thesis Editora Científica.

Editor Chefe: Felipe Cardoso Rodrigues Vieira.

Diagramação, Projeto Gráfico e Design da Capa: Thesis Ed. Cien.

Revisão: O autor.

Catálogo na Publicação

Elaborada por Bibliotecária Janaina Ramos – CRB-8/9166

O48b

Oliveira Filho, Francisco Paulo Alves de

Uma breve análise de métodos quantitativos de avaliação de políticas públicas: aplicações ao Programa Bolsa Família, desigualdade e transferências de renda no Brasil / Francisco Paulo Alves de Oliveira Filho. – Teresina-PI: Thesis Editora Científica, 2025.

Livro em PDF

ISBN 978-65-83199-82-9

1. Políticas públicas - Avaliação. 2. Programa Bolsa Família (Brasil). 3. Distribuição de renda - Brasil. I. Oliveira Filho, Francisco Paulo Alves de. II. Título.

CDD 320.6

Índice para catálogo sistemático

I. Políticas públicas - Avaliação

Thesis Editora Científica
Teresina – PI – Brasil
contato@thesiseditora.com.br
www.thesiseditora.com.br

APRESENTAÇÃO

Este livro transforma o estudo empírico sobre desigualdade e repasses do Programa Bolsa Família em uma obra de métodos quantitativos aplicada à avaliação de políticas públicas. O estudo original permanece integralmente incorporado como caso empírico, inclusive suas tabelas, mapas, gráficos, regressões descontínuas, testes de quebra estrutural e modelos em painel. A ampliação não procura substituir o artigo por uma exposição genérica: ela parte das decisões efetivamente tomadas na pesquisa e mostra como diferentes estratégias de identificação podem ser mobilizadas quando a pergunta muda, quando surgem novas fontes de variação ou quando se deseja testar a robustez de um resultado.

A primeira parte apresenta a lógica do contrafactual e desenvolve, de forma progressiva, regressão linear, painéis, diferenças-em-diferenças, estudo de evento, pareamento por escore de propensão, variáveis instrumentais, séries temporais interrompidas, controle sintético, regressão descontínua, testes de quebras e GMM dinâmico. A segunda parte revisa de maneira extensa a literatura empírica sobre o Bolsa Família, distinguindo resultados relativos a pobreza e desigualdade, focalização, educação, saúde e nutrição, oferta de trabalho, formalização e efeitos econômicos locais. Em seguida, o artigo que originou o livro é reproduzido como estudo de caso, preservando suas figuras e resultados, e a parte final integra os métodos em um roteiro de escolha, replicação e exercícios.

O livro não trata métodos quantitativos como uma coleção de comandos de software. A preocupação central é mostrar a pergunta que cada desenho consegue responder, a hipótese que sustenta a inferência e o tipo de viés que permanece quando essa hipótese não é plausível. O estudo do Bolsa Família é utilizado como laboratório empírico porque combina uma política pública de grande escala, dados longitudinais, heterogeneidade entre unidades federativas e mudanças institucionais capazes de gerar variação útil para análise.

SOBRE O AUTOR



Francisco Paulo Alves de Oliveira Filho é bacharel em Finanças pela Universidade Federal do Ceará, mestre em Economia pelo CAEN/UFC e doutorando em Economia pela mesma instituição. Foi aprovado em primeiro lugar na seleção pública para professor temporário de 40 horas da UECE, em 2025, e atua como professor substituto do Departamento de Teoria Econômica da UFC. Sua trajetória de ensino e pesquisa concentra-se em métodos quantitativos aplicados à Economia e às Finanças, com ênfase em Econometria Aplicada e Teórica, Estatística, Microeconomia, Teoria Econômica e Teoria Moderna das Finanças e Estudos Quantitativos Aplicados a Demografia e Políticas Públicas.

SUMÁRIO

PARTE I - FUNDAMENTOS E ESTRATÉGIAS DE AVALIAÇÃO	11
1. AVALIAÇÃO QUANTITATIVA DE POLÍTICAS PÚBLICAS: PERGUNTAS, CONTRAFCTUAIS E IDENTIFICAÇÃO.....	12
1.1 Viés de seleção e comparabilidade	12
1.2 Validade interna e validade externa	12
1.3 Decompondo a comparação observada	13
2. DADOS, MENSURAÇÃO E PREPARAÇÃO PARA AVALIAÇÃO.....	14
2.1 Unidade de análise, frequência e painel	14
2.2 Desagregação temporal e Denton-Chollette	14
2.3 Estatística descritiva antes da causalidade.....	14
2.4 Qualidade de dados, deflacionamento e rastreabilidade.....	15
3. REGRESSÃO LINEAR, CONTROLES OBSERVÁVEIS E COMPARAÇÕES BÁSICAS	16
3.1 Antes-depois e diferença de médias.....	16
3.2 Controles, forma funcional e erros-padrão	16
3.3 OLS, interpretação parcial e inferência	16
4. DADOS EM PAINEL: POOLED OLS, EFEITOS FIXOS E EFEITOS ALEATÓRIOS	18
4.1 Pooled OLS, FE e RE	18
4.2 Tendências e dependência temporal	18
4.3 Transformação within e interpretação da variação temporal.....	18
5. DIFERENÇAS-EM-DIFERENÇAS.....	20
5.1 Tendências paralelas e composição	20
5.2 DiD com covariáveis	20
5.3 DiD moderno, adoção escalonada e efeitos heterogêneos.....	20
6. ESTUDOS DE EVENTO E EXTENSÕES DO DID	22
6.1 Adoção escalonada	22
6.2 Leitura dos coeficientes dinâmicos.....	22
7. PAREAMENTO E PROPENSITY SCORE MATCHING	24
7.1 Estimação e diagnóstico do escore	24
7.2 Formas de matching e weighting.....	24
7.3 PSM combinado com DiD.....	24
7.4 Diagnóstico de balanceamento e limites do matching.....	24
7.5 Medida padronizada de balanceamento.....	25
8. VARIÁVEIS INSTRUMENTAIS.....	26
8.1 Instrumentos fracos e interpretação local	26
8.2 Relevância, exclusão e interpretação local	26

9. SÉRIES TEMPORAIS INTERROMPIDAS	28
9.1 ITS controlada	28
9.2 Nível, inclinação, sazonalidade e autocorrelação	28
10. CONTROLE SINTÉTICO.....	29
10.1 Inferência por placebos.....	29
10.2 Construção do contrafactual e inferência por placebos	29
11. REGRESSÃO DESCONTÍNUA: DESENHO, ESTIMAÇÃO E DIAGNÓSTICO	31
11.1 RDD temporal e pontos de corte endógenos	31
11.2 Bandwidth, kernel e polinômios locais.....	31
11.3 Testes de manipulação e continuidade	31
11.4 RDD local, largura de banda e interpretação do salto	31
11.5 Condição de continuidade.....	32
12. QUEBRAS ESTRUTURAIS, PELT, CHOW E BAI-PERRON	33
12.1 Quebra estatística não é automaticamente evento causal	33
12.2 Validação de datas e multiplicidade de quebras	33
13. GMM EM PAINEL DINÂMICO.....	35
13.1 Diagnósticos e proliferação de instrumentos.....	35
13.2 GMM como robustez, não como substituto do desenho	35
13.3 Condições de momento, instrumentos e diagnóstico.....	35
13.4 Condições de momento e teste J.....	36
PARTE II - REVISÃO DA LITERATURA: BOLSA FAMÍLIA E AVALIAÇÃO QUANTITATIVA	37
14. TRANSFERÊNCIAS CONDICIONADAS DE RENDA E O PROBLEMA DE AVALIAÇÃO	38
15. POBREZA, DESIGUALDADE E FOCALIZAÇÃO.....	40
16. EDUCAÇÃO, SAÚDE E NUTRIÇÃO	42
17. MERCADO DE TRABALHO, CONSUMO E EFEITOS ECONÔMICOS LOCAIS	44
18. SÍNTESE METODOLÓGICA DA LITERATURA E LACUNAS DE PESQUISA	46
PARTE III - ESTUDO DE CASO: BOLSA FAMÍLIA, DESIGUALDADE E REPASSES ESTADUAIS.....	49
19. ESTRATÉGIAS COMBINADAS E TESTES DE ROBUSTEZ.....	75
19.1 Robustez por desenho	75
19.2 Placebos e falsificações	75
19.3 Heterogeneidade de efeitos.....	75
20. COMO ESCOLHER O MÉTODO DE AVALIAÇÃO	77
20.1 Árvore decisória	77
20.2 O método mais sofisticado nem sempre é o melhor.....	77
20.3 Checklist mínimo de identificação	77

21. ROTEIRO DE REPLICAÇÃO EM R E PYTHON.....	79
21.1 Estrutura de pastas	79
21.2 Exemplo conceitual em R.....	79
21.3 Exemplo conceitual em Python	79
21.4 Validação automática	79
22. EXERCÍCIOS, QUESTÕES DE REVISÃO E PROPOSTAS DE PESQUISA.....	80
22.1 Questões conceituais.....	80
22.2 Exercícios empíricos.....	80
22.3 Propostas de pesquisa	81
23. CONSIDERAÇÕES FINAIS DO LIVRO.....	82
REFERÊNCIAS COMPLEMENTARES	83
APÊNDICE A - FORMULÁRIO DE LEITURA DE ARTIGOS DE AVALIAÇÃO	85

PARTE I - FUNDAMENTOS E ESTRATÉGIAS DE AVALIAÇÃO

1. AVALIAÇÃO QUANTITATIVA DE POLÍTICAS PÚBLICAS: PERGUNTAS, CONTRAFCTUAIS E IDENTIFICAÇÃO

Avaliar uma política pública significa comparar o resultado observado com uma referência que represente, de forma crível, o que teria ocorrido na ausência da intervenção. Essa referência é o contrafactual. A dificuldade fundamental é que a mesma unidade não pode ser observada simultaneamente em dois estados do mundo: tratada e não tratada. Por isso, toda estratégia de avaliação procura construir um grupo, uma trajetória ou uma função que aproxime o resultado potencial não observado. O problema é conceitualmente simples, mas suas consequências são profundas: uma correlação entre participação em um programa e renda, emprego, escolaridade ou desigualdade não é automaticamente um efeito causal.

Em notação de resultados potenciais, $Y_i(1)$ representa o resultado que a unidade i teria sob tratamento e $Y_i(0)$ o resultado sob ausência de tratamento. O efeito individual é a diferença $Y_i(1) - Y_i(0)$, mas apenas um desses resultados é observado. A avaliação quantitativa procura estimar médias de efeitos, como o efeito médio do tratamento (ATE) ou o efeito médio entre os tratados (ATT), usando hipóteses capazes de substituir o resultado potencial faltante por informação observada em outras unidades ou períodos.

$$ATE = E[Y_i(1) - Y_i(0)], \quad ATT = E[Y_i(1) - Y_i(0) | D_i = 1]$$

O primeiro passo de uma boa avaliação é formular a pergunta corretamente. Perguntas descritivas tratam de níveis, tendências e desigualdades; perguntas associativas investigam como variáveis se movem conjuntamente; perguntas causais perguntam quanto um resultado mudou por causa de uma intervenção. Cada nível exige uma estratégia diferente. Estatísticas descritivas e regressões simples podem ser excelentes para caracterizar um fenômeno, mas não substituem um desenho de identificação quando o objetivo é afirmar impacto.

1.1 Viés de seleção e comparabilidade

O viés de seleção surge quando unidades tratadas e não tratadas diferem em fatores que também afetam o resultado. Em programas sociais focalizados, esse problema é esperado: beneficiários costumam ser mais vulneráveis antes do tratamento. Comparar diretamente beneficiários e não beneficiários pode, portanto, atribuir ao programa diferenças que existiam antes dele. O mesmo ocorre quando governos ampliam recursos justamente em localidades que atravessam choques adversos. A direção do viés não é necessariamente óbvia e deve ser discutida em função do mecanismo institucional.

Há três famílias de soluções. A primeira explora seleção em observáveis, como regressão com controles e matching. A segunda usa variações quase experimentais, como diferenças-em-diferenças, variáveis instrumentais e regressão descontínua. A terceira constrói contrafactuais temporais ou agregados, como séries temporais interrompidas e controle sintético. Os métodos não competem apenas por sofisticação; eles respondem a desenhos de dados diferentes e dependem de hipóteses distintas.

1.2 Validade interna e validade externa

Validade interna é o grau em que o efeito estimado pode ser interpretado causalmente para a população e o contexto analisados. Validade externa é a capacidade de generalização. Um RDD pode ter validade interna muito alta próximo ao ponto de corte e, ao mesmo tempo, validade externa limitada para observações distantes. Um painel nacional com efeitos fixos pode cobrir todos os estados, mas depender de hipóteses mais fortes para interpretar seu coeficiente como efeito causal. A avaliação deve declarar explicitamente essa troca.

Regra prática: antes de escolher o estimador, escreva em uma frase qual é o contrafactual e por que ele seria comparável ao observado. Se essa frase não for convincente, a sofisticação algébrica dificilmente resolverá o problema de identificação.

1.3 Decompondo a comparação observada

Uma maneira útil de visualizar o problema de seleção consiste em decompor a diferença observada entre tratados e não tratados. A diferença bruta contém o efeito médio sobre os tratados, mas também incorpora a diferença que já existiria entre os grupos na ausência do programa. Em políticas focalizadas, esse segundo termo tende a ser relevante porque a própria regra de seleção procura unidades em situação distinta. A finalidade dos desenhos quase experimentais é criar condições para que o componente de seleção seja removido, anulado por diferenciação, controlado por observáveis ou substituído por uma fonte de variação exógena.

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = ATT + \{E[Y_i(0) | D_i = 1] - E[Y_i(0) | D_i = 0]\}$$

Essa decomposição ajuda a entender por que resultados de regressões diferentes não devem ser comparados apenas pela magnitude do coeficiente. Um estimador pode produzir efeito menor porque corrige um viés positivo presente na comparação simples; outro pode produzir efeito maior porque restringe a análise a uma população local com resposta mais intensa. A interpretação começa pelo parâmetro identificado e pela população a que ele se refere. ATE, ATT, LATE e efeitos locais de RDD são objetos distintos, mesmo quando todos são medidos na mesma unidade.

2. DADOS, MENSURAÇÃO E PREPARAÇÃO PARA AVALIAÇÃO

A qualidade de uma avaliação depende da qualidade da mensuração. Variáveis administrativas, pesquisas amostrais e indicadores construídos podem possuir frequências, unidades e definições incompatíveis. No estudo empírico que dá origem a este livro, os repasses do Bolsa Família são mensais, enquanto o índice de Gini estadual tem origem em informação anual. A transformação para uma frequência comum permite explorar variação temporal, mas também introduz hipóteses sobre a trajetória intranual. Por isso, a preparação dos dados faz parte do método, e não deve ser tratada como etapa puramente mecânica.

2.1 Unidade de análise, frequência e painel

A unidade de análise determina a interpretação dos coeficientes. Em um painel de estados por mês, cada observação corresponde a uma unidade federativa em um período. O número de observações pode ser grande, mas a quantidade de unidades independentes continua limitada. Esse detalhe afeta a escolha dos erros-padrão, a estrutura de dependência e a capacidade de identificar efeitos. Em séries mensais, sazonalidade, autocorrelação e choques nacionais comuns também exigem atenção.

O pesquisador deve construir um dicionário de dados com nome, definição, unidade, fonte, frequência, transformação e faixa esperada de cada variável. Valores monetários precisam ser convertidos para preços constantes quando o objetivo é comparar poder de compra ao longo do tempo. Índices limitados entre zero e um devem ser inspecionados quanto a erros de escala. Datas precisam ser representadas de modo consistente, e mudanças metodológicas nas fontes devem ser documentadas.

2.2 Desagregação temporal e Denton-Cholette

Quando um indicador anual precisa ser compatibilizado com uma série mensal, a desagregação temporal pode ser realizada por métodos que preservam a coerência com o agregado. O procedimento Denton-Cholette procura gerar uma trajetória de alta frequência que respeite a restrição anual e minimize oscilações artificiais. Sem indicador mensal auxiliar, a versão proporcional distribui o movimento de forma suave. Essa escolha aumenta a quantidade de observações, mas não cria informação econômica nova: os meses interpolados continuam dependentes dos valores anuais observados.

$$\min_{\{x_t\}} \sum_t \left(\frac{\Delta x_t}{x_{t-1}} \right)^2 \quad \text{sujeito a} \quad \frac{1}{12} \sum_{m=1}^{12} x_{t,m} = X_t$$

Essa limitação deve aparecer na interpretação. A desagregação pode ser útil para alinhar frequências e implementar modelos que exigem uma grade temporal comum, porém não deve ser apresentada como se revelasse choques mensais do Gini que não existiam na fonte original. Uma análise de sensibilidade recomendável compara resultados com dados anuais, com alternativas de interpolação e com especificações que utilizem apenas a variação efetivamente observada.

2.3 Estatística descritiva antes da causalidade

Antes de estimar qualquer modelo, é necessário examinar distribuição, dispersão, tendência, valores extremos, quebras e correlações. Médias podem esconder forte heterogeneidade entre estados; desvios-padrão absolutos precisam ser comparados com a escala da variável; mapas ajudam a localizar padrões regionais; séries temporais permitem identificar

choques comuns. Uma avaliação causal consistente começa por uma boa leitura descritiva, porque diagnósticos visuais frequentemente revelam violações que testes formais confirmam apenas depois.

2.4 Qualidade de dados, deflacionamento e rastreabilidade

Em avaliações aplicadas, uma parte importante do trabalho econométrico ocorre antes da estimação. Séries monetárias, cadastros administrativos e indicadores socioeconômicos frequentemente são produzidos por instituições distintas, com calendários de divulgação, conceitos e unidades que mudam ao longo do tempo. A compatibilização exige registrar cada transformação, inclusive mudanças de base, deflacionamento, tratamento de ausências e decisões sobre agregação. Uma regra útil é conservar os dados brutos sem qualquer edição e produzir uma base analítica por meio de scripts reproduzíveis. Esse procedimento reduz a possibilidade de que correções manuais se tornem invisíveis e permite reconstruir cada número apresentado em tabela ou figura.

O deflacionamento é indispensável quando se pretende comparar valores monetários em diferentes datas. Se o valor nominal cresce apenas porque o nível geral de preços aumentou, uma regressão em níveis nominais pode atribuir significado econômico a uma variação que corresponde sobretudo à inflação. No caso de transferências públicas, a escolha do deflator deve ser coerente com a pergunta. O IPCA é apropriado para expressar poder de compra agregado das famílias, mas análises setoriais ou regionais podem exigir índices mais específicos. É igualmente importante informar o mês-base e evitar misturar valores reais calculados com bases diferentes.

Dados faltantes merecem tratamento explícito. Interpolar uma observação ausente, carregar o último valor disponível ou excluir a unidade são decisões que alteram a informação utilizada pelo estimador. Em vez de adotar uma regra automática para todos os casos, convém distinguir ausência por falha de registro, ausência porque a variável não existia e ausência associada ao próprio fenômeno estudado. Quando a imputação é inevitável, a análise de sensibilidade deve comparar os resultados com e sem os valores imputados. Essa disciplina é especialmente relevante em painéis de políticas sociais, nos quais alterações administrativas podem modificar simultaneamente a cobertura do programa e a qualidade do registro.

$$ValorReal_t = \frac{ValorNominal_t}{IPCA_t/IPCA_{base}}$$

3. REGRESSÃO LINEAR, CONTROLES OBSERVÁVEIS E COMPARAÇÕES BÁSICAS

A regressão linear é o ponto de partida de grande parte da econometria aplicada. Em sua forma mais simples, relaciona um resultado Y a uma variável explicativa X por meio de uma função linear. Em avaliação, pode ser usada para comparar grupos controlando por características observáveis. O coeficiente de tratamento mede uma diferença condicional, isto é, a diferença entre tratados e controles depois de manter constantes as covariáveis incluídas. A interpretação causal, entretanto, exige que não existam fatores omitidos correlacionados simultaneamente com tratamento e resultado.

$$Y_i = \beta_0 + \tau D_i + X_i' \beta + \varepsilon_i$$

Se D_i indica exposição ao programa, τ é a diferença média ajustada. O estimador OLS escolhe os coeficientes que minimizam a soma dos quadrados dos resíduos. Seu uso é simples, transparente e útil como benchmark. A fragilidade está na hipótese de exogeneidade condicional: $E[\varepsilon_i | D_i, X_i] = 0$. Se vulnerabilidade não observada, capacidade administrativa, choques locais ou expectativas influenciam simultaneamente o tratamento e o resultado, τ pode incorporar esses efeitos.

3.1 Antes-depois e diferença de médias

A comparação antes-depois calcula a mudança no resultado de um grupo entre dois períodos. É informativa quando o objetivo é descrever evolução, mas não separa o efeito da política de tendências macroeconômicas, sazonalidade ou choques simultâneos. A diferença de médias entre tratados e controles também é descritiva se os grupos não forem comparáveis. Esses dois procedimentos são importantes porque ajudam a construir a intuição do contrafactual, mas devem ser apresentados como desenhos básicos, não como prova definitiva de causalidade.

3.2 Controles, forma funcional e erros-padrão

A inclusão de controles deve seguir uma lógica causal. Variáveis determinadas depois do tratamento podem ser mediadores e sua inclusão pode remover parte do próprio efeito que se deseja estimar. Variáveis pré-tratamento relacionadas ao resultado podem aumentar precisão e reduzir viés sob seleção em observáveis. Polinômios excessivos, interações não justificadas e seleção automática de variáveis podem produzir resultados instáveis. A forma funcional deve ser guiada por teoria, visualização e testes de robustez.

Erros-padrão robustos tratam heterocedasticidade, mas não resolvem dependência serial ou correlação dentro de grupos. Em dados estaduais mensais, erros agrupados por unidade federativa são frequentemente mais adequados do que erros independentes. Quando há poucos clusters, correções para amostras pequenas ou inferência por bootstrap de clusters podem ser consideradas. O pesquisador deve separar claramente a estimativa pontual do coeficiente da incerteza usada para julgar sua precisão.

3.3 OLS, interpretação parcial e inferência

A regressão linear é simples apenas na aparência. O coeficiente associado ao tratamento possui interpretação causal somente se, condicionadas as covariáveis incluídas, não restarem fatores não observados correlacionados simultaneamente com tratamento e resultado. Em políticas focalizadas por renda, essa condição pode ser particularmente exigente, porque vulnerabilidade econômica, estrutura familiar, inserção ocupacional e acesso a serviços influenciam tanto a probabilidade de receber o benefício quanto diversos desfechos. Por isso, uma regressão com muitos controles não deve ser apresentada automaticamente como avaliação

de impacto. Seu papel pode ser descritivo, servir de benchmark ou explorar associações condicionais que depois serão confrontadas com desenhos de identificação mais fortes.

O estimador de mínimos quadrados ordinários escolhe o vetor de coeficientes que minimiza a soma dos quadrados dos resíduos. Na forma matricial, a solução torna evidente que a identificação depende da variação existente em X e da ausência de colinearidade perfeita. Em amostras com grande heterogeneidade de escala, transformações logarítmicas podem ser úteis, mas mudam a interpretação: em um modelo log-log, o coeficiente é uma elasticidade; em um modelo log-nível, mede a variação percentual aproximada em Y provocada por uma unidade adicional de X . A forma funcional deve ser escolhida a partir da teoria, da distribuição dos dados e de diagnósticos, e não apenas pelo maior R^2 .

A inferência exige atenção à estrutura dos erros. Heterocedasticidade pode ser tratada por matrizes robustas, mas dados repetidos por estado, município ou indivíduo costumam apresentar correlação dentro da unidade. Nesse caso, erros agrupados são mais adequados porque permitem dependência arbitrária no interior do cluster. Quando o número de clusters é pequeno, aproximações assintóticas podem ser imprecisas, e procedimentos como wild cluster bootstrap constituem alternativa. A mensagem prática é que o tamanho do erro-padrão não é um detalhe de apresentação: ele incorpora uma hipótese sobre quais observações oferecem informação independente ao estimador.

$$\hat{\beta} = (X'X)^{-1}X'Y$$

4. DADOS EM PAINEL: POOLED OLS, EFEITOS FIXOS E EFEITOS ALEATÓRIOS

Dados em painel acompanham as mesmas unidades em vários períodos e permitem controlar heterogeneidade não observada que seja constante no tempo. Essa estrutura é especialmente útil para avaliação de políticas públicas porque estados, municípios, escolas ou indivíduos possuem características persistentes difíceis de mensurar: capacidade institucional, geografia, cultura administrativa e composição econômica. O modelo de efeitos fixos remove essas diferenças invariantes e identifica o coeficiente a partir de mudanças dentro da própria unidade.

$$Y_{it} = \alpha_i + \lambda_t + \beta X_{it} + \tau D_{it} + \varepsilon_{it}$$

O termo α_i representa o efeito específico da unidade, enquanto λ_t pode capturar choques comuns de cada período. Ao incluir efeitos fixos de unidade e tempo, o pesquisador compara a unidade consigo mesma, descontando fatores permanentes e choques nacionais. A hipótese necessária é que, condicionados aos efeitos fixos e controles, os regressores não estejam correlacionados com choques idiossincráticos contemporâneos ou futuros de forma que viole a exogeneidade requerida.

4.1 Pooled OLS, FE e RE

O pooled OLS ignora a estrutura longitudinal e trata as observações como se fossem independentes. Pode ser útil como referência, mas tende a misturar variação entre unidades e dentro das unidades. O modelo de efeitos fixos permite correlação arbitrária entre α_i e os regressores. O modelo de efeitos aleatórios, por sua vez, assume que o componente específico da unidade é não correlacionado com as variáveis explicativas. Quando essa hipótese é plausível, RE pode ser mais eficiente; quando não é, o estimador é inconsistente.

O teste de Hausman é frequentemente usado para comparar FE e RE, mas sua decisão não deve substituir a reflexão substantiva. Em políticas sociais, é comum que características permanentes das unidades estejam relacionadas à intensidade de tratamento, o que torna a hipótese de efeitos aleatórios forte. FE costuma ser uma escolha conservadora para relações dentro da unidade, embora não estime o efeito de variáveis invariantes no tempo.

4.2 Tendências e dependência temporal

Adicionar efeitos fixos de tempo controla choques nacionais comuns, mas não tendências específicas por estado. Em algumas aplicações, tendências lineares por unidade são usadas para absorver trajetórias diferenciadas. Essa prática pode melhorar comparabilidade, mas também remover parte do efeito se a política mudar lentamente. A decisão precisa ser justificada e acompanhada de especificações alternativas. Autocorrelação serial exige inferência robusta, e choques espaciais comuns podem exigir erros com estruturas de dependência mais amplas.

