

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EM GESTÃO E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA APLICADA

Bruna Sineta Scarpioni

**UMA INVESTIGAÇÃO EMPÍRICA DA RELAÇÃO ENTRE ESCRAVIDÃO
E VIOLÊNCIA LETAL NO BRASIL CONTEMPORÂNEO**

Sorocaba

2026

Bruna Sineta Scarpioni

**UMA INVESTIGAÇÃO EMPÍRICA DA RELAÇÃO ENTRE ESCRAVIDÃO
E VIOLÊNCIA LETAL NO BRASIL CONTEMPORÂNEO**

Dissertação apresentada ao Programa de Pós-Graduação em Economia Aplicada da Universidade Federal de São Carlos para obtenção do título de Mestre em Economia.
Área de concentração: Desenvolvimento Econômico, Sociedade e Meio Ambiente.

Orientação: Prof. Dr. Geraldo Edmundo Silva Júnior.

Financiamento: CAPES/DS

Sorocaba

2026

Scarpioni, Bruna Sineta

Uma investigação empírica da relação entre escravidão e violência letal no Brasil contemporâneo / Bruna Sineta Scarpioni -- 2023.
131f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus Sorocaba, Sorocaba

Orientador (a): Geraldo Edmundo Silva Júnior

Banca Examinadora: Andrea Rodrigues Ferro, Jorge Paulo de Araújo

Bibliografia

1. Escravidão. 2. Violência. 3. Causalidade. I. Scarpioni, Bruna Sineta. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Maria Aparecida de Lourdes Mariano -
CRB/8 6979

BRUNA SINETA SCARPIONI

UMA INVESTIGAÇÃO EMPÍRICA DA RELAÇÃO ENTRE ESCRAVIDÃO
E VIOLÊNCIA LETAL NO BRASIL CONTEMPORÂNEO

Dissertação apresentada ao Programa de Pós-Graduação em Economia Aplicada da Universidade Federal de São Carlos para obtenção do título de Mestre em Economia. Desenvolvimento Econômico, Sociedade e Meio Ambiente Sorocaba, 11 de setembro de 2023.

Orientador

Dr. Geraldo Edmundo Silva Júnior
UFSCar

Examinadora

Dra. Andrea Rodrigues Ferro
UFSCar

Examinador

Dr. Jorge Paulo de Araújo
UFRGS

Para Sandra Padula, a primeira a me
ouvir.

Para Carlos André, que fez/faz tudo
valer a pena.

Para Danilo Canato, que me levou pelo
mundo com ele

Para Flávia e Uillian, que passaram por
isso comigo.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela proteção, pela misericórdia e pela inspiração deste trabalho, surgida em outubro de 2018. Agradeço, sobretudo, pelo papel de cada um dos citados aqui na minha vida e na elaboração deste trabalho

Agradeço ao Prof. Dr. Geraldo Edmundo Silva Júnior pela orientação, e sobretudo, pela imensa paciência e pela sabedoria. Agradeço por ter mantido sua integridade e seu coração puro diante do esfriamento do amor do mundo. Agradeço a Prof^{ta} Dra^a Andrea Rodrigues Ferro, também pela paciência e pelos conselhos, por deixar a porta sempre aberta nos momentos mais confusos...A pesquisa tinha problemas complexos, mas para realiza-la, todos tiveram que lidar com os problemas que eu, Bruna, tinha e tenho como ser humano. Agradeço ao Prof^o Jorge Paulo de Araújo pelas contribuições valiosas, através de perguntas cruciais para entender os limites e o significado deste estudo.

Agradeço imensamente às irmãs que a vida me deu, Luísa Fancelli Coelho e Laura Grimaldo Hidalgo. Obrigada por me enxergarem apesar de mim mesma, por me ensinarem tanto, por tornarem a vida tão mais leve e prazerosa. Obrigada por serem modelos pra mim.

Agradeço à minha mãe, Marisa Sineta, e à minha avó, Anésia Maria Sineta, por serem o norte para onde eu sempre posso olhar e encontrar força. Obrigada pela coragem, pela força, pelo amor e pela sabedoria que me transmitiram, por todos os sacrifícios que fizeram para que agora este trabalho estivesse aqui. Ele é nosso. É uma vitória nossa. Se eu venci, então é porque vocês venceram muito mais.

Agradeço a Sandra Padula, minha neuropsicóloga, por ter me escutado e me ajudado a entender e a sobreviver, pelos olhos com que me viu. Agradeço a Flávio Geraldес Alves, meu neurologista, sempre celebrando cada pequena vitória desde o início do meu tratamento. Agradeço a Danilo Bortoli, decisivo nos momentos mais difíceis e absurdos, conseguindo vitórias que eu jamais imaginaria ter.

Agradeço a Carlos André Sanches de Carvalho. O seu amor tornou tudo possível. Tudo. Obrigada por eu te contar toda a verdade sobre tudo e ainda assim, permanecer ao meu lado e me fazer sentir digna, além de amada. O mundo se tornou seguro quando você chegou porque agora eu tinha um lugar pra onde sempre voltar, um motivo pra continuar e uma certeza de que no final, eu ia conseguir. Obrigada por ter tanta força pra ser bom, pra ser doce e pra ser exatamente como você é. Te amo.

Agradeço a Danilo Canato, por sempre ver em mim o que eu podia ser, por alguma razão quase que mágica. Agradeço a Carlos por ter me trazido o encontro com aquele que mudou a

minha história para sempre, por vários motivos. Pelo seu interesse nesta pesquisa, pelo seu entusiasmo contagiante sobre a vida e a verdade. Obrigada por me mostrar que é legítimo ser a gente mesmo, e o espaço pra isso, a gente é que constrói. Por mostrar que não existe um grande teste depois do qual se está autorizado a decidir, ser, fazer, pensar o que se pensa. Obrigada por me dar uma chance, e com essa chance, por passar comigo por tantos problemas e dores que impediam, inclusive (ou principalmente) a publicação desta pesquisa. Obrigada pela sua sabedoria, liderança e pelo tanto que você se importa com as pessoas que lidera, como seres humanos e como profissionais.

Agradeço a Flávia Mendonça, madrinha desta dissertação. Obrigada por passar comigo por tantas questões e nunca ter deixado de acreditar em mim. Obrigada pela dignidade com que você sempre me viu e pela torcida incansável pela Dirce. Conseguimos, apipa. Obrigada a Uilian Mendonça, pela nossa equipe, pela compreensão e pela inspiração que levaram à conclusão da revisão deste estudo. Obrigada por ter nos protegido e ouvido tantas vezes, por tudo o que nos ensinou e pela referência de profissional e pessoa que você é.

Agradeço a Raphael Gouvêa e a Cláudio Amitrano, por acreditarem em mim, pela paciência, pelo exemplo, pela humanidade. Obrigada por mudarem a academia de Economia no Brasil.

Obrigada a todos os meus alunos, cuja necessidade do que eu podia lhes fornecer foi a base da minha autoestima por tantos anos em que essa pesquisa não teve resultado. Pela imagem de mim que eu via nos seus olhos enquanto lhes ajudava a aprender algo que poderia mudar suas vidas. Obrigada por terem tornado a minha existência significativa neste mundo. Obrigada Nicolly, Tatiane, Bruna, Guilherme, Rafaelle, Lívia, Ana Carolina, Raphael, Ana Luiza e tantos presentes que a vida me deu. Ensinar vocês foi meu grande privilégio.

Agradeço a André Kërr e àqueles que me ensinaram tanto sobre mim, sobre a vida, e sobre o Caminho Eterno.

Agradeço a todos os autores citados nas Referências Bibliográficas. Agradeço a todos que se dedicaram a estudar a escravidão e a violência no Brasil diante de tantos obstáculos. Agradeço aos estudiosos de causalidade com tratamentos contínuos, sem quem este trabalho jamais existiria. Quando começou a ser escrito, em 2019, nem mesmo as bibliotecas do R estavam disponíveis para fazê-lo. O primeiro resultado surgiu em maio de 2021, em uma noite, e eu tenho o print da tela até hoje – e ainda seriam necessários muitos, e muitos, e muitos ajustes e correções. Agradeço a todos aqueles que me ajudaram com a pesquisa de diferentes formas,

que enviaram dados, responderam e-mails e doaram seu tempo e paciência. Agradeço a Noah Greifer pelas contribuições inestimáveis, decisivas para próximos trabalhos sobre o tema.

Agradeço imensamente ao Conselho Nacional de Pesquisa (CNPq) pelo apoio financeiro para realização deste estudo. Agradeço à UFSCar e ao PPGEc. Esta pesquisa não poderia ter sido realizada em nenhum outro centro. Obrigada por todas as experiências e pela oportunidade que o mestrado me proporcionou e proporciona. Obrigada, inclusive, pelas decisões, pressões e punições iminentes ou definitivas que me tiraram da paralisia e do terror de seguir com este trabalho.

Agradeço a todos que contribuíram, em maior ou menor medida, para que este trabalho fosse publicado e para que eu me tornasse quem me tornei. Obrigada por tudo o que me ensinaram e por fazerem ou terem feito parte dessa história.

E disse Deus: Que fizeste? A voz do
sangue do teu irmão clama a mim desde a
terra.

E agora maldito és tu desde a terra, que
abriu a boca para receber da tua mão o sangue
do teu irmão.

Quando lavrares a terra, não te dará
mais a sua força; fugitivo e vagabundo serás
na terra.

(...)

O Senhor, porém, disse-lhe: Portanto,
qualquer um que matar a Caim, sete vezes será
castigado. E pôs um sinal em Caim, para que o
não ferisse qualquer que o achasse.

(BÍBLIA SAGRADA. Gênesis, 4:10-
12; 15. Tradução: Almeida Corrigida Fiel)

RESUMO

SCARPIONI, Bruna Sineta. Uma investigação empírica da relação entre escravidão e violência letal no Brasil contemporâneo. 2023. Dissertação (Mestrado em Economia no Programa de Pós-Graduação em Economia Aplicada) – Universidade Federal de São Carlos, Sorocaba, 2023.

Desde a década de 1980, a escalada da criminalidade tem motivado investigações sobre as causas da violência no Brasil. Enquanto a maioria dos estudos quantitativos se refere a fatores contemporâneos ao fenômeno, a literatura sociológica aponta a escravidão como determinante histórico da estrutura social que ensejaria a violência no país. Este trabalho preenche a lacuna de estudos empíricos sobre as causas históricas da violência no Brasil, ao verificar a existência de correlações positivas, significativas e robustas entre o percentual de pessoas escravizadas nos municípios brasileiros em 1872 e as taxas municipais de homicídios no ano 2000. Como a variável que representa a escravidão assume diferentes valores entre os municípios do grupo tratado, utiliza-se o método de ponderação das observações por *propensity score* generalizado, que gera estimativas correlacionais robustas da relação entre o tratamento contínuo e a variável resultante. A amostra de municípios do Nordeste apresentou correlações positivas, mas não significantes. Amostras referentes ao Sudeste e à união das regiões Sudeste e Nordeste apresentaram correlações consistentemente positivas, cujos níveis de significância variaram ao longo das especificações dos modelos. Os estimadores são positivos, significantes e robustos para a amostra que reúne todo o território nacional, controlando efeitos fixos por estado. Os resultados obtidos foram robustos à inclusão do controle da desigualdade de renda e de sua interação com a renda média domiciliar *per capita*. Devido às diferentes dinâmicas entre desigualdade, renda e violência nos municípios brasileiro, a interação foi altamente significativa. A correlação positiva, significativa e robusta entre a intensidade da escravidão em 1872 (mensurada em seu período de declínio) e a violência letal em 2000 para a amostra que reuniu todas as regiões geográficas é uma evidência de transbordamento da influência da escravidão sobre a violência contemporânea. Os resultados sugerem que o efeito pernicioso da escravidão sobre a violência “transbordou” sensivelmente dos municípios escravagistas, os únicos ocupados em 1872, para todo o território nacional.

Palavras-chave: escravidão; violência; *propensity score weighting*; causalidade.

ABSTRACT

SCARPIONI, Bruna Sineta. An empirical inquiry into the relationship between slavery and lethal violence in contemporary Brazil. 2023. Thesis (Master's in Economics at Economics Graduate Course – Federal University of São Carlos, Sorocaba, 2023).

Since the 1980s, the escalation of crime has motivated investigations into the causes of violence in Brazil. While most quantitative studies focus on contemporary factors, sociological literature points to slavery as a historical determinant of the social structure that would engender violence in the country. This study fills the gap in empirical research on the historical causes of violence in Brazil by verifying the existence of positive, significant, and robust correlations between the percentage of enslaved people in Brazilian municipalities in 1872 and municipal homicide rates in 2000. Since the variable representing slavery assumes different values among municipalities in the treated group, the generalized propensity score weighting method is employed, which generates robust correlational estimates of the relationship between continuous treatment and the outcome variable. The sample of municipalities in the Northeast region showed positive but non-significant correlations. Samples from the Southeast region and the combined Southeast and Northeast regions exhibited consistently positive correlations, whose significance levels varied across model specifications. The estimates are positive, significant, and robust for the sample encompassing the entire national territory, controlling for state fixed effects. The results obtained were robust to the inclusion of income inequality controls and its interaction with average household per capita income. Due to different dynamics between inequality, income, and violence in Brazilian municipalities, the interaction term was highly significant. The positive, significant, and robust correlation between slavery intensity in 1872 (measured during its period of decline) and lethal violence in 2000 for the sample comprising all geographic regions provides evidence of spillover effects of slavery's influence on contemporary violence. The results suggest that the pernicious effect of slavery on violence "spilled over" significantly from slave-holding municipalities—the only ones occupied in 1872—to the entire national territory.

Keywords: slavery; violence; propensity score weighting; causality.

LISTA DE FIGURAS

Figura 1 - Distribuição da intensidade da escravidão (%) em 1872	46
Figura 2 - Intensidade da escravidão em 1872 por região geográfica	47
Figura 3 - Correlação de Pearson entre controles e tratamento (<i>sld</i>).....	48
Figura 4 - Descrição resumida das variáveis	63
Figura 5 - Correlação de Pearson na pseudoamostra: Sudeste	68
Figura 6 - Correlação de Pearson na pseudoamostra: Nordeste	69
Figura 7 - Correlação de Pearson na pseudoamostra: Nordeste e Sudeste.....	71
Figura 8 - Correlação de Pearson na pseudoamostra: Brasil.....	73
Figura 9 - Relação entre escravidão e violência: Sudeste	78
Figura 10 - Gráficos de Johnson-Neyman para a região Sudeste.....	80
Figura 11 - Descrição dos intervalos de Johnson-Neyman para a região Sudeste	82
Figura 12 - Intervalos de Johnson-Neyman para a região Sudeste.....	82
Figura 13 - Relação entre escravidão e violência: Nordeste	84
Figura 14 - Relação entre escravidão e violência: Nordeste	87
Figura 15 - Descrição dos intervalos de Johnson-Neyman para a região Nordeste.....	88
Figura 16 - Intervalos de Johnson-Neyman para a Região Sudeste	88
Figura 17 - Relação entre escravidão e violência: Nordeste e Sudeste	89
Figura 18 - Gráficos de Johnson-Neyman para as regiões Nordeste e Sudeste	92
Figura 19 - Descrição dos intervalos de Johnson-Neyman para as regiões Nordeste e Sudeste	93
Figura 20 - Intervalos de Johnson-Neyman para as regiões Nordeste e Sudeste	93
Figura 21 - Efeito da escravidão sobre a violência: Brasil.....	95
Figura 22 - Gráficos de Johnson-Neyman para o Brasil	98
Figura 23 - Descrição dos intervalos de Johnson-Neyman para o Brasil	99
Figura 24 - Intervalos de Johnson-Neyman para o Brasil	99

LISTA DE ABREVIATURAS E SIGLAS

AMCs	Áreas Mínimas Comparáveis
ATE	<i>Average Treatment Effect</i>
CAC	Caçadores, Atiradores e Colecionadores de Armas de Fogo
CID-10	Classificação Internacional de Doenças (10ª Revisão)
DATASUS	Base de Dados do Sistema Único de Saúde
ENEM	Exame Nacional do Ensino Médio
FCO	Fundo Constitucional de Financiamento do Centro-Oeste
FNO	Fundo Constitucional do Norte
FBSP	Fórum Brasileiro de Segurança Pública
GBM	<i>Generalized Boosted Model</i>
GPS	<i>Generalized Propensity Scores</i>
GPS-IPW	<i>Generalized Propensity Scores- Inverse Probability Weighting</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH-M	Índice de Desenvolvimento Humano Municipal
IEP	<i>Institute for Economics and Peace</i>
INPC	Índice Nacional de Preços ao Consumidor
IPW	<i>Inverse Probability Weighting</i>
MS	Ministério da Saúde
MUNIC	Pesquisa de Informações Básicas Municipais (antes Perfil dos Municípios Brasileiros)
NEV-USP	Núcleo de Estudos da Violência da Universidade de São Paulo
PIB	Produto Interno Bruto
PNAD	Pesquisa Nacional por Amostra de Domicílios
PNUD	Programa das Nações Unidas para o Desenvolvimento
PSG	<i>Propensity Score</i> Generalizado
PSW	<i>Propensity Score Weighting</i>
SIM	Sistema de Informações sobre Mortalidade
SUSP	Sistema Único de Segurança Pública
UF	Unidade(s) Federativa(s)

SUMÁRIO

1 INTRODUÇÃO.....	13
2 REVISÃO DE LITERATURA E FUNDAMENTAÇÃO TEÓRICA.....	23
2.1 A PERSISTÊNCIA DO LEGADO DA ESCRAVIDÃO.....	23
2.2 A VIOLÊNCIA COMO LEGADO DA ESCRAVIDÃO NO BRASIL	30
3 METODOLOGIA.....	45
3.1 PECULIARIDADES DA AMOSTRA: IMPLICAÇÕES METODOLÓGICAS	45
3.2 GPS-IPW POR <i>BOOSTING</i> E FUNÇÃO DOSE-RESPOSTA	50
3.3 DADOS E MODELOS	59
4 RESULTADOS E DISCUSSÃO.....	67
4.1 DIAGNÓSTICO DE EQUILÍBRIO DAS AMOSTRAS PONDERADAS.....	67
4.1.1 Sudeste	68
4.1.2 Nordeste	69
4.1.3 Nordeste e Sudeste	71
4.1.4 Brasil	73
4.2 RESULTADOS DAS REGRESSÕES PONDERADAS.....	77
4.2.1 Sudeste	78
4.2.2 Nordeste	84
4.2.3 Nordeste e Sudeste	89
4.2.4 Brasil	95
5 CONSIDERAÇÕES FINAIS.....	103
REFERÊNCIAS	112
APÊNDICE A – RESULTADOS COM VARIÁVEIS BINÁRIAS DE UF.....	123

1 INTRODUÇÃO

No Índice Global de Paz de 2023, o Brasil ficou em 132º lugar em relação aos 163 países analisados (sendo o 1º o mais pacífico, e o 163º, o mais turbulento) (INSTITUTE FOR ECONOMICS & PEACE, 2023), superior apenas a países da América Central, da África Central e do Leste Europeu envolvidos em conflitos paramilitares ou em guerras internacionais. Em relação ao custo econômico da violência como percentual do Produto Interno Bruto (PIB), o país ocupa o 30º lugar. O indicador descreve “segurança social” como o estado caracterizado por baixas taxas criminais, relações harmoniosas com países vizinhos, cenário político estável e baixo percentual de migração doméstica ou de refugiados entre a população. Neste quesito, o Brasil ficou em 151º lugar em 2023, atrás de países como México, Sudão e África do Sul (INSTITUTE FOR ECONOMICS & PEACE, 2023).

Em relação ao Índice de Percepção de Segurança de 2023, o Brasil obteve taxas de preocupação com a violência superiores às taxas de vitimização autodeclarada. Observada pelos pesquisadores desde 2022, essa diferença foi apontada como reflexo de esforços de individuais ou coletivos de proteção para lidar com riscos de violência realmente altos. Na ausência desses esforços, as taxas de ocorrência seriam muito maiores. Contudo, a prevenção de ocorrências efetivas não é capaz de reduzir a percepção de insegurança que motivou a adoção das medidas de proteção. Uma evidência a favor dessa hipótese é o alto gasto com segurança privada no país (INSTITUTE FOR ECONOMICS & PEACE, 2022, 2023). Além disso, apesar da queda das mortes violentas observada desde 2018, o aumento de outros crimes violentos alimenta a percepção de insegurança dos brasileiros.

Em 2022, o Brasil chegou à menor taxa de homicídios em doze anos (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2023): 23,4 a cada cem mil habitantes. Em relação ao perfil das vítimas, 91,4% são homens; 76,9% são negras e 50,2% têm entre 12 e 29 anos de idade. Entretanto, no mesmo ano, foi registrado o maior número de estupros da série histórica, dos quais 68,3% ocorreram na residência da vítima. Os indicadores de violência contra criança e adolescente, de violência doméstica e de violência contra a mulher cresceram significativamente. O número de munições vendidas aumentou em 147% desde 2017, e o número de pessoas registradas como Caçadores, Atiradores e Colecionadores (CAC) aumentou sete vezes desde 2018 (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2023).

Nos últimos anos, ações atomizadas, de curto prazo e intensivas em uso da força passaram a ser consideradas alternativas à segurança pública mesmo pelas autoridades. É o que

mostra a flexibilização da posse de arma de fogo, do excludente de ilicitude e do rastreamento de munições. Nos últimos três anos, episódios de sedição entre as forças de segurança pública sugerem a naturalização de ações extraoficiais entre os quadros institucionais (LIMA; BUENO; ALCADIPANI, 2021), justificados como respostas à falência da segurança pública. Grupos paramilitares, que buscam se afirmar como reação à ineficácia do Estado, ganharam força e passaram a disputar território e o monopólio e diversas atividades ilegais com facções criminosas mais antigas (CANO; DUARTE, 2012; MANSO, 2020).

Por outro lado, políticas de desincentivo ao crime baseadas em tecnologia e inteligência perderam espaço, apesar de gerarem retornos de longo prazo significativos em relação ao investimento necessário. É o caso do Sistema Único de Segurança Pública (SUSP), criado em 2018 com o objetivo de aproximar as esferas municipal, estadual e federal na articulação do combate ao crime. Até a revisão deste trabalho, o projeto enfrenta diversos desafios para sua implantação (LIMA; BUENO; ALCADIPANI, 2021): por exemplo, a falta de engajamento dos gestores municipais, que seguem tratando a segurança pública como atribuição exclusivamente estadual – ao contrário do que determina a Lei nº 13.675/2018, que criou o SUSP (BRASIL, 2018).

Desde que o Brasil começou a enxergar-se como um país violento, ações com objetivos de curto prazo e altamente dependentes do uso da força prevaleceram sobre políticas integradas de prevenção e combate à violência. Essa percepção se deu no início da década de 1980 (AMORIM, 1993, 2003; CERQUEIRA, 2010), quando a escalada da violência urbana levou imprensa e população a tomarem conhecimento das primeiras facções criminosas, formadas no fim dos anos 1970 (AMORIM, 1993, 2003). A emergência do crime organizado e do narcotráfico levou a um salto na complexidade da violência urbana no Brasil (AMORIM, 2003; MANSO; DIAS, 2018). Contudo, mesmo após a Redemocratização (1985), as políticas estatais pouco se alteraram: continuavam baseadas na repressão pontual e arbitrária da criminalidade urbana (BUENO, 2014; SPANIOL; MORAES JÚNIOR; RODRIGUES, 2020), e em alguns casos, acabavam sendo contraproducentes.

Um exemplo foi a interiorização do sistema penitenciário nas décadas de 1990 e 2000. Considerada uma medida pacífica e estratégica, a interiorização do sistema prisional visava desarticular o crime organizado afastando seus integrantes. Contudo, acabou expandindo a área de atuação do narcotráfico e facilitando a penetração das facções em áreas de baixa criminalidade relativa, que passaram a conviver com os mesmos distúrbios das regiões metropolitanas (SALLA, 2007).

Em geral, é possível observar que a abordagem de segurança pública no Brasil desconsidera eventuais causas históricas da violência. É razoável que fatores históricos não sejam amplamente considerados ao elaborar políticas e medidas específicas. Contudo, o contexto histórico determinou as instituições e estruturas que geraram incentivos e desincentivos à violência e à criminalidade ao longo do tempo. Ignorá-lo pode levar à elaboração de políticas que não alteram os incentivos estruturais ao comportamento violento, contribuindo para o caráter inócuo ou contraproducente frequentemente observado nas medidas de segurança pública. Com isso, a ação do Poder Público perde credibilidade, o que desgasta ainda mais a legitimidade das próximas políticas de segurança, mantendo ou agravando a violência que se buscava combater.

O estudo da violência no Brasil foi significativamente comprometido pela vigência da Ditadura Militar (1964-1985), cuja natureza autoritária levou ao distanciamento entre pesquisadores e formuladores de políticas públicas. Por exemplo: a eclosão do crime organizado no início da década de 1980 não foi acompanhada por ações governamentais de coleta de dados padronizados sobre a criminalidade no Brasil. Tal foi particularmente prejudicial para estudiosos de Ciências Sociais Aplicadas, que só teriam dados quantitativos mais estruturados da década de 2000 em diante (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2007).

Nessas condições, desde a Redemocratização (1985) até os dias atuais, é possível observar duas vertentes da análise da violência no Brasil. A primeira consiste na análise sociológica da trajetória e das causas da violência no Brasil (ADORNO, 1996; BUENO, 2014; MANSO; DIAS, 2018; MANSO, 2020). Destaca-se o Núcleo de Estudos da Violência da Universidade de São Paulo (NEV-USP), criado em 1987, como centro de pesquisa pioneiro em investigar a criminalidade no Brasil ainda durante a Ditadura Militar (1964-1985).

A segunda vertente é composta essencialmente pela Economia do Crime e áreas correlatas, severamente limitadas pela disponibilidade de estatísticas sobre a criminalidade no Brasil. Sob esta restrição, a Economia do Crime se ocupou em analisar os padrões de violência observados desde o final dos anos 1990, quando há maior disponibilidade de dados (CERQUEIRA, 2010; SANTOS, 2012; JUSTUS; KAHN; CERQUEIRA, 2017). Com a consolidação da democracia no país, a comunidade acadêmica pôde estabelecer novos centros de pesquisa dedicados à coleta, tratamento, análise e divulgação das estatísticas sobre criminalidade no país. Fundado em 2006, o Fórum Brasileiro de Segurança Pública (FBSP) conta pesquisadores de ambas as vertentes, entre outros atores relacionados à segurança pública,

e ocupa um papel fundamental no monitoramento das políticas de segurança municipais, estaduais e federais (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2007).

A divisão entre as vertentes é marcada pela ausência de uma abordagem empírica das causas históricas da violência no Brasil, lacuna que este trabalho visa preencher, como veremos ao fim desta seção. A literatura sociológica afirma que a escravidão foi a base da estrutura social que proporcionou os incentivos socioeconômicos que alimentam grande parte da violência no Brasil, naturalizando-a nas relações sociais a ponto de invisibilizá-la. A análise de outros países submetidos ao sistema escravagista sugere que os prejuízos da escravidão persistem ao longo do tempo. Engerman e Sokoloff (1997, 2002) notaram que, durante o período colonial, áreas climaticamente aptas à agricultura de exportação foram exploradas pelas metrópoles através da escravidão, que lhes proporcionou lucros com o baixo custo da mão-de-obra e com o tráfico transatlântico de pessoas. Após a descolonização, as instituições legadas pelas metrópoles asseguraram a manutenção de altos níveis de desigualdade de renda e de terras, o que contribuiu para o subdesenvolvimento dessas nações. Nunn (2008) verificou que a escravidão estava correlacionada a prejuízos no desenvolvimento econômico dos Estados Unidos, das ex-colônias ibéricas e dos países africanos cujos habitantes foram escravizados.

Acemoglu, Johnson e Robinson (2002) obtiveram evidências empíricas de que, nas ex-colônias de exploração, as instituições extrativas e predatórias características do período escravagista sobreviveram à abolição e à descolonização, perpetuando os privilégios das elites e prejudicando o desenvolvimento econômico. Acemoglu, García-Jimeno e Robinson (2012) encontraram relação direta entre a escravidão empregada na extração de ouro e maiores taxas de pobreza, menor escolaridade, menor taxa de vacinação e menor acesso a bens públicos nos municípios da Colômbia no início do século XX. Ao analisarem o sul dos Estados Unidos, Bertocchi e Dimico (2012, 2014) identificaram que a desigualdade racial na educação entre 1940 e 2000 estava significativamente correlacionada com a escravidão em 1860.

No caso do Brasil, diversos aspectos socioeconômicos vêm sendo empiricamente relacionados à escravidão. Maloney e Valencia-Caicedo (2016) encontraram uma relação significativa e decrescente entre escravidão e renda *per capita* municipal em 2000 utilizando dados do Censo Imperial de 1872 sobre a população escravizada nos municípios. Fujiwara, Valencia-Caicedo e Laudares (2017) concluíram que a escravidão teve impacto significativo e positivo sobre as atuais desigualdades de renda e de educação no país.

Em comparação com outros aspectos socioeconômicos, a relação entre escravidão e violência é relativamente pouco analisada no Brasil do ponto de vista quantitativo. Os Estados

Unidos são o principal objeto destas investigações, o que se deve principalmente à sua maior disponibilidade de dados sobre o período escravagista e também sobre diferentes tipos de crimes. Gouda e Rigterink (2017) verificaram que a proporção de escravos nos municípios estadunidenses em 1860 tem relação crescente e significativa com as taxas de crime violento (homicídio, estupro, roubo e agressão) nos Censos de 1970, 1980, 1990 e 2000. No século XVIII, a violência no Sul do país era muito superior à do Norte, o que persiste até hoje (INSTITUTE FOR ECONOMICS & PEACE, 2011). A principal diferença entre as duas regiões é a escravidão, “aspecto mais importante da história do Sul dos Estados Unidos” (CARDEN, 2008).

De acordo com o banco de dados *Trans-Atlantic Slave Trade Database* (ELTIS; BEHRENDT; DOMINGUES DA SILVA; KLEIN; LACHANCE; MISEVICH; RICHARDSON, 2021), elaborado pela Universidade de Emory com apoio da Universidade de Harvard, os Estados Unidos receberam cerca de quatrocentos mil pessoas escravizadas, enquanto o Brasil recebeu cerca de 4,9 milhões. A Abolição americana ocorreu em 1865, e a brasileira, em 1888, sob forte pressão internacional. Diante destes dados e das conclusões de Gouda e Rigterink (2017) para os Estados Unidos, é razoável esperar que a relação entre escravidão e violência contemporânea no Brasil seja mais pronunciada: as dificuldades de mensurar seus efeitos podem ser atribuídas à subrepresentação da sua magnitude nos dados disponíveis para análise.

No Brasil, as configurações sociais, políticas e espaciais permaneceram semelhantes às do período colonial mesmo após a Abolição da Escravatura, em 1888 (RIBEIRO, 1995; SCHWARCZ, 2012; SOUZA, 2017). Estabelecida durante o período escravagista, a autoridade local dos senhores de terras representou um obstáculo à autoridade da República (1889). A primazia dos poderes locais e a baixa legitimidade dos poderes oficiais favoreceram o uso arbitrário da força, a apropriação das instituições pelas elites e as soluções extraoficiais em detrimento do devido processo legal, práticas que persistiram no período republicano.

Após a Abolição (1888), os trabalhadores libertos enfrentaram a marginalização econômica, pois as principais atividades econômicas do início do século XX, a indústria e a agricultura cafeeira, se baseavam no trabalho livre dos imigrantes europeus. Não houve ação estatal voltada para a inserção dos libertos e de seus descendentes na economia, o que perpetuou as desigualdades e os prejuízos legados pela escravidão. Na prática, isso significa que a maior parte da população brasileira não gozava do pleno exercício da cidadania e da liberdade, o que

criou uma sociedade pós-colonial segregada, tradicionalista e pouco coesa, cujo caráter individualista e autoritário persistia sob um verniz moderno (SCHWARCZ, 2012).

Mesmo após a inserção do Brasil no capitalismo internacional e o fim da colonização, a política econômica e institucional seguiu desconsiderando a herança escravagista, o que contribuiu para sua persistência. A violência intrínseca à escravidão permaneceu “invisível”, tamanha sua normalização (FRANCO, 1983; SOUZA, 2017; ROSENTHAL, 2018). A escravidão baseada no tráfico transatlântico realizado pela metrópole portuguesa vigorou por pelo menos 350 anos até sua Abolição em 1888. Sendo assim, a economia brasileira dependeu de um sistema produtivo inerentemente violento durante a maior parte da existência da nação. Apesar de mais de três séculos de violência institucionalizada, a percepção de que o Brasil era um país violento só se consolidou entre a imprensa e a população após a emergência da violência urbana organizada e do narcotráfico no início dos anos 1980 (AMORIM, 1993, 2003).

Até o momento da revisão deste trabalho, não foi encontrado nenhum estudo que verificasse a relação entre escravidão e violência letal contemporânea nos municípios brasileiros. Este estudo tem o objetivo de preencher esta lacuna, investigando a associação entre a escravidão em 1872 e as taxas de homicídio municipais em 2000, tendo como referência o trabalho de Gouda e Rigterink (2017) sobre os Estados Unidos. Embora a hipótese subjacente seja de natureza causal – amplamente fundamentada na literatura sociológica e empírica (como demonstra a Seção 2) – este estudo adota uma postura metodologicamente prudente ao estabelecer como objetivo primário a identificação de uma correlação positiva e significativa após controle rigoroso do confundimento – isto é, da correlação entre a variável de tratamento contínua que representa a escravidão e os controles observados.

Antes de enumerar as razões desta escolha, é útil definir a terminologia que será empregada ao longo deste trabalho. Apesar de não visar a obtenção de estimadores causais, utiliza-se a denominação de grupo de controle e grupo de tratamento porque a escravidão como experimento natural atingiu os municípios brasileiros de forma não aleatória.

A variável de interesse é denominada intensidade da escravidão (*sld*), dada pelo percentual de pessoas escravizadas na população total de um município em 1872. Neste ano, todos os municípios já ocupados possuíam pessoas escravizadas entre a população total. Portanto, foram atingidos pelo experimento natural e formam o grupo de tratamento (que conta com 640 municípios sob a divisão municipal atualizada pelo IBGE). O resto do território ainda não havia sido ocupado, de modo que municípios com intensidade da escravidão nula em 1872 formam o grupo de controle. A distribuição da população escravizada se concentrou próxima

ao litoral das regiões Sudeste e Nordeste devido aos interesses da metrópole portuguesa, então a seleção dos grupos tratado e não tratado foi não aleatória.

Pertencer ou não ao grupo de tratamento em 1872 está altamente correlacionado aos valores observados das variáveis de controle e da variável resultante na amostra observada, gerando o confundimento. Por isso, este trabalho adota técnicas de balanceamento amostral características de estimações causais para corrigi-lo, exploradas detalhadamente na Seção 3. Essas técnicas devem levar em consideração que, em 1872, a distribuição da população escravizada variava por município. Logo, a intensidade da escravidão não se distribuiu igualmente entre as unidades tratadas, mas em diferentes “doses”. As técnicas de balanceamento amostral adotadas consideram, portanto, o caráter contínuo da variável de interesse ao reduzir e controlar o confundimento observável.

Devido às razões expressas nos parágrafos anteriores, utiliza-se o método de ponderação pelo inverso dos escores de propensão generalizados (mais conhecido pelas siglas em inglês, *Inverse Probability Weighting* (IPW) e *Generalized Propensity Scores* (GPS)) (HIRANO; IMBENS, 2004), uma extensão da técnica criada para tratamentos binários tradicionalmente empregada em estimações causais com tratamentos contínuos. Originalmente, o método GPS-IPW tem o objetivo de gerar estimadores causais, mas este trabalho o emprega com a finalidade de reduzir o confundimento e produzir correlações robustas entre *sld* em 1872 e violência no período contemporâneo.

Em linhas gerais, estima-se a probabilidade condicional de cada município apresentar dado valor de *sld* com base nos valores observados dos controles – estes são os GPSs. A ponderação dos municípios pelo inverso dos GPS leva a uma amostra na qual *sld* não está correlacionada com os controles dentro de estratos homogêneos de propensão. Sob a condição de que *sld* é condicionalmente independente dos controles a cada estrato de propensão (denominada ignorabilidade fraca (Seção 3)), estimações baseadas na amostra ponderada por GPS-IPW levam a estimadores causais do efeito da variável de tratamento sobre a variável resultante.

Este trabalho, no entanto, opta por uma alternativa mais prudente ao considerar o estimador de *sld* gerado por regressões ponderadas como uma correlação robusta, evidência de uma correlação positiva e significativa entre escravidão histórica e violência contemporânea após controle rigoroso de confundidores observáveis. Essa escolha se deve às limitações nos dados disponíveis sobre escravidão, violência e sobre controles tradicionalmente empregados

pela literatura de Economia do Crime, que dificultam o cumprimento da condição de ignorabilidade fraca e a reivindicação de causalidade (estatística, não sociológica) definitiva.

A única fonte de dados observados sobre a população escravizada é o Censo Imperial de 1872, cuja divisão municipal foi atualizada pelo IBGE (IBGE, 2012). Vinte e dois anos antes, em 1850, a Lei Eusébio de Queiróz proibiu o tráfico transatlântico, diminuindo consideravelmente a oferta de trabalhadores escravizados no Brasil. Dezesesseis anos depois do Censo Imperial de 1872, aconteceria a Abolição (1888). A população escravizada apresentava maior mortalidade devido às condições desumanizantes de trabalho e de subsistência a que estava sujeita, então o fim do tráfico em 1850 levou a uma diminuição considerável de pessoas escravizadas na população total do Brasil em 1872.

Conclui-se que o Censo Imperial de 1872 retrata um cenário de decadência da escravidão no Brasil. Por isso, tomar seus dados como representativos do experimento natural escravidão em sua totalidade levaria à subestimação da sua magnitude e dos seus efeitos no Brasil contemporâneo. Por exemplo, nesta seção, vimos que o total de pessoas traficadas para o Brasil até 1850 foi de cerca de 4,9 milhões. Entretanto, o Censo Imperial de 1872 registrou cerca de 1,5 milhões de pessoas escravizadas numa população absoluta de cerca de 9,9 milhões de habitantes – ou seja, aproximadamente 30% do total de pessoas traficadas. Se tivessem sido realizados, Censos referentes a períodos anteriores tenderiam a apresentar menor população absoluta e maior percentual de pessoas escravizadas em relação ao total, fornecendo um cenário mais representativo da dependência do Brasil em relação ao sistema escravagista.

Gouda e Rigterink (2017) contam com dados representativos sobre a escravidão no sul dos Estados Unidos, e por isso, realizaram estimações lineares não ponderadas para obter a correlação entre intensidade da escravidão e diversas taxas de criminalidade observadas em 1970, 1980, 1990 e 2000. Dadas as limitações de dados a que este trabalho está sujeito, não é prudente interpretar o estimador de *sld* como estatisticamente causal – mesmo que técnicas tradicionalmente associadas à causalidade sejam empregadas para reduzir o confundimento na amostra. Não é possível garantir que o método GPS-IPW conseguirá atender a condição de ignorabilidade fraca em uma amostra de dados tão irregular (Seção 4).

Gouda e Rigterink (2017) também dispunham de dados municipais sobre agressão não letal, estupro, roubo e homicídio nos municípios contemporâneos de 1970 a 2000. Em relação ao Brasil, o principal indicador de criminalidade comparável entre municípios é a taxa de homicídios a cada cem mil habitantes, que conta com maior regularidade na coleta do que os outros crimes analisados pelos autores do estudo de referência.

A especificação das regressões ponderadas se baseia nos modelos utilizados por Gouda e Rigterink (2017) para os Estados Unidos e Buonanno e Vargas (2019) para a Colômbia. Ambos os trabalhos ressaltam a importância de controlar a qualidade das instituições municipais ao analisar a relação entre escravidão e violência.

No Brasil, indicadores de qualidade institucional municipal coletados esporadicamente e raramente atualizados com regularidade. Este trabalho utiliza dados da Pesquisa de Informações Básicas Municipais (MUNIC) de 2001, cujo tema foi gestão pública, para criar um indicador municipal de justiça e segurança pública (*cps*) baseado na quantidade de órgãos relacionados ao tema localizados no município no ano de referência (IBGE, 2003).

Dados semelhantes não estavam disponíveis para outros anos quando este estudo foi concluído. Por isso, optou-se pela análise de dados em *cross-section* para o ano 2000, o ano mais “próximo” de 1872 para o qual há dados suficientes de um controle tão essencial como a gestão pública. Esta é mais uma razão pela qual este trabalho busca chegar a um estimador correlacional e não causal (mesmo empregando métodos tradicionalmente relacionados à causalidade). Mesmo que a condição de ignorabilidade chegasse a se cumprir, estudos causais tendem a analisar mais de um corte transversal para garantir a robustez das suas conclusões. Alternativas para contornar a limitação de dados municipais sobre gestão pública serão discutidas na Seção 5.

Até aqui, este trabalho contribuiu para preencher a lacuna de estudos quantitativos sobre fatores históricos relacionados à violência no Brasil contemporâneo; para difundir uma metodologia de balanceamento de amostras desequilibradas quando o experimento natural é representado por uma variável de tratamento contínua (o método GPS-IPW); e para a criação de um indicador municipal de justiça e segurança pública que, como demonstrará a Seção 4, se mostrou robusto ao longo de diversas estimações, modelos e áreas analisadas. A última contribuição metodológica é a especificação das diferentes relações entre renda e desigualdade nos municípios brasileiros como forma de reduzir ainda mais o confundimento entre a amostra analisada.

A desigualdade de renda foi um dos primeiros fenômenos relacionados empiricamente à escravidão (FUJIWARA; LAUDARES; VALENCIA-CAICEDO, 2017). Ao mesmo tempo, a literatura de Economia do Crime a reconhece como fator fundamental na explicação das diferentes taxas de criminalidade (SOARES, 2004; SOARES; NARITOMI, 2010). Como está relacionada tanto à variável de tratamento que representa o experimento natural como à variável resultante analisados neste estudo, a desigualdade de renda é uma variável de confusão

extremamente relevante, considerada inclusive um meio pelo qual os efeitos da escravidão persistem após seu fim (GOUDA; RIGTERINK, 2017; FUJIWARA; LAUDARES; VALENCIA-CAICEDO, 2017; BUONANNO; VARGAS, 2019).

As estimações deste trabalho contra renda, desigualdade e também a interação entre essas variáveis, como forma de identificar se a relação da desigualdade com as taxas de homicídio varia ao longo dos diferentes níveis de renda dos municípios brasileiros em 2000. Os resultados (Seção 4.2) confirmam esta hipótese e são robustos para dois índices de desigualdade diferentes, Gini e Theil. A análise das diferentes dinâmicas entre renda, desigualdade e violência foi realizada através da técnica de Johnson-Neyman para investigar interações em modelos de regressão nos quais a mediação é feita por uma variável contínua (a desigualdade de renda) (LONG, 2021, 2022).

Este estudo está distribuído da seguinte forma: a atual Seção 1 apresenta a importância e os objetivos da análise; a Seção 2 revisa a literatura sociológica e quantitativa, nacional e internacional, sobre os temas abordados; a Seção 3 descreve os dados e a estratégia empírica utilizada; a Seção 4 apresenta e discute os resultados obtidos. Por fim, a Seção 5 apresenta as conclusões e considerações finais.

2 REVISÃO DE LITERATURA E FUNDAMENTAÇÃO TEÓRICA

2.1 A PERSISTÊNCIA DO LEGADO DA ESCRAVIDÃO

Por que faz sentido supor que a escravidão colonial afeta o Brasil contemporâneo? Esta seção busca responder esta pergunta. Esta hipótese é razoável porque há estudos sociológicos e empíricos que ligam a escravidão a fenômenos contemporâneos em várias nações, relacionadas em maior ou menor grau com essa instituição. Aspectos do desenvolvimento econômico, desvantagens raciais na transmissão de capital humano e discriminação político-institucional são os efeitos de longo prazo mais comumente observados. A maior parte da literatura empírica se refere aos Estados Unidos, às ex-colônias espanholas e aos países africanos. Contudo, análises empíricas dos efeitos da escravidão no Brasil vêm crescendo em número e complexidade.

Estudos de trajetória da dependência (NORTH, 1990; ENGERMAN; SOKOLOFF, 1997, 2002; ACEMOGLU; JOHNSON; ROBINSON, 2002) observaram que áreas geográfica e climaticamente aptas ao cultivo de produtos valorizados no período colonial receberam mão-de-obra escrava, como o Brasil e o Sul dos Estados Unidos. Nestas regiões, as metrópoles instalaram instituições extrativas para obter benefícios a curto prazo, o que gerou altos índices de desigualdade de renda nas colônias. Estas instituições não se alteraram após a abolição da escravidão ou o fim do colonialismo, garantindo os privilégios das elites, agravando as desigualdades e dificultando a superação dos atrasos no desenvolvimento.

Analisando empiricamente esta hipótese, Nunn (2008) encontrou evidências de que a escravidão prejudicou o desenvolvimento econômico dos Estados Unidos, das ex-colônias ibéricas e dos países africanos onde ocorrera o tráfico transatlântico. No entanto, não foi possível detectar que esses prejuízos foram causados pela desigualdade de renda ou de terras.

Ao estudarem os países africanos, Nunn e Wantchekon (2011) verificaram que grupos étnicos cujos ancestrais foram mais atingidos pelo tráfico no passado apresentavam menores níveis de confiança nos poderes institucionais e em outros grupos sociais. Tal sugere que a escravidão comprometeu a integridade do tecido social nos países vitimados pelo tráfico, com consequências significativas para a instabilidade social e política que se seguiu ao fim do colonialismo.

Estudos empíricos sobre o impacto do trabalho compulsório nas ex-colônias espanholas sugerem que o sistema escravagista comprometeu o desenvolvimento destas regiões. Acemoglu, García-Jimeno e Robinson (2012) encontraram uma relação significativa e positiva

entre o trabalho compulsório indígena (empregado na extração de ouro durante a colonização espanhola) e maiores taxas de pobreza, menor escolarização, menores taxas de vacinação e menor acesso a bens públicos nos municípios colombianos.

Ao analisarem diversos países da América Latina, Maloney e Valencia-Caicedo (2016) notaram que a densidade populacional contemporânea está correlacionada com a atividade econômica durante o período colonial. Além disso, utilizando dados municipais do Censo Imperial de 1872, os autores encontraram uma relação significativa e negativa entre escravidão e renda *per capita* entre os municípios brasileiros no ano de 2000.

A escravidão nos Estados Unidos é objeto da maioria dos estudos empíricos, o que pode ser atribuído à maior disponibilidade de dados sobre o período no país. Bertocchi e Dimico (2012, 2014) concluíram que desigualdade escolar entre brancos e negros¹ nos Estados Unidos pode ser atribuída à herança escravagista. A desigualdade educacional sustenta a desigualdade de renda ao longo do tempo, levando a uma desvantagem de transmissão e absorção de capital humano entre os descendentes da população escravizada. Segundo os autores, a exclusão política após a liberdade impediu a correção dessas distorções.

Em relação à defasagem de capital humano, Sacerdote (2005) comparou a população negra liberta com a população negra livre e seus respectivos descendentes antes e depois da Abolição Americana (1865). Observando os índices de alfabetização, de frequência escolar e de emprego, o autor concluiu que foram necessárias duas gerações de descendentes da população negra liberta para que estes tivessem as mesmas condições de vida que os descendentes da população negra livre. Com isso, foi possível concluir que os prejuízos causados pela escravidão chegaram pelo menos a 1920.

Utilizando dados municipais, Jung (2019) encontrou evidências empíricas de que o incentivo para que a população liberta absorvesse capital humano foi reduzido pela legislação trabalhista discriminatória, que impedia o acesso ao mercado de trabalho livre. Ao estimar o retorno da educação para os municípios americanos em 1940, o autor verificou que, em regiões altamente dependentes da escravidão no passado, pessoas negras tinham maior dificuldade de converter sua educação em renda e mobilidade social.

¹ Segundo a metodologia do IBGE, a categoria "negro" não constitui uma opção de autodeclaração nos censos, mas representa a união analítica das pessoas que se autodeclararam pretas e pardas (IBGE, 2013). As cinco categorias oficiais de autodeclaração do IBGE são: branca, preta, parda, amarela e indígena. Uma vez que a diferenciação entre pretos e pardos não compromete as premissas deste trabalho, a denominação "negros" foi adotada ao longo de todo o texto para designar o conjunto dessas duas categorias.

As atitudes políticas da população americana também foram influenciadas pela escravidão, o que contribuiu para a persistência de seus efeitos. Através de questionários respondidos por cerca de quarenta mil pessoas, Acharya, Blackwell e Sen (2016) concluíram que habitantes brancos de cidades com maior parcela de população escravizada em 1860 tendiam a manifestar visões negativas sobre a população negra e a não apoiar ações afirmativas.

Segundo os autores, as heranças da escravidão são transmitidas por instituições sociais formais e informais. As leis de segregação “Jim Crow” e a criminalização do desemprego após a Abolição Americana (1865) são exemplos de instituições formais. A última permitiu pagar baixos salários aos libertos, o que com frequência levava ao trabalho análogo ao escravo. Dentre as instituições informais, estão a punição social (velada ou não) a relacionamentos interraciais e a desconfiança das autoridades em relação à população preta (associada à pobreza e considerada propensa à criminalidade). O fato de que as penas para o mesmo crime são maiores para perpetradores negros do que para perpetradores brancos, por exemplo, sugere que a desconfiança tem implicações concretas. Os dois mecanismos descritos correm paralelamente na história, ao mesmo tempo em que interagem e se reforçam (ACHARYA; BLACKWELL; SEN, 2016).

No caso do Brasil, o papel da escravidão na realidade nacional vem sendo analisado pelos principais nomes da Sociologia desde o início do século XX. Freyre (1973), Holanda (1984), Fernandes (1965, 1978) e Ribeiro (1995) afirmaram que a escravidão é essencial para a compreensão do Brasil contemporâneo. Os dois primeiros foram expoentes do chamado paradigma culturalista, para o qual a cultura é o principal fator a ser analisado para compreender a identidade nacional. Mesmo reconhecendo o peso da escravidão, Holanda enfatizava a herança ibérica, enquanto Freyre afirmava o caráter híbrido da população brasileira, a “mestiçagem”, a “harmonia dos contrários” como principal traço nacional. Para Freyre, a ausência de conflitos raciais abertos e a miscigenação teriam levado a uma sociedade pacífica após a Abolição (1888). A primazia das relações pessoais sobre as regras do espaço público, tema central da obra de Holanda, também contribuiria para essa visão de harmonia social.

Em meados do século XX, estudiosos com abordagens materialistas se afastaram do culturalismo buscando compreender as consequências da escravidão brasileira de forma mais precisa. Em *A integração do negro na sociedade de classes*, Fernandes (1965) observou que o sistema escravagista colonial foi determinante para a acumulação de capital que financiava a industrialização da Europa e de suas ex-colônias, mesmo que em ritmos diferentes. Segundo

ele, o Brasil vivera suas “revoluções burguesas” ao realizar a Abolição em 1888 e instaurar a República em 1889, uma vez que essas transformações não poderiam romper (e não romperam) com as dinâmicas sociais e econômicas do sistema escravagista. A manutenção das condições de trabalho possibilitadas pela escravidão e pelo tráfico transatlântico permitiu a urbanização, o investimento na grande lavoura moderna e o início da industrialização no Brasil. O Brasil não rompeu de fato com as estruturas escravagistas, mas perpetuou-as sob (e para) o capitalismo nascente no país, o que subverteu as transformações e liberdades adquiridas (QUEIROZ, 2021).

Esse processo histórico marcado pela dissonância cognitiva² gerou o mito da democracia racial, no qual a igualdade jurídica adquirida com a Abolição (1888) era tomada como igualdade de fato. O mito naturaliza e invisibiliza a exclusão social dos libertos e de seus descendentes, o que contribui decisivamente para que esta se perpetue e chegue ao século XX. Na obra *O negro no mundo dos brancos*, escrita durante a Ditadura Militar, Fernandes constata que:

Em suma, a expansão urbana, a revolução industrial e a modernização ainda não produziram efeitos bastante profundos para modificar a extrema desigualdade racial que herdamos do passado. [...] Essa afirmação contraria o que se costuma dizer sobre a democracia racial que imperaria no Brasil. É que se confundem padrões de tolerância estritamente imperativos na esfera do decoro social com igualdade propriamente dita. (FERNANDES, 1972, p. 67).

Aliando um rigoroso trabalho empírico ao método materialismo histórico, Florestan Fernandes foi categórico ao afirmar que as mazelas sociais do Brasil no século XX remontam ao sistema escravagista: a Abolição (1888) e outras modernizações institucionais não romperam de fato com a estrutura socioeconômica engendrada pela escravidão. Após estas reformas, as populações libertas e seus descendentes não foram contempladas por medidas que de fato afirmariam sua cidadania, como o acesso a bens essenciais e a inclusão nos setores mais modernos da economia, e por isso, o Brasil contemporâneo se veria cercado pela herança perniciosa da escravidão.

² A Teoria da Dissonância Cognitiva é uma das mais importantes da Psicologia Social. Segundo seu criador, Leon Festinger (1957), os seres humanos ajustam ações e comportamentos às suas convicções e vice-versa. Caso haja inconsistência entre convicção e ação (ou seja, caso haja dissonância cognitiva), surge um desconforto. Para remediá-lo, ocorrem tentativas conscientes ou inconscientes de reduzir essa distância, ajustando o comportamento ou as convicções envolvidas. No caso da República nascente, a realidade da Abolição e da dependência das estruturas escravagistas não correspondia às modernizações emergentes. Para equilibrar essa distância sem alterar as bases da economia, ocorre um esforço de reconciliação formal (e não factual nem estrutural) que remete à condição de dissonância cognitiva. É possível observá-lo entre os intelectuais que minimizam o efeito pernicioso da escravidão, como Freyre (1973), e nos esforços institucionais para criar uma identidade nacional modernizada que desconsiderasse as diferenças e desvantagens da população liberta (como mostram as campanhas de destruição das favelas em nome da saúde pública exploradas a seguir, na Seção 2.2).

A abordagem materialista-histórica empregada por Fernandes (1965, 1972) tem paralelos significativos com os estudos de trajetória da dependência explorados no início desta seção (NORTH, 1990; ENGERMAN; SOKOLOFF, 1997, 2002; ACEMOGLU; JOHNSON; ROBINSON, 2002). Ambos permitem concluir que, no século XIX, a economia brasileira colonial – agroexportadora e profundamente dependente do sistema escravagista – não estava preparada para acompanhar o capitalismo industrial nascente nas principais economias mundiais. A tentativa de alcançá-las passou por modernizações institucionais que, por si sós, não confrontaram o caráter extrativista e predatório legado pelo domínio metropolitano às instituições coloniais e pós-coloniais. Com isso, a transformação estrutural da sociedade e da economia brasileiras que permitiria superar de fato o legado da escravidão não se concretizou.

Em sua obra *O povo brasileiro*, o sociólogo Darcy Ribeiro destaca o abismo que separava as condições de vida de brancos e de negros durante a escravidão, e afirma que “nessas condições é que se deve procurar a explicação da gritante discrepância entre a expansão do contingente branco e do negro no desenvolvimento da população brasileira” (RIBEIRO, 1995, p. 233). Por fim, entre os trabalhos mais recentes, Souza (2017) reafirmou que a escravidão é a base sobre a qual a sociedade brasileira se assenta:

No Brasil, desde o ano zero, a instituição que englobava todas as outras era a escravidão, que não existia em Portugal, a não ser de modo muito tópico e passageiro. Nossa forma de família, de economia, de política e de justiça foi toda baseada na escravidão (p.40).

O número de trabalhos empíricos sobre o impacto contemporâneo da escravidão brasileira vem aumentando nas últimas duas décadas. Summerhill (2010) utilizou dados disponíveis de cinquenta mil fazendas paulistas para testar a validade da hipótese de Engerman e Sokoloff no Brasil. Os modelos de mínimos quadrados ordinários e de variável instrumental não permitiram afirmar que a escravidão estava significativamente relacionada ao desenvolvimento econômico de longo prazo no estado de São Paulo.

Nos últimos anos, a sofisticação dos trabalhos sobre o tema cresceu significativamente. Fujiwara, Laudares e Valencia-Caicedo (2017) chegaram à conclusão oposta ao analisar o total de municípios brasileiros em 2000 e 2010. O Tratado de Tordesilhas foi utilizado como referência na elaboração de um modelo de regressão descontínua, uma vez que a porção portuguesa do Brasil recebeu pessoas escravizadas em volume muito superior à porção espanhola. Os autores verificaram que as áreas com maior parcela de escravos na população em 1872 são mais ricas atualmente, porém possuem maior desigualdade de renda e de educação, além de instituições públicas de pior qualidade. Em um minucioso levantamento de dados sobre

a indústria têxtil e de algodão, Palma, Papadia, Pereira e Weller (2021) sugerem que a escravidão está relacionada a um atraso na industrialização brasileira.

Define-se capital social como a rede de relações entre pessoas que vivem e trabalham juntas para formar uma sociedade funcional. Partindo da premissa de que nações pós-escravagistas tem menores níveis de capital social, Uttermark (2020) utilizou regressões logísticas para testar duas hipóteses sobre a relação entre escravidão e capital social (mensurado pela confiança interpessoal) e no Brasil e nos Estados Unidos. A primeira é a de que a primazia da monocultura escravagista teria gerado desigualdades econômicas que, por sua vez, diminuiriam o capital social a longo prazo; e a segunda diz que a escravidão teria influenciado as posições políticas das pessoas, as quais, transmitidas ao longo do tempo, prejudicaram o capital social. O autor concluiu que a escravidão está negativamente correlacionada com o capital social: a confiança interpessoal tende a ser menor em regiões onde o sistema escravagista estava concentrado. Os testes revelaram que a segunda hipótese foi a mais adequada para explicar esse resultado.

Utilizando variáveis instrumentais, Seyler (2021) observou que a escravidão e o apoio a instituições coercitivas no século XIX estão associados a menores níveis de desenvolvimento econômico nos municípios brasileiros. O apoio a instituições tirânicas foi medido pelos votos dos legisladores do final do século XIX em questões relativas à Abolição (1888), e o desenvolvimento econômico municipal, pelo Produto Interno Bruto (PIB), pela pobreza e pela desigualdade de renda. O autor verificou que, nos municípios com maior presença da escravidão e forte apoio a instituições coercitivas no passado, os habitantes exibiram menores níveis de confiança interpessoal, maior probabilidade de não apoiar a democracia e de crer que a corrupção é aceitável em algumas circunstâncias. Esse conjunto de comportamentos é parte do capital social do município: para o autor, este foi o canal pelo qual os malefícios da escravidão foram transmitidos para o desenvolvimento econômico ao longo do tempo.

Papadia (2019) ressalta as dificuldades de determinar os efeitos da escravidão sobre o Brasil contemporâneo, tanto pela escassez de fontes de dados, como pela distribuição irregular da população escravizada ao longo do território nacional. Além dos dados municipais do Censo Imperial de 1872, o autor utiliza os quilombos, locais de refúgio para trabalhadores escravizados em fuga, para mensurar a escravidão. Parte-se do pressuposto que locais com maior número de quilombos tiveram maior penetração do sistema escravagista.

Utilizando métodos de paramento entre observáveis para analisar os estados de Minas Gerais, Rio de Janeiro e São Paulo, o autor verificou que a escravidão teve impacto negativo

sobre o desenvolvimento econômico durante seu período de vigência e nos próximos trinta anos após sua Abolição (1888). Municípios onde a escravidão era mais intensa apresentaram piores índices de alfabetização e gestão pública. Em relação aos salários e à densidade populacional, não foi possível atribuir diferenças entre municípios diretamente à escravidão ao analisar dados de 1872 (p.36). Contudo, a persistência de disparidades no período de 1888 a 1920, após a Abolição, sugere que o legado da escravidão continuou afetando o desenvolvimento econômico ao inibir o surgimento de setores nascentes, como a indústria e outras atividades não agrícolas, seja por prejudicar o desenvolvimento de mercados eficientes de trabalho livre, seja por reduzir a acumulação de capital humano.

Lambais (2020) identificou que os quilombos têm uma relação positiva e robusta com o desenvolvimento econômico local no Brasil. Segundo ele, os quilombos permitiram a transmissão intergeracional de capital humano e cultural (por exemplo, de crenças religiosas, conhecimentos de metalurgia e outras habilidades trazidas pelos imigrantes africanos escravizados) que beneficiou o desenvolvimento econômico a longo prazo nessas regiões. A estratégia do autor para contornar as dificuldades metodológicas que o estudo da escravidão e dos quilombos coloca é digna de nota.

Para obter efeitos fixos mais precisos, dividiu-se o território brasileiro em células de aproximadamente 121 quilômetros quadrados. Utilizando a iluminação noturna como variável *proxy* da atividade econômica, observou-se que áreas com maior número de quilombos apresentavam uma economia mais dinâmica. Configurações espaciais alternativas de “quilombos contrafactuais” foram obtidas através de inferência aleatória, e mostraram que a proximidade com os quilombos está relacionada a maior proporção de empregos em metalurgia ou em outras áreas intensivas em capital humano, a maiores níveis de atividade cultural, confiança interpessoal e ação coletiva organizada.

Esta seção procurou demonstrar que há evidências quantitativas de que os efeitos da escravidão persistem no tempo e no espaço mesmo após sua abolição formal. Estudos empíricos de outras nações pós-escravagistas verificaram que a escravidão está relacionada a menores índices de desenvolvimento econômico e institucional no passado recente. Em relação ao Brasil, a exploração do tema vem crescendo (em volume de publicações e complexidade de métodos), com conclusões semelhantes às dos estudos sobre outras ex-colônias. Estudos recentes concluíram que a escravidão prejudicou a industrialização, o desenvolvimento econômico, a educação, a equidade de renda, a qualidade das instituições e até mesmo os níveis de confiança

interpessoal no Brasil. Sendo assim, é razoável supor que os níveis de violência no Brasil contemporâneo também podem ter sido significativamente afetados pela escravidão.

2.2 A VIOLÊNCIA COMO LEGADO DA ESCRAVIDÃO NO BRASIL

Esta seção tem o objetivo de embasar a hipótese de que a escravidão contribuiu para a violência no Brasil contemporâneo, fornecendo argumentos e evidências que lhe dão suporte. Existem estudos empíricos sobre o impacto da escravidão nos níveis de violência do sul dos Estados Unidos, uma sociedade escravagista e pós-escravagista comparável ao Brasil, mas com maior disponibilidade de dados históricos sobre escravidão e violência. Ambas as regiões foram colônias de exploração cuja economia se baseava na monocultura agrícola voltada para exportação, ou seja: em unidade produtiva com alta demanda de mão-de-obra. A escravidão era fundamental para garantir o lucro da atividade, mantendo os custos de trabalho baixos enquanto permitia que as demandas de exportação fossem atendidas.

A escravidão brasileira, no entanto, foi consideravelmente maior em volume de pessoas escravizadas, em tempo de duração e em dependência econômica do sistema escravagista (pois setores mais favoráveis ao trabalho livre, como a indústria, se desenvolveram mais tarde, prologando a dependência econômica da agricultura de exportação e da mão-de-obra de baixo custo proporcionada pela escravidão) (PAPADIA, 2019; PALMA; PAPADIA; PEREIRA; WELLER, 2021). Sendo assim, se foram observados efeitos significativos da escravidão sobre a violência contemporânea no sul dos Estados Unidos (onde a magnitude da escravidão foi menor), é razoável esperar que esses efeitos sejam tão ou mais pronunciados ao analisar o Brasil – ainda que limitações na disponibilidade de dados históricos dificultem sua estimação e comprometam sua representatividade.

Gouda e Rigterink (2017) encontraram uma correlação significativa e positiva entre a escravidão nos Estados Unidos em 1860 e as taxas de criminalidade violenta (homicídio, estupro, roubo e agressão não letal) em 1970, 1980, 1990 e 2000. A amostra abrangeu todos os municípios do país, sendo que a escravidão se localizou majoritariamente nos do Sul. Entre a literatura de Economia do Crime, não foi encontrado nenhum outro estudo semelhante para qualquer outra nação até o momento. Este trabalho é a principal referência para a estruturação do presente estudo.

Os autores exploram os mecanismos pelos quais a escravidão afetou os níveis de violência mesmo após a Abolição (1865) nos Estados Unidos, classificando-os em três grandes grupos (em linhas gerais): (1) desigualdade de renda e menor grau de desenvolvimento

econômico nas regiões pós-escravagistas, que tornariam o custo de oportunidade da atividade criminosa/violenta mais baixo para diversos grupos sociais (e não apenas para os descendentes de escravizados); (2) cultura de normalização da violência e discriminação institucional (formal e informal); (3) segregação espacial, e a consequente disparidade de acesso a recursos.

A Seção 2.1 apresentou alguns estudos que se enquadram no grupo 1, principalmente, e outros, no grupo 3. O grupo 2, referente aos mecanismos culturais pelos quais a escravidão afeta os níveis de violência mesmo após ser abolida, receberá atenção especial mais adiante nesta seção. Por sua vez, o grupo 3 é útil para demonstrar como esses mecanismos não anulam uns aos outros: pelo contrário, reforçam-se.

Considerando a escalada da violência racial ocorrida entre 1870 e 1940 como um experimento natural, Cook (2014) observou que esta reduziu significativamente o número de patentes registradas no período, comprometendo a inovação e o desenvolvimento econômico dos Estados Unidos. A intensificação dos linchamentos, do terrorismo racista (como o Massacre de Tulsa (1921)) e das leis de segregação “Jim Crow” estão correlacionadas com menor produtividade de inventores brancos e negros, mas foram significativas apenas entre os últimos. Portanto, as instituições formais e informais escravagistas comprometeram o desenvolvimento econômico: a violência fora tanto o resultado da persistência destas instituições após a Abolição (1865) como o meio pelo qual a escravidão levou a reduzir o desenvolvimento científico e econômico.

Logan e Parman (2017) verificaram que, entre 1880 e 1940, a segregação residencial entre brancos e negros aumentou, tanto nas regiões urbanas como nas rurais. Enquanto o volume de pessoas adquirindo casas próprias aumentou no período, a segregação espacial dobrou nos Estados Unidos. Os autores desenvolveram uma medida de segregação baseada na probabilidade de uma família negra ter vizinhos brancos (e a compara com a probabilidade em caso de segregação nula, ou completa integração racial). Contrariando estudos anteriores, os autores notaram que as áreas urbanas (e não as rurais) do sul dos Estados Unidos eram as mais segregadas do país até meados do século XX. A probabilidade de uma família negra ter vizinhos brancos diminuiu mais de 15% ao longo do período observado. Neste exemplo, vê-se que, mesmo após a Abolição (1865), a segregação espacial característica da sociedade escravagista não só se manteve, como se aprofundou nos Estados Unidos.

Enquanto Cook (2014) analisou o impacto da violência legada da escravidão sobre o desenvolvimento econômico e Logan e Parman (2017) detectaram a persistência e a intensificação de segregação espacial mesmo após a Abolição (1865), Cook, Logan e Parman

(2018) detectaram que a segregação residencial analisada pelos últimos estava altamente correlacionada à ocorrência de linchamentos (em áreas rurais ou urbanas) analisados pela primeira. Logo, o ciclo se fecha: a segregação legada da escravidão alimenta a violência racial, que por sua vez, reforça os prejuízos no desenvolvimento econômico (para além do prejuízo que a própria violência representa).

Os três mecanismos listados por Gouda e Rigterink (2017) como transmissores dos efeitos da escravidão sobre os níveis de violência contemporâneos podem ser relacionados através da dinâmica descrita por Rosenthal (2018) ao analisar livros contábeis de fazendas escravagistas do sul dos Estados Unidos. Considerando que o modo de produção escravagista colonial exigia o controle dos corpos escravizados que formavam sua força de trabalho, a autora concluiu que classes médias livres controlavam a produtividade do trabalho através da violência infligida à população escravizada. Isto é: a violência era o instrumento de trabalho das camadas livres, determinante para obter os níveis de produto necessários e garantir os lucros da atividade agroexportadora.

Como resultado da violência inerente às relações de trabalho, toda a sociedade passa a normalizá-la. Esta normalização se manifesta como uma espécie de invisibilização que permitiu à violência escravagista ultrapassar a Abolição (1865) e chegar à contemporaneidade. Da mesma forma, as condições de precariedade econômica (SACERDOTE, 2005; BERTOCCHI; DIMICO, 2012, 2014; COOK, 2014; JUNG, 2019), a segregação institucional (ACHARYA; BLACKWELL; SEN, 2016) e a segregação espacial (LOGAN; PARMAN, 2017; COOK; LOGAN; PARMAN, 2018) características das sociedades escravagistas e pós-escravagistas persistem e contribuem para retroalimentar a violência engendrada pela escravidão.

Conclui-se que diversos estudos empíricos identificaram correlações entre a escravidão no sul dos Estados Unidos e os níveis de violência da região mesmo após a Abolição (1865). Esta literatura também descreveu e mensurou os mecanismos pelos quais esses efeitos se perpetuaram e se reforçaram. Os mecanismos culturais e antropológicos pelos quais a escravidão afeta os níveis de violência mesmo após ser abolida, identificados com o grupo 2 a que nos referimos anteriormente, será explorado a seguir.

A visão de Rosenthal (2018) em relação à violência como instrumento de controle da produtividade e sua consequente normalização nas sociedades escravagistas e pós-escravagistas é notavelmente semelhante à de Franco (1983) ao analisar a população branca e livre do Vale do Paraíba, região do estado de São Paulo onde a escravidão foi particularmente mais intensa. A primeira, no entanto, analisou livros contábeis datados do período escravagista nos Estados

Unidos, enquanto a segunda procurava analisar as classes médias brancas do fim do século XIX, após a Abolição (1888) no Brasil, o que, por si só, sugere que a hipótese na qual este trabalho se baseia é razoável.

O trabalho de Maria Sylvia Carvalho Franco merece ser explorado com maior profundidade por descrever mecanismos culturais de normalização e transmissão da violência escravagista ao longo do tempo (o grupo 2 da classificação de Gouda e Rigterink (2017)) e entre grupos sociais que não foram diretamente vitimizados pela escravidão. Como veremos a seguir, os conceitos de “desafio”, de violência costumeira e de violência institucionalizada estão muito próximos das conclusões de Rosenthal (2018). Além disso, a descrição da violência como imperativo moral feita por Franco (1983) tem paralelo com a cultura da honra e a *southernness* observadas no sul dos Estados Unidos por autores que, no entanto, se apressaram em descartar a escravidão americana como fator cultural determinante, como veremos mais à frente (GASTIL, 1971; NISBETT; COHEN, 1996).

Franco tinha o objetivo de analisar aspectos gerais da vida das camadas médias da população, e para isso, utilizou cerca de trezentos processos-crime registrados nos municípios do Vale do Paraíba, interior de São Paulo, durante o século XIX. Diferente do Oeste Paulista (cuja modernização fora motivada pela cafeicultura), o Vale do Paraíba estagnou economicamente após o esgotamento do solo causado por práticas agrícolas predatórias características da monocultura agroexportadora. Por isso, representava fielmente a ordem social pré-republicana mesmo após a Abolição (1888). A autora relata que os padrões revelados pelas suas fontes primárias tornaram a violência impossível de ignorar:

(...) ao passo que a pesquisa ia progredindo, a violência aparecia por toda parte, como um elemento constitutivo das relações mesmas que se visavam conhecer. Assim, não cabe a arguição de que a violência ressaltou porque esquadrinhei uma documentação especializada nela. O contrário é verdadeiro: foi a violência entranhada na realidade social que fez a documentação, nela especializada, expressiva e válida (FRANCO, 1983, p. 17).

Analisando os documentos, a autora notou que atos violentos entre homens livres eram súbitos, aparentemente sem motivo, e ocorriam entre pessoas que geralmente tinham uma relação amistosa. A princípio, a resolução era verbal, mas as partes envolvidas passavam a “medirem-se” ou a crerem que estavam sendo provocadas ou desafiadas (consideradas fracas, incapazes de cumprirem com sua palavra ou suas ameaças) e então a tensão escalava para a violência física. Franco concluiu que, “na quase totalidade dos casos, há uma desproporção

entre os motivos imediatos que configuraram um determinado contexto de relações e o seu curso violento” (FRANCO, 1983, p.26).

Cotidianamente, conflitos reconhecidos como fúteis pelos próprios perpetradores evoluíam para o que Franco denominou “violência costumeira”, aquela que não chocava os envolvidos, a comunidade ou mesmo as autoridades. O conceito evidencia a normalização da violência descrita por Rosenthal (2018) entre as camadas médias livres no Vale do Paraíba. Nesse contexto, a vizinhança era vista como um ambiente integrador devido à obrigatoriedade tácita da contraprestação de serviços. Ao contrário do esperado nessas condições, a resolução dos conflitos era costumeiramente violenta, por mais simples que fossem:

(...) A violência está incorporada com uma regularidade, eclodindo de circunstâncias que não comprometem as probabilidades de sobrevivência e apresentando um caráter costumeiro suficientemente arraigado (...). Pode-se prosseguir nessa linha de interpretação e propor, mesmo, que a violência seja uma forma rotinizada de ajustamento nas relações de vizinhança. Isso se confirma quando a troca de facadas e bordoadas resulta de contatos passageiros, aguçados sem que nenhum incidente de importância tenha ocorrido. (FRANCO, 1983, p. 30)

Além da violência costumeira, Franco definiu a “violência institucionalizada” como uma característica do trabalho e do lazer das camadas livres. O papel periférico e substituível dessas populações na economia as levava a migrar constantemente, o que impedia a formação de vínculos comunitários que reduziriam os atritos, além de dificultar o desenvolvimento de habilidades voltadas ao trabalho moderno - impessoal, racional, eficiente (WEBER, 2004). A baixa especialização profissional as levava a atividades mal remuneradas e pouco especializadas, sem liderança clara ou rotina definida.

Nessas condições, o único controle efetivo do rendimento do trabalho entre a população livre era o “desafio” (FRANCO, 1983, p.37). Quando desafiados, os participantes têm um incentivo para trabalharem da melhor forma possível. Os observadores também acabavam motivados a trabalhar, seja por se entreterem com a competição ou para não serem os próximos a serem desafiados. Portanto, o conflito era inerente à vida laboral dos homens livres, e com frequência, terminava em violência física (que poderia inclusive se estender aos seus espectadores). A semelhança do desafio de Franco (1983) com a violência como instrumento de trabalho descrita por Rosenthal (2018) é evidente, e é mais um indício de que mecanismos de transmissão da violência escravagista observados nos Estados Unidos podem ser tão ou mais pronunciados na realidade brasileira.

Apesar da conotação lúdica, o desafio pode ser considerado uma forma de ameaça. Sem possibilidades de ascensão econômica, sem laços comunitários fortes e sem bens culturais tanto

antes como depois da Abolição (1888)³, a reputação era o principal bem entre as camadas livres, muito dependentes da boa vontade de “conhecidos”, como sugerem as relações de contraprestação de serviços na vizinhança. Sendo assim, até mesmo superiores não recusavam desafios propostos por subordinados, pois isso colocaria sua imagem em risco (FRANCO, 1983). Recusar ou perder um desafio poderia gerar prejuízos concretos. Portanto, a melhor estratégia era estar sempre preparado para reagir de forma drástica, caso fosse necessário.