4.3 Transformação within e interpretação da variação temporal

O estimador de efeitos fixos pode ser entendido pela transformação within. Para cada unidade, subtrai-se de cada variável sua média temporal; com isso, desaparecem todos os fatores invariantes no tempo, observados ou não. A identificação passa a utilizar apenas mudanças dentro da própria unidade. Essa característica é valiosa quando estados diferem persistentemente em estrutura produtiva, geografia, cultura administrativa ou composição demográfica, mas implica uma limitação: variáveis praticamente constantes ao longo do tempo terão pouca informação útil para o coeficiente. Em painéis curtos, mudanças pequenas também

podem amplificar ruído de mensuração.

Efeitos fixos de tempo complementam o controle de heterogeneidade ao absorver choques comuns a todas as unidades, como recessões nacionais, inflação, alterações federais de regra e crises sanitárias. Contudo, se a política muda para todos os estados no mesmo mês e não existe heterogeneidade de exposição, uma especificação com efeitos fixos de tempo pode absorver grande parte da variação de interesse. A pergunta de identificação deve, então, explorar intensidade diferencial, elegibilidade, exposição ou outra dimensão que varie entre unidades. Esse ponto é central para distinguir um painel meramente correlacional de um desenho que produz comparação contrafactual.

O modelo de efeitos aleatórios utiliza também a variação entre unidades, mas requer que o efeito específico não observado seja não correlacionado com os regressors. Em avaliações de políticas sociais essa hipótese pode ser difícil de sustentar, já que características permanentes frequentemente influenciam a participação ou intensidade da política. A comparação entre FE e RE por teste de Hausman é informativa, porém a decisão final deve ser substantiva. Mesmo quando o teste não rejeita diferenças sistemáticas, a plausibilidade institucional da hipótese de exogeneidade continua necessária.

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + \tau(D_{it} - \bar{D}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

5. DIFERENÇAS-EM-DIFERENÇAS

Diferenças-em-diferenças (DiD) é uma das estratégias mais utilizadas em avaliação quase experimental. O método compara a mudança ao longo do tempo no grupo tratado com a mudança observada em um grupo de controle. A ideia é descontar fatores permanentes que diferenciam os grupos e choques temporais comuns a ambos. O contrafactual é a trajetória que o grupo tratado teria seguido caso sua evolução após a intervenção mantivesse a mesma diferença de tendência em relação ao controle.

$$Y_{it} = \alpha_i + \lambda_t + \tau(\text{Tratado}_i \times \text{Pos}_t) + \varepsilon_{it}$$

O coeficiente τ é a diferença entre duas diferenças: a mudança dos tratados menos a mudança dos controles. A hipótese central é a de tendências paralelas na ausência do tratamento. Ela não exige que os níveis sejam iguais; exige que, sem a intervenção, a diferença entre os grupos evoluísse de forma estável. Como essa trajetória contrafactual não é observável depois do tratamento, o pesquisador usa períodos pré-tratamento, conhecimento institucional e análises de sensibilidade para avaliar plausibilidade.

5.1 Tendências paralelas e composição

Gráficos de médias pré-tratamento são essenciais. Coeficientes de leads em estudos de evento também ajudam a detectar diferenças sistemáticas antes da intervenção, embora ausência de significância não prove tendências paralelas. Mudanças na composição dos grupos, antecipação do tratamento e políticas concomitantes podem comprometer a identificação. Em programas sociais, deve-se verificar se critérios de elegibilidade, cadastros ou cobertura mudaram ao mesmo tempo que a intervenção estudada.

5.2 DiD com covariáveis

Covariáveis podem ser incluídas para melhorar precisão ou ajustar diferenças observáveis, desde que não sejam afetadas pelo tratamento. Em painéis, a forma com efeitos fixos de unidade e tempo é natural. Em amostras repetidas, pesos e modelos de composição podem ser necessários. O uso de erros agrupados no nível em que o tratamento varia é regra importante: se a política é definida por estado, a inferência precisa refletir que unidades dentro do mesmo estado compartilham o mesmo choque.

Uma leitura correta de DiD separa três elementos: o efeito estimado, a hipótese de tendências paralelas e o nível em que a política varia. Ignorar qualquer um deles pode produzir intervalos de confiança excessivamente estreitos ou uma narrativa causal mais forte do que o desenho permite.

5.3 DiD moderno, adoção escalonada e efeitos heterogêneos

A formulação clássica de duas unidades e dois períodos é uma excelente introdução ao diferenças-em-diferenças, mas aplicações modernas frequentemente envolvem tratamentos iniciados em datas distintas e efeitos que variam entre grupos. Nessa situação, a regressão convencional com efeitos fixos de unidade e tempo pode combinar comparações muito diferentes, inclusive usando unidades já tratadas como controle para unidades tratadas mais tarde. Goodman-Bacon (2021) mostra que o estimador pode ser decomposto em médias ponderadas de comparações 2x2; quando os efeitos são heterogêneos, alguns pesos produzem uma média de interpretação difícil.

Uma alternativa é estimar efeitos por coorte e período. Callaway e Sant'Anna (2021) definem parâmetros $ATT(g,t)$, em que g identifica o primeiro período de tratamento e t o período em que o resultado é medido. As unidades ainda não tratadas ou nunca tratadas formam grupos

de comparação adequados, e os efeitos são posteriormente agregados segundo critérios transparentes. Esse procedimento separa a heterogeneidade temporal da heterogeneidade entre coortes e evita que a especificação esconda quais comparações sustentam o resultado agregado.

A hipótese de tendências paralelas continua sendo o núcleo do desenho. Ela deve ser defendida com conhecimento institucional e evidência pré-tratamento, mas testes formais de pré-tendências possuem poder limitado. Grupos podem apresentar trajetórias visualmente próximas por poucos períodos e divergir logo depois por razões não relacionadas à política. Além disso, antecipação altera o momento efetivo do tratamento: se famílias ou gestores ajustam comportamento antes da data oficial, períodos classificados como pré podem já conter efeito. Por isso, janelas de antecipação, tendências alternativas e outcomes placebo ajudam a avaliar robustez.

Na interpretação, é importante separar o efeito da política para as unidades efetivamente expostas do efeito de uma mudança nacional concomitante. O DiD identifica um contraste relativo. Se tratado e controle sofrem simultaneamente um choque macroeconômico, o método pode removê-lo desde que o choque afete os grupos de forma paralela. Se a exposição ao choque é diferencial e correlacionada ao tratamento, a hipótese falha. Uma boa aplicação documenta a composição dos grupos, apresenta resultados por tempo relativo e explica por que o grupo de comparação representa a trajetória contrafactual plausível.

$$ATT(g, t) = E[Y_t(1) - Y_t(0) \mid G = g]$$

6. ESTUDOS DE EVENTO E EXTENSÕES DO DID

O estudo de evento generaliza diferenças-em-diferenças ao estimar efeitos em vários períodos relativos ao início do tratamento. Em vez de uma única dummy pós, são criados indicadores para tempos relativos, como dois anos antes, um ano antes, o primeiro ano de tratamento e os anos seguintes. O resultado permite visualizar antecipação, dinâmica do impacto e persistência.

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \beta_k \mathbf{1}\{\text{tempo relativo} = k\} + \varepsilon_{it}$$

O período imediatamente anterior ao tratamento costuma ser a categoria de referência. Coeficientes anteriores ao tratamento ajudam a examinar se tratados e controles já apresentavam trajetórias divergentes. Coeficientes posteriores mostram como o efeito evolui. A interpretação exige cautela quando unidades adotam o tratamento em momentos diferentes, porque o estimador tradicional de efeitos fixos de duas vias pode usar unidades já tratadas como controles de unidades tratadas posteriormente.

6.1 Adoção escalonada

Quando o tratamento ocorre em datas distintas, métodos modernos de DiD estimam efeitos por coorte e período, construindo comparações com unidades ainda não tratadas ou nunca tratadas. Essa abordagem evita ponderações negativas e combinações difíceis de interpretar que podem surgir no estimador convencional. O princípio geral é simples: definir claramente quem é um controle válido em cada momento e agregar efeitos heterogêneos de forma transparente.

Para um programa nacional com mudanças comuns a todos os estados, um estudo de evento tradicional pode não ter grupo de controle nunca tratado. Nesse caso, a estratégia precisa explorar heterogeneidade na intensidade da exposição, regras de elegibilidade, limiares ou outras fontes de variação. Quando todas as unidades sofrem o mesmo choque simultaneamente, uma série temporal interrompida ou um desenho de controle sintético pode ser mais adequado do que DiD.

6.2 Leitura dos coeficientes dinâmicos

O estudo de evento permite observar como o efeito evolui antes e depois da intervenção. Os coeficientes anteriores ao tratamento são frequentemente usados como diagnóstico de tendências paralelas, enquanto os posteriores descrevem velocidade, persistência e eventual dissipação do impacto. Entretanto, a leitura de um gráfico não deve ser reduzida a procurar estrelas de significância. Intervalos amplos podem esconder diferenças economicamente relevantes, e coeficientes pré-tratamento individualmente não significativos podem, em conjunto, sugerir trajetória sistemática. Testes conjuntos e inspeção da magnitude são complementares.

Em adoção escalonada, a especificação tradicional de estudo de evento com efeitos fixos pode sofrer o mesmo problema de comparações contaminadas do DiD convencional. Sun e Abraham (2021) propõem estimadores que constroem efeitos por coorte e tempo relativo antes de agregá-los, reduzindo viés quando os efeitos diferem entre grupos ou ao longo do tempo. A escolha entre métodos deve ser informada pelo calendário da política e pela existência de unidades nunca tratadas ou ainda não tratadas.

A dinâmica também ajuda a avaliar mecanismos. Um efeito que surge imediatamente após uma transferência monetária pode ser compatível com resposta de consumo ou renda disponível; um efeito educacional pode demandar vários períodos; resultados de saúde podem

aparecer em horizontes distintos conforme o desfecho. Essa interpretação exige prudência, porque o padrão temporal não identifica sozinho o mecanismo. Ainda assim, combinar conhecimento institucional, resultados intermediários e estudo de evento produz uma narrativa empírica mais convincente do que um único coeficiente pós-tratamento.

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \beta_k \mathbf{1}\{t - T_i = k\} + X'_{it}\gamma + \varepsilon_{it}$$

7. PAREAMENTO E PROPENSITY SCORE MATCHING

Métodos de pareamento procuram construir um grupo de comparação semelhante ao grupo tratado em características observáveis. A ideia é aproximar uma situação em que, condicionadas às covariáveis, as diferenças de resultado possam ser atribuídas ao tratamento. O propensity score é a probabilidade de receber o tratamento dado um conjunto de características pré-tratamento. Rosenbaum e Rubin mostraram que, sob hipóteses específicas, parear pelo escore pode equilibrar a distribuição das covariáveis.

$$e(X_i) = P(D_i = 1 | X_i)$$

A primeira hipótese é a independência condicional: depois de controlar por X , os resultados potenciais são independentes do tratamento. A segunda é suporte comum ou positividade: para combinações relevantes de X , deve existir probabilidade positiva de estar em ambos os grupos. O PSM não corrige seleção por fatores não observados. Sua utilidade depende, portanto, de uma base rica em covariáveis anteriores ao tratamento e de boa sobreposição entre tratados e controles.

7.1 Estimação e diagnóstico do escore

O escore pode ser estimado por logit, probit ou métodos flexíveis, mas a meta não é maximizar poder preditivo. O objetivo é obter equilíbrio. Depois do pareamento, devem ser examinadas diferenças padronizadas de médias, razões de variâncias e distribuição dos escores. Unidades fora do suporte comum podem ser descartadas, alterando a população para a qual o efeito é estimado. Essa mudança deve ser explicitada.

7.2 Formas de matching e weighting

Nearest neighbor seleciona controles com escore próximo; caliper impõe distância máxima; kernel utiliza vários controles com pesos; weighting pelo inverso da probabilidade cria uma pseudo-população balanceada. Métodos de balanceamento direto e entropy balancing são alternativas úteis. Em todos os casos, inferência deve considerar que o escore foi estimado e que o procedimento de pareamento altera a amostra efetiva.

7.3 PSM combinado com DiD

A combinação PSM+DiD é comum quando grupos diferem em observáveis e também se dispõe de dados antes e depois. O matching procura aproximar a composição inicial, enquanto DiD remove diferenças permanentes não observadas sob tendências paralelas. Essa dupla estratégia não torna a análise automaticamente causal; ela apenas desloca as hipóteses. Ainda é necessário demonstrar sobreposição, equilíbrio, tendências prévias e ausência de choques específicos que atinjam apenas um grupo.

7.4 Diagnóstico de balanceamento e limites do matching

O propensity score é uma ferramenta de redução de dimensionalidade: sob independência condicional, unidades com o mesmo escore possuem distribuição semelhante das covariáveis utilizadas no modelo de tratamento. O objetivo do PSM, portanto, não é prever perfeitamente quem recebe a política, mas construir comparabilidade em características pré-tratamento. Um modelo com elevada acurácia preditiva pode produzir pouco suporte comum se separar quase perfeitamente tratados e controles; nesse caso, a ausência de unidades comparáveis limita a inferência e deve levar à redefinição da população de interesse.

Depois do pareamento, o diagnóstico principal é o balanceamento. Diferenças padronizadas de médias, razões de variância e gráficos de distribuição ajudam a verificar se as

covariáveis se tornaram semelhantes. Reestimar o propensity score até obter um p-valor desejado não é uma boa prática; o processo deve ser orientado por variáveis determinadas antes do tratamento e por conhecimento do mecanismo de seleção. Variáveis posteriores ao tratamento podem introduzir viés ao controlar canais pelos quais a própria política afeta o desfecho.

Mesmo com excelente balanceamento, o matching não resolve confusão por fatores não observados. Por essa razão, sua força aumenta quando combinado com estrutura temporal, como PSM seguido de diferenças-em-diferenças. O pareamento aproxima os grupos em níveis observáveis e o DiD remove diferenças invariantes no tempo, mas a estratégia conjunta ainda exige tendências paralelas condicionais. Estudos de sensibilidade, como bounds de Rosenbaum, podem indicar quanto uma variável não observada teria de influenciar a seleção para alterar a conclusão, sem converter a análise em prova definitiva de ausência de viés.

$$ATT = E[Y_i(1) - Y_i(0) | D_i = 1] \approx \frac{1}{N_1} \sum_{i:D_i=1} \left(Y_i - \sum_{j:D_j=0} w_{ij} Y_j \right)$$

7.5 Medida padronizada de balanceamento

Uma medida simples para comparar covariáveis antes e depois do matching é a diferença padronizada de médias. Como ela divide a diferença de médias por uma medida de dispersão, não depende da unidade em que a variável foi medida. Valores próximos de zero indicam maior semelhança; limites como 0,1 são referências práticas, não testes universais. O diagnóstico deve ser realizado para todas as covariáveis relevantes e, quando possível, para a distribuição completa, porque médias iguais podem coexistir com caudas muito diferentes.

$$SMD(X) = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(s_T^2 + s_C^2)/2}}$$

8. VARIÁVEIS INSTRUMENTAIS

Variáveis instrumentais (IV) são utilizadas quando a variável de tratamento é endógena e existe uma fonte de variação externa capaz de afetar o tratamento sem influenciar diretamente o resultado por outros canais. O instrumento Z precisa ser relevante para explicar D e satisfazer uma restrição de exclusão: uma vez controladas as demais variáveis, Z deve afetar Y apenas por meio do tratamento.

$$D_i = \pi_0 + \pi_1 Z_i + X_i' \pi + v_i$$

$$Y_i = \beta_0 + \tau \hat{D}_i + X_i' \beta + \varepsilon_i$$

O estimador de dois estágios (2SLS) usa a parcela do tratamento induzida pelo instrumento. A validade não depende somente de correlação estatística entre Z e D . Um instrumento forte, mas diretamente relacionado ao resultado, produz estimativa viesada. Por isso, a justificativa institucional da exclusão é mais importante do que um p-valor.

8.1 Instrumentos fracos e interpretação local

Instrumentos fracos tornam a inferência instável e podem aproximar o estimador IV do OLS. Estatísticas do primeiro estágio e intervalos robustos a instrumentos fracos devem ser reportados. Sob monotonicidade e heterogeneidade de efeitos, IV identifica um efeito médio local do tratamento para unidades cuja participação é alterada pelo instrumento, os compliers. Esse efeito pode diferir do ATE e do ATT.

Em programas focalizados, possíveis instrumentos podem surgir de regras administrativas, variação exógena de elegibilidade ou implementação. Contudo, distância a postos, disponibilidade local ou capacidade municipal frequentemente afetam resultados por canais próprios, violando exclusão. A estratégia IV exige uma narrativa institucional particularmente cuidadosa.

8.2 Relevância, exclusão e interpretação local

Variáveis instrumentais são úteis quando o tratamento é endógeno e existe uma fonte de variação que altera a exposição sem afetar diretamente o resultado. Três condições merecem ser separadas. A relevância requer correlação entre instrumento e tratamento; a independência exige que o instrumento seja tão bom quanto aleatório, possivelmente condicionado a controles; e a restrição de exclusão determina que o instrumento afete o resultado somente por meio do tratamento. A última condição é substantiva e não pode ser validada apenas por testes estatísticos.

Em instrumentos binários e tratamento binário, sob monotonicidade, o estimador de Wald identifica um efeito médio local para os compliers, isto é, unidades cujo tratamento muda em resposta ao instrumento. Esse LATE pode diferir do efeito médio na população. A interpretação local não é uma deficiência do método, mas precisa ser explicitada. Em programas sociais, uma regra administrativa, um limiar de elegibilidade ou uma variação de acesso pode alterar a participação apenas para grupos específicos, de modo que extrapolar o resultado para todos os beneficiários seria inadequado.

Instrumentos fracos representam outro risco. Quando o primeiro estágio é pequeno, estimativas 2SLS podem ser imprecisas e apresentar viés relevante em amostras finitas. Estatísticas do primeiro estágio e intervalos robustos a instrumentos fracos devem ser relatados. Se houver múltiplos instrumentos, testes de sobreidentificação avaliam compatibilidade conjunta com as condições de momento, mas não demonstram a exclusão. Uma defesa convincente combina institucionalidade, gráficos, primeiros estágios transparentes e análises

de falsificação.

$$\tau_{LATE} = \frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{E[D \mid Z = 1] - E[D \mid Z = 0]}$$

9. SÉRIES TEMPORAIS INTERROMPIDAS

Séries temporais interrompidas (Interrupted Time Series, ITS) avaliam uma intervenção que ocorre em um momento definido e para a qual existe uma série suficientemente longa antes e depois. O método estima mudanças de nível e/ou de inclinação no ponto da intervenção. O contrafactual é a continuação da tendência prévia, ajustada por componentes sazonais e autocorrelação quando necessário.

$$Y_t = \beta_0 + \beta_1 tempo_t + \tau D_t + \delta(tempo_t \times D_t) + \varepsilon_t$$

O coeficiente τ mede uma mudança imediata de nível e δ mede a mudança na tendência após a intervenção. Se a política produz efeito gradual, a inclinação pode ser mais informativa que o salto. Uma ITS é mais convincente quando a data da intervenção é externa à série analisada e não coincide com outros choques capazes de alterar o resultado.

9.1 ITS controlada

A inclusão de uma série de controle não exposta à intervenção fortalece o desenho. O controle ajuda a descontar choques macroeconômicos, sazonalidade e mudanças de mensuração comuns. Em contexto estadual, pode-se utilizar uma variável de resultado relacionada, mas não afetada pelo programa, ou unidades com menor exposição, desde que a comparabilidade seja defendida. Essa lógica aproxima a ITS de um DiD em séries agregadas.

Diagnósticos de autocorrelação são indispensáveis. Erros Newey-West, modelos ARIMA com intervenção ou regressões segmentadas com erros autocorrelacionados podem ser usados. A escolha do ponto de quebra não deve ser feita apenas porque maximiza significância; quando a data é descoberta nos dados, a inferência precisa reconhecer o processo de seleção e buscar validação externa ou testes de robustez.

9.2 Nível, inclinação, sazonalidade e autocorrelação

A série temporal interrompida é apropriada quando uma intervenção ocorre em data bem definida e existe histórico suficiente antes e depois. O modelo segmentado separa mudança imediata de nível e mudança de inclinação. O primeiro parâmetro responde se a série sofreu um salto no ponto de intervenção; o segundo indica se a trajetória passou a crescer ou cair a ritmo diferente. Essa distinção evita interpretar uma alteração gradual como se fosse efeito instantâneo.

Como observações próximas no tempo são dependentes, erros independentes raramente são adequados. A inspeção da função de autocorrelação, testes de resíduos e erros HAC, como Newey-West, são alternativas simples; modelos ARIMA com intervenção podem ser úteis quando a dinâmica é mais complexa. Sazonalidade também precisa ser considerada em séries mensais. Se pagamentos, matrículas ou indicadores de saúde têm padrão sistemático por mês, dummies sazonais ou componentes harmônicos reduzem o risco de atribuir ao evento uma oscilação recorrente.

A principal ameaça à ITS é um evento concorrente na mesma data. Quando outra mudança econômica, administrativa ou legal coincide com a intervenção, a quebra não distingue suas causas. Uma série de controle não afetada ou um grupo comparável pode fortalecer a identificação, formando uma ITS controlada. Em estudos de repasses sociais, mudanças cadastrais, reajustes de benefício e choques macroeconômicos podem ocorrer em períodos próximos, motivo pelo qual a interpretação causal deve relacionar a data estimada a evidências institucionais independentes.

$$Y_t = \beta_0 + \beta_1 t + \tau D_t + \delta(t - c)D_t + \varepsilon_t$$

10. CONTROLE SINTÉTICO

O método de controle sintético é útil quando uma unidade ou pequeno conjunto de unidades recebe uma intervenção e existe um conjunto de potenciais controles não tratados. Em vez de escolher um único controle, o método combina ponderadamente várias unidades para reproduzir o comportamento pré-tratamento da unidade tratada. O contrafactual pós-intervenção é a trajetória da combinação ponderada.

$$Y_t^N = \sum_{j=1}^J w_j Y_{jt}, \quad w_j \geq 0, \quad \sum_{j=1}^J w_j = 1$$

Os pesos são escolhidos para aproximar resultados anteriores e covariáveis preditoras. A qualidade do ajuste pré-tratamento é central: se o sintético não reproduz bem a trajetória antes da política, a discrepância posterior é difícil de interpretar. O método é transparente porque os pesos mostram quais unidades compõem o contrafactual.

10.1 Inferência por placebos

Como o número de unidades tratadas costuma ser pequeno, a inferência frequentemente usa placebos: aplica-se artificialmente a mesma data de intervenção às unidades doadoras e compara-se a magnitude do gap observado. Placebos no tempo também ajudam a verificar se efeitos semelhantes apareceriam antes da política. O objetivo não é produzir um p-valor convencional a qualquer custo, mas mostrar se a trajetória tratada é excepcional em relação ao que ocorre quando o procedimento é repetido em unidades não tratadas.

Para mudanças nacionais do Bolsa Família, um controle sintético internacional exigiria dados comparáveis e hipóteses fortes. Entretanto, o método pode ser adaptado a políticas estaduais, expansões municipais ou regras diferenciadas. Sua principal contribuição didática neste livro é mostrar uma forma explícita de construir o contrafactual como combinação de trajetórias observadas.

10.2 Construção do contrafactual e inferência por placebos

O controle sintético foi desenvolvido para situações em que uma unidade agregada recebe uma intervenção e existe um conjunto de unidades doadoras não tratadas. Em vez de escolher um único controle, o método determina pesos não negativos que fazem uma combinação convexa dos doadores reproduzir, no período pré-tratamento, os preditores e a trajetória do resultado da unidade tratada. A qualidade do ajuste antes da intervenção é parte da credibilidade do desenho: se o contrafactual sintético não acompanha a unidade tratada no pré, a diferença posterior possui interpretação fraca.

A matriz V determina a importância relativa dos preditores na função de distância, enquanto W contém os pesos das unidades doadoras. Em aplicações transparentes, devem ser reportados os pesos, o ajuste pré-tratamento, a sensibilidade à exclusão de doadores e a razão entre erro pós e pré-tratamento. O método não elimina a necessidade de conhecimento substantivo; doadores expostos a políticas semelhantes, spillovers ou choques específicos podem contaminar o contrafactual e devem ser excluídos ou analisados separadamente.

A inferência costuma usar placebos espaciais e temporais. No placebo espacial, cada unidade doadora é tratada como se tivesse recebido a intervenção e sua diferença em relação ao respectivo sintético é comparada à diferença da unidade realmente tratada. Em amostras pequenas, essa distribuição fornece uma noção de quão incomum é o efeito observado. Placebos com ajuste pré muito pior que o da unidade tratada não são comparáveis e devem ser filtrados segundo regras definidas previamente. A robustez aumenta quando o resultado persiste sob

diferentes conjuntos de preditores e doadores.

$$W^* = \underset{W}{\operatorname{argmin}} (X_1 - X_0 W)' V (X_1 - X_0 W) \quad \text{sujeito a} \quad w_j \geq 0, \sum_j w_j = 1$$

11. REGRESSÃO DESCONTÍNUA: DESENHO, ESTIMAÇÃO E DIAGNÓSTICO

A regressão descontínua (RDD) explora uma regra de tratamento que muda de forma discreta quando uma variável de ordenação cruza um limiar. Em um desenho sharp, a probabilidade de tratamento salta de zero para um. Em um desenho fuzzy, a probabilidade muda de forma descontínua, mas não perfeitamente. A identificação decorre da suposição de que unidades muito próximas ao ponto de corte são comparáveis, exceto pela mudança na exposição ao tratamento.

$$Y_i = \alpha + \tau \mathbf{1}(X_i \geq c) + f(X_i - c) + \varepsilon_i$$

O parâmetro τ representa o salto local no resultado no ponto c . Estimação moderna privilegia polinômios locais de baixa ordem, kernels que dão maior peso a observações próximas e seleção de largura de banda orientada por erro quadrático médio. Polinômios globais de ordem alta podem gerar oscilações artificiais e devem ser usados com cautela.

11.1 RDD temporal e pontos de corte endógenos

Quando a variável de ordenação é o tempo, a interpretação requer cuidado adicional. Observações imediatamente antes e depois de uma data podem diferir devido a eventos simultâneos, tendências ou sazonalidade. Se o ponto de corte é escolhido pelo próprio algoritmo com base na significância do salto, existe risco de seleção pós-hoc. Uma estratégia mais rigorosa combina detecção de quebras com validação institucional, testes em datas placebo e modelos que controlem tendência e sazonalidade.