A violência institucionalizada também se estendia ao tempo livre das camadas livres. Tensões originadas durante o trabalho poderiam ser “resolvidas” nos momentos de lazer, propícios à realização de façanhas que proporcionavam senso de valor pessoal aos cidadãos livres. Aqueles reconhecidos como “valentões” gozavam de status e recebiam o apoio do “público”, que prefere estar ao lado de quem tem maior probabilidade de vencer para evitar futuras represálias. A indigência cultural transformava disputas violentas em momentos de entretenimento e de integração social:

Compreende-se que os ambientes de lazer sejam propícios para reacender antigas disputas ou deflagrar antagonismos, quando se descobre o vivo espírito de provocação que está na base dos divertimentos. O desafio faz sua reaparição em cena, agora como forma básica de expressão de relações lúdicas. A passagem do gracejo para a agressão é rápida e contínua: o espicaçamento zombador diretamente leva ao revide do sujeito atingido, cuja afirmação se faz já a sério e com animosidade. (...) O desafio está presente, aparecendo novamente como elo entre diversão e agressão. (FRANCO, 1983, p.40)

A autora conclui ainda que a violência entre as classes livres era um imperativo moral, pois o valor dos indivíduos estava centrado na coragem pessoal. Assim, “os riscos de assalto não são evitados, mas ousadamente enfrentados” e “não há outro recurso socialmente aceito senão o revide hábil para restabelecer a integridade do agravado” (FRANCO, 1983, p.51). O trecho a seguir é especialmente revelador da violência como imperativo moral entre as populações brancas pobres do Vale do Paraíba no final do século XIX:

A visão de si mesmo e do adversário como homens integrais impede que as desavenças sejam conduzidas para lutas parciais, mas faz com que tendam a transformar-se em lutas de extermínio. Em seu mundo vazio de coisas e falta de regulamentação, a capacidade de preservar a própria pessoa contra qualquer violação aparece como a única maneira de ser: conservar intocada a independência e ter a coragem necessária para defendê-la são condições de que o caipira não pode abrir mão, sob pena de perder-se. A valentia constitui-se pois, como o valor maior de suas vidas. (1983, p. 63).

³ No início do século XX, a indústria nascente e a cafeicultura moderna se basearam na mão-de-obra imigrante europeia. A pobreza relativa que persistiu entre os brancos livres no Vale do Paraíba dá a dimensão das dificuldades desta população em se inserir na economia após as modernizações republicanas (FRANCO, 1983).

Reações violentas a conflitos rotineiros dificilmente suscitavam algum senso de injustiça ou de desproporção, mesmo que levassem a desfechos trágicos. A violência contava com a admissão tácita de suas testemunhas, que a consideravam inofensiva, corriqueira ou de foro íntimo, e por isso, não criminosa ou passível de punição. A evasão dos acusados era frequente. Intervenções não se justificavam e, caso ocorressem, poderiam prejudicar quem “se metesse” em uma situação “privada”. O envolvimento das autoridades era visto como uma traição do pacto de silêncio entre os membros da comunidade. A primazia da “privacidade” sobre o cumprimento das regras do espaço público pode ser considerada um traço da “cordialidade” brasileira descrita por Holanda (1984). As pessoas consideravam que nenhuma forma de justiça formal seria capaz de compensar os prejuízos que a perda da reputação acarretava, tidos como piores do que a vitimização em si.

Franco (1983) ressalta que não havia nenhuma admissão do descrito acima nem qualquer defesa consciente de práticas violentas nos depoimentos às autoridades registrados nos boletins de ocorrência. Pelo contrário: os depoentes ressaltavam a futilidade dos conflitos e a desproporção entre ofensa e retaliação. A autora conclui que as pessoas agiam de forma semiautomática, sob a premissa de que a violência era natural, inofensiva ou justificada.

O cenário descrito por Franco (1983) ao analisar as camadas brancas livres do Vale do Paraíba após a Abolição (1888) sugere que a violência intrínseca à escravidão de fato se naturalizou na sociedade brasileira. Surpreendentemente, este cenário se assemelha ao descrito pelos primeiros estudos quantitativos sobre a violência no sul dos Estados Unidos.

Para o pioneiro Gastil (1971), a violência sulista apresentava uma qualidade particular, a *southernness*, mensurada por um índice desenvolvido pelo autor. Simpson (1985) verificou que os índices de *southernness* foram mais importantes do que aspectos socioeconômicos para explicar as taxas de contemporâneas de homicídio, roubo, estupro e agressão não letal na região. Nisbett e Cohen (1996), por sua vez, afirmaram que a *culture of honor* ou cultura da honra explicaria a violência na região, definindo-a como "uma causa singular de violência entre homens – o senso do perpetrador de que uma de suas mais valiosas posses está ameaçada, (...) sua reputação de força e dureza". Aqui, vê-se uma semelhança com a violência enquanto imperativo moral descrita por Franco (1983) anteriormente.

Em uma sociedade marcada pela cultura da honra, o acesso aos recursos é proporcional à confiança que os membros da comunidade têm na capacidade de ataque ou defesa de um indivíduo. Isto é: a reputação (justificada ou não) de implacabilidade e eficácia no uso da violência é fundamental para garantir a segurança e a manutenção do padrão de vida. Esses

traços culturais são comuns em sociedades nas quais não há poderes centralizados com legitimidade suficiente para estabelecer regras e punir violações, o que desestimularia a decisão de cometê-las ou de retaliá-las violentamente.

A colonização de exploração supunha um afastamento entre poderes metropolitanos e coloniais, como se verificou tanto no Brasil como no sul dos Estados Unidos. Após a Independência (1776), os sulistas seguiram considerando que os poderes centrais (agora federais) seriam incapazes de induzir ao cumprimento da Lei devido a seu distanciamento da realidade local. Assim, os proprietários de terras definiram o código de conduta das suas esferas de influência: a retaliação direta era a principal forma de punição, e se sua crueldade se tornasse suficientemente notória, se tornaria a principal forma de desencorajar adversários.

Esse cenário, por sua vez, comprometeu ainda mais a legitimidade do Estado em face dos poderes locais. Normas legais e decisões judiciais eram preteridas, e os membros das comunidades favoreciam soluções extraoficiais por entender que renunciar a retaliações violentas levaria a prejuízos reais, mas o descumprimento de ordens oficiais, não. O poder de fato era das autoridades locais, que não abdicaram das práticas violentas para se submeter ao Estado de Direito. Com isso, punir ou desencorajar violações de forma racional (isto é, utilizando a mínima força necessária) e imparcial se tornava cada vez mais difícil, e a reputação violenta seguia como principal forma de defesa. A dinâmica de reforço das autoridades locais arbitrariamente violentas e comprometimento da legitimidade do Estado de Direito em um contexto de colonização de exploração se assemelha muito à que ocorreu no Brasil.

Utilizando dados de questionários e censos, Nisbett e Cohen (1996) concluíram que os sulistas valorizavam a violência para proteger a propriedade privada, para retaliar insultos e para educar crianças. As instituições sulistas tendiam a ser mais tolerantes com indivíduos que cometiam crimes “de honra”, ou seja, associados à proteção da reputação, como duelos e linchamentos. Os autores destacam que as crianças, principalmente os meninos, eram socializadas para demonstrarem agressividade e para considerarem a violência como forma de entretenimento (p.18). Essas características remetem à violência costumeira e à violência institucionalizada descritas por Franco (1983) sobre as relações sociais no Vale do Paraíba após a Abolição (1888).

Tanto Gastil como Nisbett e Cohen alegaram que a *southernness* e a cultura da honra não seriam consequências da escravidão no sul dos Estados Unidos, atribuindo-as à influência da economia pastoril, da cultura britânica e à ausência de legitimidade na aplicação da lei. Os autores não se propuseram a examinar a influência da escravidão como principal determinante

da violência no sul dos Estados Unidos, mas é notável que a tenham minimizado em face de outros fatores com menor disponibilidade de dados, entre outras fragilidades. A ênfase de Nisbett e Cohen (1996) na “herança pastoril britânica” como origem da cultura de honra sulista parece evadir a questão central: por que essa herança não produziu os mesmos níveis de violência no Norte, que também recebeu colonizadores britânicos?

A concentração espacial da violência contemporânea justamente nas regiões historicamente escravagistas sugere que a escravidão desempenhou papel mais central na formação desses padrões do que os autores estavam dispostos a reconhecer (ou a investigar). Como vimos anteriormente, diversos estudiosos que exploraram a mesma questão – as particularidades da violência no sul dos Estados Unidos – demonstraram que a semelhança com o Brasil (FRANCO, 1983) não seria mero acaso, mas fruto da influência escravagista em ambas as sociedades (ROSENTHAL, 2018; COOK; LOGAN; PARMAN, 2018). Os trabalhos de Gastil (1971) e Nisbett e Cohen (1996) sugerem que mesmo aqueles que atribuíram um papel secundário à escravidão conseguem observar traços culturais que revelam a sua influência persistente em sociedades pós-escravagistas.

Para encerrar esta seção, abordaremos os efeitos do racismo na sociedade brasileira à luz de evidências empíricas. O racismo é, ao mesmo tempo, uma consequência da escravidão e um reforço de outros dos seus malefícios. A discriminação racial é frequentemente vista como subjetiva ou abstrata. Contudo, o avanço das técnicas de estimação causal vem demonstrando que populações negras enfrentam prejuízos socioeconômicos diretamente causados pelo racismo (institucional, formal, informal, espacial, entre outros) (BOTELHO; MADEIRA; RANGEL, 2015; TRUZZI; LIRIO; CERQUEIRA; COELHO; 2022; LEUNG-GAGNÉ, 2023). Em contrapartida, esses prejuízos reduzem os custos de oportunidade de diversos comportamentos (como abandono escolar, por exemplo) associados a maiores níveis de violência, retroalimentando esse círculo vicioso.

No Brasil, a Abolição (1888) não foi acompanhada por políticas de reparação, acesso à terra, inclusão no mercado de trabalho livre ou outras formas de promoção da cidadania de fato. Os libertos e seus descendentes foram sistematicamente excluídos dos setores mais modernos da economia da época: a cafeeicultura e a indústria nascentes, que priorizaram a força de trabalho de imigrantes europeus brancos (FERNANDES, 1965, 1972; QUEIROZ, 2021). Sem ações diretas de reparação, essa exclusão se cristalizou, fomentando desigualdades raciais transmitidas às próximas gerações por mecanismos formais e informais, como discriminação institucional, dificuldade de acesso a bens essenciais e segregação espacial. Dados do passado

recente apontam para a persistência dessas desigualdades, e a relação destas com a discriminação racial foi explorada em diversos estudos empíricos analisados a seguir.

Em 2021, embora representasse 56,1% da população brasileira e 55,2% da força de trabalho, a população negra⁴ ocupava apenas 29,5% dos cargos gerenciais. Brancos, que somavam 43,8% da população total, ocupavam 69% desses cargos. Negros eram maioria entre trabalhadores desocupados (64%) e informais (IBGE, 2022). A renda domiciliar *per capita* dos brancos foi quase o dobro da dos negros. Do total de pessoas vivendo em pobreza extrema (menos de US\$ 5,50/dia), 72,8% eram negras, contra 18,6% brancas (IBGE, 2022). O rendimento médio real do trabalho foi maior para brancos em todos os níveis de ensino. Com ensino superior completo, brancos receberam em média 50% a mais que pretos e 40% a mais que pardos (IBGE, 2022).

Além das desvantagens no mercado de trabalho, a população negra enfrenta desigualdades de acesso à educação, agravadas pela pandemia de coronavírus, em 2020. Neste ano, alunos negros foram maioria entre aqueles que não receberam atividades escolares, que tiveram menos de 5 dias semanais dedicados à escola e que dedicaram menos de duas horas diárias aos estudos. As taxas de comparecimento ao Exame Nacional do Ensino Médio (ENEM) foram cerca de 10% menores entre negros do que entre brancos (IBGE, 2022). Desigualdades de acesso à educação e ao emprego estão associadas à pobreza, que, por sua vez, está associada a maiores níveis de violência e vitimização.

Em 2020, a taxa de homicídios entre brancos foi de 11,5 vítimas a cada cem mil habitantes; 21,9 entre pardos e 34,1 entre pessoas pretas. Analisando as taxas de homicídios da população masculina (cerca de 12,5 vezes maior que da população feminina), é possível observar diferenças raciais ainda mais pronunciadas. A taxa de homicídios entre homens pretos foi aproximadamente o triplo (64,3 vítimas a cada cem mil habitantes) da taxa entre brancos (21,2). Entre pardos, a taxa de homicídios foi aproximadamente o dobro (41,4) desta última. A maioria das vítimas de homicídio no país foram homens de 15 a 29 anos. Neste grupo, pretos representaram 94,4 vítimas a cada cem mil habitantes; pardos, 136,5; e brancos, 41,6 (IBGE, 2022).

As estatísticas de violência contra a mulher também exibem disparidades raciais flagrantes: em 2021, mulheres negras foram 62% das vítimas de feminicídio, 70,7% das vítimas de mortes violentas intencionais e 52,2% das vítimas de estupro e estupro de vulnerável (FBSP, 2023).

⁴ 47,0% de pardos e 9,1% de pretos (IBGE, 2022).

Em relação à vitimização letal, dados recentes mostram que as disparidades não apenas persistem, como tendem a se agravar: segundo o 17º Anuário Brasileiro de Segurança Pública, entre 2021 e 2022, homicídios de brancos diminuíram 26,5%, enquanto entre negros aumentaram 7,5% (FBSP, 2023).

Analisadas em conjunto, todas essas estatísticas sugerem que existem relações entre diferentes formas de discriminação racial e os obstáculos ao desenvolvimento socioeconômico enfrentados pelas populações negras (e vice-versa). Isso, por sua vez, corrobora a premissa de que o sistema escravagista – que instituiu a negação da humanidade dos trabalhadores negros, naturalizando sua exploração e a precariedade decorrente – afeta as sociedades pós-escravagistas a longo prazo, com efeitos concretos observáveis e mensuráveis. Desde a última década, a literatura de causalidade vem produzindo evidências robustas da presença do racismo em diversos âmbitos da sociedade brasileira, como o mercado de trabalho, o sistema educacional e a segurança.

Com uma amostra de cerca de 277 mil alunos de escolas públicas do estado de São Paulo, Botelho, Madeira e Rangel (2015) verificaram que os professores atribuíam notas mais baixas para alunos negros do que para alunos brancos que apresentavam o mesmo desempenho. Os autores ressaltaram que este viés pode levar à maior evasão escolar entre estudantes negros. Leung-Gagné (2023) observou que a discriminação racial nas escolas brasileiras se assemelha à da Carolina do Norte, nos Estados Unidos. Analisando os dados da Prova Brasil para alunos do 5º e do 9º ano, o autor concluiu que o país padece de “segregação ao acaso”, isto é: a distribuição dos alunos ocorre sem intervenções diretas que busquem segregá-los. Contudo, ao não agirem para corrigir as desvantagens raciais naturalizadas na sociedade brasileira, as escolas acabam por reproduzi-las, e o resultado acaba sendo muito semelhante ao que seria se medidas de segregação racial explícitas tivessem sido implantadas.

As desigualdades raciais observadas na escola seguem para o mercado de trabalho. Diferenças salariais que não podem ser explicadas por diferenças de capital humano ou de contribuição para os resultados econômicos apontam para o racismo como sua causa mais plausível. Analisando a Pesquisa Nacional por Amostra de Domicílios (PNAD), Arcand e D’Hombres (2004) foram alguns dos primeiros a correlacionar diferenças salariais entre negros e brancos à discriminação racial. Em seu estudo sobre o mercado de trabalho feminino, Fernandes (2015) observou que a diferença salarial entre as trabalhadoras mais eficientes foi alta, e ressaltou que políticas de equidade salarial precisam considerar explicitamente raça e cor

para serem eficazes. Ambos os trabalhos utilizaram a decomposição de Oaxaca-Blinder para observar os componentes da renda do trabalho de pessoas brancas e negras.

Fiuza-Moura, Maia e Souza (2018) encontraram evidências de discriminação racial no setor manufatureiro entre 2003 e 2013, ainda que a diferença salarial entre negros e brancos tenha diminuído ao longo do período, especialmente nas indústrias mais intensivas em tecnologia. Ao analisarem o processo de abertura e liberalização econômica após a Redemocratização (1988-95), Barros e Santos Silva (2021) também observaram progresso. Os autores concluíram que os salários de mulheres não brancas aumentaram em relação aos dos homens e mulheres brancos, e que a discriminação racial diminuiu entre os níveis salariais mais altos. Contudo, as reformas econômicas também estão relacionadas ao aumento da discriminação racial nos níveis salariais mais baixos.

Sobre o comportamento das firmas brasileiras, Gerard, Lagos, Severini e Card (2021) observaram que pessoas não brancas (isto é, negras, indígenas, asiáticas, entre outros) tinham menor probabilidade de trabalhar em estabelecimentos que pagam mais para todos os grupos raciais, assim como de receberem bonificações financeiras. Tal sugere que trabalhadores não brancos tendem a ter dificuldades de acesso a empregadores que não praticam discriminação salarial por cor/raça. Cerca de 20% das diferenças de salário entre brancos e não brancos podiam ser atribuídas à menor probabilidade de não brancos trabalharem em estabelecimentos livres de discriminação salarial. A menor probabilidade de receberem bônus respondia por 5%.

Em geral, cerca de dois terços da sub-representação de pessoas não brancas em empresas que pagam altos salários podia ser atribuída a diferenças de capital humano. Contudo, os autores encontraram evidências de discriminação racial na contratação e nos salários entre trabalhadores com ensino superior completo. A discriminação de trabalhadores não brancos com educação superior leva a concluir que o racismo tem um custo relativamente alto para o mercado de trabalho e a economia brasileira.

Miller e Schmutte (2021) elaboraram um modelo para determinar se há discriminação racial nas contratações por indicação, comuns no mercado de trabalho brasileiro. Analisando a demanda de trabalho, os autores concluíram que os empregadores tendiam a priorizar contratações de pessoas com a mesma cor ou raça, mas abandonavam este comportamento à medida que as firmas aumentam de tamanho. Além disso, notaram que o empreendedorismo era cerca de duas vezes maior entre brancos do que entre não brancos. Essas duas conclusões sugerem que contratações por indicação são uma desvantagem agregada para os trabalhadores não brancos, o que ajuda a explicar por que eles têm maior probabilidade de serem demitidos,

menores chances de chegar a cargos mais altos e por que tendem a procurar trabalho em empresas maiores (p. 28).

Por fim, estudos empíricos recentes observaram que a discriminação racial de fato agrava a violência no Brasil. Truzzi, Lírio, Cerqueira, Coelho e Cardoso (2022) calcularam o diferencial de vitimização racial entre brancos e negros em relação a homicídio e agressão não letal. Com base nos dados da PNAD e do Sistema de Informações sobre Mortalidade (SIM-DATASUS), os autores utilizam a decomposição de Oaxaca-Blinder para concluir que, para ambos os crimes, 40% desse diferencial se deve à discriminação racial estrutural. Os maiores níveis de discriminação foram encontrados nas regiões Norte e Nordeste, caracterizadas por maiores níveis de desigualdade de renda e de violência.

Esta revisão de literatura demonstrou que estudos empíricos apoiam a hipótese de que a escravidão afeta os níveis de violência contemporâneos no Brasil. Análises do sul dos Estados Unidos – sociedade pós-escravagista comparável ao Brasil, mas cuja escravidão foi relativamente menor – encontraram correlações significativas entre escravidão e violência, bem como entre a primeira e diversos mecanismos pelos quais seu legado persistiu. Sobre esta região, a literatura aponta a desigualdade de renda, a normalização da violência e a segregação institucional e espacial como principais transmissores.

Estudos sobre outras nações pós-escravagistas na África e na América Latina também encontraram relações significativas entre escravidão e desigualdade de renda, segregação institucional (na educação e no mercado de trabalho), baixa confiança e baixa legitimidade das instituições. O efeito causal entre escravidão e fenômenos que contribuem para a violência já foi estimado para o Brasil, mas o efeito causal direto sobre a violência ainda representa uma lacuna na literatura empírica.

Paralelamente à análise empírica, a sociologia foi fundamental para revelar outras evidências e identificar dinâmicas que explicam o transbordamento da violência escravagista e sua persistência no tempo. Estudos sociológicos sobre o Brasil e sobre o sul dos Estados Unidos após o fim da escravidão em cada uma dessas sociedades revelaram semelhanças notáveis em relação aos padrões de violência, mesmo quando seus autores descartam a escravidão como determinante. Dentre os estudos que reconhecem o papel crucial da escravidão, a hipótese de transbordamento da violência inerente ao sistema produtivo para todas as classes sociais emerge nas análises de ambas as regiões.

Por fim, estudos empíricos recentes concluíram que a discriminação racial tem impactos significativos na educação, no mercado de trabalho e no sistema judiciário brasileiros. O

racismo é um dos traços culturais diretamente relacionados à experiência escravagista, e a mensuração dos seus impactos contribui para as evidências de que a abolição formal da escravidão não foi suficiente para conter a transmissão do seu legado.

Diante do exposto, a hipótese deste trabalho é que a escravidão abolida em 1888 tem impacto direto sobre os níveis de violência nos municípios brasileiros em 2000. Contudo, limitações na oferta de dados dificultam significativamente a estimação desses efeitos. Para contornar essas dificuldades, este trabalho emprega uma estratégia metodológica descrita na seção seguinte, que visa estimar o efeito causal da escravidão em 1872 sobre as taxas de homicídios municipais em 2000.

3 METODOLOGIA

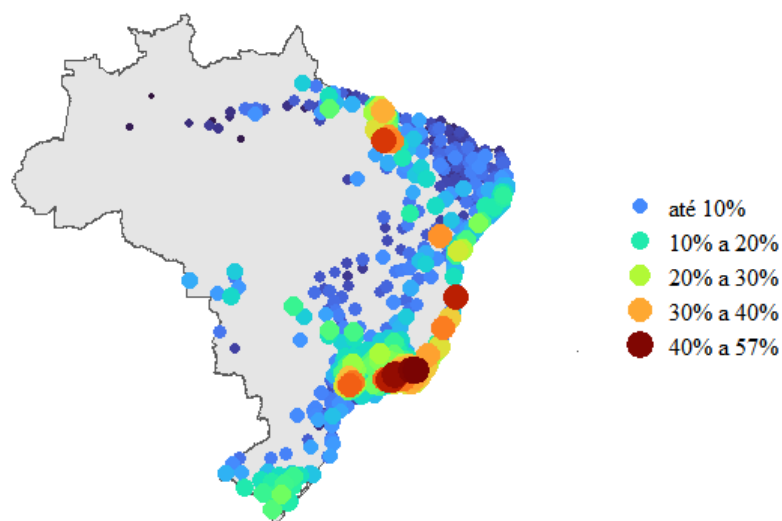
3.1 PECULIARIDADES DA AMOSTRA: IMPLICAÇÕES METODOLÓGICAS

A vigência do sistema escravagista no Brasil pode ser considerada um experimento natural, isto é: um evento exógeno que altera o ambiente em que os agentes estudados operam (SHADISH; COOK; CAMPBELL, 2002). O povoamento se concentrou no litoral (PALMA; PAPADIA; PEREIRA; WELLER, 2021) devido à lucratividade do tráfico transatlântico, o que tornava esforços de interiorização comparativamente mais custosos para a metrópole portuguesa (FUJIWARA; LAUDARES; VALENCIA-CAICEDO, 2017). A divisão entre municípios atingidos e não atingidos pela escravidão ocorreu de forma “natural” – isto é, independente da ação de um pesquisador, e de acordo com fatores históricos e geográficos que caracterizaram o experimento escravagista no Brasil.

Ao longo de todo o período escravagista, o Censo Imperial de 1872 foi o único a coletar dados sobre a população escravizada. Utilizando Áreas Mínimas Comparáveis (AMCs), o Instituto Brasileiro de Geografia e Estatística (IBGE) aplicou a divisão municipal do Censo Demográfico de 2010 a todos os Censos anteriores (IBGE, 2011). Com isso, foi possível estabelecer equivalência entre a localização da população escravizada em 1872 e as taxas de homicídios municipais em 2000 – variável resultante utilizada neste trabalho, detalhada na Seção 3.3 – uma vez que ambos os Censos de 1872 e 2000 passaram a utilizar a mesma definição de limites municipais (IBGE, 2010).

A equivalência geográfica permitiu obter os valores da variável de tratamento que representa a escravidão brasileira em 1872, análoga à empregada por Gouda e Rigterink (2017) ao estudar a relação entre escravidão e violência nos municípios dos Estados Unidos. A intensidade da escravidão (identificada pela sigla *sl_d*) é dada pelo percentual de pessoas escravizadas na população total de um município em 1872 (sob os limites municipais de 2010 (IBGE, 2010)). Devido às limitações na disponibilidade de dados sobre escravidão e violência no Brasil, utilizar a mesma definição de variável de tratamento não é suficiente para que os resultados deste trabalho sejam comparáveis aos encontrados por esses autores. Contudo, utilizar sua definição contribui para testar sua representatividade em relação à escravidão brasileira. A Figura 1 ilustra a distribuição da variável de tratamento ao longo dos municípios brasileiros em 1872.

Figura 1 - Distribuição da intensidade da escravidão (%) em 1872



Fonte: elaboração própria com o software *RStudio*.

Na Figura 1, os pontos coloridos representam todos os municípios ocupados em 1872. Neste ano, apenas 640 dos 5.565 municípios listados em 2010 já existiam e estavam ocupados. Todos contavam com uma parcela de habitantes escravizados entre a população total (IBGE, 2011), o que ilustra a dependência da economia brasileira em relação ao sistema escravagista (mesmo que decadente). Portanto, os pontos coloridos representam o grupo de tratamento: municípios cuja intensidade da escravidão é maior que zero. As áreas desocupadas em 1872, em cinza, representam o grupo de controle deste estudo: sua intensidade da escravidão é nula, uma vez que não estavam povoadas em 1872.

Na Figura 1, as bolhas menores (que vão do azul escuro aos tons de verde) representam municípios com até 30% de habitantes escravizados em 1872. Estas bolhas estão concentradas nas atuais regiões Nordeste e Sudeste. Áreas mais interiores do território também apresentam bolhas pequenas, mas com menor concentração: a dispersão aumenta à medida que se avança para o interior do país. Por outro lado, bolhas laranjas e vermelhas indicam municípios cuja população escravizada representa mais de 30% do total de habitantes: estas se concentram maciçamente no litoral da região Sudeste e nas fronteiras da região Nordeste. Não existem centros de intensidade da escravidão igual ou superior a 30% no interior do país e nem nas regiões Norte, Sul e Centro-Oeste. A Figura 2 representa o total de pessoas escravizadas, o total de pessoas residentes e o valor da intensidade da escravidão (*sl_d*) em 1872 por região geográfica (IBGE, 2010).

Figura 2 - Intensidade da escravidão em 1872 por região geográfica

	População Escravizada (hab.)	População Total (hab.)	Intensidade da escravidão (<i>sld</i>) (% pop.)
Norte	30.410	370.446	8,21%
Nordeste	480.884	4.641.793	10,36%
Sudeste	891.548	4.029.974	22,12%
Sul	93.335	722.327	12,92%
Centro-Oeste	15.346	183.213	8,38%
Brasil	1.511.523	9.947.753	15,19%

Fonte: elaboração própria com o auxílio do software RStudio.

A Figura 2 mostra que, em 1872, as regiões Sudeste e Nordeste apresentavam os maiores valores de população absoluta e de população escravizada, evidenciando a importância do sistema escravagista na sua história e na sua economia. As Figuras 1 e 2 indicam que a seleção dos grupos de tratamento e controle não foi aleatória: municípios das regiões Sudeste e Nordeste tem maior probabilidade de estarem povoados em 1872, quando todos os municípios brasileiros possuíam pessoas escravizadas entre os residentes (IBGE, 2011). Logo, tem maior probabilidade de pertencerem ao grupo de tratamento.

Como mencionado no início desta seção, a explicação para a divisão dos grupos se encontra nos interesses da metrópole portuguesa: a proximidade entre o litoral do Sudeste e do Nordeste com as costas da África e da Europa reduzia os custos do tráfico transatlântico de pessoas escravizadas. Além disso, o clima nessas regiões era mais propício para a agricultura de exportação característica do Brasil-colônia. Por isso, as regiões Sudeste e Nordeste foram os maiores centros econômicos e populacionais do período observado, nas quais a infraestrutura do sistema escravagista estava concentrada.

A Figura 2 mostra que os maiores valores de *sld* são os das regiões Sudeste e Sul, respectivamente (em vermelho). Tal ocorre porque esta última é relativamente menos populosa que o Nordeste, e seu volume de pessoas escravizadas é proporcionalmente maior. Contudo, a região Sul foi uma colônia de povoamento, cuja relação com a escravidão é radicalmente diferente da observada nas regiões Sudeste e Nordeste. Os fatores históricos que explicam a distribuição dos grupos de controle e de tratamento, o contexto geográfico e econômico e os valores absolutos de população escravizada em 1872 justificam a escolha de analisar as regiões Sudeste e Nordeste individual e conjuntamente, bem como todo o território nacional. Com isso, busca-se manter coerência entre estratégia de estimação e narrativa histórica, para garantir a explicabilidade dos resultados obtidos.

Como a divisão dos grupos de tratamento e controle não foi aleatória e reflete o contexto histórico e econômico da colonização do Brasil, é razoável assumir que pertencer a um ou outro grupo tem impacto significativo sobre os valores das variáveis de controle referentes aos municípios observados. Por exemplo: municípios do grupo de tratamento tendem a ter maiores indicadores de renda, emprego, densidade populacional (e até de violência) pois foram povoados antes dos municípios do grupo de controle. Quando estes últimos estavam começando a ser povoados, os primeiros já concentravam infraestrutura produtiva, capital e contingentes populacionais.

Esse cenário se reflete nos valores de renda, emprego e densidade populacional – alguns dos controles utilizados neste trabalho – no passado recente. Sendo assim, como é possível afirmar que a intensidade da escravidão em 1872 tem efeito causal e positivo sobre as taxas de homicídio municipais em 2000? Como ter certeza de que a causa é a intensidade da escravidão e não outro fator que leve o grupo de tratamento a ter níveis de violência significativamente maiores ou menores que os do grupo de controle?

A Figura 3 exibe a correlação de Pearson entre a variável de tratamento *sld* e as variáveis de controle utilizadas neste trabalho, descritas detalhadamente na Seção 3.3 (ver Figura 4). A amostra consiste nos 5.498 municípios para os quais os valores de tratamento e controles estão completos. Todos os controles se referem ao ano 2000, e *sld*, a 1872.

Figura 3 - Correlação de Pearson entre controles e tratamento (*sld*)

Variável de Controle	Correlação com <i>sld</i>
Renda domiciliar <i>per capita</i> média (<i>ypc</i>)	0,139
Densidade populacional (<i>pd</i>)	0,060
Desemprego (<i>unmp</i>)	0,149
Educação Básica (<i>be</i>)	0,186
Homens Jovens (<i>ym</i>)	0,018
Índice Municipal de Justiça e Segurança Pública (<i>cps</i>)	0,244
Longitude (<i>long</i>)	0,105
Latitude (<i>lat</i>)	-0,009
Desigualdade de Renda - Índice de Gini (<i>gini</i>)	0,075
Desigualdade de Renda - Índice de Theil (<i>theil</i>)	0,084

Fonte: elaboração própria com auxílio do software *RStudio*

Na Figura 3, é possível observar que a correlação entre a intensidade da escravidão e diversas variáveis de controle é relevante, com destaque para a renda domiciliar *per capita* média, o desemprego, a educação básica e o índice municipal de justiça e segurança pública – criado neste estudo para capturar a qualidade institucional dos municípios em 2000 (ver Seção 3.3). Logo, o valor dos controles difere significativamente entre os grupos de tratamento e de controle, de modo que diferenças nas suas taxas de homicídio (médias) não podem ser diretamente atribuídas ao tratamento. Utilizar apenas municípios suficientemente semelhantes de cada um dos grupos reduziria consideravelmente a amostra, comprometendo a representatividade dos resultados. Para preservar ao máximo o tamanho da amostra disponível, emprega-se o método de ponderação por escores de propensão – em inglês, *propensity score weighting* (PSW), como é mais conhecido.

Segundo o método PSW, a probabilidade de uma observação pertencer ao grupo de tratamento representa uma estimativa da correlação entre variável de tratamento e variáveis de controle. Ao ponderar as observações de acordo com essa probabilidade, a correlação entre tratamento e controles passa a ser considerada pelo modelo através dos pesos, e deixa de gerar confundimento. Após a ponderação, a amostra ponderada é dita equilibrada se a variável de tratamento passar a ser independente dos controles. Em estudos causais, isso quer dizer que, com base na amostra ponderada equilibra, as diferenças dos grupos de tratamento e de controle em relação à variável resultante podem ser atribuídas diretamente ao experimento natural. Através da ponderação, o método PSW corrige as limitações da amostra sem comprometer seu tamanho.

Como visto na Seção 1, o presente estudo não visa obter estimadores causais. Entretanto, o emprego do método PSW é necessário para obter correlações não espúrias (verdadeiras) entre a variável que representa a escravidão em 1872 e as taxas de homicídio municipais em 2000 porque o controle do confundimento leva em consideração as particularidades da variável de tratamento. A intensidade da escravidão é contínua: assume zero para o grupo de controle, e varia no intervalo de 6,87% a 57,41% entre o grupo de tratamento. Originalmente, o método PSW foi desenvolvido para abordar variáveis de tratamento binárias (ROSENBAUM; RUBIN, 1983). Hirano e Imbens (2004) estenderam-no para gerar estimadores causais quando o tratamento é contínuo, concebendo a chamada função dose-resposta.

A função dose-resposta pode ser melhor compreendida através de um exemplo comum nas Ciências da Saúde que é análogo ao cenário definido pela escravidão brasileira enquanto experimento natural. Imagine-se um experimento em que um grupo de pessoas não ingeriu um

medicamento experimental, enquanto o outro grupo ingeriu diferentes doses deste medicamento. Considerando que seu efeito pode ser mais forte para alguns e mais fraco para outros, a função dose-resposta fornece o efeito médio do medicamento sobre a população observada (conhecido como *average treatment effect* (ATE)). Na presente análise, o primeiro grupo representa os municípios não existentes em 1872 (cuja intensidade da escravidão é nula), e o segundo, os municípios existentes em 1872 (cuja intensidade da escravidão varia em um intervalo). Contudo, dado que o objetivo não é obter estimadores causais, este trabalho não visa determinar o ATE (Seção 1), mas sim uma evidência robusta da relação entre escravidão em 1872 e violência letal em 2000, através de um estimador correlacional positivo e significativo de *sld*.

Para obter estimadores adequados, é necessário corrigir as limitações da amostra: não-aleatoriedade, correlação entre variável de tratamento e variáveis de controle. Por isso, as unidades amostrais são ponderadas pelo inverso dos seus escores de propensão generalizados (*generalized propensity scores* (GPS), como são mais conhecidos), que estimam a probabilidade de pertencer ao grupo de tratamento levando em conta que a intensidade da escravidão assume diversos valores em um intervalo (ou seja: que o tratamento é contínuo). Portanto, ainda que os objetivos não sejam causais, técnicas de equilíbrio da amostra tradicionalmente aplicadas para obter causalidade de tratamentos contínuos levarão a estimadores correlacionais mais robustos de *sld*. A ponderação pelo inverso dos GPSs é conhecida como *inverse probability weighting* (IPW), de modo que, para simplificação, o método de ponderação utilizado será representado pela GPS-IPW (HEISS, 2020). Utiliza-se um algoritmo de aprendizado de máquina, o *boosting*, para estimar os GPSs com precisão (ZHU; COFFMAN; GHOSH, 2015).

3.2 GPS-IPW POR *BOOSTING* E FUNÇÃO DOSE-RESPOSTA

Esta seção explora a literatura de causalidade para ilustrar a lógica e a forma pela qual o método GPS-IPW é empregado neste estudo. Pelas razões descritas na Seção 1, o objetivo deste estudo não é chegar a um estimador causal da relação entre escravidão em 1872 e violência letal em 2000. A ótica da causalidade empregada nesta seção tem finalidade didática: o método GPS-IPW foi criado e descrito para atender obter estimadores causais de tratamentos contínuos caso haja confundimento, isto é, caso pertencer ao grupo tratado afete significativamente os valores observados das variáveis de controle.

Este trabalho analisa a escravidão como um experimento natural cuja variável de tratamento é contínua e cuja amostra observada apresenta confundimento entre a intensidade da escravidão municipal em 1872 e os valores das variáveis de controle observadas em 2000. Para abordar estas particularidades e obter estimativas robustas da relação entre intensidade da escravidão em 1872 e violência letal em 2000, emprega-se um método tradicionalmente utilizado em estudos causais. Como este não é o objetivo deste estudo, menções a causalidade nesta subseção devem ser interpretadas como recursos didáticos e ilustrativos.

Nesta seção, a taxa de homicídios municipal em 2000 $homr$ e a intensidade da escravidão municipal em 1872 sld serão representadas respectivamente por Y e T . \mathbf{X} é um vetor de dimensão p que contém as variáveis de controle (definidas na seção seguinte) e as variáveis binárias referentes aos efeitos fixos da estimação, representando vinte e cinco unidades federativas (UFs)⁵. De (Y, T, \mathbf{X}) , extrai-se uma amostra aleatória (Y_i, T_i, \mathbf{X}_i) , $i = 1, \dots, n$, que representa os dados observados referentes a n municípios.

A variável T é aleatória e está definida no intervalo τ , ao qual pertence o valor t . Para cada município i , existe um conjunto de potenciais valores de Y obtidos em função do valor de tratamento t aplicado ao município, isto é: $Y_i(t)$ para $t \in \tau$. O objetivo do método GPS-IPW é obter a função dose-resposta $\mu(t) = E[Y_i(t)]$, segundo a qual o estimador da relação entre sld e $homr$ é dado pelo valor esperado dos resultados potenciais para os níveis de tratamento aplicados às n observações. Assume-se que $Y_i(t)$ está definido para todo $t \in \tau$, pois é necessário que os resultados potenciais estejam definidos para todos os níveis de tratamento observados na amostra⁶.