11.2 Bandwidth, kernel e polinômios locais

A largura de banda determina quão local é a comparação. Bandas estreitas reduzem viés, mas aumentam variância; bandas largas aumentam precisão e risco de extrapolação. Resultados devem ser reportados para bandas alternativas. O kernel triangular é comum porque atribui maior peso às observações próximas ao corte. A ordem do polinômio também deve ser variada em análises de sensibilidade, com preferência por especificações simples.

11.3 Testes de manipulação e continuidade

O teste de McCrary examina descontinuidade na densidade da variável de ordenação e é particularmente relevante quando agentes podem manipular sua posição em torno do limiar. Em RDD temporal, a data em si não é manipulável por unidades da mesma forma, mas testes de composição, continuidade de covariáveis e inspeção de eventos concorrentes são essenciais. A lógica permanece: nada além do tratamento deveria mudar discretamente no corte.

Placebos são uma ferramenta poderosa. Estimar falsos cortes antes e depois do evento, usar resultados que não deveriam responder à política e verificar covariáveis predeterminadas ajudam a avaliar se o salto é específico. A robustez não deve ser resumida a uma coleção de p-valores; ela deve mostrar que a conclusão resiste a mudanças razoáveis na especificação e falha quando o desenho deveria falhar.

11.4 RDD local, largura de banda e interpretação do salto

A regressão descontínua explora uma regra em que a probabilidade de tratamento muda de forma descontínua em um limiar conhecido. Em um desenho sharp, o tratamento muda deterministicamente no corte; em um desenho fuzzy, a elegibilidade altera a probabilidade de tratamento e o efeito é obtido por uma razão do salto no resultado pelo salto na participação. A identificação é local: unidades imediatamente abaixo e acima do limiar são consideradas

comparáveis porque pequenas diferenças na running variable não deveriam gerar uma mudança abrupta no resultado potencial na ausência da política.

A estimação local linear evita extrapolar polinômios de alta ordem para regiões distantes do corte. Uma largura de banda h define quais observações entram na regressão, e kernels podem atribuir maior peso às unidades mais próximas. Bandwidths muito pequenos reduzem viés, porém aumentam variância; bandwidths grandes usam mais dados, mas tornam a comparação menos local. Procedimentos como os de Calonico, Cattaneo e Titiunik ajustam viés e fornecem inferência robusta, devendo ser acompanhados de especificações alternativas de h e gráficos que mostrem a forma dos dados de ambos os lados.

O pressuposto de continuidade dos resultados potenciais no limiar não é diretamente testável, mas suas implicações podem ser investigadas. Covariáveis pré-tratamento não deveriam apresentar saltos sistemáticos; a densidade da running variable não deveria indicar manipulação precisa; e falsos pontos de corte não deveriam reproduzir o efeito principal. O teste de McCrary é especialmente conhecido para detectar descontinuidade na densidade, mas sua interpretação em RDD temporal exige cautela, porque o tempo não pode ser manipulado pelas unidades. Em desenhos temporais, o desafio central é excluir choques simultâneos e modelar a dependência serial.

No estudo empírico do Bolsa Família, a busca endógena por datas de quebra funciona como mecanismo de detecção de descontinuidades na série, mas esse procedimento não é equivalente a uma regra de elegibilidade clássica. Por isso, a evidência deve ser lida como uma combinação entre detecção de mudanças e estimação local ao redor das datas encontradas. A interpretação causal fica mais forte quando a data coincide com uma mudança institucional bem documentada, quando os resultados resistem a janelas alternativas e quando não existem outros eventos plausíveis capazes de gerar o mesmo salto.

$$Y_i = \alpha + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + \varepsilon_i, \quad |X_i - c| \leq h$$

11.5 Condição de continuidade

A condição de identificação da RDD pode ser expressa como continuidade do resultado potencial sem tratamento no ponto de corte. Intuitivamente, se nada além da regra de tratamento muda de forma descontínua em c , a diferença entre os limites laterais do resultado observado pode ser atribuída ao tratamento. Em aplicações reais, essa afirmação deve ser sustentada por desenho institucional, covariáveis, testes de densidade e análise de eventos concorrentes.

$$\lim_{x \uparrow c} E[Y_i(0) | X_i = x] = \lim_{x \downarrow c} E[Y_i(0) | X_i = x]$$

12. QUEBRAS ESTRUTURAIS, PELT, CHOW E BAI-PERRON

Testes de quebra estrutural investigam se os parâmetros de uma série ou regressão mudam em determinados pontos. O teste de Chow é adequado quando a data da quebra é conhecida a priori. Ele compara o ajuste de uma regressão única com o ajuste de regressões separadas antes e depois do ponto. A hipótese nula é estabilidade dos coeficientes.

Quando a data é desconhecida, procedimentos de múltiplas quebras, como Bai-Perron, procuram segmentos em que coeficientes permanecem relativamente estáveis. O PELT, por sua vez, é um algoritmo eficiente de detecção de pontos de mudança que minimiza uma função de custo com penalização pelo número de quebras. Critérios como MBIC evitam que pequenas flutuações sejam tratadas como mudanças estruturais relevantes.

$$\min_{\tau_1, \dots, \tau_m} \left\{ \sum_{j=0}^m C(y_{\tau_j+1:\tau_{j+1}}) + \beta m \right\}$$

12.1 Quebra estatística não é automaticamente evento causal

Uma quebra detectada estatisticamente indica alteração de padrão, mas não identifica sua causa. A associação com uma mudança institucional exige triangulação: calendário normativo, evidência administrativa, sinais coerentes em variáveis relacionadas e ausência de eventos alternativos dominantes. No estudo do Bolsa Família, a coincidência entre quebras e momentos de expansão ou contração deve ser apresentada como evidência contextual, não como prova isolada de mecanismo.

A combinação de PELT, Chow e RDD é metodologicamente interessante porque separa tarefas. A detecção de mudança procura localizar rupturas; o teste de estabilidade confirma alterações em parâmetros; a RDD estima o salto local em torno da data. Quanto mais endógena for a seleção da data, maior deve ser o esforço de validação fora da mesma amostra que gerou o candidato.

12.2 Validação de datas e multiplicidade de quebras

Testes de quebra estrutural respondem a uma pergunta diferente da avaliação causal: eles procuram determinar se parâmetros estatísticos mudam ao longo da série. O teste de Chow compara o ajuste de uma regressão única com o de regressões estimadas em subamostras separadas por uma data conhecida. Quando a data é escolhida após inspeção dos dados, o p-valor convencional pode subestimar a incerteza da busca. Métodos de múltiplas quebras, como Bai-Perron, tratam explicitamente a possibilidade de vários regimes e estimam conjuntamente posições que minimizam a soma dos quadrados residuais sujeita a comprimentos mínimos de segmento.

O PELT adota uma lógica de otimização penalizada e pode localizar múltiplas mudanças com custo computacional eficiente. A penalidade controla o equilíbrio entre ajuste e complexidade: penalidades pequenas produzem muitas quebras; penalidades grandes preservam poucos regimes. Critérios como MBIC ajudam a disciplinar a seleção, mas não substituem validação substantiva. Em políticas públicas, uma quebra estatística precisa ser confrontada com calendário institucional, regras de benefício, revisão cadastral, execução orçamentária e choques macroeconômicos.

Combinar métodos pode ser produtivo. Uma data sugerida pelo PELT pode ser examinada por Chow, RDD local, ITS e placebos temporais. Se diferentes procedimentos, com pressupostos distintos, identificam uma mudança semelhante e a documentação institucional aponta para um evento compatível, a evidência de mudança estrutural se fortalece. Ainda assim,

convergência entre métodos não transforma automaticamente associação temporal em causalidade. O passo decisivo continua sendo demonstrar por que outras causas não explicam a ruptura.

$$F = \frac{(SSR_p - SSR_1 - SSR_2)/k}{(SSR_1 + SSR_2)/(n_1 + n_2 - 2k)}$$

13. GMM EM PAINEL DINÂMICO

Modelos dinâmicos incluem a variável dependente defasada entre os regressores. Em painéis com efeitos fixos e horizonte temporal moderado, OLS e o estimador within podem ser viesados porque a defasagem do resultado se correlaciona com o erro transformado. O Generalized Method of Moments (GMM) usa condições de momento e instrumentos internos, como defasagens da própria variável, para obter estimação consistente sob hipóteses específicas.

$$Y_{it} = \rho Y_{i,t-1} + \beta X_{it} + \alpha_i + \varepsilon_{it}$$

No Difference-GMM de Arellano e Bond, a equação é diferenciada para remover α_i , e defasagens em nível são usadas como instrumentos das variáveis endógenas em diferenças. O System-GMM acrescenta equações em nível instrumentadas por diferenças, podendo aumentar eficiência quando as séries são persistentes. A validade depende da ausência de correlação serial de segunda ordem no erro em diferenças e da validade dos instrumentos.

13.1 Diagnósticos e proliferação de instrumentos

Os testes AR(1) e AR(2) são fundamentais. Correlação de primeira ordem nos resíduos diferenciados é esperada; correlação de segunda ordem compromete instrumentos baseados em defasagens. Testes de Sargan ou Hansen avaliam restrições de sobreidentificação, mas podem perder poder quando há instrumentos demais. A proliferação de instrumentos é um dos principais riscos do GMM aplicado.

Para reduzir esse problema, o pesquisador pode limitar o intervalo de defasagens, colapsar a matriz de instrumentos e comparar especificações mais parcimoniosas. O número de instrumentos deve ser reportado e avaliado em relação ao número de grupos. Coeficientes muito sensíveis a pequenas alterações na matriz de instrumentos indicam fragilidade.

13.2 GMM como robustez, não como substituto do desenho

Em avaliação de políticas, GMM pode tratar endogeneidade dinâmica e heterogeneidade não observada, mas não substitui uma fonte clara de variação exógena quando a pergunta é causal. Um painel dinâmico robusto pode confirmar que a relação persiste sob outra estrutura econométrica; ainda assim, a interpretação depende das condições de momento e da plausibilidade de que as defasagens sejam instrumentos válidos. O método deve complementar, e não encobrir, a lógica de identificação.

13.3 Condições de momento, instrumentos e diagnóstico

Painéis dinâmicos incluem defasagens do resultado entre os regressores, o que torna o estimador de efeitos fixos enviesado em T pequeno porque a variável defasada permanece correlacionada com o erro transformado. O Difference-GMM de Arellano e Bond elimina o efeito fixo por primeira diferença e utiliza níveis defasados como instrumentos, desde que as condições de momento sejam válidas. O System-GMM acrescenta equações em nível instrumentadas por diferenças, podendo melhorar eficiência quando as séries são persistentes, mas exige condições adicionais sobre as correlações iniciais.

A validade empírica depende menos do número de instrumentos do que de sua qualidade. Usar todas as defasagens disponíveis pode gerar proliferação de instrumentos, sobreajustar as variáveis endógenas e enfraquecer testes de Hansen. Limitar janelas de defasagem e colapsar a matriz de instrumentos são práticas recomendadas. O pesquisador deve relatar explicitamente número de instrumentos, número de grupos, AR(1), AR(2), Hansen ou Sargan, além de justificar quais variáveis são tratadas como exógenas, predeterminadas ou endógenas.

No Difference-GMM, autocorrelação de primeira ordem nos resíduos diferenciados é

esperada, enquanto autocorrelação de segunda ordem ameaça a validade dos instrumentos em níveis defasados. O teste AR(2), portanto, é central. Já o teste de Hansen avalia se o conjunto de condições de momento é compatível com os dados, mas p-valores muito altos também podem sinalizar baixa potência causada por instrumentos em excesso. GMM deve ser tratado como parte de uma estratégia de robustez e dinâmica, não como substituto de um desenho de identificação bem fundamentado.

$$\Delta Y_{it} = \rho \Delta Y_{i,t-1} + \beta \Delta X_{it} + \Delta \varepsilon_{it}$$

13.4 Condições de momento e teste J

A lógica do GMM pode ser resumida pelas condições de momento: instrumentos válidos devem ser ortogonais ao erro. O estimador escolhe os parâmetros de forma a tornar os momentos amostrais tão próximos de zero quanto possível, ponderando-os por uma matriz que, em versões eficientes, aproxima a inversa da variância dos momentos. A estatística J avalia a magnitude conjunta dessas violações quando há mais instrumentos do que parâmetros identificados. Sua utilidade depende de um conjunto de instrumentos parcimonioso e substantivamente defensável.

$$E[Z'_{it} \Delta \varepsilon_{it}] = 0$$

$$J = \hat{g}(\hat{\theta})' \hat{W} \hat{g}(\hat{\theta})$$

PARTE II - REVISÃO DA LITERATURA: BOLSA FAMÍLIA E AVALIAÇÃO QUANTITATIVA

14. TRANSFERÊNCIAS CONDICIONADAS DE RENDA E O PROBLEMA DE AVALIAÇÃO

O Programa Bolsa Família ocupa posição central na literatura brasileira de avaliação porque combina grande escala, focalização em famílias de baixa renda, condicionalidades associadas à educação e à saúde e ampla capilaridade territorial. Essas características geram perguntas distintas. Uma primeira família de estudos investiga focalização e incidência distributiva: quem recebe, quanto recebe e como a transferência altera pobreza e desigualdade. Uma segunda pergunta é comportamental e procura saber se o programa modifica matrícula, frequência, progressão escolar, utilização de serviços de saúde, consumo de alimentos ou oferta de trabalho. Uma terceira linha analisa efeitos agregados sobre municípios e estados, incluindo atividade econômica, emprego e indicadores de saúde pública. Cada pergunta exige unidade de análise, contrafactual e estratégia de identificação próprios.

A literatura internacional sobre transferências condicionadas de renda mostrou cedo que a simples comparação entre beneficiários e não beneficiários é inadequada, porque os programas são desenhados justamente para selecionar famílias mais pobres. No Brasil, esse problema é agravado por diferenças regionais persistentes e por heterogeneidade na capacidade administrativa dos municípios. Estudos que utilizam microdados domiciliares tendem a aplicar matching, ponderação, regressões com controles, variáveis instrumentais ou descontinuidades de regras de elegibilidade. Trabalhos com escolas e municípios exploram adoção temporal, intensidade de cobertura e painéis com efeitos fixos. A diversidade metodológica é uma vantagem porque permite observar o programa por ângulos distintos, mas também significa que coeficientes não são diretamente comparáveis sem considerar população, período e parâmetro estimado.

Ribeiro, Shikida e Hillbrecht (2017), em uma revisão extensa da literatura acumulada até meados da década de 2010, organizaram evidências sobre focalização, consumo, educação, saúde, oferta de trabalho e outros possíveis efeitos. Os autores destacam que alguns resultados aparecem de maneira recorrente, como melhora em indicadores de participação escolar e expansão do consumo de alimentos, ao passo que outras dimensões permaneciam mais incertas. Essa revisão é importante metodologicamente porque evidencia que a conclusão sobre um programa não deve ser resumida por um único desfecho: a mesma transferência pode melhorar frequência escolar, não alterar aprendizagem no curto prazo e produzir respostas diferentes na alocação de trabalho dentro do domicílio.

A revisão de escopo de Magalhães et al. (2024) amplia a perspectiva ao reunir 94 estudos sobre impactos econômicos, sociais e de saúde publicados desde a criação do programa até 2021. A síntese identifica evidência consistente de redução da pobreza e resultados favoráveis em vários indicadores de saúde e educação, mas também aponta desfechos com base empírica mais estreita ou resultados mistos. A existência de uma literatura numerosa não elimina a necessidade de avaliar desenho e qualidade de cada estudo. Revisões agregam evidências produzidas com estratégias diferentes; portanto, a força de uma conclusão depende da convergência entre métodos plausíveis e da consistência dos resultados em períodos, populações e bases de dados distintas.

Para um livro de métodos quantitativos, o Bolsa Família é particularmente instrutivo porque oferece exemplos reais de quase todas as dificuldades clássicas de avaliação. Há seleção por renda, mensuração imperfeita da renda autodeclarada, mudanças de regras, alterações de benefício, cobertura heterogênea, potenciais spillovers, outcomes em frequências variadas e choques macroeconômicos simultâneos. Em vez de buscar um estimador universal, a literatura mostra que a estratégia mais produtiva é alinhar método e pergunta. Uma RDD baseada em uma

regra de elegibilidade identifica efeito local; um painel municipal pode captar respostas agregadas; PSM aproxima famílias observacionalmente semelhantes; DiD explora variação temporal relativa; e modelos de decomposição medem contribuição contábil para pobreza e desigualdade sem necessariamente constituir avaliação causal.

15. POBREZA, DESIGUALDADE E FOCALIZAÇÃO

A relação entre Bolsa Família, pobreza e desigualdade é uma das áreas em que a literatura apresenta maior acumulação. Trabalhos de decomposição da renda mostram que uma transferência pode ter impacto distributivo superior à sua participação no total da renda quando é fortemente concentrada na base da distribuição. Esse mecanismo ajuda a explicar por que o orçamento do programa, relativamente pequeno em proporção ao PIB, pode gerar contribuição mensurável para a redução da pobreza extrema e do índice de Gini. Souza, Osorio, Paiva e Soares (2019), utilizando PNAD e PNAD Contínua de 2001 a 2017, reavaliam os primeiros quinze anos do programa e enfatizam justamente a combinação entre boa focalização e capacidade redistributiva.

Segundo Souza et al. (2019), as transferências associadas ao Bolsa Família são fortemente progressivas e contribuem para reduzir pobreza e extrema pobreza. A interpretação deve distinguir efeito aritmético imediato de efeito comportamental de longo prazo. Quando se compara a renda domiciliar antes e depois da transferência, mede-se quanto o benefício desloca famílias em relação às linhas de pobreza e como altera a distribuição observada. Essa decomposição é informativa para incidência e redistribuição, mas não incorpora necessariamente respostas de trabalho, composição familiar ou preços locais. Por isso, estudos de impacto comportamental complementam, e não substituem, a análise distributiva.

A literatura anterior já havia documentado a relevância das transferências focalizadas no processo de queda da desigualdade brasileira dos anos 2000. Barros et al. (2006), Soares e colaboradores e os volumes organizados pelo Ipea discutiram o papel relativo de mercado de trabalho, previdência, Benefício de Prestação Continuada e transferências condicionadas. A conclusão recorrente é que a renda do trabalho explicou parcela importante da queda do Gini, enquanto o Bolsa Família apresentou elevada efetividade distributiva por unidade de recurso devido à focalização. Essa distinção evita atribuir ao programa toda a mudança observada na desigualdade nacional e, ao mesmo tempo, reconhece que sua contribuição marginal pode ser expressiva apesar do peso orçamentário limitado.

Focalização, contudo, não é uma propriedade imutável. Erros de inclusão e exclusão dependem da qualidade do cadastro, da atualização de renda e das regras de elegibilidade. Corrêa et al. (2023) discutem critérios de elegibilidade e mostram como modelos baseados em características observáveis podem ser utilizados para avaliar cobertura e vazamento do programa. Do ponto de vista metodológico, a renda autodeclarada é uma variável suscetível a erro de mensuração e volatilidade. Famílias próximas ao limiar podem alternar elegibilidade sem que sua condição estrutural mude de forma equivalente, o que exige cuidado ao interpretar avaliações que usam somente a renda corrente como medida de vulnerabilidade.

A auditoria do Tribunal de Contas da União sobre o desenho do novo Bolsa Família em 2023 acrescenta uma dimensão institucional recente. O TCU avaliou focalização, equidade per capita, cobertura regional e custo-efetividade, identificando tensões entre benefício mínimo por família e componentes pagos por pessoa. Esse tipo de análise não é substituto de uma avaliação causal, porque trabalha também com simulações, indicadores administrativos e critérios de desenho, mas oferece informações importantes sobre mecanismos distributivos. Para pesquisa quantitativa, relatórios de auditoria podem auxiliar na formulação de hipóteses, na identificação de mudanças de regra e na escolha de variáveis de heterogeneidade.

O problema de focalização também interfere na escolha do contrafactual. Beneficiários e não beneficiários podem diferir substancialmente em renda, número de crianças, localização, escolaridade e inserção ocupacional. Regressões sem boa sobreposição extrapolam para regiões do suporte em que quase não existem controles comparáveis. O matching procura reduzir esse

problema em observáveis, enquanto regras de corte podem produzir RDD local. Painéis agregados contornam parte da seleção individual, mas introduzem outros desafios, como composição populacional e correlação entre cobertura e condições econômicas locais. Assim, o desenho que melhor responde à pergunta distributiva depende do nível em que se mede a política.

O estudo empírico que originou este livro se insere nessa literatura por uma pergunta distinta: em vez de estimar somente quanto o programa reduz o Gini, investiga se unidades federativas com maior desigualdade tendem a receber repasses maiores e como essa relação se comporta em torno de mudanças temporais. A abordagem utiliza Gini estadual, repasses reais e modelos locais e de painel. A comparação com a literatura de focalização é útil, mas deve preservar a diferença entre focalização em famílias e associação entre agregados estaduais. Um estado pode receber volume maior porque possui população maior, maior número de elegíveis ou benefício médio distinto; por isso, controles, escalas per capita e análises de robustez são dimensões naturais para extensões futuras.

16. EDUCAÇÃO, SAÚDE E NUTRIÇÃO

As condicionalidades educacionais fizeram da escolarização um dos desfechos mais estudados. Glewwe e Kassouf (2012) exploram dados do Censo Escolar e diferenças na adoção do Bolsa Escola/Bolsa Família para comparar mudanças em matrícula, abandono e progressão. Os resultados indicam aumento de matrícula, redução de abandono e melhora na promoção escolar. O valor metodológico do estudo está no uso de painel de escolas e na exploração de variação temporal na presença do programa, em vez de comparar apenas alunos beneficiários e não beneficiários em um único momento.

Resultados posteriores mostram que efeitos educacionais são heterogêneos. Estudos com dados domiciliares e métodos de ponderação por propensity score encontraram impactos mais fortes para determinados grupos, especialmente meninas, adolescentes e áreas rurais em algumas amostras. O trabalho publicado em *World Development* em 2015, por exemplo, documenta diferenças por sexo, idade e localização na participação escolar e progressão. Heterogeneidade desse tipo é substantivamente importante: uma média nacional pequena pode combinar efeitos relevantes em grupos com maior risco de evasão e efeitos próximos de zero em grupos que já apresentavam frequência elevada.

A literatura também distingue frequência e aprendizagem. A revisão sistemática de Santos et al. (2019) encontrou evidências de melhora em frequência e redução do abandono, mas resultados menos claros para proficiência e desempenho acadêmico. Essa diferença é coerente com o mecanismo institucional: a condicionalidade atua diretamente sobre presença, enquanto aprendizagem depende da qualidade da escola, recursos pedagógicos, ambiente familiar e tempo de exposição. Avaliações que usam nota em teste como desfecho precisam lidar ainda com seleção de quem realiza o exame e com diferenças de composição entre escolas.

Trabalhos mais recentes continuam examinando desempenho. Evidência publicada em 2023 com Prova Brasil, Censo Escolar e Sistema Presença utiliza propensity score matching para comparar alunos beneficiários com estudantes economicamente desfavorecidos não beneficiários e encontra associação positiva, porém pequena, com notas de língua e matemática. A magnitude reduzida reforça a importância de separar significância estatística de relevância substantiva. Grandes bases podem detectar diferenças muito pequenas; cabe ao pesquisador expressar o efeito em desvios-padrão ou unidades interpretáveis e discutir se a mudança é educacionalmente relevante.

Na saúde, um dos estudos mais conhecidos é Rasella et al. (2013), que utiliza painel de municípios para analisar mortalidade infantil e condições específicas associadas à pobreza. Os autores relacionam expansão da cobertura do programa a reduções em mortalidade, especialmente por causas sensíveis a condições socioeconômicas e acesso a cuidados. O desenho incorpora controles municipais e estrutura longitudinal, mas, como em todo estudo ecológico, o parâmetro se refere a efeitos agregados da cobertura e não pode ser convertido diretamente em efeito individual para cada família beneficiária.

A literatura sobre nutrição e segurança alimentar é mais heterogênea. Revisões sistemáticas encontraram aumento na aquisição de alimentos e indícios de melhora em segurança alimentar, mas resultados de estado nutricional variam conforme idade, indicador e desenho. Isso é plausível porque renda adicional pode elevar quantidade consumida sem garantir qualidade da dieta; preços locais, oferta de alimentos e preferências domésticas mediam o efeito. Para avaliação, é importante especificar se o outcome é gasto, diversidade alimentar, insegurança alimentar, antropometria ou micronutrientes, pois cada medida responde a um estágio diferente do mecanismo.

A revisão de escopo de Magalhães et al. (2024) sintetiza evidências de saúde, pobreza,

educação e outros domínios, destacando resultados favoráveis em mortalidade infantil e adulta e em vários indicadores de utilização e proteção social. Ao mesmo tempo, a revisão registra áreas em que poucos estudos sustentam conclusões ou em que os resultados são mistos. Esse padrão ilustra um princípio de meta-ciência aplicado à avaliação: quantidade de publicações não equivale a certeza homogênea. Desfechos amplamente replicados com desenhos distintos merecem confiança maior do que resultados baseados em poucas amostras ou especificações muito particulares.

Para pesquisas futuras, a integração de bases administrativas oferece oportunidades e riscos. Vincular Cadastro Único, pagamentos, presença escolar, registros de saúde e emprego formal pode reduzir erro de memória e ampliar tamanho amostral, mas exige atenção a mudanças de sistemas, chaves de ligação e cobertura diferencial. O fato de uma variável estar em um cadastro não garante comparabilidade ao longo do tempo. Auditorias de qualidade, documentação de versões e análise de missingness tornam-se parte do desenho causal, sobretudo quando a própria política altera a probabilidade de uma pessoa aparecer no sistema.

17. MERCADO DE TRABALHO, CONSUMO E EFEITOS ECONÔMICOS LOCAIS

Uma crítica recorrente a programas de transferência é a possibilidade de desincentivo ao trabalho. A literatura brasileira, porém, não sustenta uma conclusão simples de redução generalizada da oferta laboral. Foguel e Barros (2010), utilizando painel de municípios, analisam oferta de trabalho adulta e encontram efeitos limitados em diversas margens. Estudos posteriores com microdados exploram horas, participação, formalização e alocação entre membros do domicílio. A ausência de grande efeito sobre participação agregada não implica ausência de respostas: famílias podem reorganizar setores, horas e responsabilidades sem reduzir o total de trabalho.

De Brauw et al. (2015) utilizam regressão ponderada por propensity score e encontram pouca evidência de desincentivo sobre a oferta de trabalho agregada do domicílio, mas identificam realocação entre trabalho formal e informal em áreas urbanas e mudanças na distribuição de horas entre homens e mulheres em áreas rurais. A interpretação proposta relaciona parte da resposta ao uso da renda formal nos critérios de elegibilidade. O estudo é um exemplo importante de como um programa pode afetar composição ocupacional mesmo quando o efeito sobre horas totais é pequeno.

Barbosa e Corseuil (2014) exploram uma descontinuidade relacionada à idade das crianças para identificar efeito local do Bolsa Família sobre escolha entre ocupações formais e informais. Os resultados não indicam alteração estatisticamente relevante na escolha ocupacional dos adultos no entorno da regra utilizada. A diferença entre esse resultado e outros trabalhos não deve ser tratada como contradição automática. RDD identifica uma população local de compliers próxima ao limiar, enquanto PSM e painéis podem estimar parâmetros para grupos mais amplos. População, período e definição do outcome mudam o objeto estimado.

Estudos sobre consumo mostram que a transferência aumenta a capacidade de aquisição das famílias, com evidências frequentes de maior gasto em alimentos. O desafio empírico é separar efeito renda de mudanças simultâneas em preços, composição do domicílio ou oferta local. Pesquisas com orçamento familiar e surveys específicos permitem examinar categorias de gasto, mas normalmente observam beneficiários em condições socioeconômicas distintas dos não beneficiários. Matching, ponderação e controles ricos reduzem diferenças observáveis; instrumentos ou regras de elegibilidade podem produzir identificação mais forte quando disponíveis.

Uma linha mais recente desloca a unidade de análise da família para a economia local. Transferências recebidas por domicílios podem gerar demanda em comércio e serviços e, por esse canal, afetar renda e produção de não beneficiários. Esses spillovers violam a hipótese de não interferência usada implicitamente em muitas avaliações individuais, porque o tratamento de uma família pode afetar o outcome de outra. Em áreas com alta cobertura, comparar beneficiários a não beneficiários do mesmo município pode, portanto, subestimar o efeito total se ambos se beneficiam do aumento da demanda local.

Mendes, Miyamoto, Nguyen, Pennings e coautores, em trabalho recente sobre efeitos macroeconômicos das transferências no Brasil, exploram variação do Bolsa Família entre estados por uma estratégia instrumental do tipo Bartik. O estudo reporta respostas de PIB estadual e composição do emprego e representa uma mudança de escala analítica: a pergunta deixa de ser apenas o efeito sobre famílias elegíveis e passa a incluir multiplicadores locais. Como se trata de working paper revisado, seus resultados devem ser lidos no contexto da estratégia de identificação e da discussão em andamento, mas o desenho mostra como choques distributivos podem ser estudados também em nível macroregional.

A coexistência de efeitos micro e macro reforça a necessidade de especificar o estimand. Uma política pode não reduzir oferta de trabalho individual e, ao mesmo tempo, alterar a composição setorial do emprego; pode aumentar consumo dos beneficiários e também vendas de estabelecimentos locais; pode reduzir pobreza diretamente e ainda produzir efeitos indiretos por preços e atividade econômica. Avaliações que ignoram spillovers estimam um efeito parcial, enquanto modelos espaciais, desenhos com saturação ou instrumentos de exposição podem capturar dimensões adicionais. Nenhum estimador isolado resume todos esses mecanismos.

18. SÍNTESE METODOLÓGICA DA LITERATURA E LACUNAS DE PESQUISA

A literatura do Bolsa Família evoluiu paralelamente ao desenvolvimento da econometria de avaliação. Trabalhos iniciais frequentemente combinaram regressões, decomposições e matching; depois ganharam espaço desenhos que exploram variação temporal, regras administrativas e painéis de grande dimensão. Essa evolução não significa que métodos recentes tornem os anteriores inúteis. Decomposições continuam apropriadas para medir incidência distributiva; regressões descritivas ajudam a caracterizar heterogeneidade; PSM é útil quando a seleção em observáveis é defensável; e RDD oferece forte validade interna quando existe um limiar genuíno. A questão é evitar que o nome do método seja confundido com a força da identificação.

Uma lacuna recorrente é a distância entre elegibilidade, recebimento e intensidade do benefício. O fato de uma família ser elegível não garante que receba; receber não implica valor idêntico; e o valor pode mudar com composição familiar ou regra vigente. Desenhos sharp podem ser inadequados quando a adesão é imperfeita, situação em que uma RDD fuzzy ou variável instrumental representa melhor o mecanismo. Em painéis agregados, cobertura percentual e valor per capita podem conter informação distinta do repasse total. Testar diferentes medidas de exposição ajuda a separar escala populacional de intensidade programática.

Outra lacuna é a heterogeneidade. Médias nacionais podem ocultar respostas diferentes por sexo, idade, região, urbanização, composição familiar, situação ocupacional e nível inicial de pobreza. A literatura educacional mostra isso claramente. Entretanto, explorar dezenas de subgrupos após observar os resultados eleva o risco de descobertas ocasionais. Análises heterogêneas mais convincentes especificam hipóteses antes da estimação, corrigem multiplicidade quando necessário e relacionam as diferenças a mecanismos observáveis.

A mensuração temporal também merece atenção. Indicadores anuais de desigualdade não contêm automaticamente variação mensal. Métodos de desagregação, como Denton-Cholette, preservam coerência com o agregado, mas geram uma trajetória suavizada construída matematicamente. Quando essa série é usada como covariável em modelos mensais, a interpretação deve reconhecer que a alta frequência foi interpolada. Uma alternativa é trabalhar em frequência anual, incorporar indicadores mensais auxiliares ou realizar análise de sensibilidade comparando especificações. O livro mantém a desagregação do artigo original, mas explicita essa limitação para que o leitor diferencie informação observada de informação distribuída por procedimento estatístico.

Mudanças institucionais oferecem oportunidades quase experimentais, mas também risco de confusão. Reajustes, revisões cadastrais, transições entre programas e crises econômicas podem ocorrer em períodos próximos. Uma quebra encontrada por PELT ou uma descontinuidade temporal não atribui sozinha causalidade a um evento. A evidência melhora quando a data é documentada, outcomes placebo permanecem contínuos, resultados são estáveis a janelas alternativas e um grupo de controle não exposto apresenta trajetória semelhante antes da mudança.

A literatura recente de DiD também traz implicações para avaliações subnacionais. Se municípios ou estados aumentam cobertura em momentos diferentes, uma especificação de efeitos fixos de duas vias pode misturar comparações inadequadas quando os efeitos são heterogêneos. Estimadores por coorte e tempo, como Callaway-Sant'Anna ou Sun-Abraham, tornam a composição das comparações explícita. Para políticas federais, ainda é necessário identificar de onde vem a heterogeneidade de exposição: capacidade de implementação,

estrutura demográfica, linha de elegibilidade, intensidade de expansão ou outro mecanismo.

A crescente disponibilidade de dados administrativos permite combinações metodológicas. Um caminho promissor é parear unidades em covariáveis pré-tratamento e depois aplicar DiD; outro é utilizar RDD para identificar efeito local e painéis para examinar validade externa; controle sintético pode estudar mudanças concentradas em uma unidade enquanto estudos de evento avaliam adoção escalonada em várias. Triangulação não significa escolher o resultado mais favorável, mas comparar estimandos e verificar se conclusões qualitativas resistem a hipóteses diferentes.

Por fim, transparência e replicabilidade tornaram-se parte da qualidade metodológica. Um estudo deve documentar fontes, transformações, critérios de inclusão, código, especificações principais e alternativas. Tabelas precisam informar erros-padrão, clusters, efeitos fixos, número de observações e instrumentos quando aplicável. Gráficos de RDD devem mostrar dados e ajuste; PSM deve mostrar suporte e balanceamento; DiD deve apresentar trajetória pré; GMM deve relatar testes de autocorrelação e sobreidentificação. Essas práticas não garantem resultado correto, mas tornam visível a cadeia de decisões que liga dados brutos à conclusão.

Quadro 1 - Estudos selecionados sobre o Bolsa Família e suas estratégias quantitativas

Estudo	Estratégia	Desfecho	Síntese
Souza et al. (2019)	Decomposição/distribuição de renda	Pobreza e desigualdade	Alta progressividade e contribuição relevante para redução de pobreza e desigualdade.
Glewwe e Kassouf (2012)	Painel de escolas / adoção temporal	Matrícula, abandono e promoção	Melhoras em matrícula, abandono e progressão escolar.
Rasella et al. (2013)	Painel municipal	Mortalidade infantil	Associação da expansão do programa com redução de mortalidade, sobretudo em causas relacionadas à pobreza.
de Brauw et al. (2015)	Ponderação por propensity score	Oferta e alocação de trabalho	Pouco efeito no total de trabalho, com realocações entre setores e membros do domicílio.
Barbosa e Corseuil (2014)	RDD / elegibilidade por idade	Formalidade e escolha ocupacional	Sem evidência de mudança relevante na escolha formal/informal no efeito local estudado.
Santos et al. (2019)	Revisão sistemática	Educação	Evidência mais consistente para frequência e abandono do que para aprendizagem.
Magalhães et al. (2024)	Scoping review	Pobreza, saúde, educação e outros	Síntese ampla de 94 estudos, com resultados positivos em várias dimensões e lacunas em outras.
TCU (2024)	Auditoria/indicadores administrativos	Focalização e equidade	Identificação de trade-offs de desenho e problemas de cobertura/dados no período auditado.

Fonte: elaboração do autor com base na literatura citada.

Quadro 2 - Relação entre desenho, hipótese de identificação e principal ameaça

Método	Hipótese-chave	Estimando típico	Ameaça principal
Diferença de médias / OLS	Comparabilidade condicional	Associação ajustada; benchmark	Confusão por não observáveis.
PSM / weighting	Seleção em observáveis + suporte comum	ATT ou efeito na população sobreposta	Não corrige fatores não observados.
DiD / event study	Tendências paralelas	Efeito relativo ao grupo de comparação	Choques diferenciais, antecipação, adoção escalonada.
RDD sharp/fuzzy	Continuidade no limiar; regra de tratamento	Efeito local no corte / LATE	Validade externa e manipulação/choques no corte.
IV	Relevância, independência e exclusão	LATE para compliers	Instrumento fraco ou exclusão implausível.
ITS	Continuidade da trajetória sem intervenção	Mudança de nível/inclinação	Eventos concorrentes e autocorrelação.
Controle sintético	Combinação de doadores reproduz o pré	Efeito para unidade tratada	Ajuste pré ruim, spillovers e poucos doadores.

Painel GMM	Condições de momento e instrumentos válidos	Dinâmica/endogeneidade em painel	Proliferação e fragilidade de instrumentos.
------------	---	----------------------------------	---

Fonte: elaboração do autor com base na literatura citada.

PARTE III - ESTUDO DE CASO: BOLSA FAMÍLIA, DESIGUALDADE E REPASSES ESTADUAIS

A partir deste ponto é reproduzido e reorganizado o núcleo empírico do estudo que motivou o livro. O objetivo é manter os resultados, tabelas, figuras e diagnósticos originais, agora contextualizados pelos capítulos metodológicos anteriores. O leitor pode interpretar a RDD, os testes PELT e Chow e os modelos de painel GMM como componentes de uma estratégia de triangulação, distinguindo associação, quebra estrutural e identificação local.

OS MAIORES REPASSES DO BOLSA FAMÍLIA SÃO DIRECIONADOS PARA OS ESTADOS MAIS DESIGUAIS? REVISITANDO O MAIOR PROGRAMA DE TRANSFERÊNCIA DE RENDA DO BRASIL

*ARE THE LARGEST BOLSA FAMÍLIA TRANSFERS DIRECTED TO THE MOST UNEQUAL
STATES? REVISITING BRAZIL'S LARGEST INCOME TRANSFER PROGRAM*

Francisco Paulo Filho

Resumo: Este artigo analisa para o período que começa em Janeiro de 2013 e termina em Dezembro de 2020 como o índice de Gini da renda domiciliar per capita afeta os valores reais repassados para os estados do Programa Bolsa Família (PBF), nos 27 estados brasileiros. Primeiro, foi estimada uma Regressão Descontínua (RDD) com seleção de quebras endógena e estimada uma RDD global, os testes de PELT e Chow foram realizados para validação das quebras estatisticamente significantes, modelos do tipo painel GMM foram estimados para dar robustez aos resultados, o testes de McCrary foi realizado para a RDD local. O algoritmo PELT-MBIC confirma saltos estruturais em junho de 2014, expansão, e julho de 2016, redução, em todos os estados, enquanto testes de Chow validam as mesmas datas na maioria das séries do PBF estadual. Os resultados indicam que os estados com mais desigualdade tendem a receber repasses maiores do PBF, os modelos estimados via GMM confirmaram a robustez.

Abstract: This article analyzes, for the period beginning in January 2013 and ending in December 2020, how the Gini coefficient of per capita household income affects the real amounts transferred to the states under the Bolsa Família Program (PBF) in the 27 Brazilian states. First, a Regression Discontinuity (RDD) with endogenous break selection and a global RDD were estimated. PELT and Chow tests were performed to validate statistically significant breaks. Panel GMM models were estimated to provide robustness to the results. McCrary's test was performed for the local RDD. The PELT-MBIC algorithm confirms structural jumps in June 2014, expansion, and July 2016, reduction, in all states, while Chow tests validate the same dates in most of the state PBF series. The results indicate that states with greater inequality tend to receive larger PBF transfers; the models estimated via GMM confirmed the robustness.

1. INTRODUÇÃO

Criado em 2003, o Programa Bolsa Família (PBF) consolidou-se no Brasil como a principal política de transferência de renda do século XXI, alcançando mais de 19,6 milhões de famílias em Julho de 2025 com um investimento na ordem dos R\$ 13 bilhões segundo o Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome, além disso o PBF é bastante citado como referência internacional no combate à pobreza e estudos realizados com dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), revelaram que o PBF

reduziu tanto a pobreza quanto a desigualdade no Brasil na última década (Barros, *et al* 2006; Hoffmann, 2006; Soares, Ribas e Osório, 2010, Dantas *et al*, 2018. Lindert *et al* (2007) destaca que o desenho inovador e simples junto com o alcance amplo do Bolsa Família o tornaram modelo referência para formuladores de políticas sociais em todo o mundo, e segundo Souza et al. (2019) as transferências do PBF foram responsáveis por até 17% da queda da desigualdade entre 2001 e 2006 no Brasil, retirando mais de 3,4 milhões de pessoas da pobreza extrema em 2017; o estudo comparou indicadores de pobreza e de desigualdade antes e depois das transferências do Programa Bolsa Família, utilizando dados da PNAD tradicional de 2001 a 2015, e da PNAD Contínua de 2016 a 2017, aplicando técnicas de decomposição índice de Gini por fontes de renda, permitindo calcular a contribuição relativa do PBF para a redução da desigualdade ao longo do tempo no Brasil. Nessa linha, Sátyro *et al.*(2009) ressalta que o PBF transfere os maiores montantes para os que são realmente mais pobre, sendo um dos principais causadores da distribuição mais igualitária de renda no Brasil. Soares et al. (2010) utilizando os dados do período de 1999 a 2009 observou que 16% da redução da pobreza no Brasil poderia ser atribuída aos montantes de transferências do PBF para os estados mais desiguais, linha de pobreza utilizada foi de R\$ 100,00 a preços de janeiro de 2004. Através do RDD *fuzzy* Barbosa e Corseuil (2013) avalia o impacto do programa de transferência condicionada de renda, o Programa Bolsa Família (PBF), na composição ocupacional do mercado de trabalho brasileiro.

Soares, Ribas e Osório (2010) mostram, por meio de uma decomposição cuidadosa do Gini, que entre 2001 e 2015 o Bolsa Família reduziu a desigualdade no Brasil, algo em torno de 10 % da queda no índice de Gini se deveu diretamente ao programa segundo os autores. Resultados parecidos são destacados em Barros et al (2006). O consenso da literatura é que o programa, sobretudo em sinergia com a Previdência Rural e o BPC, foi decisivo para reduzir os níveis de pobreza extrema no Brasil a patamares inéditos desde 2003 (Soares, Ribas e Osório, 2010; Barros et al., 2006; Campello e Neri, 2013; Fiszbein, Schady e Ferreira, 2019).

Para ter direito ao Bolsa Família basta que a renda mensal por pessoa do domicílio fique abaixo de R\$ 218, ou seja, imagine um único provedor recebendo um salário mínimo de R\$ 151): se a família tiver sete membros, a renda per capita cai para R\$ 216,85, portanto dentro do limite, quem se enquadra já sai com um piso de R\$ 600 por domicílio, e se a soma dos benefícios não chegar a esse valor, o governo completa a diferença. Além disso, vale a regra por pessoa, cada integrante garante R\$ 142, mecanismo que evita que famílias numerosas fiquem descobertas. Cada criança de 0 a 6 anos recebe R\$ 150 extras, crianças e adolescentes de 7 a 18 anos, gestantes e nutrízes ganham R\$ 50 cada e se alguém da casa consegue emprego e a renda per capita sobe para algo entre R\$ 218 e meio salário-mínimo, essa pessoa entra na Regra de Proteção, o benefício continua vindo, mas pela metade durante até 12 meses para que a família não saia abruptamente do programa e consiga se planejar. Em troca, mantêm-se as condicionalidades, que seriam crianças na escola, carteirinha de vacinação em dia, pré-natal feito. Quem cumpre as regras e permanece dentro do novo teto continua amparado pelo programa.

Fica claro que o Bolsa Família é um programa inovador que já contribuiu ao longo das últimas duas décadas para a redução da desigualdade dentro dos estados, mas será que os estados mais desiguais são os que recebem mais repasses? O presente estudo tenta responder essa pergunta utilizando o método *Regression Discontinuity Designs* (RDD) *sharp* com seleção endógena de pontos de corte para a série mensal do Bolsa Família real, usando o índice de Gini como covariável observável. Para testar a robustez dos resultados foram realizados os testes PELT e Chow bem como a estimação de modelos em painel. Na sessão 2 é apresentado o método de desagregação de Denton-Cholett que o presente trabalho usa para desagregar os dados do índice de Gini dos estados de ano para mês, e também serão apresentadas as estatísticas descritivas da base de dados. Na sessão 3 a metodologia, na sessão 4 os resultados, na 5 os modelos que testaram robustez, e por fim as na sessão 6 as conclusões.

2. DESAGREGAÇÃO DE DENTON-CHOLETT E BASE DE DADOS

Os valores dos repasses do PBF para os estados brasileiros são contabilizados mensalmente, portanto qualquer análise detalhada e de baixa frequência que utilize essa base de dados deve ser mensal. Sendo assim, para responder o questionamento do trabalho a base do índice de Gini deve ter frequência mensal, somente dessa forma tem-se uma quantidade maior de dados e de graus de liberdade para testes estatísticos e estimações.

Para medir então a distribuição de renda em alta frequência utilizamos uma nova série mensal do índice de Gini para as 27 Unidades da Federação, cobrindo o período de janeiro de 2013 a dezembro de 2020, 96 observações por UF. O ponto de partida foram os valores anuais do Gini do Instituto Brasileiro de Geografia e Estatística (IBGE) e esses totais foram desagregados para a frequência mensal pelo método Denton-Cholette proporcional, de modo que a média dos doze meses reproduza exatamente o dado anual, eliminando qualquer risco de tendência espúria. A variável dependente dos modelos a serem estimados no trabalho será o valor pago pelo Bolsa Família (BPF) em cada estado, da PNAD Contínua, e convertidos a preços constantes de dezembro 2012 mediante deflacionamento pelo IPCA mensal do IBGE. A correspondência temporal entre o Gini e o BPF real, ambos mensais, permite avaliar no curto prazo até que ponto oscilações no Gini montante transferido se refletem em mudanças no BPF dentro do próprio ano, algo que análises anuais não conseguem captar. As estatísticas descritivas do Bolsa Família podem ser consultados na tabela 2, de Janeiro de 2013 a Dezembro de 2020.

Para gerar a série mensal do índice de Gini partimos do algoritmo de Denton-Cholette, e em sua formulação original de Denton (1971), depois generalizada por Cholette (1984), parte-se de um vetor anual G_t , com $t=1,...,T$, então busca-se um vetor mensal $g_{t,m}$, com $m=1,...,12$, que satisfaça duas condições simultâneas. A primeira é a restrição de identidade, ou seja, a média dos doze valores mensais deve reproduzir exatamente o dado anual. A segunda impõe variação mínima, logo entre todas as distribuições que respeitam a identidade, escolhe-se aquela que torna a série a mais suave possível. No caso proporcional, considerado mais adequado quando não existe indicador mensal auxiliar, o critério de suavidade minimiza a soma dos quadrados das primeiras diferenças relativas. A solução obtida via multiplicadores de Lagrange equivale, intuitivamente, a desagregar o valor anual por igual razão entre meses adjacentes, alterando então o mínimo necessário para que a média anual permaneça a mesma¹.

Para cada estado, o algoritmo produz um vetor de 96 pontos, 8 anos por 12 meses, cuja suavidade resulta diretamente do critério de minimização, dispensando qualquer *proxy* mensal de fora, dessa forma, preserva-se a coerência agregada do dado original, evitando a introdução de tendência espúria.

Em resumo, o método Denton-Cholette realiza a desagregação de dados anuais do índice de Gini para frequência mensal proporcional com média anual preservada. O método é oriundo dos trabalhos em interpolação de Dagum e Cholette (2006) e de Denton (1971) e os resultados das estatísticas descritivas dos dados mensais do índice de Gini da distribuição da renda per capita domiciliar dos estados brasileiros obtido a partir da tal desagregação, de 2013 a 2020, podem ser consultados na tabela 2. Optou-se por calcular as variações e em formato de mapa apresentar nas figuras 1 e 2 a evolução mensal da desigualdade nos estados brasileiros para o período estudado.

A análise descritiva do Bolsa Família real evidencia forte heterogeneidade entre os estados brasileiros em termos de volume médio de repasses e variabilidade. Bahia, São Paulo e

¹ Implementa-se esse procedimento com a função *td()* do pacote *tempdisagg*, usando o modo “*denton-cholette*” e a opção *conversion = "average"*, que garante que cada bloco de 12 observações mensais feche exatamente no valor anual.

Pará se destacam com as maiores médias mensais, superando 230 milhões de reais, refletindo tanto o tamanho populacional quanto a proporção de famílias elegíveis. Estados como Rondônia, Distrito Federal e Acre apresentam médias bastante inferiores, compatíveis com populações menores. O desvio-padrão indica que a variação ao longo do tempo é relativamente mais acentuada em unidades como Alagoas, Mato Grosso e Roraima, enquanto Rondônia e Acre mantêm séries mais estáveis. Os valores mínimos e máximos também revelam amplitudes importantes, especialmente em estados com forte crescimento ou retração no programa em determinados períodos, sinalizando que as flutuações do PBF não são uniformes e podem estar ligadas a eventos econômicos, políticos ou operacionais específicos.

Tabela 1 – Estatísticas Descritivas: PBF mensal estadual 2013 a 2020

UF	Obs	Média	Desvio Padrão	Mínimo	Máximo
Acre	96	24569,82	1357,121	20088,25	26413,21
Alagoas	96	89212,74	8292,273	78827,86	108185,6
Amapá	96	95997,19	4486,662	85825,66	104537,4
Amazonas	96	16074,54	1740,819	13403,39	18781,61
Bahia	96	379567,9	22886,1	341717	437912
Ceará	96	222203,2	17233,64	199248,6	262863,4
Distrito Federal	96	15283,93	1517,131	12630,06	18662
Espírito Santo	96	34942,04	3173,473	29184,77	41105,94
Goiás	96	58099,11	8045,61	46832,21	74374,68
Maranhão	96	232544,6	15188,45	205909,7	271810,7
Mato Grosso	96	214796,8	20397,82	187234,5	260027
Mato Grosso do Sul	96	26883,78	3320,179	21666,98	34262,24
Minas Gerais	96	32517,64	4785,089	25590,94	42430,04
Paraná	96	208422	9226,305	194232,5	237254,4
Paraíba	96	114921,1	6119,443	105560,5	132902,4
Pará	96	234183,6	13304,83	212981,6	269812,1
Pernambuco	96	105751	6708,519	95882,41	123408,6
Piauí	96	70202,11	8593,251	58361,2	86948,51
Rio Grande do Norte	96	168130,1	10924,42	145102,2	191103,3
Rio Grande do Sul	96	73211,77	5190,091	66283,23	86473,81
Rio de Janeiro	96	17847,35	4917,359	11618,68	26272,48
Rondônia	96	11184,59	701,3316	10003,87	12941,19
Roraima	96	76100,42	12032,97	59712,34	98586,76
Santa Catarina	96	24688,93	3290,343	19866,66	30906,76
Sergipe	96	54274,13	3879,721	48814,65	64602,19
São Paulo	96	271440,6	12473,06	246213,7	293033,7
Tocantins	96	27580,86	3745,062	21841,77	35242,35

Fonte: dados da pesquisa.

A estatística descritiva do índice de Gini mensal revela diferenças marcantes no nível médio de desigualdade entre os estados brasileiros no período analisado. Unidades federativas como Distrito Federal (0,572), Amapá (0,561), Sergipe (0,557) e Pará (0,558) apresentam, em média, os maiores valores, indicando estruturas distributivas mais desiguais. Em contraste, Santa Catarina (0,417), Minas Gerais (0,457) e Goiás (0,462) registram as menores médias, sugerindo padrões mais igualitários. O desvio-padrão revela maior volatilidade da desigualdade em estados como Amazonas, Bahia e Paraná, enquanto locais como Roraima e Santa Catarina apresentam séries mais estáveis. Os valores mínimos e máximos mostram que, embora alguns estados mantenham desigualdade persistentemente alta, outros experimentaram variações expressivas, o que pode refletir tanto choques econômicos quanto mudanças nas políticas públicas e no mercado de trabalho local.

Tabela 2 – Estatísticas Descritivas: Gini mensal estadual 2013 a 2020.