Estimadores causais devem ser obtidos através do cumprimento da condição de ignorabilidade, segundo a qual a correlação entre tratamento e controles é nula se não houver variáveis omitidas na especificação dos modelos. Para isso, é necessário controlar todas as variáveis de confusão: aquelas que afetam a variável de tratamento e a variável resultante simultaneamente. Contudo, quando a variável de tratamento é contínua, não é necessário que os resultados potenciais de Y sejam conjuntamente independentes dos valores de T : basta que sejam condicionalmente independentes (HIRANO; IMBENS, 2004). Esta é a chamada condição de ignorabilidade fraca, descrita pela Equação 1:

$$f(t|Y(t), \mathbf{X}) = f(t|\mathbf{X}), \quad \forall t \in \tau. \quad (1)$$

⁵ Para evitar multicolinearidade entre as variáveis binárias, uma das unidades federativas é omitida.

⁶ Como as variáveis serão expressas em relação à amostra total, o subscrito i deixa de ser empregado a seguir.

Na Equação 1, $f(t|.)$ representa a densidade condicional de um nível de tratamento $t \in \tau$. A condição de ignorabilidade fraca determina que, dadas as variáveis de controle incluídas no modelo, a intensidade da escravidão num município deve ser condicionalmente independente dos potenciais valores da taxa de homicídios do local. Sob a condição de ignorabilidade fraca, pertencer ou não ao grupo de tratamento não afeta os resultados potenciais, e é possível considerar que os controles incluídos são suficientes para garantir a independência condicional de t em relação a $Y(t)$ (ZHU; COFFMAN; GHOSH, 2015).

Em experimentos naturais, não é possível assegurar a aleatoriedade da amostra e a independência condicional entre T e \mathbf{X} . No entanto, é matematicamente possível encontrar uma função $r(\mathbf{X})$ que seja condicionalmente independente de T e utilizá-la para ponderar as observações (HULING; GREIFER; CHEN, 2021). Assim, é possível atingir a independência condicional entre tratamento e controles utilizando a amostra ponderada.

Os *propensity scores* generalizados $r(t, \mathbf{X})$ são as densidades condicionais de cada nível de tratamento t dados os valores das variáveis de controle utilizadas (ZHU; COFFMAN; GHOSH, 2015). Com base nas características expressas pelos controles, estas funções determinam a probabilidade de ocorrência de cada nível de tratamento na amostra:

$$r(t, \mathbf{X}) \equiv f_{t|\mathbf{X}}(t|\mathbf{X}) \quad (2)$$

É possível expressar a condição de ignorabilidade fraca em relação aos *propensity scores* generalizados:

$$f(t|Y(t), r(t, \mathbf{X})) = f(t|r(t, \mathbf{X})), \quad \forall t \in \tau. \quad (3)$$

A Equação 3 afirma que, para eliminar o viés de confusão, é suficiente que os níveis de tratamento sejam condicionalmente independentes dos *propensity scores* (ZHU; COFFMAN; GHOSH, 2015). Isso significa que, em cada conjunto de municípios determinado pelo mesmo score $r(t, \mathbf{X})$, a probabilidade de que T seja igual a t não depende de \mathbf{X} . Dito de outra forma, em cada grupo de municípios com a mesma probabilidade de pertencer ao grupo de tratamento, o valor do tratamento em cada um deles não depende dos valores de seus controles.

Não é possível garantir que T e \mathbf{X} sejam independentes em toda a amostra, mas a condição de ignorabilidade fraca da Equação 3 afirma que é possível garanti-la entre municípios com a mesma probabilidade de pertencer ao grupo de tratamento, isto é, com o mesmo *score*.

Portanto, na amostra ponderada de acordo com as funções de densidade condicional $r(t, \mathbf{X})$, T e \mathbf{X} podem ser condicionalmente independentes. Com base nos dados observados $(Y_i, T_i,$

\mathbf{X}_i), $i = 1, \dots, n$, o peso atribuído ao município i é dado pelo inverso dos GPSs – e por isso, chamado de *inverse probability weighting* (IPW):

$$w_i = \frac{f_{T|\mathbf{X}}(T_i)}{f_{T|\mathbf{X}}(T_i|\mathbf{X}_i)} = \frac{r(T_i)}{r(T_i, \mathbf{X}_i)}, \text{ para } i = 1, \dots, n. \quad (4)$$

A Equação 4 mostra que a obtenção dos pesos depende apenas da relação entre T e \mathbf{X} , então é imprescindível que a densidade condicional de T em relação a \mathbf{X} seja realista. Entretanto, não é razoável assumir uma relação linear entre a intensidade da escravidão e os controles observados em 2000 (Figura 3). Sendo assim, é preciso estimar os pesos através de um método não paramétrico, que admite a não linearidade entre T e \mathbf{X} .

Como a condição de ignorabilidade fraca não pode ser testada, há incentivo para incluir controles e assim, calcular os pesos com o maior rigor possível. Por isso, há risco de “maldição da dimensionalidade” (ZHU; COFFMAN; GHOSH, 2015) em modelos GPS-IPW. O vetor \mathbf{X} pode atingir uma alta dimensão p , dificultando consideravelmente a estimação sem torná-la comprovadamente mais acurada. O modelo *boosted* generalizado, mais conhecido como *generalized boosted model* (GBM), seleciona os principais controles sem comprometer significativamente a precisão dos GPSs (MCCAFFREY; RIDGEWAY; MORRAL, 2004).

Algoritmos *boosting* são técnicas de aprendizado de máquina capazes representar adequadamente um vetor \mathbf{X} de alta dimensão p (isto é, um grande número de variáveis de controle) quando sua relação com T é não linear. Os algoritmos *boosting* assumem que a relação representada pela Equação 5 é válida para a amostra (ZHU; COFFMAN; GHOSH, 2015):

$$E[T|\mathbf{X}] = m(\mathbf{X}) \quad (5)$$

Na Equação 5, m representa a média de T em função de \mathbf{X} e é desconhecida. Ao contrário do que ocorre nos modelos que presumem linearidade entre T e \mathbf{X} , \mathbf{X} não é multiplicado por um vetor de coeficientes. Como a Equação 5 não presume coeficientes lineares, está aberta a diferentes distribuições de T dado \mathbf{X} .

Algoritmos GBM aplicam funções a pequenos grupos da amostra e depois as combinam para estimar uma função suave da relação entre T e \mathbf{X} . A primeira partição divide a amostra em dois grupos segundo algum critério (por exemplo, o valor da variável X_1). Em cada grupo resultante, estima-se a média simples de T (o valor de $m(\mathbf{X})$ para o grupo) utilizando árvores de

regressão, outro algoritmo não linear. As partições continuam, e de todas as possíveis formas de dividir a amostra, são escolhidas aquelas que levam ao menor erro de previsão de T (MCCAFFREY; RIDGEWAY; MORRAL, 2004).

Feito isso, o algoritmo GBM realiza combinações lineares das árvores de regressão com menores erros de previsão de T para gerar uma função mais suave, utilizada para estimar os *propensity scores*. Para determinar a função final, o algoritmo identifica qual das combinações lineares minimiza as diferenças entre os grupos. É comum definir o modelo ótimo como aquele que minimiza a correlação absoluta média (de Pearson) entre a variável de tratamento contínua e cada um dos controles (ZHU; COFFMAN; GHOSH, 2015). Assim, obtém-se o equilíbrio dos grupos amostrais e a redução do confundimento entre variável de tratamento e variáveis de controle.

Algumas variáveis independentes podem apresentar interações cuja relação com a variável resultante pode ser significativa. Um exemplo é a interação entre renda domiciliar *per capita* e desigualdade de renda, que foi incluída nas especificações dos modelos (Seção 3.3). Entretanto, ambas ainda podem interagir com o percentual de pessoas que estudaram oito anos ou mais ou com a taxa de desemprego. Estas interações são relativamente mais claras e mais relevantes do que outras (como as que envolvem variáveis binárias referentes às UFs, por exemplo). Contudo, interações omitidas atuarão como variáveis de confusão, comprometendo a qualidade dos estimadores do efeito do tratamento. Além de identificar e ordenar as variáveis mais importantes para prever o grupo a que um município pertence, o GBM é capaz de incluir interações entre as variáveis e termos de maior ordem automaticamente (ZHOU, 2015), evitando que interações ocultas levem à persistência do viés de seleção.

O algoritmo GBM também contribui para mitigar o *trade-off* entre variância e viés característico dos modelos GPS-IPW (ZHU; COFFMAN; GHOSH, 2015). Para corrigirem o desequilíbrio amostral quando a variável de tratamento é contínua, esses modelos costumam gerar um intervalo de pesos muito grande. A estimação é dita instável quando a variância do intervalo de pesos é alta. O risco de instabilidade aumenta quando há um grande número de controles e quando os grupos são muito diferentes entre si. Neste caso, poucas observações são diretamente comparáveis, e por isso, recebem pesos muito altos, enquanto a maioria das observações recebe pesos menores, o que pode distorcer o resultado das estimações.

Quando não há valores ausentes entre os dados observados, foi demonstrado que os pesos obtidos por GBM levam a estimadores mais precisos e a um menor erro quadrático médio em comparação com os métodos lineares (ZHU; COFFMAN; GHOSH, 2015). Para 5.498 dos

5.565 municípios determinados pela Divisão Territorial de 2010 (IBGE, 2013), todas as variáveis independentes estão completas (Figura 3). Entretanto, apenas 2.699 destes municípios reportaram suas taxas de homicídio em 2000. A estratégia adotada para abordar essas limitações é descrita a seguir.

A Equação 4 mostra que a estimação dos pesos depende apenas de T e X . Essencialmente, pesos corretos são aqueles que refletem a seleção dos grupos de tratamento e controle estabelecida pelo experimento natural. Como os dados de T e X estão disponíveis para os 5.498 municípios, a ponderação foi executada sobre esses 5.498 municípios. Portanto, a pseudoamostra ponderada tem $n = 5.498$. Reduzi-la com base nos valores ausentes da variável resultante significaria excluir valores observados das variáveis de controle, o que distorceria o retrato da relação empírica entre a escravidão em 1872 e a sociedade brasileira em 2000. Pesos que ignorassem os dados disponíveis não poderiam ser considerados acurados. A teoria econométrica recente ampara esta visão.

Alternativas comuns para lidar com dados ausentes são limitar a amostra às unidades cujos valores são todos observados (casos completos), já mencionada, ou fazer a imputação dos valores ausentes. Segundo Van Buuren (2018), se apenas a variável resultante possui valores ausentes, as duas abordagens se equivalem e geram estimadores não viesados. A imputação levaria às mesmas estimativas que a análise dos casos completos, mas aumentaria sua variância, gerando estimadores inconsistentes e não robustos (VON HIPPEL, 2007).

A imputação só seria uma alternativa viável se outras variáveis independentes além das contidas em X estivessem disponíveis. Neste estudo, os preditores tradicionalmente analisados (GOUDA; RIGTERINK, 2017; BUONANNO; VARGAS, 2019) já estão incluídos, e a natureza do algoritmo GBM previne significativamente a ocorrência de viés de variável omitida. Como apenas a variável resultante padece de valores ausentes, a imputação não representaria um avanço significativo, pois levaria aos mesmos resultados que a análise dos casos completos, mas aumentaria a variância dos estimadores (VAN BUUREN, 2018).

Estas premissas embasam a decisão de calcular os GPSs com base nos 5.498 municípios que exibem dados completos para todas as variáveis independentes. Os GPSs resultantes são aplicados a esses mesmos municípios e geram uma pseudoamostra com $n = 5.498$. Esta pseudoamostra, por sua vez, é utilizada para estimar uma regressão ponderada análoga à função dose-resposta, com o objetivo de determinar a correlação entre a intensidade da escravidão em 1872 e a taxa de homicídios municipal em 2000. Como não há imputação de valores ausentes da variável resultante, a amostra utilizada para estimar as regressões se reduz aos 2.699

municípios para os quais há dados de taxa de homicídios em 2000. Sob essas condições, é possível assumir que os pesos calculados pelo algoritmo GBM são mais acurados que aqueles calculados por métodos lineares, mesmo que referentes a uma amostra mais restrita dos municípios brasileiros.

Como é não paramétrico, o algoritmo GBM não fornece coeficientes de regressão. Ainda que observáveis, as árvores de regressão executadas não permitem descrever a associação entre cada variável de controle e a probabilidade de pertencer ao tratamento, como fazem os coeficientes de uma regressão linear (logística, por exemplo, frequentemente utilizada para obter *propensity scores* entre tratamento e controles linearmente relacionados).

Por outro lado, as árvores de regressão dividem a amostra segundo diferenças acuradamente detectadas entre os controles: isso significa que a probabilidade de pertencer ao tratamento é minuciosamente construída com base nas características de cada município. Tal ocorre porque os algoritmos de aprendizado de máquina têm como principal objetivo fazer previsões corretas, e não descrever um padrão de comportamento (como, em geral, fazem os modelos econométricos tradicionais). Um modelo que descreva a relação entre T e X didaticamente não será útil se não gerar pesos que descrevam esta relação com acurácia. Por isso, considerando os objetivos do modelo GPS-IPW, a natureza de “caixa preta” do algoritmo GBM (ZHOU, 2015) não representa uma desvantagem significativa.

É útil revisar brevemente as etapas da metodologia explicadas até aqui. Inicialmente, a relação entre tratamento e controles é examinada pelo algoritmo GBM, capaz de representar relações não lineares e interações entre as variáveis, assim como de assegurar a estabilidade dos pesos obtidos. As árvores de regressão geradas são linearmente combinadas para formar uma função suave, responsável por gerar os GPS. A amostra será ponderada pelo inverso destes, o que justifica a denominação GPS-IPW atribuída a este método. Gera-se então a pseudoamostra ponderada, na qual o confundimento entre variável de tratamento e controles é menor, e a correlação entre *sld* e *homr* é verdadeira, positiva, significativa e robusta.

Para verificar a eficácia do método GPS-IPW na redução do confundimento e garantir a robustez dos estimadores de *sld*, este trabalho verifica a correlação absoluta média de Pearson entre a variável de tratamento e cada um dos controles após a ponderação das unidades amostrais (ZHU; COFFMAN; GHOSH, 2015). Os critérios de avaliação das correlações são apresentados na Seção 4.1. Se forem suficientemente atendidos, entende-se os estimadores da intensidade da escravidão são associações robustamente ajustadas – mas não devem ser

interpretados como causais pelas razões expostas na Seção 1, que prejudicam o cumprimento da condição de ignorabilidade fraca expressa pela Equação 3.

Por último, a pseudoamostra é utilizada para estimar um modelo linear ponderado análogo à função dose-resposta, descrito pela Equação 6. Neste modelo, a relação entre a variável de tratamento e a variável resultante é linear. Enquanto é fundamental considerar a não linearidade ao estimar os pesos que relacionam T e \mathbf{X} , é razoável presumir que a relação entre T e Y é linear por causa da distância entre causa e efeito⁷ estudados. À medida que o tempo passa, afastando os municípios contemporâneos dos eventos de 1872, os níveis de letalidade tornam-se relativamente mais dependentes de fenômenos mais atuais. Assim, o impacto direto de eventos históricos passados é menor em relação aos fatores contemporâneos, de modo que seus efeitos verificáveis podem ser representados por formas funcionais mais simples, como as lineares. Na Equação 6, α_1 representa o efeito causal médio da intensidade da escravidão sobre as taxas de homicídio municipais em 2000:

$$E[Y(t)] = \alpha_0 + \alpha_1 t. \quad (6)$$

Note-se que, como as informações de \mathbf{X} passam a ser expressas pelos *scores* após a ponderação, não é necessário incluir este vetor no modelo. É suficiente executá-lo sobre a pseudoamostra equilibrada, obtida após a ponderação da amostra original pelos *scores* de propensão generalizados (método GPS-IPW). Os estimadores da Equação 6 possuem erros-padrão robustos (GREIFER, 2021). Não é necessário incluir as variáveis de controle no modelo linear ponderado pra obter estimadores robustos, mas incluí-las leva a estimadores duplamente robustos do efeito médio do tratamento (GREIFER, 2021), caracterizados por serem consistentes mesmo que a Equação 4 especificada ou os pesos calculados estejam incorretos (GRAHAM; MCCOY; STEPHENS, 2015).

Para além das definições anteriores e das precauções tomadas, é impossível descobrir se a especificação da Equação 4 está correta. O estimador duplamente robusto garante a consistência dos estimadores mesmo que persistam dificuldades de especificação inerentes a experimentos naturais distantes no tempo e com severas limitações na disponibilidade de dados,

⁷ E considerando também a sub-representação da variável de tratamento na única fonte de dados disponível: o Censo Imperial de 1872 foi realizado após a proibição do tráfico transatlântico, em 1850, e por isso ilustra o período de decadência do sistema escravagista que precedeu a Abolição (1888) (PAPADIA, 2019).

como a escravidão brasileira. Com isso, o estimador da correlação entre *sld* e *homr* pode ser considerado verdadeiro e não viesado.

Portanto, executado o método GPS-IPW e ponderadas as observações, o modelo linear ponderado que conta com todas as variáveis explicativas e fornece o estimador duplamente robusto do tratamento (ZHANG; ZHOU; CAO; ZHANG, 2016) é dado pela Equação 7:

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^n g(t, X_i; \widehat{\beta}^*). \quad (7)$$

Na Equação 7, $\widehat{\beta}^*$ representa os coeficientes estimados para o modelo de regressão ponderada $g(t, X_i; \widehat{\beta}^*)$. Como visto anteriormente, t representa os níveis de tratamento, e X_i representa as variáveis de controle que compõe o vetor \mathbf{X} . Os coeficientes $\widehat{\beta}^*$ representam a correlação entre as variáveis independentes e a variável resultante na amostra ponderada. A Seção 3.3 descreve os modelos lineares ponderados $g(t, X_i; \widehat{\beta}^*)$, assim como as variáveis e as fontes de dados utilizadas para estima-los. As estimações deste estudo se basearam em Heiss (2020) e Greifer (2021), utilizando a linguagem R para estimar as funções dose-resposta sobre as amostras ponderadas por GPS-IPW.

No Brasil, o estudo empírico de tratamentos contínuos ainda é relativamente recente. Oliveira, Menezes e Resende (2015) foram pioneiros ao avaliar efeitos heterogêneos do Fundo Constitucional de Financiamento do Centro-Oeste (FCO) em Goiás. Sua função dose-resposta relacionava o valor dos empréstimos (tratamento contínuo) aos níveis de emprego e salário no estado. Silva e Siqueira (2021) empregaram uma abordagem semelhante para analisar os efeitos do Fundo Constitucional do Norte (FNO) sobre o emprego formal.

Analisando dados entre 1992 e 2018, Pereira, Cireno, Goes, Calsavara e Veras (2022) estimaram o impacto da política de Seguro-Defeso, benefício pago ao pescador artesanal quando é proibido de exercer sua atividade durante períodos de reprodução e recomposição das espécies, sobre a frequência escolar das crianças das comunidades pesqueiras, assim como sobre o nível de emprego e de escolaridade entre adultos.

Todos estes estudos empregaram GPS para analisar tratamentos contínuos (valor de empréstimos concedidos ou tempo de recebimento de um benefício). Contudo, todos utilizaram métodos paramétricos para estimar os *scores*. O presente trabalho contribui com a literatura empírica ao (i) ponderar a amostra original pelo inverso dos GPS com o objetivo de eliminar

viés de seleção, formalizando o método GPS-IPW; (ii) realizar estimação não paramétrica dos GPSs via algoritmo GBM, método de aprendizado de máquina que comporta relações não lineares entre tratamento e controles sem pressupor distribuição normal (ZHU; COFFMAN; GHOSH, 2015); e (iii) utilizar estimadores duplamente robustos, que garantem a consistência dos estimadores mesmo em caso de má especificação dos modelos de resultado ou de erro ao calcular os GPSs (GRAHAM; MCCOY; STEPHENS, 2015; GREIFER, 2021).

Até onde se tem conhecimento, este é o primeiro estudo brasileiro a combinar GPS-IPW com *boosting* (GBM), e também o primeiro a aplicar essa metodologia para analisar legados históricos de longo prazo, especificamente a relação entre escravidão (1872) e violência letal contemporânea (2000) nos municípios brasileiros.

3.3 DADOS E MODELOS

Os trabalhos de Buonanno e Vargas (2019) sobre a Colômbia, de Gouda e Rigterink (2017) para os Estados Unidos e de Justus, Khan e Cerqueira (2017) para o estado de São Paulo foram as principais referências adotadas para definir as variáveis utilizadas neste estudo.

A intensidade da escravidão é representada por *sld*, e fornece o percentual de pessoas escravizadas na população total de determinado município em 1872. A variável de tratamento foi extraída da Divisão Territorial de 2010 (IBGE, 2010), que adequa o Censo Imperial de 1872 (e todos os Censos já realizados) à divisão municipal do Censo Demográfico de 2010 utilizando Áreas Mínimas Comparáveis (AMCs) (IBGE, 2012). É útil lembrar que o Censo Imperial de 1872 é o único que dispõe de dados sobre a população escravizada. A correspondência entre as divisões territoriais permite que variável de tratamento, variável resultante e variáveis de controle sejam referentes às mesmas áreas mesmo que se refiram a diferentes períodos.

A notória carência de dados sobre os níveis de violência vem sendo suprida por órgãos como o Fórum Brasileiro de Segurança Pública, criado em 2007 (FBSP, 2007). O único indicador de violência municipal cuja série histórica é suficientemente extensa e contínua é a taxa de homicídios, definida a seguir. Dada a extensão da série histórica da taxa de homicídios, o fator limitante que definiu o período analisado foi a disponibilidade de dados referentes às variáveis de controle. O ano 2000 foi o mais próximo de 1872 a contar com maior disponibilidade de dados sobre os controles empregados nos estudos de referência. Diversas mudanças na estrutura de coleta de dados dos Censos dificultam a elaboração de alguns indicadores, como o índice municipal de justiça e segurança pública (*cps*) para Censos anteriores a 2000.

A variável resultante *homr* indica a taxa municipal de homicídios em 2000, isto é, o número de vítimas de homicídio a cada cem mil habitantes em dado município neste ano. Consiste na divisão do número de vítimas de homicídios ocorridos no município pela população residente total, multiplicada por cem mil. Foi calculada utilizando os dados do Sistema de Informações sobre Mortalidade da Base de Dados do Sistema Único de Saúde (SIM-DATASUS), órgão do Ministério da Saúde (MS) (BRASIL, 2020). A população residente total provém do Censo Demográfico de 2000 (IBGE, 2003). Tradicionalmente, as mortes por homicídio são aquelas que o SIM-DATASUS atribui aos grandes grupos X85 a X90, Y35 e Y36 da décima Classificação Internacional de Doenças (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 1993). O primeiro intervalo se refere aos óbitos por Agressão, e os dois últimos grupos, às mortes por Intervenções Legais (isto é, cujo autor é um agente público).

A renda domiciliar *per capita* média municipal, representada por *ypc*, é um dos vários controles extraídos do Censo Demográfico de 2000 (IBGE, 2003). Os valores foram corrigidos em relação ao salário mínimo de 2010, último ano da série histórica disponível, quando seu valor era de R\$ 510,00. A correção foi realizada utilizando o Índice Nacional de Preços ao Consumidor (INPC), também calculado pelo IBGE.

A densidade populacional do município em 2000, representada por *pd*, também é tradicionalmente associada aos níveis de violência, com cidades mais povoadas tendo, em geral, maiores índices de criminalidade. Enquanto a população residente total foi extraída do Censo Demográfico de 2000, a área do município em quilômetros quadrados (km²) é fornecida pela Divisão Territorial de 2010, recurso que equalizou as áreas municipais analisadas (IBGE, 2003; 2010). Assim, a densidade populacional, medida em habitantes por quilômetro quadrado (hab./km²), é dada pela população total residente no município sobre sua área ocupada (km²) em 2010.

Também foram extraídos do Censo Demográfico de 2000: a taxa de desemprego municipal *unmp*, que representa a proporção da população residente economicamente ativa e com 16 ou mais anos de idade sem trabalho na semana de referência; *be*, o percentual de pessoas com 16 ou mais anos de idade que completaram oito anos de estudo ou mais, representa o quanto um município atendia à demanda por ensino básico, determinante para explicar os níveis de pobreza, desigualdade, violência e outros fatores relacionados à escravidão; por fim, *ym* representa o percentual de homens jovens (15 a 29 anos de idade) na população residente total, grupo tido como mais propenso a adotar comportamentos de risco que possam resultar em vitimização (BUONANNO; VARGAS, 2019).

Para controlar a atuação do Estado na segurança dos municípios, Buonanno e Vargas (2019) utilizam o índice de eficiência da justiça de Fergusson, Vargas e Vela (2013), baseado na quantidade de ocorrências criminais solucionadas em relação ao total de ocorrências em cada município colombiano. Não foram encontrados dados análogos para o Brasil, então este trabalho introduz o índice municipal de justiça e segurança pública, *cps*, detalhado a seguir.

Realizada pelo IBGE, a Pesquisa de Informações Básicas Municipais (MUNIC) (que até 2015 era chamada de Perfil dos Municípios Brasileiros) traz dados municipais pormenorizados sobre diversos temas. A edição de 2001 (MUNIC-2001) foi especialmente dedicada à gestão pública (IBGE, 2003). No Brasil, ainda que justiça e segurança sejam competências estaduais e federais, os municípios possuem órgãos que podem contribuir significativamente para a redução da criminalidade. Além disso, como mudanças institucionais municipais são relativamente lentas, os dados dessa pesquisa são considerados representativos do ano de 2000.

Com base na MUNIC-2001, verifica-se a existência de Guarda Municipal, Delegacia da Mulher, Defesa Civil, Tribunal de Pequenas Causas e Conselho Tutelar em cada município. Para cada um destes órgãos que estiver estabelecido no município, atribui-se o valor um; para cada órgão que não estiver, atribui-se zero. O índice municipal de justiça e segurança pública *cps* é a soma dos valores atribuídos ao município de acordo com a presença destes órgãos, e, portanto, assume valores discretos entre um e cinco.

Note-se que as estimações podem associar maiores valores de *cps* a municípios mais violentos ou mais pacíficos. Isto é, o coeficiente de *cps* pode ser negativo ou positivo. Ambos resultados são factíveis e suscitam indagações sobre possíveis relações de causalidade: caso seja negativo, é um indício de a presença de maior aparato de segurança de fato coíbe a violência letal? Ou, caso seja positivo, é um indício de que municípios mais violentos acabam investindo mais em infraestrutura de segurança? A literatura empírica de Economia do Crime no Brasil preconiza que o primeiro cenário é o mais provável (JUSTUS; KHAN; CERQUEIRA, 2017)

Assim como Gouda e Rigterink (2017), foram incluídas latitude e longitude municipais, disponíveis na Divisão Territorial de 2010 (IBGE, 2010), como controles geográficos. Para obter modelos de efeitos fixos, foram incluídas variáveis binárias referentes à Unidade Federativa (UF) onde cada município está localizado. Naturalmente, uma das vinte e seis UFs brasileiras é excluída das especificações para evitar colinearidade perfeita.

É necessário destacar que os efeitos fixos captam fatores de difícil mensuração que atuam sobre os municípios de um mesmo estado, ou seja, municípios com história e cultura

relativamente comuns. A literatura sociológica e empírica sobre a persistência dos impactos da escravidão apoia a hipótese de transbordamento – seja da violência inerente às relações laborais e sociais (FRANCO, 1983; ROSENTHAL, 2018), das instituições predatórias (ACEMOGLU; JOHNSON; ROBINSON, 2002; NUNN, 2008; ACEMOGLU; GARCÍA-JIMENO; ROBINSON, 2012), do racismo institucional e cultural (ACHARYA; BLACKWELL; SEN, 2016; LEUNG-GAGNÉ, 2023) e da segregação espacial dos descendentes dos libertos (COOK, 2014; COOK; LOGAN; PARMAN, 2018). Sendo assim, controlar a autocorrelação espacial é fundamental para captar diferenças entre as UFs, tornando as unidades amostrais diretamente comparáveis entre si. Assim, o estimador de *sld* representará a média da correlação entre a intensidade da escravidão em 1872 e as taxas de homicídio em 2000 para todos os municípios observados, independente de terem sido diretamente afetados pela escravidão ou não.

Até aqui, foram descritas a variável resultante, a variável de interesse e os controles comuns a todos os modelos especificados. Devido à sua relevância, o controle da desigualdade de renda merece atenção especial. A literatura empírica de economia do desenvolvimento considera a desigualdade de renda como herança do sistema escravagista (BERTOCCHI; DIMICO, 2012; 2014; FUJIWARA; LAUDARES; VALENCIA-CAICEDO, 2017) e, ao mesmo tempo, como determinante dos níveis de violência (SOARES, 2004; SOARES; NARITOMI, 2010; GOUDA; RIGTERINK, 2017; BUONANNO; VARGAS, 2019). Dois índices de desigualdade de renda foram incluídos como controles em diferentes modelos, colaborando para reduzir significativamente o confundimento e para testar a robustez dos estimadores obtidos.

Utilizado com mais frequência, o índice de Gini foi extraído do Censo Demográfico de 2000 (IBGE, 2003). A variável *gini* assume valores entre zero e um, sendo que o primeiro corresponde à perfeita igualdade, e o último, à perfeita desigualdade (isto é, apenas um habitante do município possuiria toda a renda, caracterizando um município perfeitamente desigual). Para assegurar a robustez das conclusões obtidas ao incluir a desigualdade, o Índice de Theil (BUONANNO; VARGAS, 2019), disponível no Atlas de Desenvolvimento Humano do Programa das Nações Unidas para o Desenvolvimento (PNUD), substitui o índice de Gini em alguns modelos (PNUD, 2010). A variável *theil* assume o valor zero em caso de igualdade perfeita, e tende ao infinito em caso de desigualdade perfeita. A Figura 4 faz um resumo dos dados utilizados nas estimações.

Figura 4 - Descrição resumida das variáveis

<i>homr</i>	Taxa de homicídios municipal (vítimas/100 mil hab.) Sistema de Informações sobre Mortalidade (SIM-DATASUS) (BRASIL, 2020) Censo Demográfico de 2000 (IBGE, 2003)
<i>sld</i>	Intensidade da escravidão (%) Parcela de pessoas escravizadas na população total do município em 1872. Divisão Territorial de 2010 e Censo Imperial de 1872 (IBGE, 2010; 2012)
<i>ypc</i>	Renda domiciliar <i>per capita</i> média (R\$) Valores corrigidos pelo Índice Nacional de Preços ao Consumidor (INPC) com base no salário mínimo de 2010 (R\$ 510,00). Censo Demográfico de 2000 (IBGE, 2003)
<i>pd</i>	Densidade populacional (habitantes/km ²) Censo Demográfico de 2000 e Divisão Territorial de 2010 (IBGE,2003; 2010)
<i>unmp</i>	Taxa de desemprego (%) Parcela da população economicamente ativa que reside no município, tem 16 ou mais anos de idade e se encontrava sem trabalho na semana de referência. Censo Demográfico de 2000 (IBGE, 2003)
<i>be</i>	Educação básica (%) Parcela da população que reside no município, tem 16 ou mais anos de idade e completou oito (ou mais) anos de estudo. Censo Demográfico de 2000 (IBGE, 2003)
<i>ym</i>	Homens jovens (%) Parcela de homens com 15 a 29 anos de idade na população total residente no município. Censo Demográfico de 2000 (IBGE, 2003)
<i>cps</i>	Índice municipal de justiça e segurança pública (escala de 0 a 5) Presença de Guarda Municipal, Delegacia da Mulher, Defesa Civil, Tribunal de Pequenas Causas e Conselho Tutelar no município. Pesquisa de Informações Básicas Municipais/Perfil dos Municípios Brasileiros (MUNIC-2001) (IBGE, 2001)
<i>long</i>	Longitude (Graus, minutos e segundos a Leste-Oeste) Divisão Territorial de 2010 (IBGE, 2010)
<i>lat</i>	Latitude (Graus, minutos e segundos a Norte-Sul) Divisão Territorial de 2010 (IBGE, 2010)
<i>gini</i>	Índice de desigualdade de renda de Gini (intervalo de 0 a 1) Censo Demográfico de 2000 (IBGE, 2003)
<i>theil</i>	Índice de desigualdade de Theil (maior ou igual a zero) Atlas do Desenvolvimento Humano no Brasil (PNUD, 2010)

Fonte: elaboração própria.

Com estas informações, é possível especificar os modelos lineares ponderados da Equação 7. A Equação 8 representa o Modelo 1, que consiste na especificação mais simples, sem os índices de desigualdade de renda. Para dado município i , tem-se:

$$\begin{aligned}
 homr_{i,2000} = & \beta_0 + \beta_1 sld_{i,1872} + \beta_2 ypc_{i,2000} + \beta_3 pd_{i,2000} + \beta_4 unmp_{i,2000} \\
 & + \beta_5 be_{i,2000} + \beta_6 ym_{i,2000} + \beta_7 cps_{i,2000} + \beta_8 long_{i,2010} + \beta_9 lat_{i,2010} \\
 & + \sum_{j=1}^{25} UF_{ij} + \varepsilon_{i,2000}.
 \end{aligned} \tag{8}$$

Para representar a desigualdade de renda, o Modelo 2 utiliza o índice de Gini, e o Modelo 3, o índice de Theil. Os Modelos 2 e 3 são representados, respectivamente, pelas Equações 9 e 10.

$$\begin{aligned}
 homr_{i,2000} = & \beta_0 + \beta_1 sld_{i,1872} + \beta_2 ypc_{i,2000} + \beta_3 pd_{i,2000} + \beta_4 unmp_{i,2000} \\
 & + \beta_5 be_{i,2000} + \beta_6 ym_{i,2000} + \beta_7 cps_{i,2000} + \beta_8 long_{i,2010} + \beta_9 lat_{i,2010} \\
 & + \beta_{10} gini_{i,2000} + \sum_{j=1}^{25} UF_{ij} + \varepsilon_{i,2000}
 \end{aligned} \tag{9}$$

$$\begin{aligned}
 homr_{i,2000} = & \beta_0 + \beta_1 sld_{i,1872} + \beta_2 ypc_{i,2000} + \beta_3 pd_{i,2000} + \beta_4 unmp_{i,2000} \\
 & + \beta_5 be_{i,2000} + \beta_6 ym_{i,2000} + \beta_7 cps_{i,2000} + \beta_8 long_{i,2010} + \beta_9 lat_{i,2010} \\
 & + \beta_{10} theil_{i,2000} + \sum_{j=1}^{25} UF_{ij} + \varepsilon_{i,2000}
 \end{aligned} \tag{10}$$

A inclusão de termos de interação entre a renda *per capita* domiciliar média *ypc* e os índices de desigualdade de renda *gini* e *theil* visa reconhecer que as relações entre renda e desigualdade podem assumir diferentes formas em diferentes áreas observadas; essas diferenças, por sua vez, podem estar significativamente associadas aos níveis de violência municipais. Por exemplo: enquanto algumas regiões podem apresentar altos níveis de renda domiciliar, de desigualdade e de violência (como é comum nas grandes metrópoles), outras podem apresentar baixas taxas de homicídio mesmo diante de altos níveis de desigualdade, caso estes sejam motivados por mudanças econômicas que levem ao aumento da renda média ao longo do tempo, associado à redução dos incentivos econômicos à violência.

O pseudo R-quadrado ajustado é utilizado para avaliar a qualidade da especificação dos modelos lineares ponderados (NAKAGAWA; SCHIELZETH, 2013). Se aumentar após a inclusão das interações (sem que outras variáveis passem a apresentar comportamentos anômalos), tem-se uma evidência de que a modelagem da relação entre renda e desigualdade contribui para a obtenção de resultados robustos ao analisar as taxas de homicídio municipais. A análise do papel dos termos de interação será aprofundada na seção seguinte.

Sendo assim, os Modelos 4 e 5 são representados pelas Equações 11 e 12, respectivamente:

$$\begin{aligned}
 homr_{i,2000} = & \beta_0 + \beta_1 sld_{i,1872} + \beta_2 ypc_{i,2000} + \beta_3 pd_{i,2000} + \beta_4 unmp_{i,2000} \\
 & + \beta_5 be_{i,2000} + \beta_6 ym_{i,2000} + \beta_7 cps_{i,2000} + \beta_8 long_{i,2010} + \beta_9 lat_{i,2010} \\
 & + \beta_{10} gini_{i,2000} + \beta_{11}(gini_{i,2000} \times ypc_{i,2000}) + \sum_{j=1}^{25} UF_{ij} + \varepsilon_{i,2000}
 \end{aligned} \tag{11}$$

$$\begin{aligned}
 homr_{i,2000} = & \beta_0 + \beta_1 sld_{i,1872} + \beta_2 ypc_{i,2000} + \beta_3 pd_{i,2000} + \beta_4 unmp_{i,2000} \\
 & + \beta_5 be_{i,2000} + \beta_6 ym_{i,2000} + \beta_7 cps_{i,2000} + \beta_8 long_{i,2010} + \beta_9 lat_{i,2010} \\
 & + \beta_{10} theil_{i,2000} + \beta_{11}(theil_{i,2000} \times ypc_{i,2000}) + \sum_{j=1}^{25} UF_{ij} + \varepsilon_{i,2000}
 \end{aligned} \tag{12}$$

Os Modelos 1 a 5 foram estimados para as regiões Sudeste, Nordeste, Sudeste e Nordeste conjuntas, e para todo o território nacional. Os resultados das estimações são exibidos e analisados na seção seguinte.