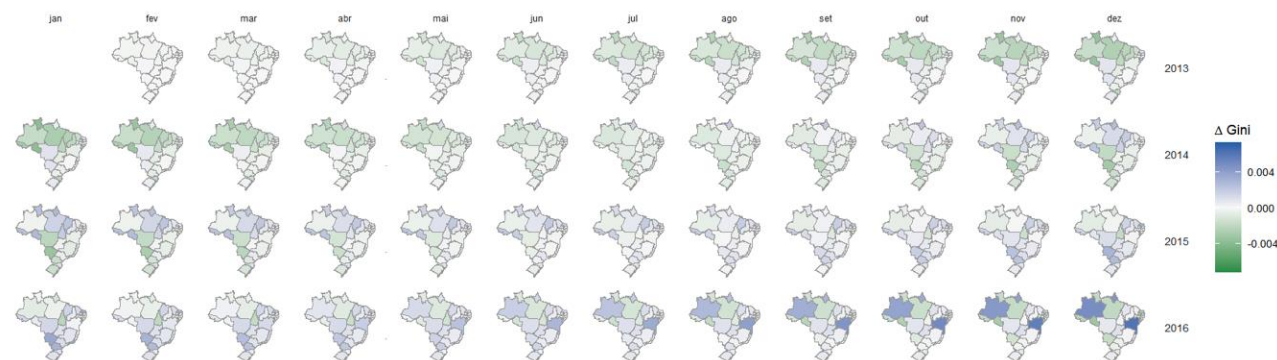
UF	Obs	Média	Desvio Padrão	Mínimo	Máximo
Acre	96	0,546375	0,016903	0,50362	0,567291
Alagoas	96	0,524625	0,012524	0,505367	0,553952
Amapá	96	0,561	0,019093	0,521695	0,599226
Amazonas	96	0,536	0,0284	0,497996	0,594961
Bahia	96	0,54475	0,021395	0,521301	0,597954
Ceará	96	0,54125	0,012153	0,519787	0,564814
Distrito Federal	96	0,57225	0,014298	0,547032	0,596127
Espírito Santo	96	0,50425	0,01064	0,473779	0,525004
Goiás	96	0,46225	0,00958	0,441565	0,478139
Maranhão	96	0,506	0,021847	0,46883	0,538754
Mato Grosso	96	0,48625	0,010754	0,453183	0,493244
Mato Grosso do Sul	96	0,47125	0,012238	0,444786	0,494815
Minas Gerais	96	0,456875	0,00719	0,440515	0,467282
Paraná	96	0,517375	0,024261	0,470723	0,570452
Paraíba	96	0,533125	0,017786	0,498095	0,567811
Pará	96	0,557625	0,015796	0,522899	0,5811
Pernambuco	96	0,521125	0,020411	0,456851	0,546608
Piauí	96	0,472375	0,012167	0,449415	0,494042
Rio Grande do Norte	96	0,527375	0,01744	0,499173	0,55255
Rio Grande do Sul	96	0,525625	0,017945	0,500343	0,560111
Rio de Janeiro	96	0,462375	0,019729	0,432938	0,504038
Rondônia	96	0,545625	0,020262	0,518445	0,586244
Roraima	96	0,47775	0,006034	0,465136	0,487735
Santa Catarina	96	0,416875	0,006552	0,405088	0,433549
Sergipe	96	0,556875	0,021166	0,509314	0,591984
São Paulo	96	0,52075	0,009856	0,506493	0,540063
Tocantins	96	0,505875	0,017667	0,473592	0,539475

Fonte: Dados pesquisa

Na versão proporcional, o método tenta manter proporcionalidade relativa entre meses consecutivos, isto é, minimizar a variação relativa de um mês para o seguinte, ao mesmo tempo em que a média dos 12 meses em cada ano reproduz exatamente o Gini anual observado. A ideia central do método é encontrar uma trajetória x_t para os períodos de alta frequência que seja consistente com os totais observados Y_s nos períodos agregados, ao mesmo tempo em que minimize as flutuações relativas dessa série. Com 96 observações por UF (12 meses $\times$ 8 anos). O método não supõe nenhuma variável explicativa (como PIB ou desemprego): trata-se de uma interpolação baseada apenas nos valores anuais, buscando suavidade relativa entre os meses.

A figura 1 revela um Brasil que andou em direções opostas num curto intervalo. Entre 2013 e 2014 predominam tons esverdeados, sinal de queda mensal no Gini; o padrão é quase generalizado, sobretudo no Norte e Nordeste, indicando o esgotamento paulatino das desigualdades herdadas dos anos 1990. Em 2015 o quadro começa a se misturar: aparecem manchas azuis em Minas, Rio Grande do Sul e São Paulo, sugerindo que a recessão já afetava o mercado de trabalho formal e, por tabela, a distribuição de renda. O ponto de virada é 2016: a paleta azul se espalha, principalmente no Sul-Sudeste, enquanto apenas bolsões do Nordeste ainda registram pequenos recuos na desigualdade. Essa inversão confirma que o choque fiscal e a retração do Bolsa Família naquele ano se refletiram em piora instantânea do índice, apagando parte dos avanços recentes.

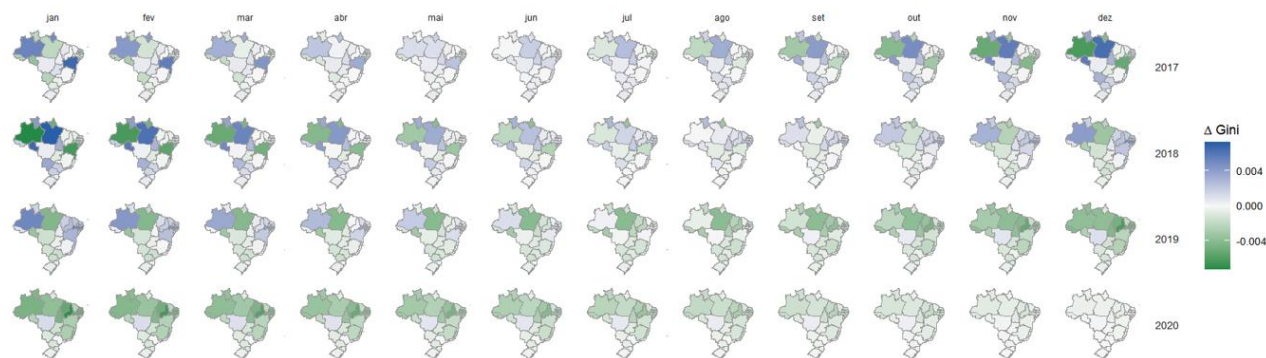
Figura 1: Evolução mensal do índice de Gini estadual ($G_{i,t} - G_{i,t-1}$) no Brasil de 2013.1 a 2016.12



Fonte: IBGE e Desagregação de Denton-Cholett.

Em 2017, a figura 2 mostra que há mais volatilidade e heterogeneidade nessa evolução, por exemplo as manchas azuis e as verdes se alternam entre estados, sinalizando meses de alta e de queda do Gini sem direção comum, e em 2018 predominam levemente quedas pontuais, e em 2019 se consolida um padrão mais difundido de redução da desigualdade, onde há vários meses verdes espalhados pelo país. Por fim, 2020 apresenta a maior recorrência de quedas mensais nos ginis estaduais, principalmente no eixo Norte-Nordeste, o que é coerente com o forte impulso redistributivo do Auxílio Emergencial, este ano todo é marcado por verde, com poucas reversões de curta duração. Em termos de magnitude, as variações mensais são pequenas (faixa de $\pm 0,004$), mas persistentes o bastante para compor um declínio agregado da desigualdade ao longo de 2019 e de 2020.

Figura 2: Evolução mensal do índice de Gini estadual ($G_{i,t} - G_{i,t-1}$) no Brasil de 2017.1 a 2020.12

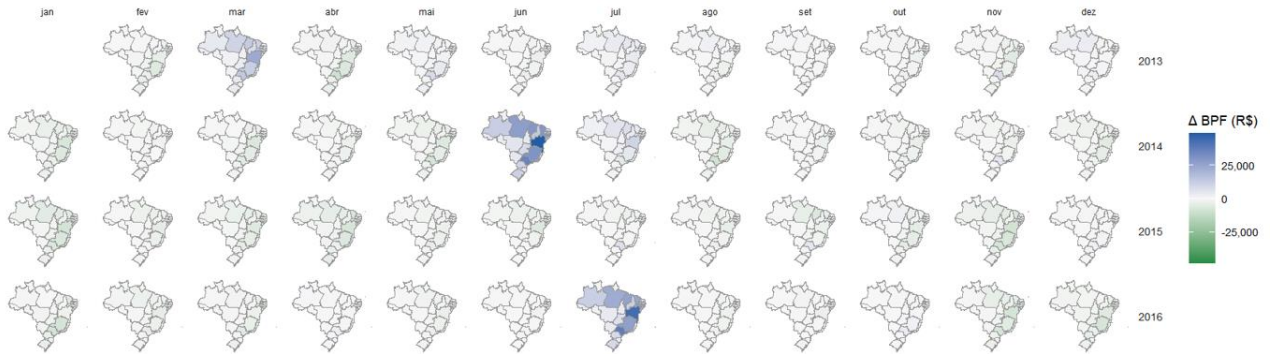


Fonte: IBGE e Desagregação de Denton-Cholett.

A mesma análise mensal foi realizada para o PBF, porém com os valores observados da PNAD contínua. A figura 3 mostra que a variação mensal dos repasses do programa em 2013 gira em torno de zero, com ajustes administrativos pequenos e esparsos, isso é normal dado que não houve alteração nos valores recebidos por faixa de renda. Em 2014, aparece o movimento mais evidente. Entre junho e julho do mesmo ano houve um salto positivo, tons de azul, especialmente no eixo Centro-Sul e parte do Nordeste, sinalizando expansão do programa no meio do ano. No fim de 2014, os mapas voltam a apresentarem variações baixas e em 2015 inteiro tem oscilações menores e sem direção aparente. Uma inflexão surge em 2016, a partir do meio do ano predominam quedas, tons de verde, em praticamente todas as UFs, compatíveis com o ajuste fiscal e o pente-fino cadastral que detectou mais de 1,1 milhão de benefícios

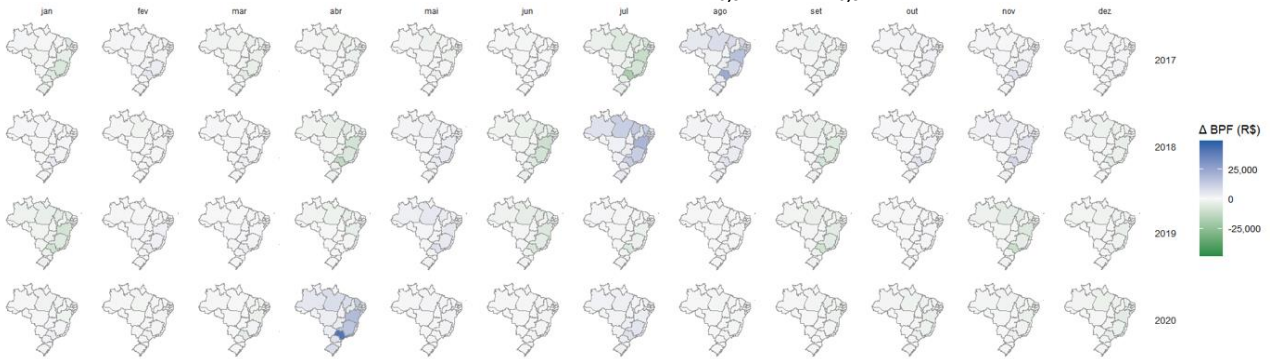
irregulares. Em suma, o mapa conta a mesma história vista nas estatísticas: pico de reforço no meio de 2014 e contração generalizada em 2016, o que ajuda a entender por que a desigualdade voltaria a subir naquele ano.

Figura 3: Evolução mensal do Bolsa Família ($bpf_{i,t} - bpf_{i,t-1}$) 2013.1 a 2016.12



Fonte: IBGE e Desagregação de Denton-Cholett.

Figura 4: Evolução mensal do Bolsa Família ($bpf_{i,t} - bpf_{i,t-1}$) 2017.1 a 2020.12



Fonte: IBGE e Desagregação de Denton-Cholett.

As séries desagregadas do índice de Gini foram utilizadas para as estimações e análises seguintes, levando em consideração a frequência mensal do PBF para cada estado brasileiro. A seguir, a metodologia do presente trabalho é apresentada.

3. METODOLOGIA

Como o objetivo do trabalho é investigar se quanto maior a desigualdade no estado maior os repasses do PBF a variável dependente dos modelos apresentados será o valor real mensal do pbf repassado ao estado i no mês t , denomina-se $pbf_{i,t}$. O índice de Gini estadual mensal será denominado por $G_{i,t}$. Com a estimação das regressões em um modelo de Delineamento de Regressão Descontínua (RDD) de desenho *sharp* com seleção endógena de pontos de corte para a série mensal do Bolsa Família real, $y_{i,t} = bpf_{i,t}$ por UF i , usando o Índice de Gini como covariável observável, $G_{i,t}$, pode-se encontrar possíveis cortes c ao longo do tempo e, para cada candidato, estimar localmente o salto em y no instante do corte via τ . O grupo de tratamento, observações que ocorrem posterior a quebra, e o grupo de controle, observações que ocorrem anterior a quebra, aqui não são definidos por uma amostra prévia, mas pela posição relativa à data de corte que o próprio algoritmo *rdrobust* do *software* R detectou. A escolha final contempla até 3 quebras não sobrepostas por UF, as que têm um menor p-valor de τ . Para cada quebra selecionada, o algoritmo implementado (i) reestima o RDD com larguras de banda ótimas do *rdrobust* e (ii) executa um OLS local para estimar τ e γ , o coeficiente do Gini. Em seguida, a validação das quebras estatisticamente significativas a 5% é

feita por meio dos testes PELT-MBIC, nível e inclinação, e Chow, na média, para as datas alvo. O teste de McCrary (2008) também é feito com objetivo de garantir que não existe manipulação nas descontinuidades.

Para cada UF, define-se para qualquer candidato c a variável $running$, $r_{i,t} = time_{i,t} - c$, que mede quantos meses faltam (ou passaram) do ponto de corte. Para cada UF, o código constrói o conjunto de candidatos c longe das bordas (-12 meses, +12 meses). Para cada c , mantém-se apenas a amostra local $|r_{i,t}(c)| \leq 12$ meses e estima-se um RDD robusto com o pacote *rdrobust* de Calonico, Cattaneo e Titiunik (2014), usando núcleo triangular, largura de banda inicial *mserd* e incluindo o gini como covariável adicional

$$y_{i,t} = \alpha + \tau D_{i,t} + \beta_1 r_{i,t} + \beta_{gini} G_{i,t} + \varepsilon_{i,t}$$

Onde $D_{i,t} = 1$ se $t \geq c$ e 0 caso contrário, τ é o salto no valor médio do Bolsa Família, controlado pelo Gini localmente. Após a seleção do menor p-valor de τ , escolhe-se c_1 e excluem-se todos os candidatos dentro de ± 12 meses dele, e repete-se o processo até no máximo 3 quebras por UF. Essa regra garante que as quebras sejam separadas e que cada janela local não se contamine com outra descontinuidade próxima.

Então faz-se a seleção de banda ótima mais um OLS local, ou seja, para cada quebra final selecionada,

- i) Reestima-se o RDD com *rdrobust* usando as larguras de banda ótimas h_l e h_r sugeridas pelo próprio estimador.
- ii) Constrói-se a amostra local filtrando $-h_l \leq r_{i,t} \leq h_r$.
- iii) Estima-se um modelo OLS local

$$y_{i,t} = \alpha + \tau D_{i,t} + \beta_1 r_{i,t} + \gamma G_{i,t} + \varepsilon_{i,t}$$

Onde $\hat{\tau}$ é o salto no intercepto ou impacto imediato e $\hat{\gamma}$ mede a relação marginal entre a desigualdade e o PBF no entorno do corte. No RDD, assume-se que o erro idiossincrático $\varepsilon_{i,t}$ possui média condicional nula em relação à variável $X_{i,t}$ e variância constante no entorno do ponto de corte. A estimação utilizou erros robustos, adequados para heteroscedasticidade e autocorrelação local.

A Regressão Discreta com Descontinuidade (RDD) global estimada neste estudo utiliza todos os dados disponíveis antes e depois de cada ponto de quebra identificado, ajustando um polinômio cúbico na variável de tempo relativa à quebra ($r_{i,t} = t - c$) e incluindo o índice de Gini (G) como covariável adicional. O modelo estimado é dado por:

$$pbf_{i,t} = \alpha + \tau D_{i,t} + \beta_1 r_{i,t} + \beta_2 r_{i,t}^2 + \beta_3 r_{i,t}^3 + \delta_1(D_{i,t} r_{i,t}) + \delta_2(D_{i,t} r_{i,t}^2) + \delta_3(D_{i,t} r_{i,t}^3) + \beta_{gini} G_{i,t} + \varepsilon_{i,t}$$

onde $D_{i,t}$ é uma *dummy* que assume valor 1 para observações posteriores à quebra e 0 caso contrário, sendo τ o coeficiente que captura a variação média instantânea no valor do Bolsa Família exatamente no ponto de corte, controlando por tendências temporais de ordem 3 e pelo nível de desigualdade.

4. RESULTADOS

Os resultados obtidos indicam que as quebras mais significativas do RDD se concentram em junho de 2014 e julho de 2016, períodos que coincidem com marcos relevantes na política social e no contexto macroeconômico brasileiro. Em junho de 2014, em ano eleitoral, o governo federal implementou reajuste médio de 10% nos benefícios do Bolsa Família, ampliando temporariamente o poder de compra dos beneficiários. Esse movimento ocorreu em

um cenário de desaceleração econômica e inflação elevada, sinalizando uma estratégia de sustentação do consumo das famílias mais pobres. Já em julho de 2016, o país atravessava a mais intensa crise econômica desde o Plano Real, com desemprego acima de 11%, inflação acumulada próxima a 9% e forte ajuste fiscal iniciado após o impeachment da presidenta Dilma Rousseff. Nesse período, a execução orçamentária do Bolsa Família sofreu restrições, com maior rigor na revisão cadastral e menor expansão da cobertura. Assim, as quebras detectadas nesses meses refletem, de um lado, a expansão pré-eleitoral e, de outro, o encolhimento ou redirecionamento do programa, ambos com impacto direto na relação entre desigualdade e volume de benefícios repassados.

A aplicação do RDD endógeno revelou que, para as quebras de junho de 2014, os estados de Minas Gerais (MG), Mato Grosso (MT) e Rio de Janeiro (RJ) apresentaram γ positivos e estatisticamente significantes a 5%, indicando que, nesses contextos, maiores níveis de desigualdade estão associados a valores mais elevados de repasses do Bolsa Família na vizinhança temporal da ruptura. Já para as quebras de julho de 2016, apenas o Distrito Federal (DF) apresentou esse padrão de associação positiva e significativa. Esses resultados sugerem que, nesses casos específicos, a alocação dos benefícios do programa foi relativamente mais intensa em períodos e locais com maior desigualdade, apontando para uma possível focalização em pessoas mais pobres ou para ajustes compensatórios do programa em resposta a aumentos na disparidade de renda.

Por outro lado, a estimação do RDD identificou diversos casos em que o coeficiente γ foi estatisticamente significativo a 5% e apresentou sinal negativo, indicando que, na vizinhança das quebras, maiores níveis de desigualdade estiveram associados a menores valores repassados pelo Bolsa Família, sendo assim esse padrão foi observado, nas quebras de junho de 2014, em estados como Amapá (AP), Ceará (CE), Maranhão (MA), Pará (PA), Pernambuco (PE), Piauí (PI), Rondônia (RO) e Roraima (RR), e para as quebras de julho de 2016 o mesmo comportamento negativo e significativo ocorreu em estados como Ceará (CE), Maranhão (MA), Pará (PA), Pernambuco (PE), Rio de Janeiro (RJ), Rondônia (RO), Santa Catarina (SC) e Tocantins (TO). Esse conjunto de resultados sugere que o programa não respondeu com aumento proporcional de benefícios diante de maiores desigualdades, podendo refletir limitações orçamentárias, critérios de focalização menos sensíveis à variação recente do Gini ou outras dinâmicas socioeconômicas locais.

Tabela 3 – Resultados da Regressão Discontinuada Local com Múltiplas Quebras por Estado: Estimativas do Efeito Médio Local (τ) e do Impacto da Desigualdade (γ) sobre o Bolsa Família, janela 12 meses.

<i>UF</i>	<i>Quebra</i>	τ	<i>p-valor τ</i>	γ	<i>p-valor γ</i>	<i>Data da quebra</i>
AC	1	-78,6404	0,86149	-165747	1,06E-07	2015-11-01
AC	2	1065,007	0,000426	4398,98	0,825896	2018-07-01
AC	3	2999,099	1,88E-13	-60831,3	7,32E-11	2014-06-01
AL	1	11798,74	2,06E-07	-398382	0,261649	2016-07-01
AL	2	13882,34	4,27E-12	297336,4	7,32E-05	2014-06-01
AL	3	4071,034	0,000584	42600,64	0,093604	2018-07-01
AM	1	13448,03	1,27E-16	30083,83	0,096037	2016-07-01
AM	2	-2529,44	0,154885	-2287,78	0,941583	2019-06-01
AM	3	1047,321	0,427513	-62147,6	0,002583	2017-08-01
AP	1	1767,521	1,29E-12	7289,968	0,06048	2016-07-01
AP	2	774,0667	0,002903	-43923,4	0,002924	2018-07-01
AP	3	2493,651	9,9E-14	-24823,5	0,016854	2014-06-01
BA	1	49719,26	1,3E-16	4076,016	0,94422	2016-07-01
BA	2	-16522,7	0,000474	-800314	2,38E-05	2019-11-01
BA	3	13124	0,004892	-128494	0,11653	2018-07-01
CE	1	9530,842	0,000142	-391157	0,006409	2018-07-01

CE	2	31599,97	1,32E-14	-582376	0,008587	2016-07-01
CE	3	34153,08	1,83E-13	-645120	4,8E-05	2014-06-01
DF	1	3320,641	1,47E-06	68665,4	0,020575	2016-07-01
DF	2	2230,392	2,74E-08	-131503	5,9E-06	2014-06-01
DF	3	-1170,88	0,050089	392883	0,000805	2019-07-01
ES	1	5952,374	4,06E-10	-20715,3	0,781545	2016-07-01
ES	2	337,3178	0,618154	27671,38	0,537482	2018-04-01
ES	3	-2864,1	0,013634	-50545,1	0,068146	2019-07-01
GO	1	-2466,36	0,004754	1133806	1,05E-05	2019-01-01
GO	2	4203,205	0,046344	-1484078	0,002088	2014-08-01
GO	3	1161,169	0,158549	-359701	1,08E-05	2017-08-01
MA	1	12371,06	2,3E-05	-620257	0,012327	2018-07-01
MA	2	31929,68	1,33E-14	-674701	0,003064	2016-07-01
MA	3	35731,15	6,98E-13	-319145	0,000283	2014-06-01
MG	1	34804,67	2,44E-12	-891175	0,112089	2016-07-01
MG	2	36555,45	3,05E-14	2424737	0,000144	2014-06-01
MG	3	-4795,63	0,093788	1022192	1,46E-05	2019-01-01
MS	1	4734,767	5,99E-12	109049,3	7,95E-08	2014-06-01
MS	2	3831,705	1,7E-11	14830,02	0,124882	2016-07-01
MS	3	660,0986	0,218033	90006,7	0,001106	2019-01-01
MT	1	4972,935	5,59E-12	-71424,1	0,367784	2016-07-01
MT	2	5962,374	2,14E-12	145845,8	1,6E-06	2014-06-01
MT	3	-850,883	0,092699	-6610,6	0,859116	2018-04-01
PA	1	2437,987	0,411128	254418,4	0,000146	2019-01-01
PA	2	24532,52	2,65E-17	-375105	0,000146	2016-07-01
PA	3	30289,72	1,29E-13	-558156	2,42E-08	2014-06-01
PB	1	14477,69	1,81E-13	-49111,1	0,197548	2016-07-01
PB	2	1838,964	0,215004	-269380	0,000612	2017-08-01
PB	3	-4537,01	0,007044	213581,9	9,1E-05	2019-01-01
PE	1	30735,41	1,21E-16	-348086	0,006992	2016-07-01
PE	2	-12128,4	0,003433	432118,8	8,79E-05	2019-01-01
PE	3	35066,85	1,87E-12	-1,8E+07	0,000126	2014-06-01
PI	1	3831,55	0,000265	-90945,5	0,119875	2018-07-01
PI	2	13333,92	7,48E-15	-18029,6	0,903338	2016-07-01
PI	3	16621,32	1,59E-13	-187078	0,001985	2014-06-01
PR	1	11570,73	4,59E-10	31,85236	0,999487	2016-07-01
PR	2	-5231,15	0,03305	-180724	0,028312	2015-04-01
PR	3	2621,173	0,106954	361678,6	0,003921	2019-01-01
RJ	1	26376,96	1,34E-13	-230245	0,000338	2016-07-01
RJ	2	23884,64	9,99E-12	1668888	4,95E-06	2014-06-01
RJ	3	5018,963	0,059695	172243,2	0,276964	2018-07-01
RN	1	9713,411	1,66E-16	-14301,9	0,171147	2016-07-01
RN	2	-3066,68	0,033401	24287,03	0,40628	2019-06-01
RN	3	1854,607	0,033453	162402,5	4,5E-07	2017-08-01
RO	1	2852,115	2,02E-10	-24507	0,031062	2016-07-01
RO	2	254,1227	0,266842	38805	3,32E-05	2019-03-01
RO	3	3529,683	4,31E-12	-32282	1,78E-05	2014-06-01
RR	1	1598,453	4,88E-08	3548,921	0,816377	2016-07-01
RR	2	-278,585	0,082867	26121,01	3,69E-07	2019-01-01
RR	3	1988,944	5,48E-12	-10036,1	0,010391	2014-06-01
RS	1	-6849,28	0,022288	-398020	0,070927	2015-04-01
RS	2	1572,447	0,233826	786456,3	0,003729	2019-01-01
RS	3	11745,7	9E-12	-525990	0,184066	2016-07-01
SC	1	4691,181	1,36E-10	-79334,1	0,038816	2016-07-01
SC	2	-958,714	0,032706	92564,28	0,332529	2018-04-01
SC	3	-469,386	0,528279	-294394	9,75E-06	2015-01-01
SE	1	-1071,34	0,078857	82106,78	1,44E-06	2019-01-01
SE	2	-2727,46	0,139676	12839,29	0,848279	2017-02-01
SE	3	1723,959	0,566176	-261920	0,057405	2015-12-01
SP	1	46613,16	2,32E-12	-584892	0,110011	2016-07-01

SP	2	21827,45	1,11E-06	-1698106	0,003047	2014-06-01
SP	3	3800,022	0,49772	578771,3	0,035112	2018-07-01
TO	1	3903,945	2,05E-13	-25274,2	0,042468	2016-07-01
TO	2	683,0395	0,037826	14341,43	0,250753	2018-07-01
TO	3	4807,623	2,09E-14	-75328,4	5,53E-07	2014-06-01

Fonte: resultados da pesquisa

Os resultados da RDD global revelam que a maioria dos estados apresenta um coeficiente de quebra estatisticamente significativo a 5% em quebras ocorridas principalmente em junho de 2014 e julho de 2016, sugerindo mudanças estruturais relevantes no programa nestes períodos. Em vários casos, o coeficiente do Gini também foi significativo a 5%, indicando que a distribuição dos benefícios do Bolsa Família está correlacionada ao nível de desigualdade estadual. Em estados como Alagoas, Amazonas, Bahia, Ceará, São Paulo e Rio de Janeiro, τ positivo e significativo a 5% aponta para aumentos expressivos do benefício no ponto de corte, enquanto em outros, como Pará e Sergipe, valores negativos sugerem reduções abruptas. Esses resultados reforçam a interpretação de que, nos períodos críticos identificados, houve ajustes substanciais na política de transferências, possivelmente relacionados a choques macroeconômicos e mudanças orçamentárias no programa.

Tabela 4 - Resultados da Regressão Discontinuada Global (RDD) com Polinômio Cúbico: Estimativas do Efeito Médio Local (τ) e do Coeficiente da Desigualdade (γ)

UF	Quebra	τ_{global}	p-valor τ_{global}	γ_{global}	p-valor γ	Data da quebra
AC	1	618,2939498	0,330747726	25834,61891	0,026097885	2015-11-01
AC	2	1410,190909	7,75191E-08	-56056,46197	1,91433E-06	2018-07-01
AC	3	3237,438181	5,3835E-20	-15203,13684	0,017521329	2014-06-01
AL	1	17281,41059	1,14766E-13	142336,2904	4,33321E-05	2016-07-01
AL	2	17277,28895	1,3757E-25	83400,52093	0,013007809	2014-06-01
AL	3	-403,1159845	0,862970103	6224,176447	0,960744719	2018-07-01
AM	1	16883,27134	3,58325E-16	-71092,47847	9,89707E-05	2016-07-01
AM	2	-5266,23535	0,059387467	32018,96808	0,246430347	2019-06-01
AM	3	-4381,374355	0,099661774	44974,47578	0,252443131	2017-08-01
AP	1	2479,736474	4,73048E-13	-42630,49424	3,9545E-09	2016-07-01
AP	2	888,4371001	3,32755E-06	8350,073523	0,279969419	2018-07-01
AP	3	2604,633345	8,06242E-22	-3750,233655	0,336179869	2014-06-01
BA	1	61608,66476	1,52828E-14	-308128,3083	3,31283E-05	2016-07-01
BA	2	-25040,48385	0,000349489	59007,78404	0,527011048	2019-11-01
BA	3	14470,65431	0,011154403	305073,5794	0,007454242	2018-07-01
CE	1	5181,725187	0,124887385	-1079087,621	0,00167375	2018-07-01
CE	2	36025,2279	2,65947E-15	-447349,3947	0,005215518	2016-07-01
CE	3	39219,0431	1,44511E-29	-282272,0069	0,024682032	2014-06-01
DF	1	3602,8121	3,90088E-16	-58459,00434	0,006920913	2016-07-01
DF	2	1654,949002	9,97866E-05	46849,4033	0,001520978	2014-06-01
DF	3	110,4441675	0,830414262	19231,4073	0,391986803	2019-07-01
ES	1	5622,553395	2,39996E-11	17458,15648	0,637855842	2016-07-01
ES	2	-302,2182073	0,676250087	-108698,5852	0,033787422	2018-04-01
ES	3	-1776,474275	0,047518823	-33852,16479	0,55858572	2019-07-01
GO	1	-2045,99889	0,126444301	117444,3713	0,221812378	2019-01-01
GO	2	-1087,635113	0,648828549	258315,7509	4,62467E-06	2014-08-01
GO	3	-2481,60444	0,223005726	-59497,12809	0,725647795	2017-08-01
MA	1	4965,276375	0,214331522	-1076290,963	0,000104475	2018-07-01
MA	2	36907,71485	5,09291E-12	-172275,618	0,133735986	2016-07-01
MA	3	39280,45886	1,88173E-25	-367096,8507	0,000103547	2014-06-01
MG	1	36343,94967	7,39098E-10	414044,2375	0,187457276	2016-07-01
MG	2	39241,45864	4,37608E-29	364893,4149	0,252493412	2014-06-01
MG	3	-6954,744123	0,172990417	498371,327	0,334103249	2019-01-01
MS	1	5451,010258	7,8142E-30	51319,06762	0,001584206	2014-06-01
MS	2	5727,438018	6,41153E-13	81645,55377	1,54874E-07	2016-07-01

MS	3	-54,23005557	0,940555878	20635,01031	0,278763126	2019-01-01
MT	1	6042,34852	6,57378E-10	120213,7683	0,012096908	2016-07-01
MT	2	7581,060285	1,69304E-23	53777,62218	0,000452958	2014-06-01
MT	3	-1380,564398	0,014009788	57335,81546	0,119434476	2018-04-01
PA	1	-17052,87812	0,002688192	-356093,2975	5,9957E-05	2019-01-01
PA	2	34972,18567	2,76021E-14	89834,8743	0,015656072	2016-07-01
PA	3	32519,19244	1,09464E-22	-48974,44382	0,210427798	2014-06-01
PB	1	18422,21968	6,32293E-10	-20235,46629	0,59087519	2016-07-01
PB	2	-3055,785473	0,193375577	79944,92937	0,134589077	2017-08-01
PB	3	-3833,588216	0,052198452	276674,5832	0,017042278	2019-01-01
PE	1	37260,04691	6,47155E-13	-80346,09557	0,104790885	2016-07-01
PE	2	-890,9135749	0,743720466	-211378,866	0,090612079	2019-01-01
PE	3	42108,26739	7,30159E-25	-106773,8643	0,008600972	2014-06-01
PI	1	-40,89755897	0,984475411	-327207,4891	0,006001588	2018-07-01
PI	2	15934,01332	1,61168E-11	-16413,66906	0,621363912	2016-07-01
PI	3	17497,4923	2,43274E-22	-61073,26396	0,014682763	2014-06-01
PR	1	16284,706	4,32027E-10	378289,279	4,16525E-08	2016-07-01
PR	2	-3508,638271	0,097747774	280071,518	0,000251285	2015-04-01
PR	3	-655,556073	0,743587152	10299,98101	0,877315682	2019-01-01
RJ	1	34652,20565	2,31546E-12	696308,3864	2,21234E-05	2016-07-01
RJ	2	26906,09886	9,16236E-19	128363,7351	0,339034065	2014-06-01
RJ	3	3917,503907	0,256553753	-135770,8469	0,354337369	2018-07-01
RN	1	12285,78052	5,87121E-11	94768,59727	0,009752468	2016-07-01
RN	2	-3096,63802	0,024636853	-81046,22358	0,054513067	2019-06-01
RN	3	-2592,420949	0,093568601	-27040,99294	0,553525122	2017-08-01
RO	1	4158,682132	6,80871E-09	17556,25917	0,000210167	2016-07-01
RO	2	-904,7775286	0,146634917	-14030,25083	0,08690117	2019-03-01
RO	3	4339,222314	1,58419E-24	5491,613633	0,159337068	2014-06-01
RR	1	2200,253932	1,00321E-08	1762,324928	0,791167496	2016-07-01
RR	2	317,3273811	0,309830685	-12090,04327	0,070326001	2019-01-01
RR	3	2114,265246	4,54689E-22	-7483,822672	0,031492784	2014-06-01
RS	1	-3595,042234	0,128013699	443624,716	0,006259716	2015-04-01
RS	2	-605,8092722	0,755215234	131222,2318	0,198274662	2019-01-01
RS	3	13821,92057	3,04296E-13	767384,4776	3,00734E-05	2016-07-01
SC	1	4712,463745	2,36626E-07	7248,858783	0,877779992	2016-07-01
SC	2	-441,7882098	0,425898137	-100016,6252	0,009407548	2018-04-01
SC	3	-2937,413167	0,002589009	-49559,36144	0,200954069	2015-01-01
SE	1	-1835,086909	0,00332758	-93656,34285	0,001199091	2019-01-01
SE	2	-4163,026619	0,007593757	-45191,06258	0,039690764	2017-02-01
SE	3	1239,465808	0,320676481	59940,96656	0,001218618	2015-12-01
SP	1	52209,30464	4,92363E-19	1523506,814	1,88692E-06	2016-07-01
SP	2	29440,78298	1,46574E-08	1198129,154	0,000392774	2014-06-01
SP	3	5703,114046	0,350925899	-503674,7907	0,402680842	2018-07-01
TO	1	4668,083139	9,3987E-12	27851,18972	0,054744092	2016-07-01
TO	2	262,8978577	0,584686355	-55373,08011	0,012491521	2018-07-01
TO	3	5145,896424	4,47225E-24	-20500,61416	0,004953245	2014-06-01

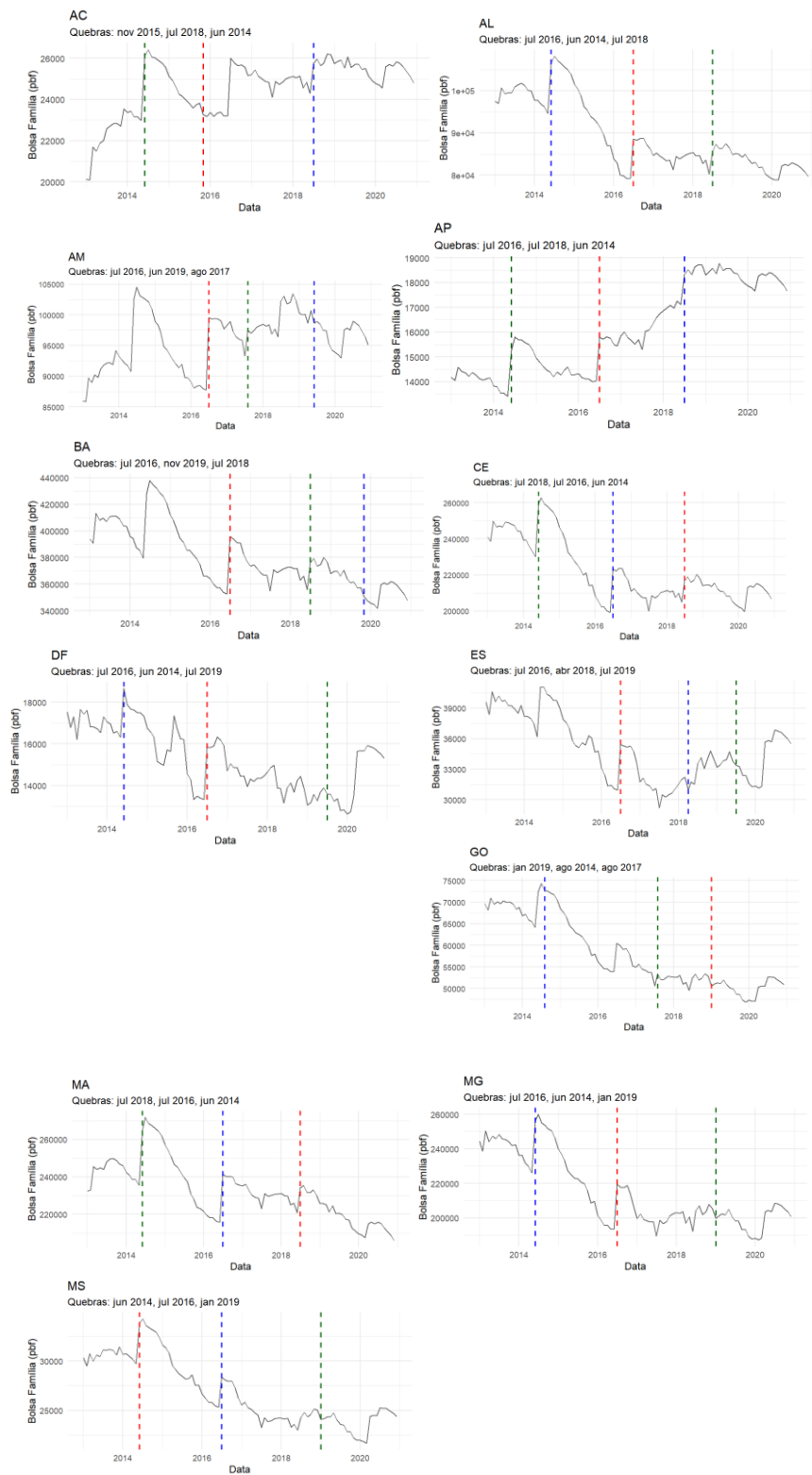
Fonte: resultados da pesquisa

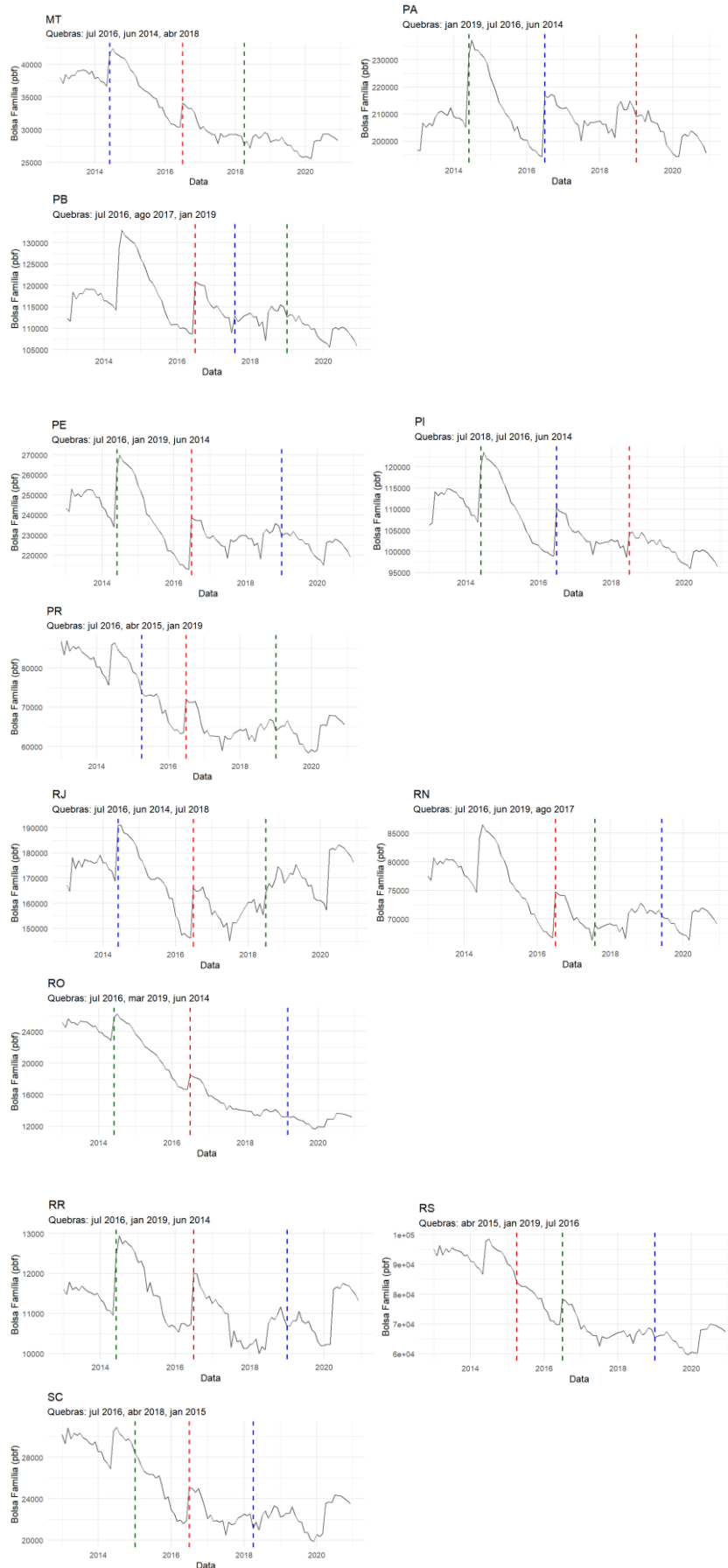
As estimações foram realizadas utilizando uma janela simétrica de ± 12 meses em torno de cada ponto de descontinuidade identificado, de forma a capturar os efeitos locais do programa Bolsa Família sobre a desigualdade (índice de Gini) com maior precisão e evitar extrapolação dos efeitos para períodos mais distantes.

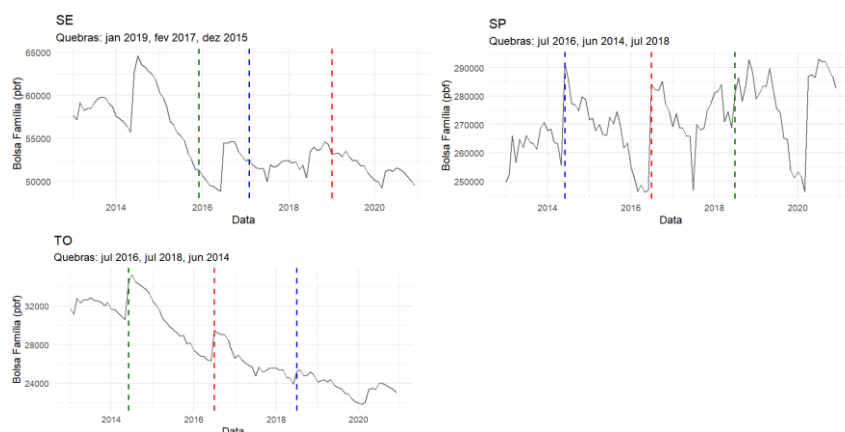
Para testar a hipótese de continuidade na densidade da variável de ordenação nos pontos de corte endogenamente detectados, foi aplicado o teste de McCrary conforme proposto por McCrary (2008), utilizando a função *dc_test()* do pacote *rddapp* no R.

Os resultados indicam que em praticamente todas as unidades da federação e para as diversas quebras detectadas, os p-valores associados à estatística que julga H_0 superaram 0,74, não havendo assim evidências estatísticas de manipulação no suporte da variável temporal em torno dos pontos de descontinuidade. Isso assegura a validade da identificação de cada ponto de corte como potencial RDD.

Figura 5 – Quebras identificadas pela RDD Local.







Fonte: dados da pesquisa.

Tabela 5 - Testes de quebra PBF – Pelt e Chow

UF	P-Valor			
	pelt jun2014	pelt 2016	chow p 2014	chow p 2016
AC	~0	1,56E-12	1,55E-15	1,71E-11
AL	2,77E-10	~0	2,77E-10	~0
AM	1,98E-08	1,55E-08	6,2E-07	5,6E-08
AP	1,37E-08	~0	1,29E-08	~0
BA	9,78E-06	1,15E-11	6,79E-07	3,01E-12
CE	2,72E-09	1,64E-14	5,65E-11	1,64E-14
DF	9,27E-08	3,9E-12	2,72E-09	2,04E-12
ES	5,26E-10	1,43E-12	1,55E-11	1,22E-12
GO	7,14E-11	~0	2,54E-12	~0
MA	0,000232	7,59E-10	0,000232	3,02E-10
MG	8,76E-11	~0	1,62E-12	~0
MS	6,25E-08	~0	2,34E-09	~0
MT	6,09E-09	~0	2,6E-10	~0
PA	0,479986	0,023546	0,974319	0,014314
PB	0,152692	6,87E-08	0,045224	1,75E-08
PE	8,81E-06	5,77E-09	5,11E-07	3,62E-09
PI	5,19E-05	9,43E-13	4,23E-06	2,37E-13
PR	4,42E-14	~0	5,55E-16	~0
RJ	0,008275	0,004462	0,001582	0,00522
RN	6,59E-07	4,77E-14	3,19E-08	2,73E-14
RO	2,13E-12	~0	8,9E-14	~0
RR	0,074381	1,5E-06	0,025858	3,59E-07
RS	8,08E-13	~0	1,71E-14	~0
SC	6,91E-13	~0	1,49E-14	~0
SE	1,29E-06	9,05E-12	6,08E-08	9,33E-12
SP	0,001017	1,76E-05	0,005585	5,31E-05
TO	5,3E-09	~0	2,52E-10	~0

Fonte: Resultados da pesquisa.

Os números reforçam a leitura de que 2016 foi um ponto de inflexão nacional no Bolsa Família. Em vinte e cinco das vinte e sete UFs o p-valor do Chow em julho de 2016 está

muito próximo de zero, indicando um corte brusco na média dos repasses, fenômeno que o PELT já havia sinalizado ao detectar quebras com p-valores igualmente ínfimos. Os estados do eixo Sul-Sudeste (PR, RS, SC, MG, SP) e do Centro-Oeste (GO, MT) seguiram esse padrão, mas o choque também se espalhou pelo eixo Norte-Nordeste (CE, BA, MA, RN, entre outros).

Em junho de 2014 a geografia das quebras foi bem mais heterogênea, pois alguns estados da Amazônia, como Acre e Amapá, do Nordeste, como Ceará e Bahia, e do Sul, como Paraná, exibem p-valores bem pequenos, sinalizando uma expansão pontual no patamar do programa às vésperas da eleição presidencial do mesmo ano. Contudo, em lugares como Pará ($p \approx 0,48$) e Roraima ($p \approx 0,07$) o teste não rejeita a estabilidade da média, sugerindo que o impulso de 2014 não foi uniforme. Até mesmo unidades mais ricas, como DF e SP, mostram quebras estatisticamente significativas, mas com p-valores na casa de 10^{-3} e 10^{-2} , bem menos contundentes do que em 2016. Em síntese, o programa passou por um choque seletivo de alta em 2014 e um rebaixamento generalizado em 2016, e esses marcos devem ser tratados como regimes distintos nas estimações de impacto, sempre lembrando que os resultados de 2014 variam muito mais de estado para estado do que o recuo observado dois anos depois. O RDD estima os pontos de corte em janelas curtas e os testes PELT e Chow mostram que esses mesmos instantes são, de fato, rupturas estruturais nas séries completas, legitimando a divisão da amostra em três regimes, pré-2014, 2014-2016 e pós-2016, para estimar o impacto do Gini no programa.

5. ROBUSTEZ VIA PAINEL COM ESTIMADOR GMM

O estudo adota diferentes especificações de modelos em painel para investigar a relação entre as transferências do PBF e o índice de nos estados brasileiros como uma forma de testar se os principais resultados das RDD são robustos, explorando tanto os efeitos dinâmicos quanto os impactos de eventos pontuais detectados nas análises de RDD. Foram estimados modelos de efeitos fixos (FE), efeitos aleatórios (RE) e duas variações do estimador GMM, o *Difference-GMM* e o *System-GMM*, conforme os trabalhos de Arellano e Bond (1991), de Arellano e Bover (1995) e Blundell e Bond (1998). Essa estratégia permite lidar com a endogeneidade de variáveis defasadas, heterogeneidade não observada e possíveis problemas de autocorrelação serial (Baltagi, 2021; Hayashi, 2000). A identificação utiliza *dummies* para quebras em junho de 2014 e em 2016, que foram validadas pelo modelo RDD local e global, além de variáveis contínuas como o índice de Gini, e defasagens instrumentais para isolar efeitos causais, garantindo desse modo a consistência sob as condições de momento (Hansen, 1982). A inclusão da variável defasada do próprio Bolsa Família captura a persistência temporal dos repasses, característica típica de políticas públicas com execução contínua.

Tabela 6 - Resultados – System GMM 1

<i>Variável</i>	<i>Coefficiente</i>	<i>Erro padrão</i>
$pbfi_{i,t-1}$	0,738***	0,099
$G_{i,t}$	42705*	19142
$D_{jun/2014}$	5354,4***	778,2
D_{2016}	-2885,6**	1103,9

Fonte: resultados da pesquisa. Nota: *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$. Instrumentos: lag(gini, 3:4) e lag(bpf, 3:4).

O System GMM 1 foi estimado em um painel balanceado, cobrindo 27 unidades federativas ao longo de 96 meses, totalizando 2.592 observações, às quais se somam defasagens utilizadas como instrumentos, resultando em 5.049 observações efetivas.

A validade dos instrumentos foi avaliada pelo teste de Sargan, que, com p-valor igual a

0,00062, rejeitou a hipótese nula de exogeneidade conjunta. Essa rejeição, contudo, deve ser interpretada com cautela por alguns motivos. Primeiro, porque o teste de Sargan clássico pressupõe homoscedasticidade, e painéis mensais heterogêneos, como o presente, tendem a super rejeitar mesmo com instrumentos corretos (Arellano e Bond, 1991; Hayashi, 2000). Segundo, os diagnósticos dinâmicos centrais se mantêm satisfatórios: o teste de autocorrelação de segunda ordem (AR(2)) não rejeita a hipótese nula, confirmando a adequação das defasagens instrumentais. Terceiro, estimativas alternativas dos instrumentos, restringindo o conjunto de *lags*, preservam sinal e magnitude dos coeficientes principais.

Os testes de autocorrelação mostram, para a primeira ordem, uma estatística normal de -2,6293 com valor-p = 0,00856, rejeita ausência de autocorrelação de primeira ordem, o que é esperado, e para a segunda ordem, estatística normal de 1,7643 com p-valor igual a 0,07769, não rejeita H0 ao nível de 5%, indicando ausência de autocorrelação de segunda ordem, o que é também o correto. Por fim, o teste de Wald para o conjunto de coeficientes é significativo a 5%, o que sugere que em conjunto as variáveis explicativas são estatisticamente significativas no modelo.

Tabela 7 - Resultados do System GMM 2

<i>Variável</i>	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>Estatística z</i>
$pbf_{i,t-1}$	0,96153***	0,00341	282,2486
$G_{i,t}$	4753,6**	1744	2,7256

Fonte: resultados da pesquisa. Nota: *** p<0,001, ** p< 0,01, * p < 0,05 e instrumentos são lag(gini, 2) e lag(bpf, 2:9).