4 RESULTADOS E DISCUSSÃO

4.1 DIAGNÓSTICO DE EQUILÍBRIO DAS AMOSTRAS PONDERADAS

A Seção 3.1 demonstrou a necessidade de abordar o confundimento entre intensidade da escravidão e valores observados dos controles, assim como o caráter contínuo da variável de tratamento, ao analisar relação entre escravidão em 1872 e taxas de homicídios municipais em 2000. Para fazê-lo, a Seção 3.2 apresentou o método GPS-IPW. Zhu, Coffman e Ghosh (2015) observaram que, se a correlação média entre tratamento e controles na amostra ponderada for suficientemente baixa, então o método foi bem sucedido. Isto é: estudos que visem obter estimadores causais do efeito do tratamento podem interpretá-los como tal. Do contrário, o estimador do tratamento é uma correlação robusta (ajustada para o confundimento) entre a variável de tratamento e a variável resultante

Este estudo tem o objetivo de encontrar uma evidência robusta, significativa e positiva da correlação entre a intensidade da escravidão em 1872 e as taxas de homicídios municipais em 2000, pelos motivos expostos na Seção 1. Entretanto, a análise da eficácia do método GPS-IPW e dos pesos calculados com o auxílio de aprendizado de máquina (Seção 2.2) ajuda a entender as fontes de confundimento dos modelos, e assim, a encontrar alternativa que permitam, no futuro, a obtenção de estimadores causais do efeito da escravidão sobre a violência. Por isso, esta seção analisa a correlação entre cada variável de controle e a variável de tratamento nas amostras ponderadas de cada modelo estimado. Para tratamentos contínuos, o índice de correlação de Pearson é capaz de representar adequadamente a eficácia dos pesos e o nível de confundimento na amostra ponderada (ZHU; COFFMAN; GHOSH, 2015).

As Figuras 5 a 8 representam os índices de correlação de Pearson entre *sld* e os controles dos Modelos 1 a 5 para cada uma das áreas analisadas neste estudo. Na primeira coluna, estão os controles X_i . As células mostram o valor do índice de Pearson para a correlação entre *sld* e o controle na linha correspondente.

Os limites definidos por Zhu, Coffman e Ghosh (2015) para a média da correlação entre as variáveis foram aplicados como referência heurística para avaliar a correlação de cada controle com o tratamento após a ponderação. Para maior interpretabilidade, utiliza-se a legenda de cores descrita a seguir. Em verde, estão as correlações cujo valor absoluto ficou entre 0 e 0,10, mais próximas da independência condicional entre tratamento e o dado controle na amostra ponderada. Valores absolutos entre 0,10 e 0,30 estão em amarelo, faixa em que há

confundimento médio. Valores absolutos acima de 0,30, estão em vermelho, pois mesmo na amostra ponderada, a correlação entre *sld* e o controle correspondente é alta.

É preciso destacar que, mesmo haja confundimento alto para alguns controles, isso não compromete os objetivos deste estudo: obter uma evidência positiva, significativa e robusta da relação entre intensidade da escravidão em 1872 e as taxas de homicídio municipais em 2000. A Seção 3.2 mostrou que, para experimentos naturais com amostras desequilibradas e tratamentos contínuos, o método GPS-IPW gera estimadores mais robustos que as estimações lineares. A robustez do estimador será testada através do aumento progressivo da amostra de municípios observados e da inclusão de controles de desigualdade de renda (e de efeitos fixos, à medida que o número de UFs analisadas aumenta até abranger todo o território nacional).

4.1.1 Sudeste

Figura 5 - Correlação de Pearson na pseudoamostra: Sudeste

X_i	(1)	(2)	(3)	(4)	(5)
<i>ypc</i>	0,259	0,259	0,259	0,257	0,256
<i>pd</i>	0,027	0,028	0,027	0,027	0,026
<i>unmp</i>	0,133	0,132	0,132	0,133	0,133
<i>be</i>	0,276	0,277	0,276	0,275	0,275
<i>ym</i>	-0,046	-0,048	-0,048	-0,048	-0,049
<i>cps</i>	0,338	0,341	0,340	0,338	0,338
<i>long</i>	0,110	0,109	0,110	0,111	0,111
<i>lat</i>	-0,182	-0,181	-0,182	-0,182	-0,183
<i>gini</i>		0,119		0,120	
<i>theil</i>			0,130		0,131
<i>uf_ES</i>	0,044	0,043	0,042	0,042	0,042
<i>uf_MG</i>	-0,140	-0,139	-0,139	-0,140	-0,140
<i>uf_RJ</i>	0,293	0,295	0,296	0,298	0,300
<i>uf_SP</i>	-0,012	-0,013	-0,013	-0,014	-0,014
<i>n</i>	1.665	1.665	1.665	1.665	1.665

X_i	(1)	(2)	(3)	(4)	(5)
n -Adj	1.661	1.662	1.662	1.661	1.661

Fonte: elaboração própria com o apoio do software *RStudio*

A Figura 5 mostra que, para a região Sudeste, a ponderação das amostras gerou quatro controles suficientemente independentes de *sld*: a densidade populacional, o percentual de homens jovens e as variáveis binárias de efeitos fixos referentes aos estados de São Paulo e Espírito Santo. Há confundimento médio em relação a *sld* e os principais controles socioeconômicos: renda média domiciliar *per capita*, desemprego, educação básica e ambos os índices de desigualdade de renda. Controles geográficos também apresentaram correlação média com *sld* na amostra ponderada: foi o caso da latitude, da longitude e das variáveis binárias que representam municípios do Rio de Janeiro e de Minas Gerais. Em todos os modelos, o índice municipal de justiça e segurança pública apresentou alta correlação com *sld*, superior a 0,30. No Modelo 5, os efeitos fixos referentes ao Rio de Janeiro apresentaram alta correlação com *sld*.

A comparação entre o número de unidades amostrais originais n e o número de amostras ponderadas n -Adj mostra que o algoritmo GBM de fato consegue preservar o tamanho da amostra original. A predominância de amarelo na Figura 5 indica que as estimações referentes à região Sudeste apresentam, em geral, confundimento médio. Para reduzi-lo, controlar a qualidade institucional de toda a amostra e a autocorrelação espacial nos municípios cariocas de forma mais eficiente é prioridade. Em segundo lugar, é necessário controlar os fatores socioeconômicos de forma mais eficiente, testando outras formas funcionais e outros indicadores das mesmas grandezas especificadas e comuns à literatura empírica.

4.1.2 Nordeste

Figura 6 - Correlação de Pearson na pseudoamostra: Nordeste

X_i	(1)	(2)	(3)	(4)	(5)
<i>ypc</i>	0,131	0,131	0,131	0,127	0,120
<i>pd</i>	0,064	0,064	0,064	0,062	0,059
<i>unmp</i>	0,111	0,111	0,110	0,107	0,102
<i>be</i>	0,176	0,176	0,176	0,171	0,164

X_i	(1)	(2)	(3)	(4)	(5)
<i>ym</i>	0,075	0,075	0,075	0,074	0,073
<i>cps</i>	0,126	0,126	0,126	0,123	0,119
<i>long</i>	-0,064	-0,064	-0,065	-0,071	-0,080
<i>lat</i>	0,000	0,000	0,000	0,003	0,008
<i>gini</i>		0,092		0,092	
<i>theil</i>			0,100		0,096
<i>uf_AL</i>	0,004	0,004	0,004	0,003	0,001
<i>uf_BA</i>	0,025	0,025	0,025	0,025	0,023
<i>uf_CE</i>	-0,047	-0,047	-0,047	-0,048	-0,051
<i>uf_MA</i>	0,173	0,173	0,174	0,183	0,198
<i>uf_PB</i>	-0,094	-0,094	-0,094	-0,096	-0,099
<i>uf_PE</i>	0,021	0,021	0,021	0,019	0,017
<i>uf_PI</i>	-0,068	-0,068	-0,068	-0,069	-0,072
<i>uf_RN</i>	-0,073	-0,073	-0,073	-0,074	-0,077
<i>uf_SE</i>	0,070	0,070	0,070	0,070	0,069
<i>n</i>	1.784	1.784	1.784	1.784	1.784
<i>n-Adj</i>	1.764	1.764	1.764	1.760	1.753

Fonte: elaboração própria com o apoio do software *RStudio*

Ao contrário do Sudeste, a região Nordeste não apresentou nenhuma correlação de Pearson acima de 0,30 na amostra ponderada. O método GPS-IPW foi mais relativamente eficaz na amostra ponderada de municípios nordestinos. A maioria dos controles apresenta baixos níveis de confundimento com *sld*. Apenas um controle relacionado ao espaço apresentou confundimento médio: a variável binária que representa municípios do Maranhão.

Entre os controles socioeconômicos, apenas densidade populacional, percentual de homens jovens no município e índice de desigualdade de Gini apresentaram baixos níveis de confundimento com *sld* após a ponderação. Logo, o padrão de confundimento médio entre os controles socioeconômicos também se observa no Nordeste. O baixo confundimento *sld* e *gini*

representa um avanço significativo, pois a desigualdade de renda foi amplamente relacionada à escravidão em diversos estudos quantitativos (Seções 1 e 2).

Em relação ao tamanho da amostra ponderada, a diferença entre n e n -Adj foi um pouco maior no Nordeste, que perdeu cerca de vinte unidades amostrais, em média, após realizar a ponderação. Tal significa que o algoritmo GBM não conseguiu atribuir uma probabilidade desses municípios pertencerem ao grupo de tratamento com base nos valores dos controles disponíveis. A ampliação do tamanho da amostra ou a inclusão de outros controles podem ajudar a reduzir a perda. A primeira alternativa funcionará caso os novos controles consigam abordar características desses municípios que não puderam ser representadas pelos controles atuais.

Dado o menor nível de confundimento médio dos modelos baseados na amostra ponderada do Nordeste, os estimadores de sld dos modelos representados na Figura 6 tendem a ser mais robustos que os obtidos pelos modelos da Figura 5 referentes à região Sudeste.

4.1.3 Nordeste e Sudeste

Figura 7 - Correlação de Pearson na pseudoamostra: Nordeste e Sudeste

X_i	(1)	(2)	(3)	(4)	(5)
<i>ypc</i>	0,169	0,170	0,169	0,181	0,165
<i>pd</i>	0,031	0,031	0,032	0,033	0,028
<i>unmp</i>	0,130	0,129	0,130	0,130	0,127
<i>be</i>	0,206	0,206	0,206	0,214	0,201
<i>ym</i>	0,010	0,009	0,010	0,004	0,011
<i>cps</i>	0,236	0,237	0,236	0,251	0,233
<i>long</i>	0,004	0,004	0,004	0,001	0,004
<i>lat</i>	-0,062	-0,062	-0,063	-0,072	-0,062
<i>gini</i>		0,101		0,095	
<i>theil</i>			0,125		0,124
<i>uf_AL</i>	-0,002	-0,002	-0,002	-0,002	-0,002
<i>uf_BA</i>	0,014	0,014	0,014	0,008	0,016
<i>uf_CE</i>	-0,037	-0,037	-0,037	-0,037	-0,038
<i>uf_ES</i>	0,043	0,043	0,043	0,041	0,036
<i>uf_MA</i>	0,086	0,085	0,086	0,076	0,091

X_i	(1)	(2)	(3)	(4)	(5)
<i>uf_MG</i>	-0,060	-0,060	-0,060	-0,060	-0,060
<i>uf_PB</i>	-0,064	-0,064	-0,064	-0,064	-0,065
<i>uf_PE</i>	0,000	0,000	0,000	-0,001	-0,001
<i>uf_PI</i>	-0,049	-0,049	-0,049	-0,050	-0,051
<i>uf_RJ</i>	0,216	0,216	0,218	0,237	0,216
<i>uf_RN</i>	-0,050	-0,050	-0,050	-0,050	-0,051
<i>uf_SE</i>	0,036	0,036	0,036	0,034	0,035
<i>uf_SP</i>	0,003	0,004	0,003	0,008	0,005
<i>n</i>	3.449	3.449	3.449	3.449	3.449
<i>n-Adj</i>	3.426	3.427	3.425	3.432	3.419

Fonte: elaboração própria com o apoio do software *RStudio*

A análise da Figura 7 mostra que, quando municípios das regiões Sudeste e Nordeste são analisados juntos, o padrão de correlação da amostra ponderada remete mais ao da região Nordeste. Nenhum valor absoluto ultrapassou 0,30. A inclusão dos municípios do Sudeste, cujas correlações tenderam a ser mais altas quando analisadas isoladamente, não prejudicou ou alterou significativamente o padrão observado individualmente na região Nordeste.

Os controles geográficos e de efeitos fixos apresentaram baixa correlação com *sld* – com exceção dos referentes ao Rio de Janeiro. Assim como na amostra ponderada do Nordeste, o índice municipal de justiça e segurança pública permaneceu com confundimento médio – apesar de ter apresentado confundimento alto com *sld* na amostra do Sudeste. A comparação entre o tamanho das amostras original e ponderada mostra a perda mais alta até agora: em média, 30 unidades amostrais por modelo. Isto é, o algoritmo GBM não conseguiu calcular a probabilidade de cerca de 30 municípios pertencerem ao grupo de tratamento.

Analisando os resultados das Figuras 5 a 7, nota-se que os controles socioeconômicos apresentam, majoritariamente, confundimento médio com a variável de tratamento. É notável como a densidade populacional e o percentual de homens jovens na população apresentaram independência condicional de *sld* para todas as áreas analisadas até aqui. Por outro lado, os índices de desigualdade de renda passaram a apresentar confundimento médio. Apenas no Modelo 4, *gini* apresentou baixo confundimento com *sld*. Na análise individual do Nordeste, observou-se maior independência condicional entre *sld* e desigualdade de renda, de modo que

o confundimento médio apresentado na Figura 7 pode ser majoritariamente atribuído à inclusão dos municípios da região Sudeste.

A análise conjunta das duas regiões representou um teste de robustez para os resultados da ponderação referentes ao Nordeste, uma vez que a inclusão do Sudeste não alterou significativamente os padrões observados na análise individual. Em geral, a análise das Figuras 5 a 7 permite concluir que, para as regiões que apresentavam maior população absoluta de trabalhadores escravizados em 1872, é necessário abordar outros aspectos socioeconômicos além dos analisados neste estudo, de modo que a correlação entre a escravidão e esses controles diminua significativamente na amostra ponderada, o que permitirá estimações causais do efeito do tratamento no futuro.

Considerando as limitações dos dados históricos e o tempo decorrido desde a Abolição (1888), é possível considerar que o algoritmo GBM foi eficiente em abordar a maior parte do confundimento entre tratamento e controles nas regiões com maior população escravizada absoluta em 1872. Em geral, as amostras ponderadas destas regiões tendem a apresentar confundimento baixo a médio. Os desafios metodológicos foram abordados com sucesso pelo método GPS-IPW aliado ao algoritmo GBM. Os objetivos deste trabalho podem ser atingidos através deste método de abordagem das peculiaridades da amostra e do experimento natural analisado. Estudos que busquem obter estimadores causais podem aplica-lo com sucesso especialmente para tratamentos contínuos com maior disponibilidade de dados. Entretanto, estudos que busquem definir o efeito causal da escravidão sobre a violência contemporânea devem priorizar a especificação de controles socioeconômicos cujo confundimento com a escravidão é baixo para serem bem sucedidos.

4.1.4 Brasil

Figura 8 - Correlação de Pearson na pseudoamostra: Brasil

X_i	(1)	(2)	(3)	(4)	(5)
<i>ypc</i>	0,112	0,114	0,114	0,111	0,111
<i>pd</i>	0,051	0,051	0,051	0,049	0,049
<i>unmp</i>	0,149	0,148	0,148	0,148	0,147
<i>be</i>	0,165	0,166	0,166	0,164	0,163
<i>ym</i>	0,026	0,026	0,026	0,025	0,024

X_i	(1)	(2)	(3)	(4)	(5)
<i>cps</i>	0,213	0,216	0,215	0,216	0,214
<i>long</i>	0,109	0,109	0,109	0,111	0,110
<i>lat</i>	0,001	0,001	0,001	0,000	0,000
<i>gini</i>		0,080		0,080	
<i>theil</i>			0,088		0,086
<i>uf_AC</i>	-0,018	-0,018	-0,018	-0,018	-0,018
<i>uf_AL</i>	0,013	0,013	0,013	0,013	0,013
<i>uf_AM</i>	-0,027	-0,027	-0,027	-0,027	-0,027
<i>uf_AP</i>	-0,003	-0,003	-0,003	-0,004	-0,003
<i>uf_BA</i>	0,037	0,036	0,037	0,037	0,037
<i>uf_CE</i>	-0,013	-0,013	-0,013	-0,013	-0,013
<i>uf_DF</i>	-0,004	-0,004	-0,004	-0,004	-0,004
<i>uf_ES</i>	0,052	0,052	0,052	0,049	0,044
<i>uf_GO</i>	-0,038	-0,038	-0,038	-0,038	-0,038
<i>uf_MA</i>	0,087	0,086	0,086	0,088	0,088
<i>uf_MG</i>	-0,007	-0,007	-0,007	-0,007	-0,007
<i>uf_MS</i>	-0,028	-0,028	-0,028	-0,028	-0,028
<i>uf_MT</i>	-0,026	-0,026	-0,026	-0,027	-0,027
<i>uf_PA</i>	0,000	0,000	0,000	0,000	0,000
<i>uf_PB</i>	-0,035	-0,035	-0,035	-0,036	-0,036
<i>uf_PE</i>	0,021	0,021	0,021	0,020	0,021
<i>uf_PI</i>	-0,023	-0,022	-0,022	-0,023	-0,023
<i>uf_PR</i>	-0,065	-0,064	-0,064	-0,065	-0,065
<i>uf_RJ</i>	0,233	0,232	0,232	0,246	0,245
<i>uf_RN</i>	-0,026	-0,026	-0,026	-0,027	-0,027
<i>uf_RO</i>	-0,028	-0,028	-0,028	-0,028	-0,028
<i>uf_RR</i>	-0,015	-0,015	-0,015	-0,015	-0,015
<i>uf_RS</i>	-0,030	-0,030	-0,030	-0,031	-0,031
<i>uf_SC</i>	-0,053	-0,053	-0,053	-0,054	-0,053

X_i	(1)	(2)	(3)	(4)	(5)
uf_SE	0,045	0,045	0,045	0,045	0,045
uf_SP	0,044	0,045	0,045	0,044	0,045
uf_TO	-0,046	-0,046	-0,046	-0,046	-0,046
n	5.498	5.498	5.498	5.498	5.498
n -Adj	5.488	5.489	5.489	5.487	5.486

Fonte: elaboração própria com o apoio do software *RStudio*

A Figura 8 mostra os resultados do método GPS-IPW para todos os municípios da amostra. A princípio, é possível observar que a perda de unidades amostrais em relação à amostra original é relativamente muito baixa (em média, 9 unidades). A ampliação radical do grupo controlado facilitou a identificação de características dos grupos por parte do algoritmo GBM e aumentou a oferta de municípios controlados potencialmente comparáveis aos tratados (dados os valores dos controles). Estes municípios tendem a ter pesos maiores, pois a probabilidade de se diferenciarem dos tratados única e exclusivamente pela ação do tratamento nos últimos é maior.

O aumento da representatividade dos pesos através da ampliação dos valores observados dos controles foi fundamental para garantir a manutenção do tamanho da amostra e os baixos níveis de confundimento predominantes (em verde) na Figura 8. Observa-se que a correlação de Pearson, em valor absoluto, foi menor que 0,10 para a maioria dos controles. Não houve correlações acima de 0,30. O padrão de baixo confundimento entre *sld* e controles geográficos e espaciais observado anteriormente persistiu na amostra ponderada de todos os municípios brasileiros. O mesmo vale para a densidade populacional e o percentual de homens jovens no município.

É notável que ambos os índices de desigualdade de renda apresentaram correlações abaixo de 0,10 em todas as especificações nas quais foram incluídos: tal sugere que a correlação entre escravidão e violência foi adequadamente separada da correlação entre desigualdade de renda e violência nos modelos referentes a todo o território nacional.

O padrão de confundimento médio entre controles socioeconômicos e *sld* observado anteriormente persistiu. As Figuras 5 a 8 mostram que a correlação residual entre tratamento e controles não se deveu a limitações metodológicas ou mesmo às limitações de representatividade dos dados do tratamento, mas à alta correlação entre a intensidade da

escravidão e diversos aspectos socioeconômicos dos municípios brasileiros observados em 2000. Para reduzi-la ou eliminá-la, a principal estratégia deve ser empregar covariantes que representem esses aspectos socioeconômicos, mas apresentem maior exogeneidade em relação a escravidão – o que não é fácil, dado o papel desta na acumulação primitiva de capital e na distribuição das atividades econômicas ao longo dos municípios brasileiros.

A Figura 8 apresenta dois controles geográficos e de efeitos fixos que permaneceram apresentando confundimento médio: a longitude e a variável binária que representa municípios do Rio de Janeiro. Esta última apresentou índices de Pearson superiores a 0,23 para todas as especificações. A correlação entre *sld* e *long* foi de cerca de 0,10, em média. Como a longitude assume valores negativos para o hemisfério ocidental, menor a longitude, mais longe do litoral o município – logo, menor a probabilidade de ter *sld* não nula. Maior a longitude, maior a proximidade do município com o litoral priorizado pela metrópole portuguesa durante a colonização, onde a escravidão se instalou com mais frequência. Daí a relação proporcional entre *sld* e *long* – a qual não se verificou ao analisar o Nordeste nem Nordeste e Sudeste conjuntos. A ampliação do grupo controlado ressaltou o papel histórico da proximidade litorânea na análise dos efeitos da escravidão.

É possível analisar as mudanças de correlação de Pearson de cada controle à medida que a amostra se amplia e que os Modelos passam do 1 ao 5, mas isto foge ao principal objetivo desta seção. As Figuras 5 a 8 permitem concluir que: o método GPS-IPW conseguiu gerar amostras ponderadas nas quais o confundimento entre tratamento e controles é médio a baixo. Médio entre os controles socioeconômicos e baixo entre os controles geográficos e de efeitos fixos (numerosos da Figura 6 em diante). Sendo assim, os estimadores de *sld* analisados na Seção 4.2 tendem a ser correlações robustas entre *sld* e *homr* especialmente nos modelos baseados na amostra ponderada de todo o território nacional

Regressões múltiplas comuns como as usadas por Gouda e Rigterink (2017) não poderiam gerar estimadores correlacionais adequados devido ao caráter não-linear da distribuição de *sld* em relação aos controles e à variável resultante. Os estimadores analisados na Seção 4.2 se baseiam nas amostras ponderadas cujo confundimento foi analisado nesta seção. É seguro afirmar que esses estimadores representam adequadamente a forma que a relação entre escravidão e violência letal assume em 2000 para diferentes áreas do território brasileiro. Os estimadores referentes à amostra ponderada que reúne todas as regiões geográficas tendem a ser os mais robustos.

Por fim, esta seção demonstrou que o método GPS-IPW foi capaz de abordar dificuldades metodológicas severas geradas por viés de seleção em experimentos naturais e variáveis de tratamento contínuas. Sua aplicação em outros estudos, com maior disponibilidade de dados, tende a gerar estimadores causais representativos, e este trabalho busca contribuir para popularizá-lo.

4.2 RESULTADOS DAS REGRESSÕES PONDERADAS

Nesta seção, opta-se por omitir as variáveis binárias de efeitos fixos por UF dos resultados dos modelos. A análise dos efeitos fixos tem papel secundário em relação à análise das outras variáveis independentes, cujas relações com a violência letal podem ser explicadas com mais clareza. O Apêndice A traz os resultados completos dos modelos de todas as áreas observadas, incluindo as variáveis binárias das UFs.

Como visto nas Seções 1, 3 e 4.1, os estimadores de *sld* gerados pelos modelos analisados a seguir são correlações entre *sld* e *homr*: se forem positivos, significantes e robustos (isto é, se permanecerem positivos e significantes mesmo após a inclusão de outros controles), conclui-se que representam adequadamente a relação entre escravidão em 1872 e violência letal em 2000 para diferentes áreas do território brasileiro. Neste caso, o objetivo deste trabalho terá sido atingido.

Outro objetivo da análise realizada nesta seção é confirmar a hipótese de que o método GPS-IPW foi eficiente para gerar estimadores correlacionais cujo comportamento está de acordo com a literatura empírica. Por exemplo: se o estimador da densidade populacional *pd* assumir um valor negativo (ou seja, representa uma associação inversa entre *pd* e *homr*) e for significativo, este é um indício de que os pesos gerados pelo algoritmo GBM e aplicados à amostra original não conseguiram representar adequadamente as informações contidas nas variáveis de controle, e o método GPS-IPW teve desempenho abaixo do esperado.

Se os estimadores, contudo, apresentarem comportamento adequado e significativo – especialmente o estimador de *sld* – então o método GPS-IPW de fato teve um desempenho superior ao de métodos lineares. Nesse caso, o emprego de aprendizado de máquina no cálculo dos pesos terá sido fundamental (ZHU; COFFMAN; GHOSH, 2015), uma vez que o algoritmo GBM terá conseguido gerar pesos que representem as relações entre *sld* e os controles – sejam elas lineares ou não – e contornar limitações e especificidades da amostra descritas na Seção 3.1.

Ao interpretar os resultados a seguir, é útil lembrar o leitor de que a ponderação foi realizada com todos os municípios das áreas observadas para os quais havia dados completos da variável de interesse e de todos os outros controles. Entretanto, as amostras ponderadas utilizadas para executar os modelos analisados a seguir são menores: isso se deve à limitação de dados observados sobre a variável resultante *homr*. A Seção 3.2 justificou as vantagens desta escolha (VAN HIPPEL, 2007; VON BUUREN, 2018).

A Divisão Territorial de 2010, segundo a qual todos os Censos anteriores a 2010 foram atualizados (IBGE, 2010), conta com 5.565 municípios, dos quais apenas 5.498 apresentaram dados completos das variáveis independentes e por isso, estavam aptos a participar das amostras ponderadas (resultados apresentados na Seção 4.1). Destes últimos, apenas 2.799 municípios apresentaram dados observados de taxas de homicídio em 2000 – 50,9% do total de municípios que participaram da ponderação. Os resultados dos modelos lineares ponderados desta seção foram obtidos com base neste último grupo de municípios para o qual há dados de *homr* em 2000. A Seção 3.2 justificou as vantagens desta escolha (VAN HIPPEL, 2007; VON BUUREN, 2018).

4.2.1 Sudeste

A Figura 9 exibe os resultados dos modelos lineares generalizados os municípios da região Sudeste no ano 2000.

Figura 9 - Relação entre escravidão e violência: Sudeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
<i>sld</i>	0,127* (0,069)	0,130* (0,069)	0,128* (0,069)	0,137** (0,069)	0,133* (0,069)
<i>ypc</i>	0,002 (0,008)	0,006 (0,009)	0,006 (0,009)	-0,142*** (0,028)	-0,061*** (0,015)
<i>pd</i>	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)
<i>unmp</i>	0,860*** (0,167)	0,932*** (0,176)	0,914*** (0,173)	0,812*** (0,168)	0,768*** (0,168)
<i>be</i>	-0,180 (0,147)	-0,247 (0,156)	-0,239 (0,154)	-0,134 (0,139)	-0,088 (0,142)
<i>ym</i>	0,684	0,532	0,556	0,350	0,471

	(0,670)	(0,662)	(0,664)	(0,625)	(0,631)
<i>cps</i>	-1,864*** (0,571)	-1,777*** (0,570)	-1,810*** (0,571)	-1,441** (0,560)	-1,409** (0,559)
<i>long</i>	-0,125 (0,387)	-0,096 (0,386)	-0,110 (0,387)	0,131 (0,386)	0,047 (0,387)
<i>lat</i>	-2,187*** (0,424)	-2,091*** (0,438)	-2,097*** (0,437)	-1,889*** (0,436)	-2,022*** (0,433)
<i>gini</i>		-22,439 (14,194)		-137,166*** (28,056)	
<i>theil</i>			-10,570 (6,846)		-62,325*** (12,139)
<i>ypc x gini</i>				0,249*** (0,045)	
<i>ypc x theil</i>					0,105*** (0,018)
Intercepto	-47,241** (19,973)	-30,236 (22,762)	-38,353* (20,984)	50,861* (26,729)	-0,774 (20,723)
<i>n-Adj</i>	776	776	776	776	776
Pseudo R ²	0,296	0,299	0,299	0,322	0,322
Pseudo R ² Ajustado	0,285	0,287	0,287	0,309	0,310

Nota:

*p<0,10 **p<0,05 ***p<0,01

Fonte: elaboração própria com o apoio do software *RStudio*

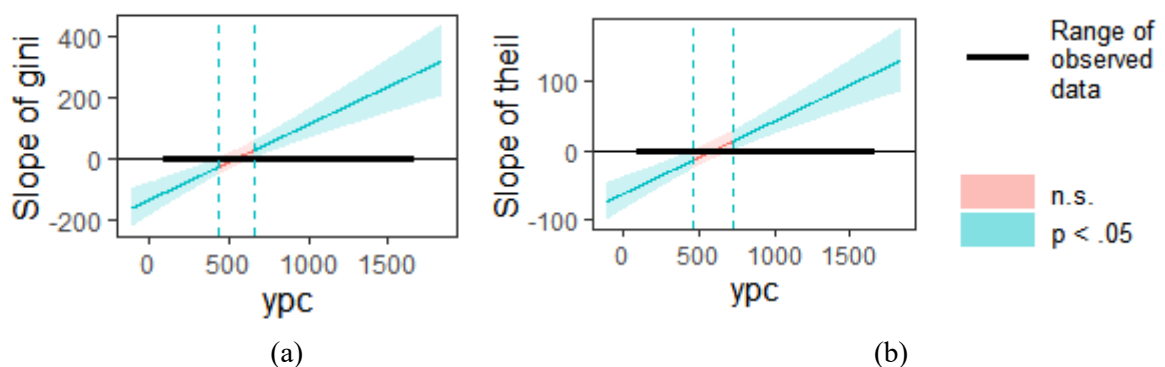
Os resultados dos Modelos 1, 2 e 3 mostram que a intensidade da escravidão em 1872 teve relação positiva e significativa (a 10%) com as taxas de homicídio municipais em 2000 para os municípios da região Sudeste. Os Modelos 2 e 3 incluíram dois indicadores de desigualdade de renda, cuja relação com a escravidão foi analisada por extensa literatura empírica (Seção 1). Nos Modelos 2 e 3, o estimador de *sld* permaneceu positivo e significativo a 10%, robusto em relação à desigualdade de renda. A relação entre escravidão e violência é intuitiva e está de acordo com a literatura especializada.

Nos primeiros três modelos, a renda média domiciliar *per capita ypc*, a educação básica *be* e o percentual de homens jovens na população do município *ym* não foram significativos, assim como os índices de desigualdade *gini* e *theil*. Desemprego *unmp* e densidade populacional *pd*, ao contrário, foram altamente significantes e apresentaram relações positivas com *homr*, em acordo com a literatura. Contudo, os estimadores de *ypc*, *gini* e *theil* apresentaram relações contraintuitivas, indicando necessidade de ajuste da especificação.

O índice municipal de justiça e segurança pública *cps*, criado para os fins deste estudo, foi significativa a 1% nos Modelos 1 a 3, e seguir significativa a 5% nos Modelos 4 a 5. Seu comportamento sugere que municípios com maiores taxas de homicídio estão associados a menor infraestrutura de justiça e segurança pública, e vice-versa. A latitude *lat* foi significativa a 1% em todos os modelos. Como é negativa no Hemisfério Sul, entende-se que latitudes maiores são as que estão mais de 0° (ou seja, da linha do Equador). O estimador negativo mostra que latitudes maiores (“menos negativas”), mais ao norte da região, estão associadas a menores níveis de violência, e vice-versa.

Os Modelos 4 e 5 incluíram os termos de interação *ypc x gini* e *ypc x theil*, respectivamente. O objetivo é verificar se a relação entre estas variáveis contribui significativamente para analisar as taxas de homicídio, dado que, isoladamente, não foram não foram significativos nas estimações anteriores – o que contraria a literatura empírica (Seção 1). Os termos de interação foram positivos e significativos a 1%. Sua inclusão levou *ypc* a apresentar o comportamento previsto pela literatura empírica: correlação negativa com *homr* e significativa a 1%. Para compreender mais detalhadamente a importância da interação entre *ypc* e desigualdade de renda, as Figuras 10a e 10b exibem resultados da técnica de Johnson-Neyman de análise de interação. O Modelo 4 é representado pela Figura 10a, e o Modelo 5, pela Figura 10b.

Figura 10 - Gráficos de Johnson-Neyman para a região Sudeste



Fonte: elaboração própria com o software *RStudio*.

Nos gráficos da Figura 10, o eixo X representa os valores que *ypc* pode assumir. A reta escura sobre o eixo representa os valores que *ypc* assumiu nas amostras utilizadas para gerar os Modelos 4 e 5 (entre R\$ 100,50 e R\$ 1638,66). O eixo Y representa o estimador dos controles de desigualdade de renda. Portanto, os gráficos de Johnson-Neyman mostram o que acontece

com o estimador da relação entre desigualdade de renda e *homr* à medida que o valor observado de *ypc* aumenta ao longo da amostra analisada.

Em linhas gerais, ambos os modelos permitem observar que, nos municípios mais pobres do Sudeste (*ypc* baixa), a relação entre desigualdade de renda e *homr* é negativa, inversa. Por exemplo: municípios com baixa *ypc* e altos índices de desigualdade estão associados a menores taxas de homicídios na amostra observada. Por outro lado, nos municípios mais ricos do Sudeste (*ypc* alta), a relação entre desigualdade de renda e *homr* é direta: municípios mais ricos e com altos índices de desigualdade estão associados a maiores taxas de homicídio nesta região.

A área em verde mostra que o estimador da relação entre desigualdade e *homr* é significativo para as menores e para as maiores faixas de *ypc*. A área em rosa representa as faixas de *ypc* para os quais esta relação não é significativa (a 5%). Para a maior parte do intervalo de valores que *ypc* assume na amostra, a desigualdade de renda é significativa para a análise da violência letal. Entretanto, a relação entre desigualdade e violência não permanece a mesma ao longo deste intervalo, sendo negativa para os municípios mais pobres e positiva para os municípios mais ricos. Essas diferenças devem ser levadas em consideração ao analisar as taxas de homicídio municipais: em outras palavras, não é possível associar maiores níveis de renda a maiores ou menores taxas de homicídios sem analisar a desigualdade – a depender desta, os padrões de violência observados podem mudar.

Nos municípios mais ricos e com alto nível de desigualdade, a relação com a violência condiz com a maior parte da literatura empírica: este grupo está associado a maiores níveis de violência. A Figura 10 mostra que, para essas faixas de *ypc*, o estimador da desigualdade é positivo, significativo e crescente. Contudo, nos municípios mais pobres, maiores desigualdades não estão associadas a mais violência: os gráficos mostram que os estimadores de *gini* e *theil* são negativos, representando uma relação inversa entre desigualdade de renda e violência letal, significativa a 5%. Esse cenário contraintuitivo foi significativo, corroborando a inclusão dos termos de interação entre *ypc* e os índices de desigualdade nos Modelos 4 e 5.

As Figuras 11 e 12 utilizam alguns valores de referência de *ypc* para ilustrar as diferentes dinâmicas entre desigualdade de renda e *homr* nos municípios do Sudeste. São eles: a média amostral de *ypc* menos um desvio padrão (R\$ 313,35 em ambos os modelos); a média amostral de *ypc* (R\$ 511,31 no Modelo 4 e R\$ 511,20 no Modelo 5) e a média amostral de *ypc* mais um desvio padrão (R\$ 709,27 e R\$ 709,05 nos Modelos 4 e 5, respectivamente). As diferenças se devem ao fato de que as médias amostrais se referem às amostras ponderadas, não à amostra

original, e por isso variam de acordo com a especificação de cada modelo. A Figura 11 exibe os valores de referência de ypc , o estimador da relação entre desigualdade e $homr$ associado a cada um deles (\hat{a}) e os intervalos de confiança correspondentes.

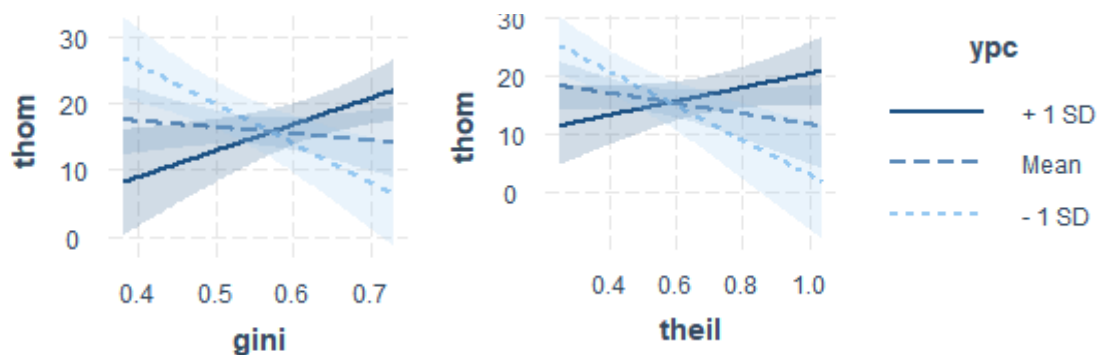
Figura 11 - Descrição dos intervalos de Johson-Neyman para a região Sudeste

<i>gini</i>				<i>theil</i>			
<i>ypc</i>	\hat{a}	2,5%	97,5%	<i>ypc</i>	\hat{a}	2,5%	97,5%
313,35	-59,04	-93,06	-25,02	313,35	-29,55	-45,46	-13,64
511,31	-9,68	-37,07	17,70	511,20	-8,86	-22,08	4,37
709,27	39,67	8,92	70,42	709,05	11,84	-1,95	25,62

Fonte: elaboração própria com o software *RStudio*.

Na Figura 12, os valores de referência ypc são representados pelas retas cheias e tracejadas em diferentes tons de azul, ao entorno das quais os intervalos de confiança se formam (áreas sombreadas no tom de azul correspondente às retas).

Figura 12 - Intervalos de Johnson-Neyman para a região Sudeste



Fonte: elaboração própria com o software *RStudio*.

Nas Figuras 12a e 12b, as retas tracejadas azul claras representam a média amostral de ypc menos um desvio-padrão para os Modelos 4 e 5, respectivamente. Em ambos os gráficos, as retas se inclinam para baixo, indicando que para este valor de ypc , a relação entre $homr$ e a desigualdade de renda é inversa. Os estimadores \hat{a} são negativos para esse valor de ypc , como mostra a Figura 11. Os intervalos de confiança (Figura 11) demonstram que \hat{a} tende a permanecer negativo para esse valor de ypc , que representa a dinâmica entre renda, desigualdade e violência observada entre os municípios mais pobres da região Sudeste.

De modo geral, é possível concluir que maiores índices de desigualdade de renda estão significativamente associados a menores taxas de homicídio entre os municípios de menor renda *per capita* domiciliar média na região Sudeste. Segue-se uma possível explicação para esse resultado, a princípio, contraintuitivo.

Em municípios mais pobres, a utilidade marginal da renda é maior. Isso significa que aumentos na renda de parte da população têm maior probabilidade de estarem associados a elevações no padrão de vida capazes de desencorajar atos violentos motivados por fatores econômicos (ou outros comportamentos de risco associados à violência, como abuso de álcool, por exemplo). Da mesma forma, quedas na desigualdade associadas à redução da renda de parte dos habitantes podem estar associadas à redução do custo de oportunidade de atos violentos. Esta hipótese pode ajudar a explicar por que maior igualdade de renda não está associada a menores, mas a maiores taxas de homicídios entre os municípios mais pobres da região Sudeste.

Na Figura 12, as retas contínuas azul escuras ilustram a dinâmica de desigualdade e *homr* associada aos municípios mais ricos da região Sudeste. Nesta reta, *ypc* é igual a sua média amostral mais um desvio-padrão. A Figura 11 mostra que, para o Modelo 4, o estimador da relação entre *gini* e *theil* assume valores positivos ao longo de todo o intervalo de confiança especificado. No caso do Modelo 5, os intervalos de confiança para esta faixa de renda consideram que \hat{a} possa chegar a um valor negativo, mas a probabilidade disso é muito baixa. No Sudeste, para municípios com valores mais altos de *ypc*, a relação entre desigualdade e *homr* é proporcional, direta, e mais consonante com a literatura empírica.

Nos municípios mais ricos, a utilidade marginal da renda é relativamente menor que nos municípios mais pobres. Por isso, aumentos de renda suficientemente capazes de desincentivar atos violentos teriam que atingir as populações de baixa renda, para quem o custo de oportunidade da violência é mais baixo – o que, por sua vez, reduziria a desigualdade. Como mostram as Figuras 11 e 12, menor desigualdade está associada a menores níveis de violência letal entre os municípios mais ricos do Sudeste, representados pela reta azul cheia na Figura 12.

A Figura 10 mostrou que municípios com valores de *ypc* próximos à média amostral não apresentam uma relação significativa entre desigualdade de renda e violência letal (área rosa das Figuras 10a e 10b). Na Figura 12, esses municípios são representados pela reta tracejada intermediária, quase plana (evidenciando a não significância do estimador). Os intervalos de confiança da Figura 11 mostram que, para este grupo, \hat{a} tem maior probabilidade de ser decrescente e, portanto, de representar uma dinâmica semelhante à observada entre os municípios de baixa renda.

Conclui-se que, na amostra de municípios da região Sudeste, a interação entre *ypc* e os índices de desigualdade de renda não pode ser ignorada sem contribuir significativamente para o aumento do viés de variável omitida, comprometendo a variância dos estimadores correlacionais. Em estimações causais, a omissão da interação aumentaria significativamente os níveis de confundimento e impediria a obtenção do efeito médio do tratamento correto.

O termo de interação foi significativo a 1% em ambos os Modelos 4 e 5. Após sua inclusão, *ypc* passou a ter correlação negativa e significativa a 1%, comportamento condizente com a literatura. Os coeficientes de *gini* e *theil* também são significantes a 1%, e a derivada parcial desses índices mostra que a correlação isolada entre desigualdade de renda e violência passou a ser positiva, também condizente com a literatura.

O aumento do pseudo R-quadrado e do pseudo R-quadrado ajustado reforçam as conclusões sobre o papel determinante da interação entre renda e desigualdade na qualidade das estimações. A inclusão dos níveis absolutos de desigualdade (Modelos 2 e 3) não elevou significativamente o pseudo R-quadrado ajustado dos modelos, o que ocorreu apenas após a inclusão das interações.

Nos Modelos 4 e 5, o estimador da correlação entre *sld* e *homr* foi positivo, de acordo com o esperado. Seu valor absoluto aumentou à medida que os controles iam sendo incluídos, e no Modelo 4, o estimador chegou a ser significativo a 5%. Nos outros modelos, manteve-se significativo a 10%. Portanto, o estimador apresentou robustez. Devido às questões analisadas na Seção 4.1, é seguro afirmar que há correlação significativa entre a intensidade da escravidão em 1872 e as taxas de homicídio municipais em 2000 na região Sudeste.

4.2.2 Nordeste

Figura 13 - Relação entre escravidão e violência: Nordeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
<i>sld</i>	0,082 (0,089)	0,093 (0,089)	0,094 (0,088)	0,091 (0,088)	0,089 (0,087)
<i>ypc</i>	0,009 (0,011)	0,015 (0,011)	0,022* (0,012)	-0,030 (0,076)	-0,001 (0,036)
<i>pd</i>	0,003** (0,001)	0,003** (0,001)	0,003** (0,001)	0,003** (0,001)	0,003** (0,001)
<i>unmp</i>	0,035 (0,102)	0,059 (0,099)	0,043 (0,101)	0,054 (0,100)	0,040 (0,102)

<i>be</i>	0,079 (0,114)	0,053 (0,111)	0,033 (0,113)	0,071 (0,117)	0,056 (0,117)
<i>ym</i>	1,082* (0,558)	0,753 (0,560)	0,739 (0,569)	0,747 (0,561)	0,745 (0,572)
<i>cps</i>	-0,675 (0,440)	-0,691 (0,439)	-0,676 (0,437)	-0,690 (0,439)	-0,672 (0,438)
<i>long</i>	0,315 (0,358)	0,133 (0,370)	0,106 (0,362)	0,150 (0,371)	0,120 (0,364)
<i>lat</i>	-0,853*** (0,274)	-0,820*** (0,273)	-0,793*** (0,274)	-0,819*** (0,274)	-0,796*** (0,275)
<i>gini</i>		-17,412* (8,910)		-29,759 (23,272)	
<i>theil</i>			-11,020*** (4,258)		-17,434 (10,768)
<i>ypc x gini</i>				0,071 (0,118)	
<i>ypc x theil</i>					0,033 (0,050)
Intercepto	0,672 (19,157)	6,832 (19,092)	1,377 (19,005)	15,114 (23,940)	5,663 (20,451)
<i>n-Adj</i>	874	874	874	874	874
Pseudo R ²	0,368	0,371	0,372	0,371	0,373
Pseudo R ² Ajustado	0,355	0,357	0,359	0,357	0,359

Nota:

*p<0,10 **p<0,05 ***p<0,01

Fonte: elaboração própria com o software *RStudio*.

A Seção 4.1.2 mostrou que, comparada ao Sudeste, a amostra ponderada dos municípios do Nordeste teve níveis de confundimento menores: o valor absoluto da correlação entre *sld* e a maioria dos controles ficou abaixo de 0,10. Na Figura 13, o estimador de *sld* mede a correlação entre a intensidade da escravidão em 1872 e as taxas de homicídios em 2000 para os municípios do Nordeste. A Figura 13 mostra que houve correlação positiva entre *sld* e *homr*, o que está de acordo com a hipótese de que a escravidão está associada a maiores níveis de violência, mas esta não foi significativa em nenhum dos modelos.

Apenas a latitude foi significativa a 1% em todas as especificações exibidas na Figura 13 – assim como observado na região Sudeste. É útil relembrar a interpretação do estimador negativo de *lat*, pois esta variável é negativa para o Hemisfério Sul, o que pode gerar confusão. Valores menos negativos (mais próximos de zero) indicam latitudes matematicamente maiores,

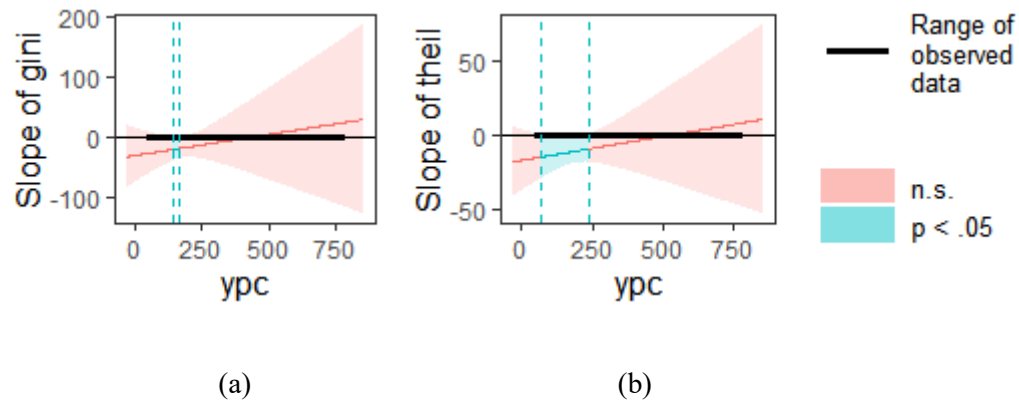
localizadas mais ao norte do Nordeste. O estimador negativo mostra que municípios mais ao norte da região (latitudes menos negativas) estão associados a menores níveis de violência, enquanto municípios mais ao sul do Nordeste (latitudes mais negativas) tendem a apresentar maiores taxas de homicídio.

Como esperado, a densidade populacional foi positivamente correlacionada à *homr*, significativa a 5%. De modo geral, a taxa de desemprego, o percentual de homens jovens e o indicador municipal de justiça e segurança pública tiveram o comportamento condizente com a literatura empírica: *unmp* e *ym* foram positivamente correlacionadas às taxas de homicídios, e *cps*, negativamente correlacionada. Os valores absolutos dos seus coeficientes foram menores que os da região Sudeste. Entretanto, a significância a 10% de *ym* não resiste aos testes de robustez representados pelos Modelos 2 a 5.

Note-se que o percentual de pessoas que estudaram pelo menos oito anos *be* contrariou o comportamento esperado, apresentando correlação positiva com *homr* em todos os modelos. A inclusão dos níveis de desigualdade absolutos nos Modelos 2 e 3 levou *ypc* a apresentar correlação negativa com *homr*. A desigualdade, por sua vez, apresentou correlações negativas com a taxa de homicídios em 2000. A necessidade do termo de interação entre renda e desigualdade fica evidente pelo acúmulo de comportamentos anômalos. Contudo, não foi capaz, por si só, de corrigir todos eles.

Nos Modelos 4 e 5, a inclusão do termo de interação permitiu que *ypc* passasse a ter correlação negativa com a violência letal. Contudo, as interações não foram significantes, assim como os níveis absolutos de desigualdade e renda. O pseudo R-quadrado ajustado aumentou muito pouco de um modelo a outro. Contudo, foi superior aos valores observados para os modelos do Sudeste em todas as especificações, sugerindo que as variáveis independentes adotadas capturam melhor os fatores que influenciam a violência letal no Nordeste do que no Sudeste.

Figura 14 - Relação entre escravidão e violência: Nordeste



Fonte: elaboração própria com o software *RStudio*.

A Figura 14 mostra que a relação entre renda e desigualdade é muito diferente da observada na região Sudeste. Em ambos os Modelos 4 e 5, a interação não é significativa para a maior parte dos valores de *ypc* observados na amostra. No caso do Modelo 5, a faixa de valores intermediários de *ypc* nas quais *ypc x theil* é significativa é um pouco maior. As linhas crescentes rosa mostram que a correlação entre desigualdade e *homr* nos municípios do Nordeste cresce à medida que os valores de *ypc* aumentam entre a amostra observada. Contudo, a linha crescente rosa fica abaixo de zero ao longo da maior parte do intervalo de valores assumido por *ypc*. Portanto, para municípios mais pobres, a correlação entre desigualdade e violência tende a ser inversa (dinâmica explorada na Seção 4.2.1), tornando-se direta (positiva) à medida que o nível de renda aumenta entre os municípios observados.

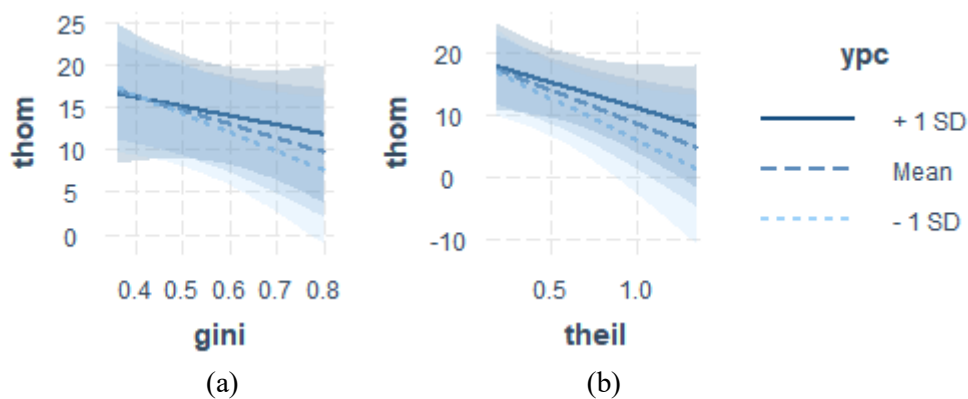
Na Figura 14, o caráter quase plano da reta rosa e a área da mesma que a acompanha indicam que a relação entre desigualdade e taxa de homicídios não muda significativamente ao longo dos valores observados de *ypc*. Tal corrobora o fato de que os índices de desigualdade e suas interações com *ypc* não foram significativos nos Modelos 4 e 5 (Figura 13), ainda que o ajuste desses modelos tenha sido maior que o dos anteriores (tanto pelo maior pseudo R^2 ajustado como porque *ypc* passou a apresentar um comportamento mais condizente com a literatura empírica).

Figura 15 – Descrição dos intervalos de Johnson-Neymann para a região Nordeste

<i>ypc</i>	<i>gini</i>			<i>theil</i>			
	\hat{a}	2,5%	97,5%	<i>ypc</i>	\hat{a}	2,5%	97,5%
106,08	-22,26	-47,06	2,55	106,15	-13,89	-26,21	-1,57
185,95	-16,61	-34,06	0,84	185,81	-11,23	-19,74	-2,72
265,81	-10,97	-37,00	15,07	265,49	-8,57	-19,28	2,14

Fonte: elaboração própria com o software *RStudio*.

Figura 16 - Intervalos de Johnson-Neyman para a Região Sudeste



Fonte: elaboração própria com o software *RStudio*.

As Figuras 15 e 16 trazem as mesmas informações que Figuras 11 e 12: valores de referência de *ypc* que representam as dinâmicas de diferentes municípios da região de acordo com seu nível de renda; estimador \hat{a} da relação entre desigualdade e taxas de homicídio municipais e intervalos de confiança correspondentes. Para todos os valores de referência de *ypc*, a correlação entre desigualdade e *homr* (\hat{a} assume valores negativos).

A aparente contradição com a reta rosa crescente da Figura 14 é, na verdade, mais um indício de que a interação entre renda e desigualdade não é significativa para analisar as taxas de homicídio no Nordeste, diferentemente do se observou ao analisar o Sudeste. Nesta última, a relação entre desigualdade e *homr* variou significativamente ao longo dos valores observados de *ypc*, o que não poderia ser ignorado sem comprometer o ajuste dos modelos. No Nordeste, por outro lado, esta relação não assumiu formas significativamente diferentes ao longo dos valores observados de renda na amostra, de modo que o termo de interação não foi significativo – apesar de colaborar para *ypc* passasse a apresentar correlação negativa com *homr*, em acordo com a literatura empírica.

4.2.3 Nordeste e Sudeste

Figura 17 - Relação entre escravidão e violência: Nordeste e Sudeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
<i>sld</i>	0,087 (0,057)	0,091 (0,056)	0,090 (0,057)	0,097* (0,056)	0,083 (0,056)
<i>ypc</i>	-0,007 (0,007)	-0,005 (0,007)	-0,005 (0,007)	-0,092*** (0,023)	-0,047*** (0,012)
<i>pd</i>	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)
<i>unmp</i>	0,336*** (0,090)	0,358*** (0,090)	0,347*** (0,090)	0,314*** (0,088)	0,296*** (0,088)
<i>be</i>	0,047 (0,092)	0,026 (0,093)	0,035 (0,093)	0,070 (0,084)	0,109 (0,085)
<i>ym</i>	0,836* (0,436)	0,688 (0,437)	0,736* (0,438)	0,526 (0,432)	0,622 (0,433)
<i>cps</i>	-1,063*** (0,355)	-1,029*** (0,352)	-1,032*** (0,354)	-0,992*** (0,345)	-0,937*** (0,349)
<i>long</i>	0,105 (0,280)	0,072 (0,281)	0,076 (0,280)	0,057 (0,281)	0,066 (0,279)
<i>lat</i>	-1,456*** (0,255)	-1,422*** (0,258)	-1,427*** (0,258)	-1,445*** (0,257)	-1,480*** (0,258)
<i>gini</i>		-12,975* (7,343)		-56,005*** (12,517)	
<i>theil</i>			-5,492 (3,522)		-25,786*** (5,831)
<i>ypc x gini</i>				0,144*** (0,036)	
<i>ypc x theil</i>					0,062*** (0,015)
Intercepto	-6,386 (14,022)	1,791 (14,308)	-3,234 (14,115)	28,336* (15,238)	9,137 (14,074)
<i>n-Adj</i>	1,650	1,650	1,650	1,650	1,650
R ²	0,312	0,314	0,313	0,323	0,321
R ² Ajustado	0,303	0,305	0,304	0,314	0,312

Nota:

*p<0,10 **p<0,05 ***p<0,01

Fonte: elaboração própria com o software *RStudio*.

A Seção 4.1.3 mostrou que a amostra ponderada que reúne os municípios das regiões Nordeste e Sudeste apresentou padrões de confundimento mais baixos, próximos dos da região Nordeste quando analisada individualmente. A Figura 17 demonstra que os modelos baseados na amostra ponderada conjunta estão melhor ajustados que os modelos referentes às regiões Nordeste e Sudeste analisadas isoladamente.

Estimadores que apresentaram comportamento contraintuitivo nas análises isoladas passaram a apresentar correlações alinhadas com a literatura empírica sobre violência. Por exemplo: o desemprego *unmp* não foi significativo na análise da região Nordeste, mas passou a ter relação positiva e significativa a 1% na análise da amostra conjunta. O comportamento de *ypc* foi mais coerente desde o Modelo 1, e com a inclusão dos controles de desigualdade, chegou a ser significativo a 1% nos Modelos 4 e 5. Latitude e densidade populacional mantiveram o comportamento esperado e altos níveis de significância.

No Sudeste, o estimador da educação básica *be* foi negativo, como esperado. Contudo, foi positivo na amostra de municípios do Nordeste e na amostra conjunta das duas regiões. Tal sugere que o percentual de pessoas com pelo menos oito anos de educação completos não foi uma boa *proxy* da educação para as estimações realizadas até aqui, com exceção daquelas referentes ao Sudeste.

O índice municipal de justiça e segurança pública *cps* apresentou estimador negativo para todas as amostras analisadas até aqui. Tal sugere que maior infraestrutura de segurança está correlacionada a menores taxas de homicídios. Contudo, *cps* não foi significativo para o Nordeste, enquanto apresentou significância a 1% para o Sudeste e na amostra conjunta de ambas as regiões. Conclui-se que, nas regiões com maior população absoluta escravizada em 1872, a existência de órgãos públicos municipais de justiça e segurança está associada a menores taxas de homicídio em 2000.

O índice *cps* apresentou baixo confundimento na amostra conjunta (Seção 4.1.3) e na amostra referente ao Nordeste (Seção 4.1.2), e médio confundimento na amostra de municípios do Sudeste (Seção 4.1.1). O indicador apresenta comportamento coerente com os padrões descritos pela literatura empírica, e vem se mostrando relevante para analisar a violência letal.

O percentual de homens jovens *ym*, ainda que positivo, não resistiu aos testes de robustez representados pelas especificações mais complexas, deixando de ser significativo a 10% nos Modelos 2, 4 e 5, que apresentam maior pseudo R^2 ajustados.

Os estimadores de *lat* foram consistentemente negativos e altamente significantes ao longo de todas as amostras analisadas. Como a latitude assume valores negativos para as regiões

Sudeste e Nordeste, os estimadores sugerem que a violência letal é menor ao norte da área estudada, e maior ao sul. No extremo sul da região Sudeste, os municípios assumem latitudes negativas de maior módulo (maior distância da linha do Equador). Multiplicadas pelo coeficiente negativo de *lat*, estas latitudes levariam um incremento de alto valor absoluto sobre a violência letal. Sendo assim, conclui-se que municípios ao sul da área estudada estão associados a maiores taxas de homicídio. O mecanismo contrário se verifica nos municípios do extremo norte da região Nordeste, mais próximos da linha do Equador e, portanto, com latitudes negativas de menor valor absoluto. Compreender o comportamento da latitude nos modelos é importante para entender o papel de *outliers* – como as grandes metrópoles do Sudeste, Rio de Janeiro e São Paulo, e suas regiões metropolitanas – nos resultados obtidos.

A longitude, por outro lado, apresentou estimadores não significantes para todas as amostras ponderadas analisadas até aqui. A leste do meridiano de Greenwich, a longitude assume apenas valores negativos. Logo, menor a longitude (valor negativo com módulo mais baixo), mais interior o município. Maior a longitude (valor negativo com módulo mais baixo, comparativamente), mais próximo do litoral está o município. O estimador de *long* foi majoritariamente positivo para todas as amostras, com exceção dos Modelos 4 e 5 referentes ao Sudeste (Seção 4.2.1). Portanto, a relação entre *long* e *homr* é *direta*, proporcional: maior a longitude, maior a proximidade com o litoral e maiores as taxas de homicídio, em média. Tal corrobora a narrativa histórica explorada na Seção 3.1, segundo a qual a proximidade com o litoral determinou a ocupação do território durante a colonização – o que tem impacto significativo sobre os padrões de densidade populacional, concentração de renda e outros fatores que contribuem para maiores níveis de violência letal ao leste do Brasil.

Na amostra conjunta das regiões Nordeste e Sudeste, o comportamento do estimador de *sld* está de acordo com o esperado, mas não foi significativo na maioria dos modelos. No Modelo 4, que inclui todos os controles disponíveis, a correlação foi negativa e significativa a 10%. Contudo, esse comportamento não resistiu ao teste de robustez representado pelo Modelo 5, que utiliza o índice de Theil ao invés do índice de Gini para mensurar a desigualdade de renda. A Seção 4.1.3, mostrou que, no Modelo 4, *gini* apresentou baixo confundimento com *sld*, enquanto *theil* apresentou confundimento médio com *sld* no Modelo 5. Por isso, é útil testar se outros índices de desigualdade que não o de Theil podem ser utilizados para testar a robustez dos estimadores obtidos no Modelo 4.

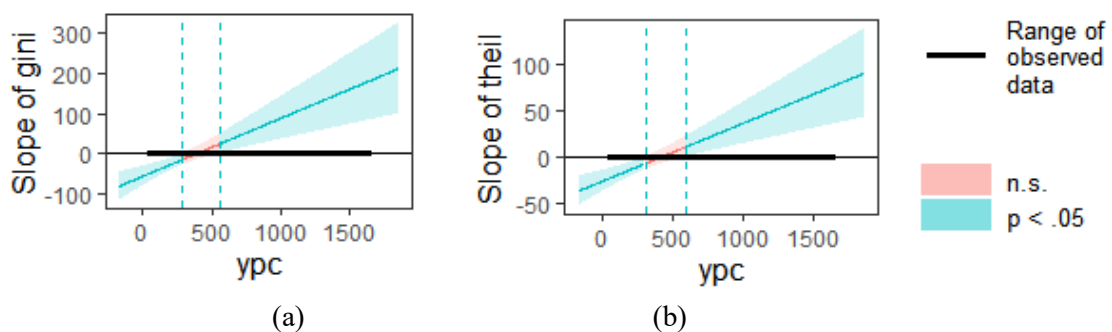
A inclusão da interação entre *ypc* e desigualdade de renda foi significativa a 1% nos Modelos 4 e 5, contribuindo para ajustar os estimadores absolutos de todas essas variáveis,

aumentando o pseudo R-quadrado ajustado. Ao contrário do que ocorreu na análise individual do Nordeste, a relação entre renda e desigualdade foi suficientemente relevante no estudo das taxas de homicídio municipais das regiões Sudeste e Nordeste como um todo, assim como fora na análise isolada dos municípios do Sudeste. Sendo assim, é razoável supor que a influência da desigualdade no Sudeste levou à necessidade de considerar o termo de interação ao analisar a amostra conjunta. Também é interessante notar que os valores absolutos dos coeficientes de desemprego, altamente significantes para o Sudeste e para a amostra conjunta, não foram altos nem significativos para o Nordeste.

A amostra conjunta fortalece a hipótese de que há transbordamento do cenário observado no Sudeste, onde o estimador de *sld* foi positivo e significativo, para outras regiões. Isso se deve ao fato de que os padrões observados na Figura 17 remetem aos observados na análise individual do Sudeste e estão mais condizentes com a literatura, mesmo com a inclusão dos municípios do Nordeste.

Na amostra ponderada de municípios do Nordeste, variáveis como *unmp*, *cps*, *ypc* e os índices de desigualdade não apresentaram comportamento condizente com a literatura empírica, que os considera fundamentais para a análise da violência. A união com os municípios do Sudeste alterou significativamente os padrões observados entre os estimadores dessas variáveis, tornando-os mais intuitivos e, ao mesmo tempo, mais próximos dos observados na amostra da região Sudeste (que não apresentou índices de confundimento tão baixos como os da amostra conjunta (Seção 4.1.2)).

Figura 18 - Gráficos de Johnson-Neyman para as regiões Nordeste e Sudeste



Fonte: elaboração própria com o software *RStudio*.

A Figura 18 mostra que, em conjunto, os municípios das regiões Sudeste e Nordeste apresentam um comportamento semelhante ao observado no Sudeste: ao longo do intervalo de valores que *ypc* observados na amostra, a taxa de homicídios varia significativa e

proporcionalmente com a desigualdade nos municípios observados. Isso corrobora os resultados referentes aos Modelos 4 e 5 (Figura 17), cujos coeficientes do efeito absoluto de *gini* e *theil* (obtidos através de derivadas parciais condicionais a *ypc* devido ao termo de interação) são positivos e significativos.

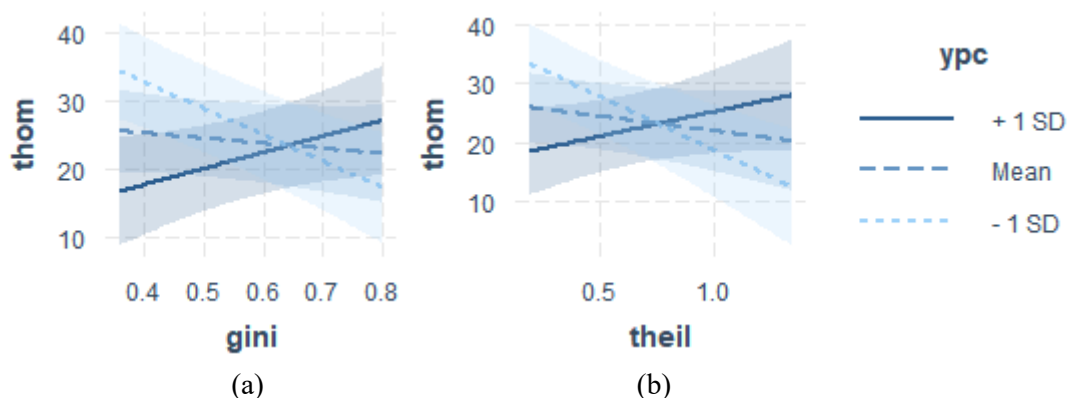
Assim como na análise individual do Sudeste, *homr* está negativamente correlacionada à desigualdade nos municípios mais pobres, e positivamente correlacionada nos municípios mais ricos da amostra conjunta. Essas dinâmicas são, ambas, significativas para analisar as taxas de homicídios municipais. As Figuras 19 e 20 fornecem mais detalhes sobre estes resultados, mostrando que a dinâmica entre renda e desigualdade nas regiões Nordeste e Sudeste é muito semelhante à observada no Sudeste isoladamente.

Figura 19 – Descrição dos intervalos de Johnson-Neyman para as regiões Nordeste e Sudeste

<i>ypc</i>	\hat{a}	<i>gini</i>		<i>theil</i>		<i>ypc</i>	\hat{a}	2,5%	97,5%
		2,5%	97,5%	2,5%	97,5%				
120,64	-38,64	-56,91	-20,38	120,77	-18,29	-27,08	-9,50		
336,73	-7,55	-22,27	7,17	335,48	-4,96	-11,88	1,95		
552,81	23,55	-0,27	47,36	550,19	8,36	-1,74	18,46		

Fonte: elaboração própria com o software *RStudio*.

Figura 20 - Intervalos de Johnson-Neyman para as regiões Nordeste e Sudeste



Fonte: elaboração própria com o software *RStudio*.

Em geral, os valores de referência são um pouco mais baixos que os observados nos modelos baseados na amostra isolada do Sudeste. Entre os municípios de baixa renda relativa na amostra, representados por *ypc* = R\$ 120 (Figura 19), o intervalo de valores assumidos por \hat{a} (inclinação da reta entre desigualdade de renda e *homr*) pode assumir é sempre negativo a 5%

de confiança (retas azul-claras tracejadas da Figura 20). Sendo assim, é possível propor que, dada a maior utilidade marginal da renda nos municípios relativamente mais pobres, maior desigualdade pode refletir aumentos de renda capazes de reduzir os incentivos a cometer atos violentos.

Por outro lado, entre os municípios com alta renda relativa, representados por $ypc = R\$ 550$ e pelas retas escuras contínuas nas Figuras 20a e 20b, $\hat{\alpha}$ assume predominantemente valores positivos, como mostra a última linha da Figura 19. Sendo assim, maiores níveis de ypc estão associados a maiores níveis de desigualdade, e estes, por sua vez, estão associados a maiores taxas de homicídio.

Como visto na Seção 4.2.1, a hipótese da utilidade marginal da renda ajuda a compreender esse cenário. Sob essa hipótese, é razoável supor que, em municípios com alta renda e alta desigualdade, aumentos de renda capazes de desincentivar a violência devem ocorrer entre as populações mais pobres – o que, por sua vez, tornaria a distribuição de renda mais igualitária. No entanto, se os aumentos de renda não estiverem associados à redução da desigualdade, então não estão associados à redução dos incentivos econômicos a cometer atos violentos. Portanto, caso a alta desigualdade seja mantida, aumentos na renda seguiriam paralelos a altas taxas de violência letal.

Municípios com renda média na amostra conjunta são representados pelas retas de cor azul intermediária e tracejado maior das Figuras 20a e 20b. Suas inclinações assumem, predominantemente, valores negativos, ao longo da amostra, sugerindo uma relação semelhante à observada entre municípios de baixa renda. Contudo, para este grupo, a relação entre desigualdade e ypc não foi significativa para analisar $homr$, como mostram as retas quase planas que os representam na Figura 20. Isso pode ser atribuído ao fato de os efeitos da utilidade marginal da renda nestes municípios é difícil de prever, pois é mais heterogêneo e depende de outros fatores.

Os resultados das estimações para os municípios das regiões Nordeste e Sudeste apresentam padrões que sugerem o transbordamento de diversas características socioeconômicas do Sudeste para o Nordeste, uma vez que se assemelharam mais aos observados na primeira. As amostras ponderadas das regiões Nordeste e Sudeste (Seções 4.1.1 e 4.1.2) apresentaram maiores níveis de confundimento que a amostra conjunta (Seção 4.1.3), indicando que o ajuste dos modelos baseados nesta última é melhor. Com a ampliação da amostra, foi possível obter resultados mais significantes e mais condizentes com a literatura empírica que os da análise isolada do Nordeste – sem que isso aumentasse os níveis de

confundimento. Contudo, na amostra conjunta, os controles socioeconômicos mais importantes para analisar as taxas de homicídio também foram os que exibiram maiores níveis de confundimento, o que evidencia a necessidade de ajustes nas especificações dos modelos analisados.

4.2.4 Brasil

Figura 21 - Efeito da escravidão sobre a violência: Brasil

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
<i>sld</i>	0,117** (0,052)	0,122** (0,052)	0,120** (0,052)	0,119** (0,052)	0,116** (0,053)
<i>ypc</i>	0,001 (0,004)	0,002 (0,004)	0,002 (0,004)	-0,050*** (0,013)	-0,016** (0,007)
<i>pd</i>	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)
<i>unmp</i>	0,194*** (0,075)	0,208*** (0,076)	0,201*** (0,075)	0,138* (0,076)	0,146* (0,076)
<i>be</i>	-0,149** (0,074)	-0,163** (0,074)	-0,159** (0,075)	-0,053 (0,075)	-0,062 (0,079)
<i>ym</i>	1,115*** (0,400)	1,054*** (0,408)	1,074*** (0,403)	1,067*** (0,403)	1,103*** (0,399)
<i>cps</i>	-1,429*** (0,321)	-1,413*** (0,318)	-1,418*** (0,319)	-1,302*** (0,315)	-1,310*** (0,317)
<i>long</i>	-0,140 (0,208)	-0,160 (0,210)	-0,157 (0,209)	-0,096 (0,210)	-0,113 (0,209)
<i>lat</i>	-0,397* (0,229)	-0,386* (0,230)	-0,388* (0,230)	-0,447* (0,231)	-0,451* (0,233)
<i>gini</i>		-5,762 (5,858)		-30,697*** (8,503)	
<i>theil</i>			-2,214 (2,676)		-8,919*** (3,243)
<i>ypc x gini</i>				0,075*** (0,018)	
<i>ypc x theil</i>					0,018*** (0,005)
Intercepto	6,314	9,294	7,043	27,942*	13,500

	(15,158)	(15,253)	(15,198)	(15,858)	(15,283)
<i>n</i> -Adj	2,699	2,699	2,699	2,699	2,699
Pseudo R ²	0,243	0,243	0,243	0,249	0,247
Pseudo R ² Ajustado	0,233	0,233	0,233	0,239	0,236

Nota: *p<0,10 **p<0,05 ***p<0,01

Fonte: elaboração própria com o software *RStudio*.

A Figura 21 mostra os resultados da amostra que reuniu todos os municípios brasileiros para os quais havia dados de *homr* disponíveis em 2000. Dos 2.799 municípios que apresentaram dados observados de taxas de homicídio em 2000, 100 ficaram fora das regressões lineares ponderadas baseadas na amostra que reúne todas as regiões geográficas. Isso significa que o algoritmo GBM não conseguiu lhes atribuir um peso que representasse sua probabilidade de pertencer ao grupo de tratamento, descartando cerca de 3,5% da amostra original. Apesar da perda, a análise da Figura 21 indica que os resultados obtidos com base na amostra ponderada são robustos.

A Seção 4.1.4 mostrou que, ainda que níveis médios de confundimento persistam entre os controles socioeconômicos, a amostra que reuniu todas as regiões geográficas foi a mais equilibrada: isto é, aquela com maior número de controles (condicionalmente) independentes de *sld* na amostra ponderada por GPS-IPW, por apresentarem baixos níveis de confundimento. A eficiência da ponderação contribuiu para que os resultados da Figura 21 fossem mais robustos que os obtidos anteriormente – não só para a intensidade da escravidão, como para a maioria dos controles.

Em geral, nenhuma variável independente apresentou comportamento contraintuitivo. Apenas o estimador da renda média domiciliar *per capita* mudou de sentido, o que ocorreu nos Modelos 4 e 5, quando foram incluídos os termos de interação entre *ypc* e desigualdade. A inclusão da interação foi fundamental para melhorar o ajuste dos modelos (aumento do pseudo R² ajustado), reduzir o confundimento e produzir estimadores correlacionais representativos. Com a inclusão da interação, a desigualdade e a renda passaram a ser significantes a 1%, enquanto o desemprego passou de significativo a 1% para significativo a 10%.

O estimador da densidade populacional permaneceu positivo e significativo a 1% em todos os modelos. O fato de ter assumido o mesmo valor em todas as especificações atesta sua robustez. Somente na amostra referente a todo país, a educação básica apresentou comportamento condizente com a literatura empírica sobre violência, com correlação negativa em todos os modelos e significativo a 5% nos Modelos 1 a 3. Após a inclusão do termo de

interação entre desigualdade e *ypc*, os estimadores de *be* deixaram de ser. Tal fortalece a hipótese de que, dada a falta de robustez do estimador, o percentual de pessoas com oito anos de estudo completos não seja a variável mais adequada para representar a educação nos municípios. O percentual de homens jovens na população do município *ym* também só passou a ser significativo, robusto e a apresentar o comportamento esperado na amostra ponderada que reuniu todas as regiões geográficas.

A redução dos níveis de significância dos controles geográficos, latitude e longitude, sugerem que a incluir todas as variáveis binárias representantes dos efeitos fixos das UFs ajudou a controlar a autocorrelação espacial de forma mais eficaz – além de proporcionar mais um teste de robustez dos resultados.