Para o System GMM 2, o teste de Sargan retorna um p-valor de 0,25472, não rejeitando a validade dos instrumentos, ou seja, não há evidência de sobreidentificação. Nos testes de autocorrelação, a primeira ordem apresenta estatística normal de -2,6661 com p-valor igual a 0,00767, sugerindo presença de autocorrelação de primeira ordem, enquanto a segunda ordem tem estatística normal de 1,97031 com p-valor igual a 0,04881, fraca evidência de autocorrelação de segunda ordem ao nível de 5%, e zero evidência para o nível de 1%. O teste de Wald para o conjunto de coeficientes é estatisticamente significativo a 5%.

Tabela 8- Resultados dos modelos Efeitos Fixos, Efeitos Aleatórios e Difference-GMM

<i>Variável</i>	<i>Efeitos Fixos</i>	<i>Efeitos Aleatórios</i>	<i>Difference-GMM</i>
intercepto	-	107754,6*** (19344)	-
G	11785,7 (10037,6)	12058,8 (10037,1)	113550** (36709)
$D_{jun/2014}$	900,9 (529,9)	902,4 (530)	-
D_{2016}	-11603,2*** (428,8)	-11606,1*** (428,8)	-
pbf_{t-1}	-	-	0,8937*** (0,0394)

Fonte: resultados da pesquisa. Nota: *** p<0,001, ** p< 0,01, * p < 0,05 e instrumentos do GMM escolhidos são lag(gini,3) e lag(bpf,3:8). (.) representa o erro padrão.

Os testes de diagnóstico do modelo *Difference-GMM* sugerem que ele é estatisticamente consistente, pois o teste de Sargan apresenta um p-valor de 0,088, o que indica que os instrumentos utilizados não estão correlacionados com os resíduos, ou seja, não há rejeição da hipótese nula de validade dos instrumentos ao nível de 5%, e o teste de autocorrelação de primeira ordem, AR(1), apresenta um p-valor de 0,0106, indicando presença de autocorrelação, o que é esperado neste tipo de modelo. O teste de autocorrelação de segunda ordem, AR(2), apresenta p-valor de 0,1229, sugerindo ausência de autocorrelação de segunda ordem, condição necessária para a validade dos instrumentos, além disso o teste de Wald indicou a significância conjunta a 5% dos regressores.

Para o modelo com efeitos fixos (FE), os testes revelam boa adequação global do modelo, com um F-estatístico de 371,33, p-valor $< 0,001$, sugerindo que as variáveis explicativas são estatisticamente relevantes no conjunto, e entre os coeficientes, apenas a *dummy* para o ano de 2016 é estatisticamente significativa a 5%, p-valor $< 0,001$, refletindo um evento pontual ou política pública relevante nesse período. Os resíduos apresentam distribuição assimétrica, o que pode indicar a presença de heterocedasticidade, tratada via erro padrão robusto HC1.

No modelo de efeitos aleatórios (RE), a estatística Qui-quadrado, 1113,7 com p-valor $< 0,001$, confirma a significância 5% global do modelo, sendo o sinal e o tamanho dos coeficientes são similares ao modelo FE, com destaque novamente para a significância a 5% da *dummy* de 2016. O intercepto é estatisticamente significativo, sugerindo a existência de efeitos fixos não capturados pelas variáveis explicativas.

6. CONCLUSÕES

Os resultados empíricos mostram que no período analisado, de Janeiro de 2013 a Dezembro de 2020, os estados brasileiros com maiores níveis de desigualdade, medidos pelo índice de Gini, tenderam a receber proporcionalmente maiores repasses do Programa Bolsa Família (PBF) segundo a RDD local, global e os painéis GMM. A análise via RDD local e global revelou que, em pontos de descontinuidade significativos, o salto médio no valor do PBF foi mais pronunciado justamente em estados com histórico de desigualdade elevada, como alguns do eixo Norte-Nordeste. Essa relação sugere que o desenho do programa ao priorizar famílias em situação de vulnerabilidade manteve a coerência distributiva ao direcionar recursos para áreas com maior necessidade, o que está alinhado à literatura sobre focalização e equidade em transferências condicionadas (Soares, Ribas e Osório, 2010; Souza et al., 2019; Hoffmann, 2006).

Para alguns estados algumas quebras da RDD local e global indicam o aumento significativo, estatisticamente, do Bolsa Família, isso condiz com o decreto 8.232/2014 que aumentou os valores do benefício básico por pessoa. Os modelos painéis via GMM confirmam os resultados de expansão e redução dos benefícios para as datadas de Junho de 2014 e o ano de 2016. Os resultados robustos concordam que em estados do Norte e Nordeste, e alguns do Sudeste, quanto maior for a desigualdade, medida pelo Gini, maior tende a ser os valores recebidos pelo estado. Esses resultados devem ser interpretados com cautela, haja vista que modelos não lineares e testes com inclinação que faça interagir na RDD a inclinação antes e depois do corte devem ser implementados para maiores evidências, além disso mais variáveis devem ser acrescentadas em análises futuras.

REFERÊNCIAS

Acosta, P. Beneficiários do Bolsa Família são preguiçosos? *Folha de S. Paulo* (Opinião), 6 abr.

2023.

- Arellano, M.; Bond, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, v. 58, n. 2, p. 277-297, 1991.
- Arellano, M.; Bover, O. Another look at the instrumental-variable estimation of error-components models. *Journal of Econometrics*, v. 68, n. 1, p. 29-51, 1995.
- BARROS, R. P.; CARVALHO, M.; FRANCO, S.; MENDONÇA, R. Macro determinantes da queda na desigualdade no Brasil. Rio de Janeiro: Ipea, 2006.
- Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, v. 18, n. 1, p. 1-22, 2003.
- Baltagi, B. H. *Econometric analysis of panel data*. 6. ed. Cham: Springer, 2021.
- Rasella, D. et al. Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. *The Lancet*, v. 382, n. 9886, p. 57-64, 2013.
- Blundell, R.; Bond, S. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, v. 87, n. 1, p. 115-143, 1998.
- Calonico, S.; Cattaneo, M. D.; Titiunik, R. rdrobust: An R package for robust inference in regression-discontinuity designs. University of Michigan, Working Paper, 2014.
- Chow, G. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, v. 28, n. 3, p. 591-605, 1960.
- Campello, T. Neri, M. C.; (Orgs.). Programa Bolsa Família: uma década de inclusão e cidadania. Brasília: Ipea; MDS, 2013.
- Dagum, E. B.; Cholette, P. A. *Benchmarking, temporal distribution, and reconciliation methods for time series*. New York: Springer-Verlag, 2006. (Lecture Notes in Statistics).
- DANTAS, Sara Verônica de Avelar Dias; MIRANDA, Maria Geralda de; DUSEK, Patricia Maria;
- Denton, F. T. Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, v. 66, p. 99-102, mar. 1971.
- Fiszbein, A.; Schady, N. Ferreira, F. *Conditional cash transfers: reducing present and future poverty*. Washington, DC: World Bank, 2009.
- FIRPO, S. et al. Evidence of eligibility manipulation for conditional cash transfer programs. *Economia*, Elsevier, v. 15, n. 3, p. 243-260, 2014.
- Foguel, M. N.; Barros, R. P. The effects of conditional cash transfer programmes on adult labour supply: an empirical analysis using a time-series-cross-section sample of Brazilian municipalities. *Economia e Sociedade*, v. 19, n. 3, p. 613-635, 2010.
- Glewwe, P.; Kassouf, A. L. The impact of the Bolsa Escola/Família conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, v. 97, n. 2, p. 505-517, 2012.
- Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica*, v. 50, n. 4, p. 1029-1054, 1982.
- Hayashi, F. *Econometrics*. Princeton: Princeton University Press, 2000.
- Haynes, K.; Eckley, I. A.; Killick, R. A computationally efficient nonparametric approach for changepoint detection. *Statistics & Computing*, v. 27, n. 5, p. 1293-1305, 2017.
- HOFFMANN, R. Transferências de renda e a redução da desigualdade no Brasil e cinco regiões entre 1997 e 2004. *Econômica*, v. 8, n. 1, p. 55-81, 2006
- Lindert, K.; Linder, A.; Hobbs, J.; De La Brière, B. The nuts and bolts of Brazil's Bolsa Família program: implementing conditional cash transfers in a decentralized context. World Bank Working Paper, Social Protection Discussion Paper No. 0709, maio 2007.
- Killick, R.; Fearnhead, P.; Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, v. 107, n. 500, p. 1590-1598, 2012.

Mccrary, J. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, v. 142, n. 2, p. 698-714, 2008.

MDS (Ministério do Desenvolvimento Social). Bolsa Família melhora a vida de milhões de pessoas há 21 anos. Brasília: MDS, 12 nov. 2024.

Porter, J.; Yu, P. Regression discontinuity designs with unknown discontinuity points: Testing and estimation. *Journal of Econometrics*, v. 189, n. 1, p. 132-147, 2015.

Shei, A. Brazil's conditional cash transfer program associated with declines in infant mortality rates. *Health Affairs*, v. 32, n. 7, p. 1274-1281, 2013.

Soares, F. V.; Ribas, R. P.; Osório, R. G. Evaluating the impact of Brazil's Bolsa Família: cash transfer programs in comparative perspective. *Latin American Research Review*, v. 45, n. 2, p. 173-190, 2010.

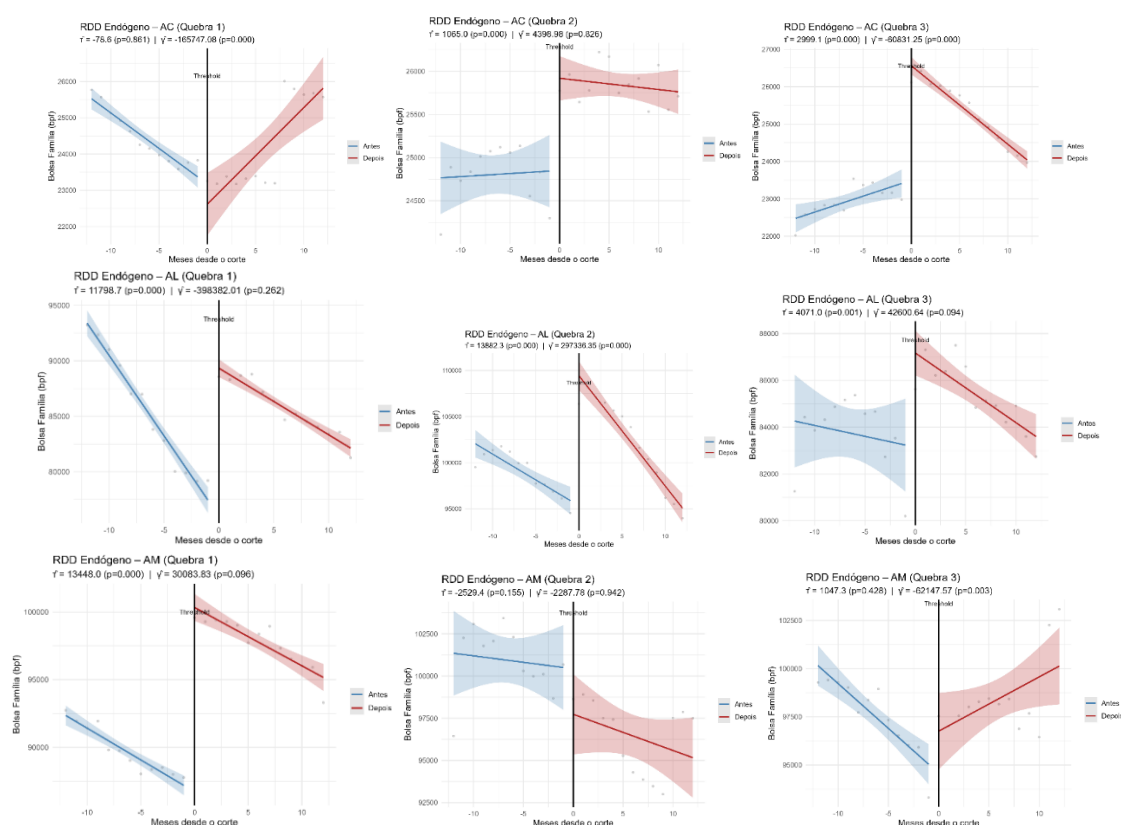
Souza, P. H. G. F.; Osório, R. G.; Paiva, L. H.; Soares, S. S. D. Os efeitos do Programa Bolsa Família sobre a pobreza e a desigualdade: um balanço dos primeiros quinze anos. Brasília: IPEA, 2019. (Texto para Discussão, n. 2499).

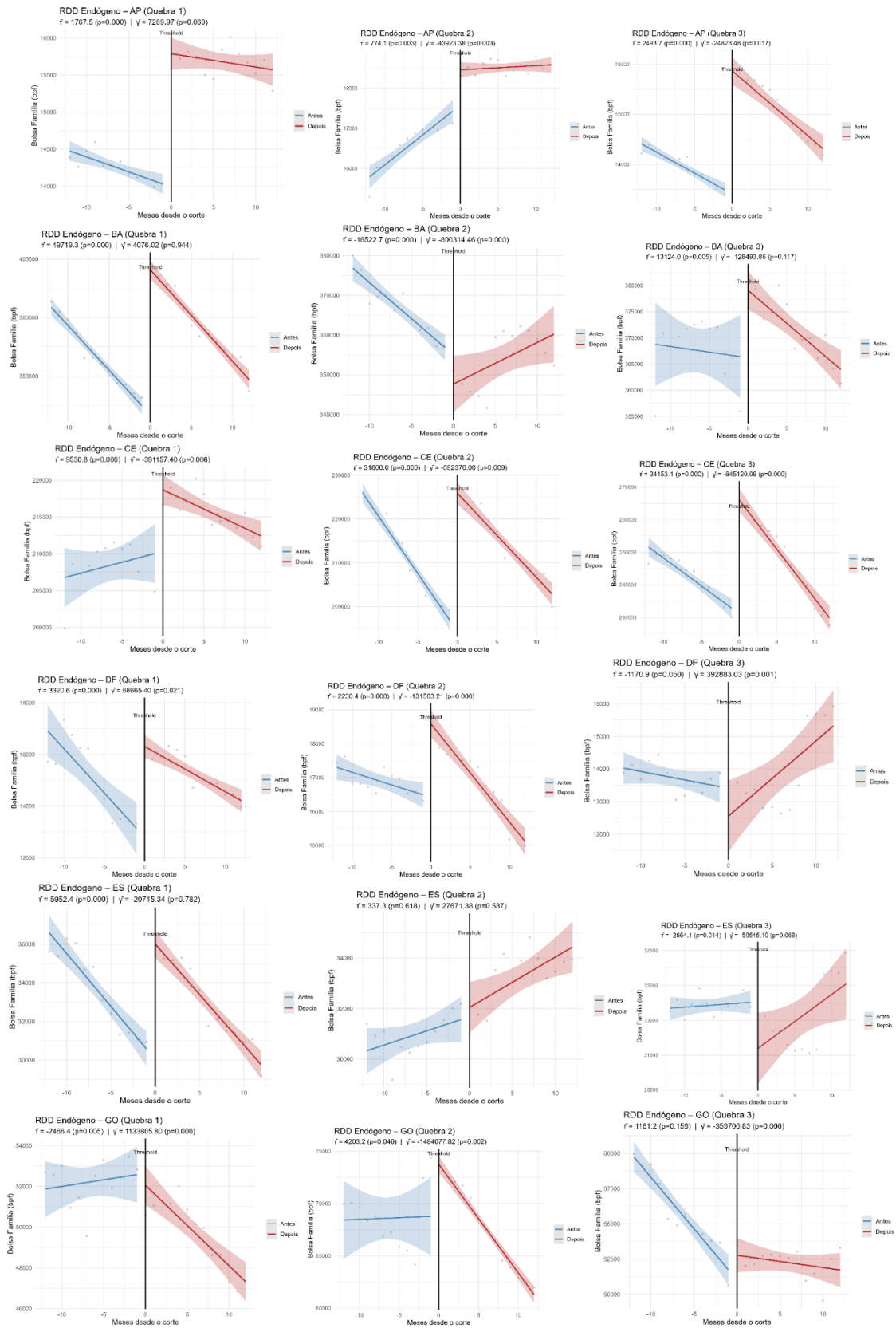
World Bank. *Brazil: Bolsa Família – 15 years of results*. Washington, DC: World Bank, 2019.

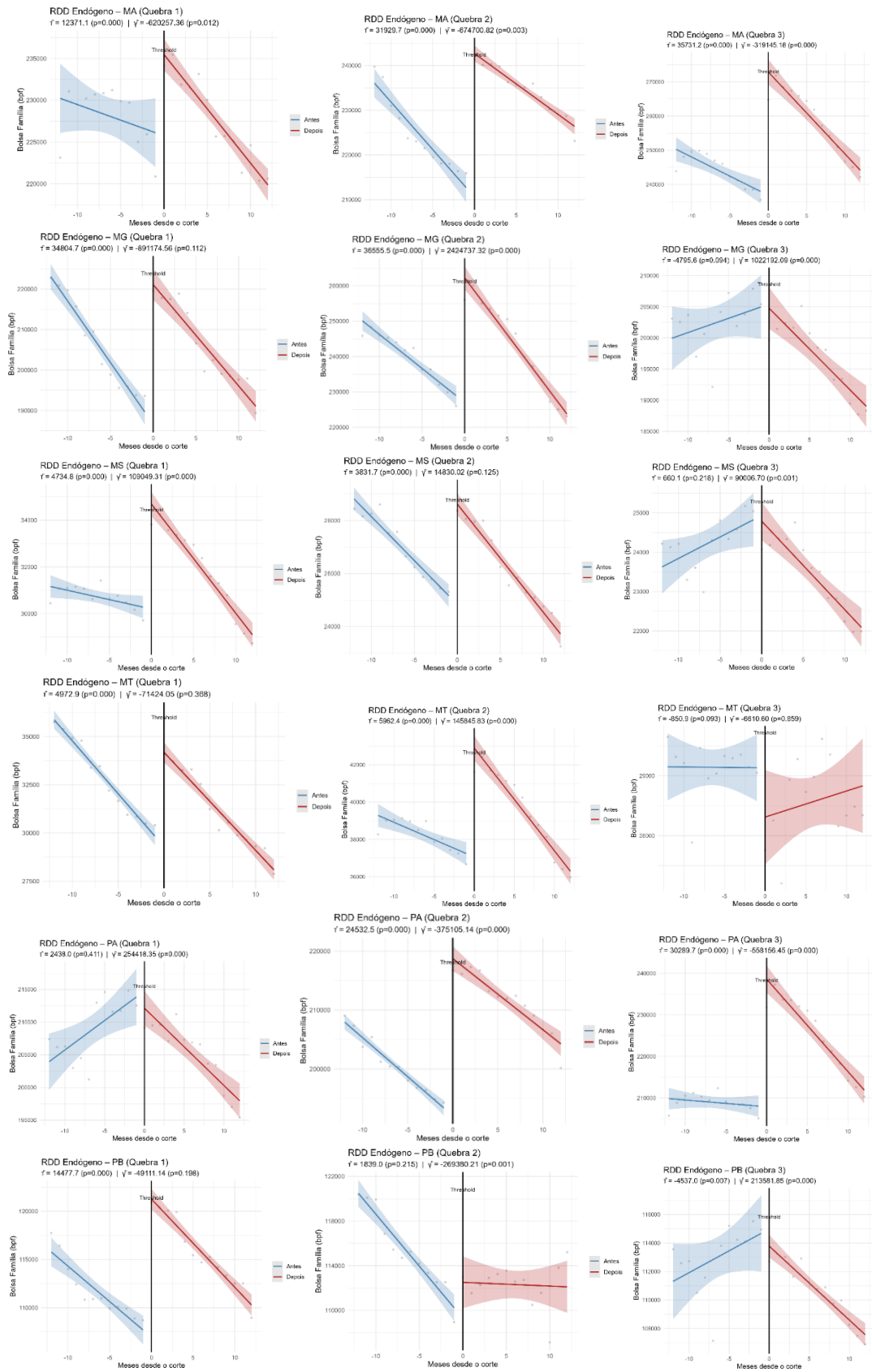
Zhang, N. R.; Siegmund, D. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, v. 63, p. 22-32, 2007.

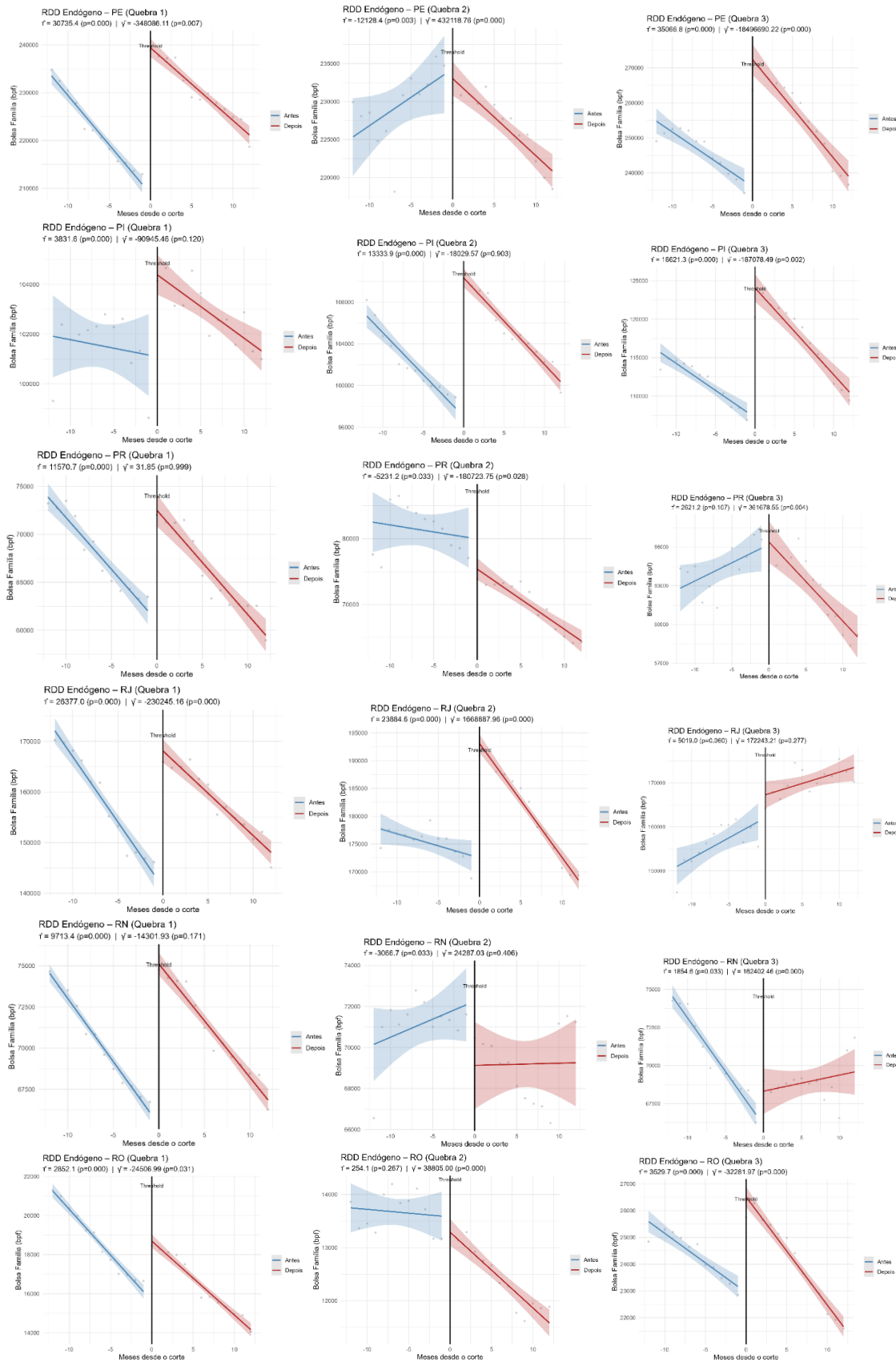
APÊNDICE

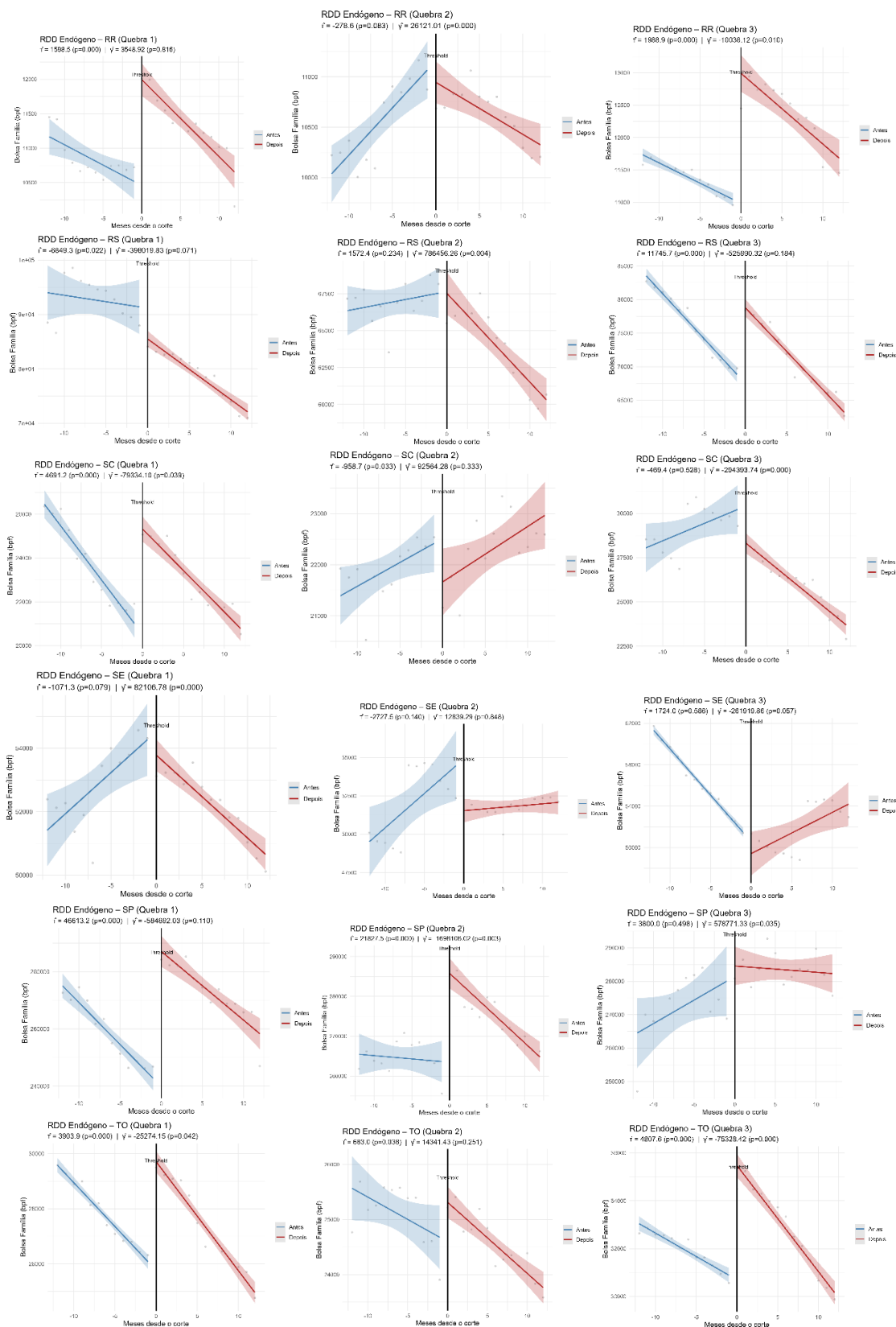
Figura 6 – Regressões Descontínuas Locais











Fonte: Elaboração própria com uso do R.