Da Seção 4.2.1 até a Seção 4.2.3, a latitude apresentou altos níveis de significância (1%). Ainda que permaneça significativa em todos os modelos da Figura 21, a redução de significância após a inclusão de todos os controles de efeitos fixos mostra que os resultados anteriores “inflavam” o papel da latitude na análise empírica da violência letal. Como o controle dos efeitos fixos do espaço foi abordado mais precisamente nas estimações da Figura 21, latitude e longitude apresentaram comportamentos compatíveis com sua natureza. Fatores socioeconômicos, históricos, culturais e de outras naturezas que estavam sendo subrepresentados agora podem ser capturados de forma mais precisa pelas variáveis binárias de UF.

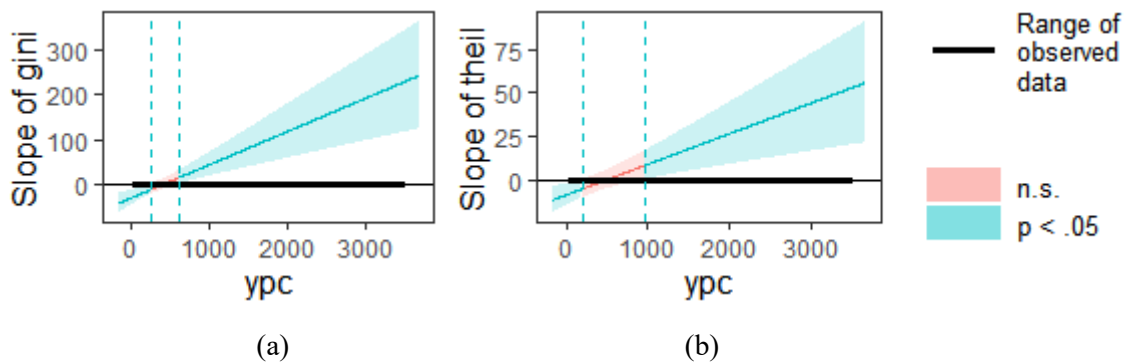
O índice municipal de justiça e segurança pública se destacou novamente, significativamente associado a menores taxas de homicídio em todo o território brasileiro observado. Caso seu coeficiente tivesse sido positivo, seria possível afirmar que municípios mais violentos apresentam maior infraestrutura de justiça e segurança pública, sugerindo que esta acompanha a “demanda” pelo controle da violência. Os resultados da Figura 21 mostram o contrário, em acordo com o observado nas amostras anteriores: nos municípios brasileiros, maior presença do poder público está significativamente associada a menores índices de violência letal.

Considerando a dificuldade de obtenção de dados oficiais, é notável que *cps* tenha sido tão consistente ao longo de todas as amostras observadas. Indicadores municipais de qualidade institucional são raros. A atualização daqueles já estabelecidos, como a MUNIC-2001, é esparsa, pois considera-se que os resultados demoram a variar significativamente. Este trabalho contribui com outros estudos que necessitem controlar a qualidade institucional ao criar o índice *cps* e testar sua representatividade (Seções 4.1 e 4.2), mas espera contribuir com o raciocínio

que se mostrou útil para criá-lo e pode ser estendido: a contagem de órgãos relacionados à justiça, segurança e gestão municipais. Apesar de sua simplicidade, o índice *cps* mostrou-se um preditor robusto da variação municipal em homicídios, indicando que dimensões institucionais muito básicas possuem correlações mensuráveis e significativas para o estudo da violência.

A Figura 21 mostrou que, até o Modelo 3, os índices de desigualdade e a renda média domiciliar *per capita* não foram significantes para a amostra ponderada de todas as regiões geográficas. Com a inclusão dos termos de interação, nos Modelos 4 e 5, ambos passaram a ser significantes a 1%. A interação entre desigualdade e renda se mostrou imprescindível para analisar as taxas de homicídios municipais em todas as regiões geográficas.

Figura 22 - Gráficos de Johnson-Neyman para o Brasil



A Figura 22 corrobora os resultados exibidos na Figura 21: a relação entre desigualdade de renda e *homr* variou ao longo do intervalo de valores de *ypc* observados na amostra (R\$ 54,18 a R\$ 3.468,20). Nos municípios mais pobres e mais ricos, essas mudanças são significativas para analisar a violência letal, o que não ocorre nos municípios com renda média. Nos primeiros, maior desigualdade está correlacionada a menores taxas de homicídio, e nos segundos, essa relação se torna direta, mais condizente com o que preconiza a literatura empírica (Seções 1 e 2.1).

Figura 23 – Descrição dos intervalos de Johnson-Neyman para o Brasil

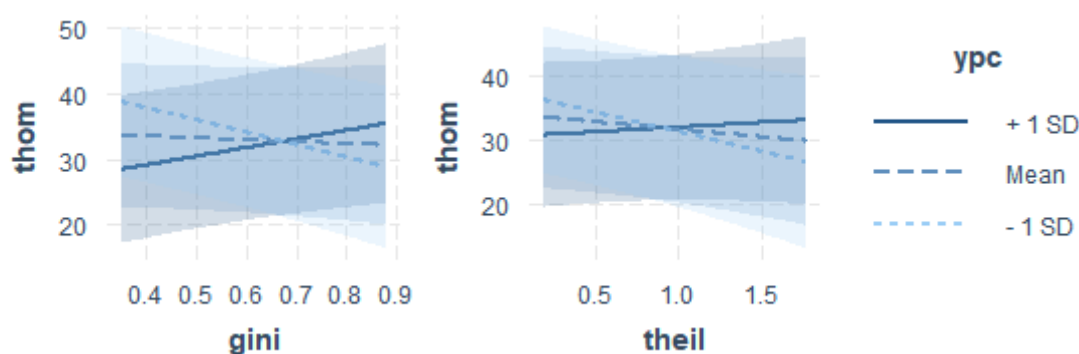
<i>ypc</i>	<i>gini</i>			<i>theil</i>			
	\hat{a}	2,5%	97,5%	<i>ypc</i>	\hat{a}	2,5%	97,5%
157,68	-18,93	-32,05	-5,81	157,59	-6,14	-11,72	-0,56
373,02	-2,86	-14,30	8,58	373,10	-2,33	-7,53	2,86
588,37	13,21	-1,23	27,65	588,60	1,47	-4,27	7,21

(a)

(b)

Fonte: elaboração própria com o software *RStudio*.

Figura 24 - Intervalos de Johnson-Neyman para o Brasil

Fonte: elaboração própria com o software *RStudio*.

A Figura 23 mostra que, na amostra ponderada de todo o Brasil, municípios mais pobres podem ser representados por uma renda média domiciliar *per capita* de aproximadamente R\$ 157, ilustrada pelas linhas tracejadas azul claras das Figuras 24a e 24b. As Figuras 22 e 23 mostram que, para essas faixas de renda, a relação entre desigualdade e *homr* assume valores negativos em todo o intervalo de confiança do seu estimador \hat{a} . A dinâmica é análoga à observada nas amostras do Sudeste e do Sudeste e Nordeste conjuntos. Isso colabora fortalece a hipótese de que a maior utilidade marginal da renda em municípios mais pobres colabora para explicar por que alto nível de desigualdade está associado a menores taxas de homicídios neste grupo.

Municípios brasileiros com renda relativa intermediária, representados por aproximadamente $ypc = R\$ 373$ na Figura 23, também apresentaram, em geral, correlação negativa entre desigualdade de renda e *homr*. A Figura 24 representa este grupo pelas linhas com tracejado maior e tom intermediário de azul: mais planas, indicam que essas correlações não são significativas.

Na amostra ponderada de todo o Brasil, municípios mais ricos podem ser representados por uma renda média domiciliar *per capita* de aproximadamente R\$ 588 (Figura 22). As Figuras

22 e 24 mostram que, para este grupo, a correlação entre desigualdade e *homr* é positiva e, em geral, significativa, comportamento condizente com a literatura empírica e também com a hipótese da utilidade marginal da renda como determinante do custo de oportunidade da violência. Na Figura 22b, que representa o Modelo 5, $ypc = R\$ 588$ pertence à área rosa, na qual a relação entre desigualdade de renda e taxa de homicídios não é significativa. Contudo, é possível observar que, para os municípios relativamente mais ricos em geral, esta relação é significativa, além de positiva ao longo da maior parte do intervalo de confiança.

O território nacional apresentou uma dinâmica similar ao Sudeste e às regiões Sudeste e Nordeste conjuntas. A relação entre desigualdade e violência letal varia de acordo com as faixas de renda dos municípios brasileiros. Estudos quantitativos da violência letal que deixem de considerar essas diferenças sofrerão com viés de variável omitida, acarretando alta variância entre os estimadores correlacionais e inviabilizando a obtenção de estimadores causais do efeito médio do tratamento, caso esse seja seu objetivo.

Por fim, analisa-se o estimador de *sld*. Como visto na Seção 4.1.4, a amostra que reúne todas as regiões brasileiras é a que apresentou os menores níveis de confundimento, medido através do número de controles com correlação de Pearson abaixo de 0,10 em relação ao tratamento. Contudo, todos os controles socioeconômicos apresentaram confundimento médio com *sld* mesmo após a ponderação. Este resultado sugere que a escravidão de fato influencia os indicadores socioeconômicos dos municípios brasileiros direta ou indiretamente atingidos pela sua existência. Para corrigir o confundimento, é necessário obter outros controles dos mesmos fatores socioeconômicos, porém capazes de apresentar maior exogeneidade em relação à escravidão.

O que a Figura 21 permite afirmar com segurança é que há correlação positiva, robusta e significativa a 5% entre o percentual de pessoas escravizadas nos municípios em 1872 e as taxas de homicídio municipais em 2000. Em todos os modelos da Figura 21, os valores do estimador se mantiveram próximos (variando entre 0,11 e 0,12), positivos e significantes.

A análise da região Sudeste tinha chegado a estimadores positivos e significantes a 10% que não resistiram aos testes de robustez empregados. Além disso, a amostra ponderada para o Sudeste não apresentou níveis de equilíbrio tão baixos como a amostra ponderada para todo o Brasil. Ainda que afirmar causalidade seja arriscado, é possível dizer que a correlação robusta, positiva e consistente entre *sld* e *homr* é uma evidência de que a escravidão está relacionada ao padrão contemporâneo de violência letal no país. Perseguir a obtenção de estimadores causais é um objetivo realista mesmo diante das limitações de dados sobre experimento natural

escravagista, sobre as diferentes faces da violência e sobre a qualidade institucional dos municípios brasileiros.

Dada a relevância dos resultados ilustrados pela Figura 21, é útil fazer um exercício de interpretação do estimador de *sld*. Os resultados indicam que um aumento de 1% na parcela de pessoas escravizadas na população de cada um dos municípios brasileiros em 1872 (tratados e não tratados) estaria associado a um aumento de, em média, 0,11 a 0,12 nas taxas de homicídios municipais em 2000. Se nos restringirmos aos 2.699 municípios que reportaram dados de vitimização letal em 2000, os óbitos por homicídio chegam a 45.433.

Como exercício interpretativo, consideremos o estimador de *sld* como causal por um momento. Devido à consistência dos seus resultados ao longo de todas as amostras, utiliza-se o Modelo 4 como especificação. Nesse caso, o aumento de 1% na parcela da população escravizada em todos os municípios brasileiros no ano de 1872 levaria a um aumento de aproximadamente 0,37% no número absoluto de homicídios dos 2.699 municípios analisados. Haveria, portanto, 168 vítimas a mais em relação ao total observado no país em 2000.

A amostra que reúne todos os municípios brasileiros levou a um aumento considerável do grupo controlado em relação ao aumento do grupo tratado, e foi a que obteve resultados mais robustos com menor nível de confundimento. Tal colabora para fortalecer a hipótese de que os efeitos da escravidão são melhor analisados sob a ótica do transbordamento da violência para além dos municípios onde o sistema escravagista se estabeleceu. As regiões nas quais a população escravizada absoluta era maior não apresentaram resultados tão robustos como estes, obtidos com a inclusão dos municípios das regiões Norte, Centro-Oeste e Sul na amostra ponderada.

Os resultados da Figura 21 se tornam ainda mais relevantes considerando que o fim do tráfico transatlântico de mão de obra escravizada ocorrera em 1850, e que a Abolição ocorreria em 1888. Como visto nas Seções 1 e 3, os dados do Censo Imperial de 1872 retratam a decadência do sistema escravagista (IBGE, 2012). Mesmo assim, os resultados obtidos com base nos seus dados foram positivos, significantes e robustos para todo o território nacional ao analisar a violência letal mais de um século depois. Tal fornece evidências de que a dificuldade para estimar os efeitos da escravidão sobre a violência se deve muito mais à disponibilidade de dados sobre suas diferentes fases do que à superação do seu legado pernicioso. É razoável inferir que as relações entre escravidão e violência seriam ainda maiores que as descritas pela Figura 21 para anos anteriores a 2000. Para estima-las, é fundamental contornar as limitações na

disponibilidade de dados sobre as variáveis analisadas, o que vem acontecendo no período recente.

Os resultados referentes a todos os municípios brasileiros que reportaram dados de vitimização por homicídio em 2000 fornecem evidências robustas de que a correlação entre escravidão e violência letal é significativa no Brasil. Por sua vez, esta é uma evidência a favor da hipótese de transbordamento da violência escravagista.

As conclusões de Franco (1983) ao analisar o cotidiano dos homens livres cultura dos homens livres do Vale do Paraíba após a Abolição (1888) e de Rosenthal (2018) ao analisar a relação entre violência e produtividade nas fazendas escravagistas do sul dos Estados Unidos também corroboram a hipótese de transbordamento. Ambas concluíram que a violência inerente às relações laborais e sociais estabelecidas pela escravidão sobrevive ao seu fim oficial. Uma evidência forte a favor da hipótese de normalização da violência escravagista é o fato de que Gastil (1971) e Nisbett e Cohen (1996) descrevem cenários muito semelhantes aos observados por Franco (1983) na sua análise do sul dos Estados Unidos – enquanto negam que a escravidão tenha contribuído para formá-los, sem, no entanto, fornecerem evidências empíricas que permitam descartar essa hipótese.

A subrepresentação da variável de tratamento, sua distância temporal da variável resultante, a dificuldade de encontrar controles adequados de qualidade institucional nos municípios, e dificuldade de obter dados municipais de violência (uma vez que apenas a taxa de homicídios é um indicador comum, reportado de forma recorrente pelo Ministério da Saúde (BRASIL, 2020)) não comprometeram a significância ou o sentido dos estimadores de *sld* exibidos na Figura 21. As estimativas foram significativas e robustas cento e doze anos depois da Abolição (1888), ultrapassando a Redemocratização (1985-1991), a proclamação da Constituição de 1988 e o sucesso do Plano Real (1994) no combate à inflação.

5 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho era encontrar correlações positivas, significativas e robustas entre o percentual de pessoas escravizadas nos municípios em 1872 (denominado intensidade da escravidão) e as taxas de homicídios municipais em 2000. Como mostrou a Seção 4.1.4, este objetivo foi atingido: controlando efeitos fixos por estado, há correlação positiva entre escravidão em 1872 e taxas de homicídios em 2000 para os municípios brasileiros.

Para a amostra que reuniu municípios de todas as regiões geográficas, a correlação entre escravidão em 1872 e taxa de homicídios em 2000 permaneceu positiva e significativa a 5%. Os modelos lineares ponderados empregados utilizaram controles considerados essenciais pela literatura empírica de Economia do Crime, como densidade populacional, desemprego, composição etária, *proxies* de renda e de educação básica, qualidade institucional, desigualdade de renda e fatores geográficos. Com isso, preenche-se a lacuna de estudos empíricos que relacionem a violência a fatores históricos e estruturais, como a escravidão no Brasil.

Como a literatura empírica frequentemente relaciona desigualdade de renda diretamente à escravidão e, ao mesmo tempo, prevê que a desigualdade está significativamente associada à violência, controlar a desigualdade foi fundamental para garantir a robustez dos resultados. Este trabalho observou que a dinâmica entre desigualdade de renda e taxa de homicídios varia entre os municípios brasileiros, e os modelos devem considerar essas diferenças para gerar estimadores representativos e correlações verdadeiras.

Observou-se que, em geral, nos municípios de maior renda média domiciliar *per capita*, maiores níveis de desigualdade de renda estão associados a maiores taxas de homicídio. Por outro lado, nos municípios com menor renda média domiciliar *per capita*, maior desigualdade está associada a menores taxas de homicídio, em média. A hipótese da utilidade marginal da renda colabora para explicar esses resultados – especialmente o último, que aparentemente contradiz a literatura tradicional sobre o tema. Sob essa hipótese, a utilidade marginal da renda é maior entre as populações mais pobres, as quais possuem mais incentivos a cometer atos violentos por falta de perspectiva (associados a abuso de álcool, falta de acesso a saúde mental, etc.) ou por interesse econômico.

Em municípios mais pobres nos quais a precariedade se distribui de forma igualitária entre a população, o aumento da desigualdade pode estar relacionado à redução da pobreza, o que, por sua vez, reduz os incentivos a cometer atos violentos. Sendo assim, a relação entre

desigualdade e violência letal é inversa nos municípios de menor renda média domiciliar *per capita*. Entretanto, isso não vale para municípios mais ricos, cuja relação entre desigualdade de renda e violência letal se mostrou direta, em acordo com o que a literatura empírica geralmente preconiza (SOARES, 2004; SOARES; NARITOMI, 2010).

Nos municípios mais ricos, a utilidade marginal da renda é menor. Caso sejam muito desiguais, o aumento da renda capaz de reduzir os incentivos à violência ocorrerá entre os mais pobres – o que reduzirá a desigualdade. Tal evidencia a relação direta entre desigualdade de renda e taxa de homicídios nos municípios mais ricos das amostras analisadas. Caso a renda seja alta, mas a desigualdade seja baixa, então o custo de oportunidade da violência provavelmente não está significativamente associado aos fatores econômicos, e a direção do efeito do aumento na renda sobre a violência não pode ser deduzida.

Os resultados obtidos demonstram a robustez da hipótese da utilidade marginal da renda para diferentes áreas analisadas, cujo tamanho da amostra aumentou progressivamente. Este estudo demonstrou que a modelagem das diferentes dinâmicas entre renda e desigualdade nos municípios brasileiros é fundamental para estudos empíricos da violência no país, seja para obter estimações correlacionais ou causais. Além disso, para garantir sua eficácia, políticas de segurança pública que visem atingir os custos de oportunidade da violência devem considerar tanto o nível de renda como o nível de desigualdade das regiões atingidas. A verificação empírica da hipótese da utilidade marginal da renda para explicar a relação entre violência letal, renda e desigualdade nos municípios brasileiros é uma das questões que este trabalho suscita.

Dada a escassez de dados periódicos sobre a qualidade da gestão pública e de segurança nos municípios brasileiros, este trabalho desenvolveu um indicador próprio: o índice municipal de justiça e segurança pública (*cps*), baseado nos dados da Pesquisa de Informações Municipais de 2001 (atual Perfil dos Municípios Brasileiros), edição dedicada à gestão pública municipal (IBGE, 2001). Para calculá-lo, verifica-se a existência de Guarda Municipal, Delegacia da Mulher, Defesa Civil, Tribunal de Pequenas Causas e Conselho Tutelar em cada município. Para cada órgão que estiver estabelecido, o município recebe o valor um; do contrário, recebe zero. O índice *cps* do município é dado pela soma desses valores e, portanto, assume valores discretos entre um e cinco.

O índice municipal de justiça e segurança pública foi o controle socioeconômico cujo comportamento mais se aproximou de variáveis como densidade populacional e latitude em consistência, significância e direção da correlação. O estimador de *cps* foi negativo para todas as áreas observadas, e significativa a 1% para as amostras de municípios do Sudeste, do Nordeste

e do Sudeste em conjunto, e de todas as regiões geográficas. Isso significa que municípios com mais órgãos públicos relacionados à justiça e à segurança pública – sejam eles voltados para a prevenção da violência, resolução pacífica de conflitos, proteção de vulneráveis, etc. – tendem a apresentar taxas de homicídios relativamente menores em 2000. A conclusão aparentemente óbvia colabora para fortalecer o papel da ação organizada do Estado diante da perda de legitimidade acarretada por políticas que eventualmente não atingem seus fins ou que visem, precisamente, diminuir o papel estatal no combate à violência.

A robustez e a significância da correlação inversa entre gestão municipal da segurança e violência letal estão de acordo com o que prevê a Lei nº 13.675/2018, que criou o Sistema Unificado de Segurança Pública (SUSP). O sistema busca integrar as esferas municipal, estadual e federal na criação de uma política de segurança pública coesa. Desde a Proclamação da República (1889), a segurança foi uma atribuição das Unidades Federativas – reforçada pela Constituição de 1946, que instituiu as Polícias Militares, sem que a Constituição de 1988 alterasse esse quadro.

Ao estabelecer o papel dos municípios na segurança, o SUSP promove uma inovação institucional cuja relevância encontra respaldo empírico neste trabalho: a correlação negativa, altamente significativa e robusta do índice *cps* foi obtida para o ano 2000, dezoito anos antes da criação desse sistema, em 2018. Atualmente, passados oito anos da sua fundação, o engajamento dos municípios ainda é relativamente baixo devido à visão cristalizada de que a segurança é atribuição das UFs. Este trabalho forneceu uma evidência empírica de que a infraestrutura municipal de segurança e justiça, por mais diversa que seja, está significativamente correlacionada a menores níveis de violência letal. Com isso, espera-se colaborar para a articulação e a legitimação da política de segurança abrangente e coesa instituída pelo SUSP. Além disso, destaca-se a importância da literatura de avaliação de impacto de políticas e órgãos municipais ou regionais na segurança pública.

A obtenção de indicadores de “gestão pública” ou “qualidade institucional” enfrenta diversas dificuldades, desde a concepção de indicadores de algo, a princípio, tão abstrato, até a coleta de dados sobre um tema relativamente sensível, sujeito a interesses eleitoreiros. Este trabalho contribuiu para a literatura empírica sobre violência e sobre gestão pública ao idealizar e testar o índice *cps*, demonstrando sua robustez estatística e conceitual. A adequação do índice *cps* mostra que soluções simples – como a criação de um *ranking* baseado no número de órgãos relacionados à justiça e segurança num município – podem atender às necessidades de representatividade de fatores socioeconômicos e correção estatística de diversos estudos. No

caso de fatores relativamente abstratos como a “qualidade institucional”, ter em mente este tipo de recurso pode ajudar a conceber novos indicadores e a preencher lacunas semelhantes.

Por fim, a principal contribuição deste trabalho – cujo objetivo não é obter estimadores causais – é com a literatura de causalidade de tratamentos contínuos, especialmente quando a amostra é não aleatória ou desequilibrada. As características do evento histórico analisado – a escravidão brasileira em 1872 – exigiram o emprego de métodos tradicionalmente usados para obter estimadores causais.

A escravidão brasileira foi um experimento natural cuja distribuição dos grupos de tratamento e controle não foi aleatória: o grupo de municípios que recebeu o tratamento – a escravidão – se concentrou no litoral das regiões Sudeste e Nordeste porque isso minimizava os custos de colonização e os custos de transporte do lucrativo tráfico transatlântico. Portanto, pertencer ao grupo de tratamento significa que um município foi ocupado há mais tempo, o que levou ao estabelecimento precoce de pessoas e de atividades econômicas em relação ao grupo de controle. É razoável supor que esses municípios têm maiores valores observados de densidade populacional, renda média domiciliar *per capita*, apresentam longitudes mais altas, etc. A alta correlação entre a variável que representa o experimento natural (a intensidade da escravidão em 1872, também chamada de tratamento) e os controles necessários para analisar a taxa de homicídios em 2000 é denominada confundimento.

Sob alto confundimento, estimar regressões multivariadas pelo método de mínimos quadrados ordinários comum levaria a correlações espúrias, estimadores viesados da relação entre escravidão em 1872 e violência letal em 2000. Isso se deve ao fato de que, neste método, o confundimento só pode ser atenuado pela inclusão de mais controles. Contudo, todos os controles tradicionalmente utilizados pela literatura empírica sobre violência (GOUDA; RIGTERINK, 2017; BUONNANO; VARGAS, 2019) já estavam incluídos. Seria necessário incluir variáveis indefinidamente, sem a certeza de que o confundimento se reduziria.

Outro problema é que o método de mínimos quadrados assume que as relações entre as variáveis analisadas são lineares, o que é bastante razoável quando se referem a um mesmo período. Contudo, a variável de interesse deste estudo, a intensidade da escravidão, se refere a 1872. Existem outras fontes de dados sobre a experiência escravagista, mas o Censo Imperial de 1872 é a única com abrangência nacional, cuja distribuição dos dados corresponde à divisão territorial dos municípios brasileiros no período recente. Portanto, utilizá-la era imprescindível para obter uma amostra municipal que abrangesse todo o território brasileiro.

Nessas condições, não é razoável assumir que a relação entre a intensidade da escravidão em 1872 e o cenário socioeconômico dos municípios em 2000 seja linear, dada a complexidade dos processos históricos, políticos e econômicos que mediarão essa longa trajetória. Além disso, como o Censo Imperial de 1872 foi realizado após a proibição do tráfico transatlântico (1850) e dezesseis anos antes da Abolição (1888), período de decadência do sistema escravagista, os dados tendem a subrepresentar a magnitude da escravidão na história e na sociedade brasileiras. Isso dificulta ainda mais o estabelecimento de relações claras, que dirá lineares, entre a intensidade da escravidão e as outras variáveis independentes.

O uso de mínimos quadrados ordinários não levaria a estimadores não viesados da intensidade da escravidão em 1872 sobre a taxa de homicídios em 2000 porque o confundimento entre a primeira e os controles é alto, dificultando a obtenção de estimadores realistas de todas as variáveis independentes, e também porque a premissa de linearidade impede que a complexidade da relação entre um evento histórico e o cenário contemporâneo seja adequadamente representada, colaborando ainda mais para o já alto nível de confundimento.

Era necessário empregar técnicas econométricas que mitigassem o confundimento e representassem adequadamente as relações não lineares entre as variáveis independentes. Para fazê-lo, é impossível ignorar que a variável de tratamento que representa o experimento natural da escravidão não é binária ou discreta, mas contínua: os municípios ocupados apresentavam diferentes níveis de população escravizada entre a sua população total em 1872. Tratá-la como discreta levaria à perda de informação (sobre um evento já subrepresentado, como explicou o parágrafo anterior).

As estimações deveriam mitigar o confundimento e representar adequadamente a não-linearidade entre as variáveis independentes sem desconsiderar o caráter contínuo da variável de tratamento. Estimadores correlacionais robustos só poderiam ser gerados por métodos que abordassem as peculiaridades da amostra observada e da variável de interesse. A solução foi definir formalmente e empregar o método sintetizado como GPS-IPW, devido à maior popularidade da sua nomenclatura em inglês: *Generalized Propensity Scores* utilizados para *Inverse Probability Weighting*.

A princípio, calcula-se a propensão de cada unidade amostral pertencer ao tratamento com base nos valores observados das variáveis independentes – os *scores* de propensão. Para fazê-lo, Hirano e Imbens (2004) elaboraram uma generalização da técnica original criada por Rosenbaum e Rubin (1983) para tratamentos binários. Como se baseiam em tratamentos

contínuos, os *scores* calculados e empregados neste estudo são chamados de generalizados. Feito isso, as unidades amostrais são ponderadas pelo inverso dos *scores* de propensão generalizados. Caso os pesos tenham sido calculados corretamente e representem as relações entre a variável de tratamento e os controles de forma adequada, a amostra ponderada deve apresentar níveis menores de confundimento.

Esse método é tradicionalmente empregado para obter estimadores causais, que exigem que o nível de confundimento seja nulo ou suficientemente baixo para serem válidos (ZHU; COFFMAN; GHOSH, 2015). No caso deste estudo, o alto nível de confundimento geraria estimadores correlacionais viesados se não fosse abordado, e mitigá-lo exige reconhecer o caráter contínuo da variável de tratamento. Além disso, para calcular pesos corretos, era imprescindível reconhecer as relações não lineares entre as variáveis de tratamento – o que pode ser feito através de algoritmos de aprendizado de máquina, como o *Generalized Boosting Model* (GBM). O cálculo de pesos por GBM como alternativa para representar não-linearidade entre variáveis independentes, preservar o tamanho da amostra original e calcular pesos acurados é outra contribuição deste trabalho para a literatura de causalidade de experimentos naturais históricos.

Algoritmos de aprendizado de máquina têm o objetivo de realizar previsões corretamente, o que é especialmente útil ao calcular os *scores* generalizados que medem a probabilidade de um município pertencer ao tratamento. O algoritmo GBM se baseia em árvores de regressão que particionam a amostra de acordo com os valores observados dos controles, com o objetivo de minimizar o erro de previsão dos valores do tratamento dados os valores do controle em cada árvore. Ao final, o algoritmo realiza combinações lineares das árvores de regressão que tiveram os menores erros de previsão de T para gerar uma função mais suave, utilizada para estimar os *scores*. A função final minimiza as diferenças entre os grupos, e é definida como aquela que minimiza a correlação absoluta média (de Pearson) entre a variável de tratamento contínua e cada um dos controles.

Como seu objetivo é minimizar o erro de previsão da probabilidade de as unidades amostrais pertencerem a um ou outro grupo e não presume linearidade entre o tratamento contínuo e os controles, o algoritmo GBM foi fundamental para chegar aos resultados robustos exibidos nas Seções 4.1.4 e 4.2.4, que representam o atingimento do objetivo deste trabalho. Outra vantagem é o fato de que o algoritmo GBM consegue identificar e incluir interações entre as variáveis e termos de maior ordem automaticamente (ZHOU, 2015), evitando que interações ocultas levem à persistência do confundimento. Tal é particularmente importante porque os

controles socioeconômicos podem ter diversas interações entre si: por exemplo, é razoável supor que há relação entre densidade populacional, desemprego e renda nos municípios brasileiros. Incluí-las indiscriminadamente é inviável (além de levar à maldição da dimensionalidade), mas omiti-las pode causar a persistência do confundimento na amostra ponderada, já que se forem significativas, distorcerão os pesos aplicados às amostras.

O emprego do método GPS-IPW, tradicionalmente empregado em estudos causais (ainda raros) e do algoritmo GBM permitiu a abordagem do confundimento reconhecendo a não-linearidade das relações entre as variáveis independentes e o caráter contínuo da variável de tratamento. A Seção 4.1 demonstrou que os níveis de confundimento se reduziram significativamente à medida que as amostras ponderadas aumentavam e que os modelos se tornavam mais representativos, com maior controle de efeitos fixos por UF.

Segundo Zhu, Coffman e Ghosh (2015), amostras ponderadas cuja média de confundimento entre todas os controles e o tratamento for menor que 0,10 (em valor absoluto da correlação de Pearson) permitem a obtenção de estimadores causais do efeito do tratamento. Este critério foi atingido para a Seção 4.2.4 durante os testes dos modelos, mas opta-se por desconsiderá-lo: não é razoável interpretar os estimadores da intensidade da escravidão, principalmente, porque as regressões da Seção 4.2 se referem apenas ao ano de 2000.

O fato de haver confundimento residual nas amostras ponderadas também impede a interpretação causal, mas é secundário em relação ao fato de que este trabalho optou por fazer uma análise transversal (*cross-section*) única (ano 2000) devido à indisponibilidade de dados longitudinais sobre qualidade institucional municipal. Também se optou por não utilizar uma métrica agregada de confundimento, como fazem os autores, dando preferência à análise de confundimento de cada controle nas amostras ponderadas. Isso permite identificar as principais fontes de confundimento e assim determinar estratégias mais eficientes, com embasamento substantivo (não apenas estatístico ou computacional) para mitigá-lo.

Dados os baixos níveis de confundimento da amostra ponderada de todos os municípios brasileiros que apresentaram dados completos para todas as variáveis em 2000, os estimadores obtidos pelos modelos de regressão ponderada descritos na Seção 4.2.4 são associações robustas e ajustadas, significativamente mais confiáveis do que aquelas derivadas de modelos lineares não ponderados. Os valores positivos e consistentemente significativos a 5% indicam que há correlação entre o percentual de pessoas escravizadas nos municípios brasileiros em 1872 e os níveis de taxa de homicídio municipal em 2000. Como o grupo controlado é muito maior que o grupo diretamente afetado pela escravidão, esse resultado fortalece a hipótese de que a

influência da escravidão sobre a violência ultrapassa as áreas afetadas e o fim do sistema escravagista.

A principal contribuição deste trabalho é empírica e metodológica. Do ponto de vista empírico, demonstra-se pela primeira vez na literatura brasileira uma associação positiva, estatisticamente significativa e robusta entre a intensidade da escravidão no século XIX e os níveis de violência letal no Brasil contemporâneo, controlando fatores socioeconômicos, institucionais e geográficos. Essa evidência preenche uma lacuna importante ao conectar legados históricos estruturais com desafios contemporâneos na segurança pública, reforçando a relevância de políticas que enfrentem desigualdades e fragilidades institucionais herdadas do passado.

Do ponto de vista metodológico, este estudo inova ao aplicar GPS-IPW com GBM em um contexto até então inexplorado na literatura nacional: a análise de legados históricos como tratamento contínuo não aleatório. Ao fazê-lo, demonstra como métodos de inferência causal podem ser adaptados para estudos observacionais complexos, mesmo quando o objetivo não é obter efeitos causais puros, mas sim associações ajustadas e interpretáveis. Essa abordagem pode servir de referência para futuras pesquisas que busquem estimar relações robustas entre eventos históricos e desfechos contemporâneos em contextos de amostras desequilibradas e confundimento observável.

A trajetória para obter essas contribuições suscita várias questões, ressalta diversas lacunas e aponta para muitos obstáculos a serem superados no estudo empírico do legado da escravidão e dos fatores históricos da violência no Brasil. Em relação à variável de tratamento, transportar a definição de Gouda e Rigterink (2017) para a realidade brasileira visava justamente testar a representatividade da sua forma de mensurar a escravidão. Os resultados mostram que definir a intensidade da escravidão de outra forma – por exemplo, número de pessoas escravizadas em 1872 por quilômetro quadrado do município sob a divisão territorial contemporânea – pode ser mais adequado para obter estimadores causais.

Para obter estimadores causais robustos, é útil incluir anos anteriores a 2000 nas análises. O principal limitante de fazê-lo é a obtenção de indicadores de qualidade institucional suficientemente representativos. Anos mais próximos a 1872 estão mais próximos do evento histórico analisado e, supostamente, são mais vulneráveis ao seu efeito direto. Contudo, na ausência de ações diretamente corretivas do legado da escravidão, anos mais recentes para os quais haja dados de qualidade institucional municipal podem ser úteis. Ampliando os cortes

transversais observados, é possível utilizar técnicas de dados em painel para obter estimadores causais do efeito médio da escravidão.

Incorporar outros controles institucionais, históricos e culturais pode contribuir significativamente para reduzir o confundimento e adequar ainda mais as amostras originais. Algumas *proxies* empregadas neste trabalho, como a de educação básica, podem ser substituídas por outras que eventualmente sejam mais representativas dos fatores socioeconômicos e mais exógenas em relação à escravidão.

A manutenção da coleta de dados municipais sobre outros tipos de crimes é fundamental, tanto para a formulação de políticas de segurança pública eficientes, como para explorar os impactos dos fatores históricos sobre os padrões de violência regionais e nacionais. Em relação à taxa de homicídios, analisar cortes transversais mais recentes pode ajudar a identificar se a correlação entre escravidão e violência letal se altera ao longo do tempo.

Em relação às contribuições paralelas deste estudo, a efetividade do índice *cps* pode ser testada em outros estudos observacionais relacionados à qualidade institucional. Além disso, a relação condicional entre renda e desigualdade observada neste trabalho demanda aprofundamento teórico e empírico, possivelmente por meio de estudos experimentais ou quase-experimentais que testem diretamente a hipótese da utilidade marginal da renda no contexto da violência brasileira.

REFERÊNCIAS

- ACEMOGLU, Daron; JOHNSON, Simon; ROBINSON, James A. Reversal of Fortune: Geography and Institutions in the Making of Modern World Income Distribution. **Quarterly Journal of Economics**, v. 117, n. 4, p. 1231-1294, 2002. Disponível em: <https://economics.mit.edu/sites/default/files/publications/reversal-of-fortune.pdf>. Acesso em: out. 2022.
- ACEMOGLU, Daron; GARCÍA-JIMENO, Camilo; ROBINSON, James A. Finding Eldorado: Slavery and Long-Run Development in Colombia. **Journal of Comparative Economics**, v. 40, n. 4, p. 534-564, nov. 2012. Disponível em: <https://economics.mit.edu/sites/default/files/publications/Finding%20Eldorado%20-%20Slavery%20and%20Long-run%20Developmen.pdf>. Acesso em: out. 2019.
- ACHARYA, Avidit; BLACKWELL, Matthew; SEN, Maya. The political legacy of American slavery. **The Journal of Politics**, v. 78, n. 3, p. 621-641, jul. 2016. Disponível em: <https://scholar.harvard.edu/files/msen/files/slavery.pdf>. Acesso em: out. 2019.
- ADORNO, Sérgio. **A gestão urbana do medo e da insegurança: violência, crime e justiça penal na sociedade brasileira contemporânea**. 1996. 281 f. Tese (Livre-Docência em Sociologia) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1996. Disponível em: <https://nev.prp.usp.br/wp-content/uploads/2015/01/down187.pdf>. Acesso em: out. 2019.
- AMORIM, Carlos. **Comando Vermelho: a história secreta do crime organizado**. Rio de Janeiro: Record, 1993.
- AMORIM, Carlos. **CV-PCC: a irmandade do crime**. 2. ed. Rio de Janeiro: Record, 2003.
- ARCAND, Jean-Louis; D'HOMBRES, Béatrice. Racial discrimination in the Brazilian labour market: wage, employment and segregation effects. **Journal of International Development**, v. 16, n. 8, p. 1053-1066, nov. 2004. Disponível em: <https://doi.org/10.1002/jid.1116>. Acesso em: jul. 2021.
- BARROS, Laura; SANTOS SILVA, Manuel. **Between sticky floors and glass ceilings: the effect of trade liberalization on double discrimination in Brazil**. Göttingen: Georg-August-Universität Göttingen, Ibero-America Institute for Economic Research, 2020. (Ibero-America Institute Discussion Papers, n. 249). Disponível em: <http://hdl.handle.net/10419/226343>. Acesso em: s
- BERTOCCHI, Graziella; DIMICO, Arcangelo. The Racial Gap in Education and the Legacy of Slavery. **Journal of Comparative Economics**, v. 40, n. 4, p. 581-595, 2012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0147596712000261>. Acesso em: out. 2019.
- BERTOCCHI, Graziella; DIMICO, Arcangelo. Slavery, Education, and Inequality. **European Economic Review**, v. 70, p. 197-209, 2014. Disponível em: <https://pure.qub.ac.uk/en/publications/slavery-education-and-inequality>. Acesso em: out. 2019.