Tabela 9 – Teste McCrary no RDD local

<i>estado</i>	<i>N. quebra</i>	<i>theta</i>	<i>z_stat</i>	<i>pval_mccrary</i>
AC	1	0,287682	0,317552	0,750825

AC	2	0,2642	0,313274	0,754073
AC	3	-	-	-
-AL	1	0,287682	0,316094	0,751931
AL	2	-	-	-
AL	3	0,2642	0,313274	0,754073
AM	1	0,287682	0,316094	0,751931
AM	2	-	-	-
AM	3	0,287682	0,306781	0,75901
AP	1	0,287682	0,316094	0,751931
AP	2	0,2642	0,313274	0,754073
AP	3	-	-	-
BA	1	0,287682	0,316094	0,751931
BA	2	-	-	-
BA	3	0,2642	0,313274	0,754073
CE	1	0,2642	0,313274	0,754073
CE	2	0,287682	0,316094	0,751931
CE	3	-	-	-
DF	1	0,287682	0,316094	0,751931
DF	2	-	-	-
DF	3	-	-	-
ES	1	0,287682	0,316094	0,751931
ES	2	0,287682	0,324396	0,745638
ES	3	-	-	-
GO	1	-	-	-
GO	2	-	-	-
GO	3	0,287682	0,306781	0,75901
MA	1	0,2642	0,313274	0,754073
MA	2	0,287682	0,316094	0,751931
MA	3	-	-	-
MG	1	0,287682	0,316094	0,751931
MG	2	-	-	-
MG	3	-	-	-
MS	1	-	-	-
MS	2	0,287682	0,316094	0,751931
MS	3	-	-	-
MT	1	0,287682	0,316094	0,751931
MT	2	-	-	-
MT	3	0,287682	0,324396	0,745638
PA	1	-	-	-
PA	2	0,287682	0,316094	0,751931
PA	3	-	-	-
PB	1	0,287682	0,316094	0,751931
PB	2	0,287682	0,306781	0,75901
PB	3	-	-	-
PE	1	0,287682	0,316094	0,751931
PE	2	-	-	-
PE	3	-	-	-
PI	1	0,2642	0,313274	0,754073

PI	2	0,287682	0,316094	0,751931
PI	3			
PR	1	0,287682	0,316094	0,751931
PR	2			
PR	3			
RJ	1	0,287682	0,316094	0,751931
RJ	2			
RJ	3	0,2642	0,313274	0,754073
RN	1	0,287682	0,316094	0,751931
RN	2			
RN	3	0,287682	0,306781	0,75901
RO	1	0,287682	0,316094	0,751931
RO	2			
RO	3			
RR	1	0,287682	0,316094	0,751931
RR	2	-	-	-
RR	3	-	-	-
RS	1	-	-	-
RS	2	-	-	-
RS	3	0,287682	0,316094	0,751931
SC	1	0,287682	0,316094	0,751931
SC	2	0,287682	0,324396	0,745638
SC	3	-	-	-
SE	1	-	-	-
SE	2	0,285237	0,330758	0,740827
SE	3	0,287682	0,324479	0,745575
SP	1	0,287682	0,316094	0,751931
SP	2	-	-	-
SP	3	0,2642	0,313274	0,754073
TO	1	0,287682	0,316094	0,751931
TO	2	0,2642	0,313274	0,754073
TO	3	-	-	-

Fonte: Elaboração própria. Nota: - representa quebras não significativas estatisticamente e a Hipótese Nula (H0): densidade da variável em execução é contínua no ponto de corte. E a Hipótese Alternativa (Ha): Uma descontinuidade na função de densidade no ponto de corte, sugerindo manipulação.

EXERCÍCIOS

A terceira parte desloca a atenção do estimador isolado para a arquitetura de uma avaliação completa. O foco passa a ser a combinação de métodos, a interpretação de resultados aparentemente divergentes, a escolha do desenho mais adequado à pergunta e a construção de rotinas reprodutíveis. O objetivo é transformar o estudo de caso em um roteiro geral que possa ser aplicado a outras políticas sociais, educacionais, de saúde, trabalho ou desenvolvimento regional.

19. ESTRATÉGIAS COMBINADAS E TESTES DE ROBUSTEZ

Nenhum método deve ser avaliado apenas pelo número de asteriscos em uma tabela. Uma estratégia empírica convincente combina desenho, diagnóstico e análise de sensibilidade. A robustez relevante não consiste em repetir a mesma regressão com pequenas alterações, mas em testar se a conclusão depende de decisões arbitrárias. Em um estudo com quebras temporais, por exemplo, resultados podem ser reestimados com janelas alternativas, tendências distintas, exclusão de períodos extremos, diferentes matrizes de covariância e datas placebo.

A triangulação é particularmente útil quando métodos possuem falhas diferentes. Um RDD temporal pode oferecer uma estimativa local forte em torno da ruptura, enquanto um painel com efeitos fixos mostra se a relação persiste no conjunto do período. Uma ITS pode esclarecer se a mudança é de nível ou de inclinação. Um teste de quebra estrutural confirma se a data realmente corresponde a uma alteração na série. A concordância entre abordagens não prova causalidade automaticamente, mas reduz a probabilidade de que a conclusão seja artefato de uma única especificação.

19.1 Robustez por desenho

Testes de robustez devem ser organizados conforme a ameaça que procuram enfrentar. Para viés de seleção observável, matching e reponderação ajudam a avaliar sensibilidade à composição. Para tendências diferentes, estudos de evento e tendências específicas podem ser informativos. Para simultaneidade e endogeneidade, IV ou GMM podem ser considerados quando seus instrumentos são defensáveis. Para choques concorrentes, controles adicionais, séries de controle ou placebos temporais são mais úteis do que simplesmente adicionar polinômios.

Uma prática recomendada é criar uma matriz de robustez. Nas linhas, listam-se ameaças: seleção, tendências, manipulação, sazonalidade, autocorrelação, composição, mensuração e especificação. Nas colunas, registram-se diagnósticos e testes. Isso obriga o pesquisador a conectar cada teste a um problema concreto e evita a chamada robustez ornamental, em que muitos resultados adicionais são apresentados sem alterar o grau de confiança na interpretação.

19.2 Placebos e falsificações

Placebos reproduzem o método em situações nas quais não deveria existir efeito. Um falso ponto de corte anterior ao evento pode revelar que a série possui saltos frequentes e, portanto, que a descontinuidade principal não é excepcional. Um resultado placebo, biologicamente ou economicamente imune à política, ajuda a detectar confusão por choques gerais. Placebos geográficos verificam se áreas não tratadas exibem o mesmo padrão. Um bom teste placebo tem uma previsão clara: se o desenho é válido, o efeito deve desaparecer.

Falsificações devem ser interpretadas com simetria. Se um resultado significativo no desfecho principal é usado como evidência, um placebo também significativo precisa ser levado a sério. Não é adequado explicar post hoc todos os placebos inconvenientes e manter apenas o resultado favorável. A transparência sobre resultados nulos e contrários melhora a credibilidade do estudo.

19.3 Heterogeneidade de efeitos

Políticas sociais raramente produzem o mesmo efeito em todas as unidades. Heterogeneidade pode ocorrer por nível inicial de pobreza, desigualdade, capacidade administrativa, urbanização, tamanho populacional ou cobertura do programa. Estimar interações pré-especificadas ou efeitos por estratos pode revelar mecanismos importantes.

Contudo, quanto maior o número de subgrupos, maior o risco de resultados fortuitos. Correções para múltiplos testes e distinção entre análise confirmatória e exploratória são recomendadas.

No caso do Bolsa Família, diferenças regionais podem refletir simultaneamente composição populacional, mercado de trabalho e estrutura de elegibilidade. Por isso, encontrar coeficientes distintos por região não implica que a região em si seja a causa da heterogeneidade. Uma análise substantiva deve procurar variáveis que expliquem por que o efeito varia, evitando transformar rótulos geográficos em mecanismos.

20. COMO ESCOLHER O MÉTODO DE AVALIAÇÃO

A escolha do método começa pelos dados e pela regra institucional, não pelo software disponível. O primeiro diagnóstico é perguntar se existe uma fonte plausível de variação exógena. Se o tratamento muda exatamente em um limiar, RDD pode ser natural. Se alguns grupos são tratados e outros não, com observações antes e depois, DiD é candidato. Se uma unidade é tratada e há bons doadores, controle sintético pode funcionar. Se existe uma data única com série longa, ITS é uma opção. Se a seleção parece depender de observáveis ricos, matching pode complementar regressões.

20.1 Árvore decisória

- Existe regra de elegibilidade com ponto de corte observável? Considere RDD sharp ou fuzzy.
- Existem tratados e controles observados antes e depois? Considere DiD e estudo de evento.
- O tratamento ocorre em datas diferentes entre unidades? Use métodos de DiD para adoção escalonada.
- Existe instrumento plausível que afeta tratamento, mas não o resultado diretamente? Considere IV.
- Existe uma unidade tratada com amplo conjunto de doadores não tratados? Considere controle sintético.
- Existe uma intervenção em data definida e longa série temporal? Considere ITS, idealmente com controle.
- Não há variação quase experimental, mas há covariáveis pré-tratamento ricas? Regressão ajustada, matching e weighting podem produzir evidência associativa mais comparável.
- O resultado é persistente e há endogeneidade dinâmica em painel? GMM pode ser usado como complemento, com forte atenção aos instrumentos.

20.2 O método mais sofisticado nem sempre é o melhor

Sofisticação estatística não compensa um contrafactual ruim. Um DiD simples com excelente grupo de controle pode ser mais persuasivo do que um GMM com dezenas de instrumentos frágeis. Um RDD local transparente pode ser preferível a um modelo global com polinômio de alta ordem. A escolha deve priorizar hipóteses plausíveis e auditáveis. O leitor precisa conseguir entender por que determinadas observações servem de comparação para outras.

Também é necessário alinhar o método ao parâmetro de interesse. RDD estima efeito local no limiar; IV geralmente identifica efeito local para compliers; PSM pode estimar ATT sob seleção em observáveis; DiD estima efeito médio sob tendências paralelas para a população comparada; controle sintético foca uma unidade ou conjunto pequeno. Métodos podem responder perguntas diferentes mesmo quando usam a mesma base.

20.3 Checklist mínimo de identificação

- Qual é a unidade tratada e qual é o período de exposição?
- Qual é o resultado e em que momento ele é medido?
- Qual é o contrafactual?

- Qual hipótese transforma esse contrafactual em uma comparação válida?
- Que evidência empírica pode avaliar essa hipótese?
- Quais choques concorrentes ameaçam a interpretação?
- Qual é a população para a qual o efeito é válido?
- Qual decisão de especificação tem maior potencial de alterar a conclusão?

21. ROTEIRO DE REPLICAÇÃO EM R E PYTHON

Reprodutibilidade significa que outra pessoa, com acesso aos mesmos dados e códigos, consegue reconstruir as tabelas e figuras. Uma estrutura simples de projeto separa dados brutos, dados processados, scripts, resultados e documentação. O pesquisador não deve editar manualmente planilhas intermediárias sem registrar o procedimento. Toda transformação relevante precisa aparecer em código ou em um log reproduzível.

21.1 Estrutura de pastas

Uma estrutura recomendada utiliza as pastas 00_raw, 01_clean, 02_code, 03_output, 04_figures e 05_tables. Dados brutos são somente leitura. Scripts são numerados na ordem de execução: 01_import, 02_clean, 03_descriptive, 04_models e 05_export. Um arquivo README descreve fontes, versões de pacotes e ordem de execução. Essa disciplina é especialmente importante em projetos com séries estaduais, porque pequenas diferenças de data ou unidade monetária podem se propagar por centenas de estimações.

21.2 Exemplo conceitual em R

```
# Painei com efeitos fixos
library(fixest)
modelo_fe <- feols(pbf_real ~ gini + controles | uf + data,
                  cluster = ~uf, data = painel)

# DiD simples
modelo_did <- feols(resultado ~ tratado:pos | unidade + tempo,
                  cluster = ~unidade, data = base)

# RDD local
library(rdrobust)
ajuste_rdd <- rdrobust(y = y, x = running, c = 0, kernel = "triangular")
```

21.3 Exemplo conceitual em Python

```
# OLS com efeitos fixos via dummies
import statsmodels.formula.api as smf
mod = smf.ols("y ~ tratamento + x1 + x2 + C(uf) + C(tempo)", data=df)
res = mod.fit(cov_type="cluster", cov_kws={"groups": df["uf"]})

# Propensity score básico
from sklearn.linear_model import LogisticRegression
ps = LogisticRegression().fit(X, d).predict_proba(X)[: , 1]

# Estudo de evento: criar dummies de tempo relativo e estimar com FE
```

21.4 Validação automática

Boas rotinas incluem verificações automáticas: número esperado de unidades, datas mínima e máxima, ausência de duplicatas unidade-tempo, faixa admissível do Gini, não negatividade de repasses, fechamento anual do Denton-Cholette, contagem de observações por estado e consistência das categorias. Em vez de descobrir um erro no final, o código deve interromper a execução quando uma condição básica é violada.

Resultados também podem ser validados. Tabelas de regressão devem registrar número de observações, efeitos fixos, tipo de erro-padrão e amostra. Gráficos devem ser gerados diretamente dos objetos de análise. Sementes aleatórias devem ser fixadas em procedimentos estocásticos. A versão de pacotes pode ser preservada com renv no R ou ambientes virtuais no Python.

22. EXERCÍCIOS, QUESTÕES DE REVISÃO E PROPOSTAS DE PESQUISA

Os exercícios a seguir foram elaborados para transformar a leitura em prática de avaliação. Eles podem ser usados em cursos de econometria aplicada, métodos quantitativos, avaliação de políticas públicas ou orientação de trabalhos de conclusão. A ênfase não está apenas no cálculo, mas na capacidade de formular o contrafactual e reconhecer as limitações de cada estratégia.

22.1 Questões conceituais

- Explique por que a diferença média entre beneficiários e não beneficiários do Bolsa Família não identifica, em geral, o efeito causal do programa.
- Construa um exemplo em que uma comparação antes-depois superestime o efeito de uma política e outro em que subestime.
- Diferencie ATE, ATT e efeito local identificado por uma RDD.
- Explique a hipótese de tendências paralelas e descreva duas formas de investigá-la empiricamente.
- Por que balanceamento de covariáveis após PSM é mais importante que a acurácia preditiva do modelo de propensity score?
- Dê um exemplo de instrumento relevante, mas inválido por violar a restrição de exclusão.
- Explique por que o teste de McCrary não é suficiente para validar uma RDD temporal.
- Qual é o problema da proliferação de instrumentos em System-GMM?
- Quando um controle sintético com ajuste pré-tratamento ruim não deve ser interpretado causalmente?
- Por que uma quebra PELT estatisticamente significativa não identifica por si só a causa da mudança?

22.2 Exercícios empíricos

- Com um painel estado-mês, estime pooled OLS, FE e RE para repasses reais e Gini. Compare os coeficientes e discuta qual variação identifica cada modelo.
- Escolha uma data institucional e estime uma ITS com mudança de nível e inclinação. Reestime com erros Newey-West e compare a inferência.
- Implemente uma RDD com três larguras de banda e dois polinômios locais. Faça um gráfico com médias por bins e ajuste de cada lado do corte.
- Crie cinco datas placebo em torno da principal quebra e compare os efeitos com o efeito real.
- Construa uma análise PSM usando covariáveis pré-tratamento. Mostre suporte comum e diferenças padronizadas antes e depois do pareamento.
- Implemente um DiD com estudo de evento. Discuta se os coeficientes pré-tratamento dão suporte à hipótese de tendências paralelas.
- Estime Difference-GMM limitando instrumentos a duas janelas de defasagem e compare com uma especificação que use muitas defasagens. Relate AR(1), AR(2), Hansen/Sargan e número de instrumentos.

- Escolha uma unidade tratada hipotética e construa um controle sintético com doadores. Avalie o RMSPE pré-tratamento e realize placebos.

22.3 Propostas de pesquisa

- Avaliar efeitos heterogêneos de transferências de renda sobre pobreza por nível inicial de informalidade.
- Explorar limiares de elegibilidade para uma RDD fuzzy com microdados, distinguindo elegibilidade de recebimento efetivo.
- Combinar PSM e DiD para comparar municípios com diferentes intensidades de expansão de cobertura.
- Aplicar controle sintético a uma mudança subnacional de gestão ou integração cadastral.
- Construir uma ITS controlada para uma reforma de benefício, usando resultado não afetado como série de comparação.
- Investigar efeitos dinâmicos com event study e métodos para adoção escalonada em políticas municipais.

23. CONSIDERAÇÕES FINAIS DO LIVRO

A avaliação quantitativa de políticas públicas é um exercício de construção de comparações críveis. O valor de um método não está apenas em sua complexidade matemática, mas na clareza com que transforma dados observados em um contrafactual defensável. Regressão linear, efeitos fixos, DiD, matching, IV, ITS, controle sintético, RDD e GMM ocupam posições diferentes nesse processo. Conhecer vários métodos amplia o repertório, mas também torna mais importante saber quando não usar cada um.

O estudo empírico do Bolsa Família mostra por que essa abordagem plural é útil. Séries estaduais, mudanças temporais, diferenças de desigualdade e alterações no padrão de repasses permitem combinar descrição, detecção de quebras, estimação local e modelos dinâmicos. Ao mesmo tempo, o caso evidencia limites: uma data de quebra pode coincidir com vários eventos; um Gini mensal desagregado não contém a mesma informação de um indicador observado mensalmente; e significância estatística não substitui identificação institucional.

O principal aprendizado é que uma boa avaliação deve deixar explícitas as hipóteses que sustentam cada resultado. Quando a conclusão depende de tendências paralelas, essa condição deve ser investigada. Quando depende de um instrumento, a exclusão precisa ser defendida. Quando depende de um corte, manipulação, continuidade e eventos concorrentes precisam ser examinados. Quando depende de instrumentos internos em GMM, a matriz de instrumentos e os testes de autocorrelação são parte do resultado, e não notas técnicas secundárias.

Em aplicações futuras, a estratégia mais promissora é combinar transparência do desenho com reprodutibilidade computacional. Bases documentadas, códigos organizados, placebos, especificações alternativas e interpretação cuidadosa tornam a evidência mais útil para pesquisa e gestão. O objetivo final não é produzir um estimador sofisticado, mas reduzir a incerteza sobre o que uma política fez, para quem, em que contexto e sob quais condições.

A revisão da literatura reforça que resultados sólidos surgem menos da complexidade isolada do estimador e mais da coerência entre regra institucional, estrutura dos dados e hipótese de identificação. No caso do Bolsa Família, estudos de pobreza, educação, saúde e mercado de trabalho respondem perguntas diferentes e, por isso, podem produzir parâmetros de magnitude distinta sem que um necessariamente invalide o outro. Uma avaliação madura explicita essa diversidade, evita extrapolações entre escalas e reconhece quando o resultado é local, agregado, descritivo ou causal.

A principal contribuição metodológica desta edição ampliada é mostrar que o RDD utilizado no estudo original pode ser colocado dentro de um repertório maior. DiD, PSM, IV, ITS, controle sintético, painéis e GMM oferecem respostas complementares quando suas hipóteses são plausíveis. Em aplicações futuras, a ampliação mais importante não é adicionar métodos por acumulação, mas selecionar aqueles capazes de explorar fontes independentes de identificação e confrontar a mesma hipótese com evidências que falham por razões diferentes.

REFERÊNCIAS COMPLEMENTARES

- ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, v. 105, n. 490, p. 493-505, 2010.
- ANGRIST, J. D.; PISCHKE, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press, 2009.
- ARELLANO, M.; BOND, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, v. 58, n. 2, p. 277-297, 1991.
- BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, v. 18, n. 1, p. 1-22, 2003.
- BARBOSA, A. L. N. H.; CORSEUIL, C. H. L. Conditional cash transfer and informality in Brazil. *IZA Journal of Labor & Development*, v. 3, art. 37, 2014. DOI: 10.1186/s40175-014-0024-0.
- BERTRAND, M.; DUFLO, E.; MULLAINATHAN, S. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, v. 119, n. 1, p. 249-275, 2004.
- BLUNDELL, R.; BOND, S. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, v. 87, n. 1, p. 115-143, 1998.
- CALLAWAY, B.; SANT'ANNA, P. H. C. Difference-in-differences with multiple time periods. *Journal of Econometrics*, v. 225, n. 2, p. 200-230, 2021.
- CALONICO, S.; CATTANEO, M. D.; TITIUNIK, R. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, v. 82, n. 6, p. 2295-2326, 2014.
- DE BRAUW, A.; GILLIGAN, D. O.; HODDINOTT, J.; ROY, S. Bolsa Família and household labor supply. *Economic Development and Cultural Change*, v. 63, n. 3, p. 423-457, 2015. DOI: 10.1086/680092.
- FOGUEL, M. N.; BARROS, R. P. The effects of conditional cash transfer programmes on adult labour supply: an empirical analysis using a time-series-cross-section sample of Brazilian municipalities. *Estudos Econômicos*, v. 40, n. 2, p. 259-293, 2010.
- GLEWWE, P.; KASSOUF, A. L. The impact of the Bolsa Escola/Família conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, v. 97, n. 2, p. 505-517, 2012. DOI: 10.1016/j.jdevco.2011.05.008.
- GOODMAN-BACON, A. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, v. 225, n. 2, p. 254-277, 2021.
- HECKMAN, J. J.; ICHIMURA, H.; TODD, P. E. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies*, v. 64, n. 4, p. 605-654, 1997.
- HERNÁN, M. A.; ROBINS, J. M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.
- IMBENS, G. W.; RUBIN, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press, 2015.
- KILLICK, R.; FEARNHEAD, P.; ECKLEY, I. A. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, v. 107, n. 500, p. 1590-1598, 2012.
- LEE, D. S.; LEMIEUX, T. Regression discontinuity designs in economics. *Journal of Economic Literature*, v. 48, n. 2, p. 281-355, 2010.
- MAGALHÃES, J. et al. Health, economic and social impacts of the Brazilian cash transfer program on the lives of its beneficiaries: a scoping review. *BMC Public Health*, v. 24, art. 2818, 2024. DOI: 10.1186/s12889-024-20046-2.
- MCCRARY, J. Manipulation of the running variable in the regression discontinuity design: a density test.

Journal of Econometrics, v. 142, n. 2, p. 698-714, 2008.

RASELLA, D.; AQUINO, R.; SANTOS, C. A. T.; PAES-SOUSA, R.; BARRETO, M. L. Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. *The Lancet*, v. 382, n. 9886, p. 57-64, 2013. DOI: 10.1016/S0140-6736(13)60715-1.

RIBEIRO, F. G.; SHIKIDA, C.; HILLBRECHT, R. O. Bolsa Família: um survey sobre os efeitos do programa de transferência de renda condicionada do Brasil. *Estudos Econômicos*, v. 47, n. 4, p. 805-862, 2017. DOI: 10.1590/0101-416147468fcr.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41-55, 1983.

SANTOS, M. C. S.; DELATORRE, L. R.; CECCATO, M. G. B.; BONOLO, P. F. Programa Bolsa Família e indicadores educacionais em crianças, adolescentes e escolas no Brasil: revisão sistemática. *Ciência & Saúde Coletiva*, v. 24, n. 6, p. 2233-2247, 2019. DOI: 10.1590/1413-81232018246.19582017.

SOUZA, P. H. G. F.; OSORIO, R. G.; PAIVA, L. H.; SOARES, S. Os efeitos do Programa Bolsa Família sobre a pobreza e a desigualdade: um balanço dos primeiros quinze anos. Rio de Janeiro: Ipea, 2019. Texto para Discussão n. 2499.

SUN, L.; ABRAHAM, S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, v. 225, n. 2, p. 175-199, 2021.

TRIBUNAL DE CONTAS DA UNIÃO. Auditoria de focalização e equidade do novo Bolsa Família. Brasília: TCU, 2024.

WOOLDRIDGE, J. M. *Econometric Analysis of Cross Section and Panel Data*. 2. ed. Cambridge, MA: MIT Press, 2010.

APÊNDICE A - FORMULÁRIO DE LEITURA DE ARTIGOS DE AVALIAÇÃO

Pergunta de pesquisa

Resuma a pergunta em uma única frase. Identifique resultado, tratamento, população e período. Reescreva a pergunta em linguagem contrafactual: o que teria acontecido com as unidades tratadas se não tivessem recebido a intervenção?

Fonte de variação

Descreva por que algumas unidades recebem mais tratamento do que outras. A variação vem de regra, tempo, geografia, elegibilidade, implementação ou escolha individual? Classifique a variação como potencialmente exógena, condicionalmente exógena ou claramente endógena.

Grupo de comparação

Identifique quem serve de controle e por quê. Verifique se controles poderiam, em princípio, representar o resultado potencial dos tratados. Registre diferenças prévias de nível, tendência e composição.

Hipótese de identificação

Escreva a hipótese central sem notação matemática. Em seguida, liste evidências que a tornam plausível e evidências que poderiam refutá-la.

Inferência

Registre nível de agregação do tratamento, estrutura dos erros-padrão, número de clusters e eventuais correções para pequenas amostras. Verifique se a inferência respeita o desenho.

Robustez

Classifique cada teste adicional conforme a ameaça que enfrenta. Diferencie robustez substantiva de mudanças cosméticas de especificação.

Generalização

Defina para quem o efeito é válido. Resultados locais, efeitos para tratados e efeitos para compliers não devem ser descritos como se fossem efeitos populacionais gerais.



9786583199829 ISBN 978-65-83199-82-9