BOTELHO, Fernando; MADEIRA, Ricardo A.; RANGEL, Marcos A. Racial discrimination in grading: evidence from Brazil. **American Economic Journal: Applied Economics**, v. 7, n. 4, p. 37-52, out. 2015. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/app.20140352>. Acesso em: jul. 2021.

BRASIL. Lei nº 13.675, de 11 de junho de 2018. **Disciplina a organização e o funcionamento dos órgãos responsáveis pela segurança pública, nos termos do § 7º do art. 144 da Constituição Federal; cria a Política Nacional de Segurança Pública e Defesa Social (PNSPDS); institui o Sistema Único de Segurança Pública (Susp); altera a Lei Complementar nº 79, de 7 de janeiro de 1994, a Lei nº 10.201, de 14 de fevereiro de 2001, e a Lei nº 11.530, de 24 de outubro de 2007; e revoga dispositivos da Lei nº 12.681, de 4 de julho de 2012.** Diário Oficial da União: seção 1, Brasília, DF, p. 4-8, 12 jun. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13675.htm. Acesso em: out. 2023.

BRASIL. **Ministério da Saúde.** Sistema de Informações sobre Mortalidade. Brasília, DF: Ministério da Saúde, 2020. Disponível em: <http://tabnet.datasus.gov.br/>. Acesso em: out. 2020.

BUENO, Samira. **Bandido bom é bandido morto:** a opção ideológico-institucional da política de segurança pública na manutenção de padrões de atuação violentos da polícia militar paulista. 2014. 145 f. Dissertação (Mestrado em Administração Pública e Governo) – Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2014. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/11990>. Acesso em: out. 2019.

BUONANNO, Paolo; VARGAS, Juan F. Inequality, crime, and the long-run legacy of slavery. **Journal of Economic Behavior and Organization**, v. 159, p. 539-552, 2019. DOI: 10.1016/j.jebo.2017.12.022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167268117303724>. Acesso em: out. 2021.

CALDEIRA, Teresa Pires do Rio. **Cidade de muros:** crime, segregação e cidadania em São Paulo. 1. ed. São Paulo: Ed. 34/Edusp, 2000.

CANO, Ignacio; DUARTE, Thais. **No sapatinho:** a evolução das milícias no Rio de Janeiro (2008-2011). Rio de Janeiro: Fundação Heinrich Böll, 2012. Disponível em: https://br.boell.org/sites/default/files/no_sapatinho_lav_hbs1_1.pdf. Acesso em: out.2019.

CARDEN, Art. **Slavery, Violence, and Law in the Nineteenth-Century South.** Rochester: Social Science Research Network (SSRN), 2008. (SSRN Working Paper). Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1319484. Acesso em: out. 2019.

CERQUEIRA, Daniel Ricardo de Castro. **Causas e consequências do crime no Brasil.** 2010. Tese (Doutorado em Economia) – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010. Disponível em: https://web.bndes.gov.br/bib/jspui/bitstream/1408/1922/2/Concurso0212_33_premiobndes_Doutorado_P.pdf. Acesso em: out. 2019.

COOK, Lisa D. Violence and economic activity: evidence from African American patents, 1870–1940. **Journal of Economic Growth**, v. 19, n. 2, p. 221-257, 2014. Disponível em: <https://doi.org/10.1007/s10887-014-9102-z>. Acesso em: jul. 2021.

COOK, Lisa D.; LOGAN, Trevon D.; PARMAN, John M. Rural segregation and racial violence: historical effects of spatial racism. **American Journal of Economics and Sociology**, v. 77, n. 3-4, p. 821-847, 2018. Disponível em: <https://doi.org/10.1111/ajes.12232>. Acesso em: jul. 2021.

ELTIS, David; BEHRENDT, Stephen D.; DOMINGUES DA SILVA, Daniel B.; KLEIN, Herbert S.; LACHANCE, Paul; MISEVICH, Philip; RICHARDSON, David. **Trans-Atlantic Slave Trade Database**. Emory University, 2021. Disponível em: <https://www.slavevoyages.org>. Acesso em: out. 2019.

ENGERMAN, Stanley L.; SOKOLOFF, Kenneth L. Factor endowments, institutions, and differential paths of growth among New World economies: a view from economic historians of the United States. In: HABER, Stephen (Ed.). **How Latin America fell behind: essays on the economic histories of Brazil and Mexico, 1800-1914**. Stanford: Stanford University Press, 1997. p. 260-304. Disponível em: <https://www.nber.org/papers/h0066>. Acesso em: out. 2022.

ENGERMAN, Stanley L.; SOKOLOFF, Kenneth L. Factor endowments, inequality and paths of development among New World economies. **Economia**, v. 3, n. 1, p. 41-109, 2002. Disponível em: <https://www.nber.org/papers/w9259>. Acesso em: out. 2022.

FERGUSON, Leopoldo; VARGAS, Juan F.; VELA, Mauricio A. **Judicial efficiency index for Colombian municipalities**. 2013. Documento de trabajo – CEDE, Universidad de los Andes, Bogotá, 2013.

FERNANDES, Florestan. **A integração do negro na sociedade de classes: o legado da "raça branca"**. São Paulo: Dominus; Ed. Universidade de São Paulo, 1965. v. 1. Disponível em: <https://professorrenato.com/wp-content/uploads/2024/12/Florestan-Fernandes-A-integracao-do-negro-na-sociedade-de-classes-Vol-I-O-legado-da-raca-branca-1.pdf>. Acesso em: out. 2023.

FERNANDES, Florestan. **O negro no mundo dos brancos**. São Paulo: Difusão Europeia do Livro, 1972. Disponível em: <https://eraju2013.files.wordpress.com/2013/09/fernandes-florestan-o-negro-no-mundo-dos-brancos-1.pdf>. Acesso em: out. 2023.

FERNANDES, Florestan. **A integração do negro na sociedade de classes: o legado da "raça branca"**. São Paulo: Dominus; Ed. Universidade de São Paulo, 1965. v. 1.

FERNANDES, Florestan. **A integração do negro na sociedade de classes: no limiar de uma nova era**. 3. ed. São Paulo: Ática, 1978. v. 2.

FERNANDES, Gustavo Andrey de Almeida Lopes. Brazilian female labor market: racial-skin color discrimination and inefficiency. **Economia Aplicada**, São Paulo, v. 19, n. 2, p. 241-259, abr./jun. 2015. Disponível em: <https://doi.org/10.1590/1413-8050/ea106033>. Acesso em: jul. 2021.

FESTINGER, Leon. **A theory of cognitive dissonance**. Stanford: Stanford University Press, 1957.

FIUZA-MOURA, Flavio Kaue; MAIA, Katy; SOUZA, Solange de Cassia Inforzato de; GOMES, Magno Rogério; MOURÃO, Paulo Reis. The luck of being of the right gender and color: a detailed discussion about the wage gaps in the Brazilian manufacturing industry. **Quality & Quantity**, v. 53, n. 3, p. 1275-1300, maio 2019. Disponível em: <https://doi.org/10.1007/s11135-018-0717-6>. Acesso em: mai. 2019.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Anuário Brasileiro de Segurança Pública 2007**. São Paulo: FBSP, 2007. Disponível em: https://forumseguranca.org.br/wp-content/uploads/2017/12/ANUARIO_2007.pdf. Acesso em: jul. 2023.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Anuário Brasileiro de Segurança Pública 2023**. São Paulo: FBSP, 2023. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2023/07/anuario-2023.pdf>. Acesso em: jul. 2023.

FRANCO, Maria Sylvia de Carvalho. **Homens livres na ordem escravocrata**. 3. ed. São Paulo: Kairós, 1983.

FREYRE, Gilberto. **Casa-grande & senzala: formação da família brasileira sob o regime da economia patriarcal**. 16. ed. Rio de Janeiro: José Olympio, 1973.

FUJIWARA, Thomas; LAUDARES, Humberto; VALENCIA-CAICEDO, Felipe. **Tordesillas, Slavery and the Origins of Brazilian Inequality**. Barcelona: Barcelona School of Economics, 2017. (Working Paper). Disponível em: <https://events.bse.eu/live/files/1869-felipe-valencia-52600pdf>. Acesso em: out. 2019.

GASTIL, Raymond D. Homicide and a regional culture of violence. **American Sociological Review**, v. 36, n. 3, p. 412-427, 1971. Disponível em: <https://www.jstor.org/stable/2093082>. Acesso em: out. 2020.

GERARD, François; LAGOS, Lorenzo; SEVERNINI, Edson; CARD, David. Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil. **American Economic Review**, v. 111, n. 10, p. 3418-3457, out. 2021. Disponível em: <https://doi.org/10.1257/aer.20181596>. Acesso em: set. 2023.

GOUDA, Moamen; RIGTERINK, Anouk. **The Long-Term Effect of Slavery on Violent Crime: Evidence from US Counties**. Rochester: Social Science Research Network (SSRN), 2017. (SSRN Working Paper). Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2358389. Acesso em: out. 2019.

GRAHAM, Daniel J.; MCCOY, Emma J.; STEPHENS, David A. **Doubly robust dose-response estimation for continuous treatments via generalized propensity score augmented outcome regression**. arXiv, 16 jun. 2015. arXiv:1506.04991. Disponível em: <https://arxiv.org/pdf/1506.04991.pdf>. Acesso em: ago. 2021.

GREIFER, Noah. **Using WeightIt to estimate balancing weights**. CRAN, 2021. Vignette do pacote WeightIt para R. Disponível em: <https://cran.r-project.org/web/packages/WeightIt/vignettes/WeightIt.html>. Acesso em: maio 2021.

HEISS, Andrew. **Generating inverse probability weights for both binary and continuous treatments**. [S. l.], 1 dez. 2020. DOI: 10.59350/1svkc-rkv91. Disponível em: <https://www.andrewheiss.com/blog/2020/12/01/ipw-binary-continuous/>. Acesso em: set. 2021.

HIRANO, Keisuke; IMBENS, Guido W. The propensity score with continuous treatments. In: GELMAN, Andrew; MENG, Xiao-Li (Ed.). **Applied Bayesian modeling and causal inference from incomplete-data perspectives**. Chichester: Wiley, 2004. p. 73-84. Disponível em: <https://www.math.mcgill.ca/dstephens/PSMMA/Articles/Hirano-Imbens-2004.pdf>. Acesso em: out. 2019.

HOLANDA, Sérgio Buarque de. **Raízes do Brasil**. 18. ed. Rio de Janeiro: José Olympio, 1984.

HULING, Jared D.; GREIFER, Noah; CHEN, Guanhua. Independence weights for causal inference with continuous exposures. arXiv, 15 jul. 2021. arXiv:2107.07086. Disponível em: <https://arxiv.org/abs/2107.07086>. Acesso em: ago. 2021.

IBGE. **Censo demográfico 2000**. Rio de Janeiro: IBGE, 2003. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9663-censo-demografico-2000.html>. Acesso em: mar. 2020.

IBGE. **Perfil dos municípios brasileiros: gestão pública 2001**. Rio de Janeiro: IBGE, 2003. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=283993>. Acesso em: mar. 2020.

IBGE. **Divisão territorial dos municípios, 2010**. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial.html>. Acesso em: mar. 2020.

IBGE. **Evolução da divisão territorial do Brasil 1872-2010**. Rio de Janeiro: IBGE, 2011. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=284481>. Acesso em: mar. 2020.

IBGE. **Recenseamento Geral do Brasil 1872**. Rio de Janeiro: IBGE, 2012. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv25477_v1_br.pdf. Acesso em: out. 2019.

IBGE. **Características étnico-raciais da população: classificações e identidades**. Rio de Janeiro: IBGE, 2013. 208 p. (Estudos e pesquisas. Informação demográfica e socioeconômica, n. 27). Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=284235>. Acesso em: jul. 2021.

IBGE. **Desigualdades sociais por cor ou raça no Brasil**. 2. ed. Rio de Janeiro: IBGE, 2022. 16 p. (Estudos e pesquisas. Informação demográfica e socioeconômica, n. 48). Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101972_informativo.pdf. Acesso em: jul. 2023.

INSTITUTE FOR ECONOMICS & PEACE. **United States Peace Index 2011**. Sydney: IEP, 2011. Disponível em: <https://www.files.ethz.ch/isn/128302/U.S.-Peace-Index-2011-3.pdf>. Acesso em: out. 2019.

INSTITUTE FOR ECONOMICS & PEACE. **Global Peace Index 2022: measuring peace in a complex world**. Sydney: IEP, 2022. Disponível em: <https://www.economicsandpeace.org/wp-content/uploads/2022/06/GPI-2022-web.pdf>. Acesso em: jul. 2023.

INSTITUTE FOR ECONOMICS & PEACE. **Global Peace Index 2023: measuring peace in a complex world**. Sydney: IEP, 2023. Disponível em: <https://www.visionofhumanity.org/wp-content/uploads/2023/06/GPI-2023-Web.pdf>. Acesso em: jul. 2023.

JUNG, Yeonha. **How the legacy of slavery survives: labor market institutions and demand for human capital**. 2019. Tese (Doutorado em Economia) – Boston University, Boston, 2019. Disponível em: <https://eh.net/eha/wp-content/uploads/2019/06/Jung.pdf>. Acesso em: out. 2020.

JUSTUS, Marcelo; KAHN, Túlio; CERQUEIRA, Daniel. The "São Paulo Mystery": the role of the criminal organization PCC in reducing the homicide in 2000s. In: ENCONTRO NACIONAL DE ECONOMIA, 45., 2017, Natal. **Anais [...]**. Natal: ANPEC, 2017. Disponível em: https://www.anpec.org.br/encontro/2017/submissao/files_I/i12-c155d0a7ce1f80953f02485b91789210.pdf. Acesso em: out. 2019.

LAMBAIS, Guilherme Berse Rodrigues. **Slave resistance, cultural transmission, and Brazil's long-run economic development**. 2020. Tese (Doutorado em Economia) – Departamento de Economia, Universidade de Brasília, Brasília, 2020. Disponível em: https://gbrlambais.github.io/resistance_longrun.pdf. Acesso em: jul. 2021.

LEUNG-GAGNÉ, Josh. Classroom segregation without tracking: chance, legitimacy, and myth in "racial paradise". **AERA Open**, v. 9, p. 1-23, 2023. Disponível em: <https://doi.org/10.1177/23328584231176456>. Acesso em: set. 2023.

LIMA, Renato Sérgio de; BUENO, Samira; ALCADIPANI, Rafael. Evolução das mortes violentas intencionais no Brasil. In: FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Anuário Brasileiro de Segurança Pública 2021**. São Paulo: FBSP, 2021. p. 21-33. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2021/07/1-evolucao-das-mortes-violentas-intencionais-no-brasil.pdf>. Acesso em: ago. 2021.

LOGAN, Trevon D.; PARMAN, John M. Segregation and homeownership in the early twentieth century. **American Economic Review**, v. 107, n. 5, p. 410-414, 2017. Disponível em: <https://doi.org/10.1257/aer.p20171081>. Acesso em: jul. 2021.

LONG, Jacob A. **jtools**: Analysis and Presentation of Social Scientific Data. R package version 2.2.2, 2022. Disponível em: <https://cran.r-project.org/package=jtools>. Acesso em: out. 2021.

LONG, Jacob A. **interactions**: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.5, 2021. Disponível em: <https://cran.r-project.org/package=interactions>. Acesso em: out. 2021.

MCCAFFREY, Daniel F.; RIDGEWAY, Greg; MORRAL, Andrew R. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. **Psychological Methods**, v. 9, n. 4, p. 403-425, 2004. DOI: 10.1037/1082-989X.9.4.403. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/15598095/>. Acesso em: ago. 2021.

MALONEY, William F.; VALENCIA-CAICEDO, Felipe. The Persistence of (Subnational) Fortune. **The Economic Journal**, v. 126, n. 598, p. 2363-2401, 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/eoj.12276>. Acesso em: out. 2019.

MANSO, Bruno Paes. **A república das milícias**: dos esquadrões da morte à era Bolsonaro. São Paulo: Todavia, 2020.

MANSO, Bruno Paes; DIAS, Camila Nunes. **A guerra**: a ascensão do PCC e o mundo do crime no Brasil. São Paulo: Todavia, 2018.

MILLER, Conrad; SCHMUTTE, Ian M. **The dynamics of referral hiring and racial inequality: evidence from Brazil**. Cambridge: National Bureau of Economic Research, 2021. (NBER Working Paper, n. 29246). Disponível em: <https://www.nber.org/papers/w29246>. Acesso em: set. 2023.

NAKAGAWA, Shinichi; SCHIELZETH, Holger. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. **Methods in Ecology and Evolution**, v. 4, n. 2, p. 133-142, fev. 2013. DOI: 10.1111/j.2041-210x.2012.00261.x. Disponível em: <https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210x.2012.00261.x>. Acesso em: mar. 2020.

NISBETT, Richard E.; COHEN, Dov. **Culture of honor**: the psychology of violence in the South. Boulder: Westview Press, 1996. Disponível em: <https://archive.org/details/cultureofhonorps0056nisb>. Acesso em: out. 2019.

NORTH, Douglass C. **Institutions, Institutional Change and Economic Performance**. Cambridge: Cambridge University Press, 1990.

NUNN, Nathan. Slavery, inequality, and economic development in the Americas: an examination of the Engerman-Sokoloff hypothesis. In: HELPMAN, Elhanan (Ed.). **Institutions and economic performance**. Cambridge: Harvard University Press, 2008. p. 148-180. Disponível em: https://mpr.ub.uni-muenchen.de/4080/1/MPRA_paper_4080.pdf. Acesso em: out. 2019.

NUNN, Nathan; WANTCHEKON, Leonard. The Slave Trade and the Origins of Mistrust in Africa. **American Economic Review**, v. 101, n. 7, p. 3221-3252, dez. 2011.

OLIVEIRA, Guilherme R.; MENEZES, Rafael T.; RESENDE, Guilherme M. **Efeito dose resposta do Fundo Constitucional de Financiamento do Centro-Oeste (FCO) no Estado de Goiás**. Rio de Janeiro: IPEA, 2015. (Texto para Discussão, n. 2133). Disponível em: <https://repositorio.ipea.gov.br/handle/11058/5759>. Acesso em: jun. 2021.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **CID-10**: classificação estatística internacional de doenças e problemas relacionados à saúde: descrições clínicas e diretrizes diagnósticas. Tradução: Dorgival Caetano. Porto Alegre: Artes Médicas, 1993.

PAPADIA, Andrea. **Slaves, migrants and development in Brazil, 1872-1923**. Fiesole: European University Institute, 2019. EUI Working Paper MWP 2019/05. Disponível em: <https://cadmus.eui.eu/handle/1814/63508>. Acesso em: jul. 2022.

PALMA, Nuno; PAPADIA, Andrea; PEREIRA, Thales; WELLER, Leonardo. Slavery and development in nineteenth century Brazil. **Capitalism: A Journal of History and Economics**, v. 2, n. 2, p. 372-426, 2021. Disponível em: <https://repec.cepr.org/repec/cpr/ceprdp/DP15495.pdf>. Acesso em: jul. 2023.

PEREIRA, Katia Cristina; CIRENO, Flávia; GOES, Gabriela; CALSAVARA, Rodrigo; VERAS, Fábio. **Impact evaluation of Seguro-Defeso**: final report. Brasília: IPEA, 2022. (Texto para Discussão, n. 2785). Disponível em: <https://repositorio.ipea.gov.br/handle/11058/11232>. Acesso em: ago. 2023.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Atlas do desenvolvimento humano no Brasil**. Brasília: PNUD, 2010. Disponível em: <http://www.atlasbrasil.org.br>. Acesso em: out. 2020.

QUEIROZ, Marcos. Clóvis Moura e Florestan Fernandes: interpretações marxistas da escravidão, da abolição e da emergência do trabalho livre no Brasil. **Revista Fim do Mundo**, São Paulo, n. 4, p. 254-280, jan./abr. 2021. Disponível em: <https://revistas.marilia.unesp.br/index.php/RFM/article/view/11139/10305>. Acesso em: out. 2021.

RIBEIRO, Darcy. **O povo brasileiro**: a formação e o sentido do Brasil. São Paulo: Companhia das Letras, 1995. Disponível em: https://arquivos.ufrj.br/arquivos/20230122460c373778628a6040b374867/RIBEIRO_Darcy_-_O_Povo_Brasileiro.pdf. Acesso em: jul. 2023.

ROSENTHAL, Caitlin. **Accounting for slavery**: masters and management. Cambridge: Harvard University Press, 2018.

ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41-55, 1983. Disponível em: <https://www.math.mcgill.ca/dstephens/SISCR2017/Articles/Rosenbaum-Rubin-Bka83.pdf>. Acesso em: out. 2019.

ROSENTHAL, Caitlin. **Accounting for slavery**: masters and management. Cambridge: Harvard University Press, 2018. Disponível em: <https://www.jstor.org/stable/j.ctv25250p7>. Acesso em: out. 2020.

SACERDOTE, Bruce. Slavery and the Intergenerational Transmission of Human Capital. **Review of Economics and Statistics**, v. 87, n. 2, p. 217-234, maio 2005. Disponível em: <https://direct.mit.edu/rest/article/87/2/217/57540/Slavery-and-the-Intergenerational-Transmission-of>. Acesso em: out. 2019.

SALLA, Fernando. De Montoro a Lembo: as políticas penitenciárias em São Paulo. **Revista Brasileira de Segurança Pública**, São Paulo, ano 1, ed. 1, p. 72-90, 2007. Disponível em: <https://revista.forumseguranca.org.br/rbsp/article/download/8/5/11>. Acesso em: out. 2019.

SANTOS, Marcelo Justus dos. **Uma abordagem econômica das causas da criminalidade: evidências para a cidade de São Paulo**. 2012. Tese (Doutorado em Economia Aplicada) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2012. Disponível em: <https://www.teses.usp.br/teses/disponiveis/11/11132/tde-14082012-095859/pt-br.php>. Acesso em: out. 2019.

SCHWARCZ, Lilia Moritz (Coord.). **A abertura para o mundo: 1889-1930**. Rio de Janeiro: Objetiva, 2012. (História do Brasil Nação: 1808-2010, v. 3).

SEYLER, François. Slavery, political attitudes and social capital: evidence from Brazil. **Journal of Historical Political Economy**, v. 1, n. 3, p. 377-409, 2021. DOI: <https://doi.org/10.1561/115.00000014>. Disponível em: <https://ideas.repec.org/a/now/jnlhpe/115.00000014.html>. Acesso em: jul. 2023.

SHADISH, William R.; COOK, Thomas D.; CAMPBELL, Donald T. **Experimental and quasi-experimental designs for generalized causal inference**. Boston: Houghton Mifflin, 2002.

SILVA, Marcelo Vasconcelos; SIQUEIRA, Marcelo Lettieri. Evidências a partir dos efeitos dose-resposta do Fundo Constitucional do Norte. **Análise Econômica**, v. 39, n. 80, p. 127-157, set. 2021. DOI: 10.22456/2176-5456.99485. Disponível em: <https://seer.ufrgs.br/index.php/AnaliseEconomica/article/view/99485>. Acesso em: ago. 2023.

SIMPSON, Miles E. Violent crime, income and regional culture: another look. **Sociological Focus**, v. 18, n. 3, p. 199-208, 1985. Disponível em: <https://www.jstor.org/stable/20831363>. Acesso em: out. 2019.

SOARES, Rodrigo R. Development, crime and punishment: accounting for the international differences in crime rates. **Journal of Development Economics**, v. 73, n. 1, p. 155-184, fev. 2004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0304387803001330>. Acesso em: out. 2019.

SOARES, Rodrigo R.; NARITOMI, Joana. Understanding High Crime Rates in Latin America: The Role of Social and Policy Factors. In: DI TELLA, Rafael; EDWARDS, Sebastian; SCHARGRODSKY, Ernesto (Ed.). **The Economics of Crime: lessons for and from Latin America**. Chicago: University of Chicago Press, 2010. p. 19-55. Disponível em: <https://www.nber.org/system/files/chapters/c11831/c11831.pdf>. Acesso em: out. 2019.

SOUZA, Jessé de. **A elite do atraso: da escravidão à Lava Jato**. Rio de Janeiro: Leya, 2017. Disponível em: https://www.mprj.mp.br/documents/20184/1330165/A_Elite_do_atraso_-_da_escravidao_%C3%A0_Lava-Jato.pdf. Acesso em: out. 2019.

SPANIOL, Marlene Inês; MORAES JÚNIOR, Martim Cabeleira; RODRIGUES, Carlos Roberto Guimarães. Como tem sido planejada a Segurança Pública no Brasil? Análise dos Planos e Programas Nacionais de Segurança implantados pós-redemocratização. **Revista Brasileira de Segurança Pública**, São Paulo, v. 14, n. 2, p. 100-127, ago./set. 2020. Disponível em: <https://revista.forumseguranca.org.br/rbsp/article/view/1035>. Acesso em: mai. 2021.

SUMMERHILL, William. **Colonial institutions, slavery, inequality, and development: evidence from São Paulo, Brazil**. Munich: University Library of Munich, 2010. (MPRA Paper, n. 22162). Disponível em: https://mpra.ub.uni-muenchen.de/22162/1/MPRA_paper_22162.pdf. Acesso em: out. 2021.

TRUZZI, Bruno; LIRIO, Viviani Silva; CERQUEIRA, Daniel Ricardo de Castro; COELHO, Danilo Santa Cruz; CARDOSO, Leonardo Chaves Borges. Racial democracy and black victimization in Brazil. **Journal of Contemporary Criminal Justice**, v. 38, n. 1, p. 13-33, fev. 2022. Disponível em: <https://doi.org/10.1177/10439862211038448>. Acesso em: set. 2023.

UTTERMARK, Matthew J. What determines social capital? evidence from slavery's legacy in the United States and Brazil. **Social Forces**, v. 98, n. 4, p. 1773-1800, 2020. Disponível em: <https://academic.oup.com/sf/article/100/4/1936/5686044>. Acesso em: jul. 2021.

VAN BUUREN, Stef. **Flexible imputation of missing data**. 2. ed. Boca Raton: CRC Press, 2018. Disponível em: <https://stefvanbuuren.name/fimd/>. Acesso em: set. 2021.

VON HIPPEL, Paul T. Regression with missing Ys: an improved strategy for analyzing multiply imputed data. **Sociological Methodology**, v. 37, n. 1, p. 83-117, ago. 2007. DOI: 10.1111/j.1467-9531.2007.00180.x. Disponível em: <https://journals.sagepub.com/doi/10.1111/j.1467-9531.2007.00180.x>. Acesso em: set. 2021.

WEBER, Max. **A ética protestante e o "espírito" do capitalismo**. Tradução José Marcos Mariani de Macedo. São Paulo: Companhia das Letras, 2004.

ZHANG, Zhiwei; ZHOU, Jie; CAO, Weihua; ZHANG, Jun. Causal inference with a quantitative exposure. **Statistical Methods in Medical Research**, v. 25, n. 1, p. 315-335, fev. 2016. DOI: 10.1177/0962280212452333. Disponível em: <https://journals.sagepub.com/doi/10.1177/0962280212452333>. Acesso em: ago. 2021.

ZHU, Yeying; COFFMAN, Donna L.; GHOSH, Debashis. A boosting algorithm for estimating generalized propensity scores with continuous treatments. **Journal of Causal Inference**, v. 3, n. 1, p. 25-40, 2015. DOI: 10.1515/jci-2014-0022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26877909/>. Acesso em: ago. 2021.

ZHOU, Jiangxiu. **Comparison of approaches for handling missingness in covariates for propensity score models**. Orientadora: Donna L. Coffman. 2015. 162 f. Tese (Doutorado em Saúde Biocomportamental) – Faculdade de Saúde e Desenvolvimento Humano, Universidade Estadual da Pensilvânia, Pensilvânia, 2015. Disponível em: https://etda.libraries.psu.edu/files/final_submissions/10917. Acesso em: set. 2021.

APÊNDICE A – RESULTADOS COM VARIÁVEIS BINÁRIAS DE UF

Na Seção 4.2, os resultados representados pelas Figuras 9, 13, 17 e 21 não apresentavam os coeficientes das variáveis binárias de UF, responsáveis por capturar os efeitos fixos por estado. Este apêndice traz as tabelas completas para consulta dos leitores.

Para evitar multicolinearidade, em todas as estimações, uma das UFs presente na amostra não está especificada.

Figura A.1 – Resultados Completos - Sudeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
Intercepto	-29,690 (19,529)	-12,307 (22,501)	-20,311 (20,662)	67,554* (26,371)	16,314 (20,317)
<i>sld</i>	0,127+ (0,069)	0,130+ (0,069)	0,128+ (0,069)	0,137* (0,069)	0,133+ (0,069)
<i>ypc</i>	0,002 (0,008)	0,006 (0,009)	0,006 (0,009)	-0,142*** (0,028)	-0,061*** (0,015)
<i>pd</i>	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)	0,005*** (0,001)
<i>unmp</i>	0,860*** (0,167)	0,932*** (0,176)	0,914*** (0,173)	0,812*** (0,168)	0,768*** (0,168)
<i>be</i>	-0,180 (0,147)	-0,247 (0,156)	-0,239 (0,154)	-0,134 (0,139)	-0,088 (0,142)
<i>ym</i>	0,684 (0,670)	0,532 (0,662)	0,556 (0,664)	0,350 (0,625)	0,471 (0,631)
<i>cps</i>	-1,864** (0,571)	-1,777** (0,570)	-1,810** (0,571)	-1,441* (0,560)	-1,409* (0,559)
<i>long</i>	-0,125 (0,387)	-0,096 (0,386)	-0,110 (0,387)	0,131 (0,386)	0,047 (0,387)
<i>lat</i>	-2,187*** (0,424)	-2,091*** (0,438)	-2,097*** (0,437)	-1,889*** (0,436)	-2,022*** (0,433)
<i>ufMG</i>	-17,551*** (3,038)	-17,929*** (3,035)	-18,043*** (3,042)	-16,693*** (3,042)	-17,088*** (3,027)
<i>ufRJ</i>	-5,925 (3,861)	-6,576+ (3,825)	-6,640+ (3,833)	-5,478 (3,821)	-5,883 (3,810)
<i>ufSP</i>	-14,708*** (3,935)	-15,549*** (3,964)	-15,723*** (3,974)	-11,890** (4,005)	-12,973** (3,986)
<i>gini</i>		-22,439 (14,194)		-137,166*** (28,056)	
<i>theil</i>			-10,570		-62,325***

			(6,846)		(12,139)
<i>ypc</i> × <i>gini</i>				0,249***	
				(0,045)	
<i>ypc</i> × <i>theil</i>					0,105***
					(0,018)
Num.Obs.	776	776	776	776	776
R2	0,296	0,299	0,299	0,322	0,322
R2 Adj.	0,285	0,287	0,287	0,309	0,310
AIC	6528,0	6524,9	6525,2	6500,4	6499,5
BIC	10092,5	10089,8	10090,4	10045,5	10044,6
Log.Lik.	-4999,694	-4994,973	-4995,306	-4969,540	-4969,083
RMSE	15,84	15,80	15,81	15,54	15,53

p < 0,1, * p < 0,05, ** p < 0.01, *** p < 0,001

Figura A.2 – Resultados Completos - Nordeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
Intercepto	4,749 (16,555)	12,849 (16,723)	7,314 (16,475)	21,193 (22,147)	11,623 (18,045)
<i>sld</i>	0,082 (0,089)	0,093 (0,089)	0,094 (0,088)	0,091 (0,088)	0,089 (0,087)
<i>ypc</i>	0,009 (0,011)	0,015 (0,011)	0,022+ (0,012)	-0,030 (0,076)	-0,001 (0,036)
<i>pd</i>	0,003* (0,001)	0,003* (0,001)	0,003* (0,001)	0,003* (0,001)	0,003* (0,001)
<i>unmp</i>	0,035 (0,102)	0,059 (0,099)	0,043 (0,101)	0,054 (0,100)	0,040 (0,102)
<i>be</i>	0,079 (0,114)	0,053 (0,111)	0,033 (0,113)	0,071 (0,117)	0,056 (0,117)
<i>ym</i>	1,082+ (0,558)	0,753 (0,560)	0,739 (0,569)	0,747 (0,561)	0,745 (0,572)
<i>cps</i>	-0,675 (0,440)	-0,691 (0,439)	-0,676 (0,437)	-0,690 (0,439)	-0,672 (0,438)
<i>long</i>	0,315 (0,358)	0,133 (0,370)	0,106 (0,362)	0,150 (0,371)	0,120 (0,364)
<i>lat</i>	-0,853** (0,274)	-0,820** (0,273)	-0,793** (0,274)	-0,819** (0,274)	-0,796** (0,275)
<i>ufBA</i>	-12,015*** (1,957)	-13,102*** (2,078)	-13,078*** (2,020)	-13,059*** (2,077)	-12,976*** (2,030)
<i>ufCE</i>	-0,582	-1,581	-1,459	-1,574	-1,339

	(2,193)	(2,252)	(2,234)	(2,255)	(2,246)
<i>ufMA</i>	-4,077	-6,017	-5,937	-6,079	-5,960
	(3,687)	(3,889)	(3,780)	(3,895)	(3,794)
<i>ufPB</i>	1,135	-0,155	-0,095	-0,217	-0,070
	(2,267)	(2,318)	(2,317)	(2,327)	(2,312)
<i>ufPE</i>	14,799***	14,253***	14,207***	14,282***	14,324***
	(1,895)	(1,898)	(1,906)	(1,901)	(1,913)
<i>ufPI</i>	-1,147	-2,739	-2,603	-2,759	-2,515
	(2,970)	(3,155)	(3,007)	(3,154)	(3,010)
<i>ufRN</i>	-0,196	-1,062	-1,330	-1,032	-1,231
	(2,169)	(2,224)	(2,219)	(2,226)	(2,230)
<i>ufSE</i>	2,556	1,639	1,767	1,687	1,930
	(2,766)	(2,635)	(2,694)	(2,653)	(2,760)
<i>gini</i>		-17,412+		-29,759	
		(8,910)		(23,272)	
<i>theil</i>			-11,020**		-17,434
			(4,258)		(10,768)
<i>ypc × gini</i>				0,071	
				(0,118)	
<i>ypc × theil</i>					0,033
					(0,050)
Num,Obs,	874	874	874	874	874
R2	0,368	0,371	0,372	0,371	0,373
R2 Adj,	0,355	0,357	0,359	0,357	0,359
AIC	6896,9	6894,6	6892,6	6898,4	6899,2
BIC	10468,8	10465,8	10462,2	10473,1	10472,4
Log,Lik,	-5170,057	-5165,190	-5163,349	-5165,450	-5165,104
RMSE	12,11	12,08	12,07	12,08	12,07

p < 0,1, * p < 0,05, ** p < 0,01, *** p < 0,001

Figura A.3 – Resultados Completos – Nordeste e Sudeste

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
Intercepto	-8,305	-1,164	-6,168	33,841+	6,991
	(18,198)	(18,840)	(18,343)	(19,761)	(18,121)
<i>sld</i>	0,145+	0,150+	0,137+	0,142+	0,129
	(0,083)	(0,083)	(0,082)	(0,082)	(0,082)
<i>ypc</i>	0,004	0,005	0,005	-0,111***	-0,048**
	(0,009)	(0,009)	(0,009)	(0,028)	(0,015)
<i>pd</i>	0,005***	0,005***	0,005***	0,005***	0,005***

	(0,001)	(0,001)	(0,001)	(0,001)	(0,001)
<i>unmp</i>	0,450**	0,465***	0,454**	0,392**	0,369**
	(0,139)	(0,139)	(0,139)	(0,134)	(0,136)
<i>be</i>	-0,021	-0,038	-0,028	0,080	0,120
	(0,131)	(0,129)	(0,128)	(0,118)	(0,123)
<i>ym</i>	0,718	0,639	0,667	0,341	0,509
	(0,661)	(0,663)	(0,661)	(0,660)	(0,655)
<i>cps</i>	-1,344*	-1,293*	-1,329*	-1,248*	-1,190*
	(0,530)	(0,521)	(0,523)	(0,510)	(0,511)
<i>long</i>	-0,045	-0,058	-0,069	-0,133	-0,136
	(0,463)	(0,463)	(0,462)	(0,462)	(0,462)
<i>lat</i>	-1,222**	-1,176**	-1,182**	-1,256***	-1,299***
	(0,370)	(0,376)	(0,374)	(0,377)	(0,375)
<i>ufBA</i>	-15,171***	-15,424***	-15,271***	-17,260***	-16,303***
	(2,843)	(2,884)	(2,847)	(2,965)	(2,894)
<i>ufCE</i>	0,751	0,311	0,497	-0,917	0,232
	(3,037)	(3,131)	(3,074)	(3,144)	(3,101)
<i>ufES</i>	-2,567	-2,849	-2,684	-4,666	-4,158
	(5,527)	(5,502)	(5,486)	(5,539)	(5,472)
<i>ufMA</i>	-3,646	-4,304	-4,067	-6,758	-5,148
	(4,705)	(4,851)	(4,752)	(4,881)	(4,773)
<i>ufMG</i>	-18,737***	-19,195***	-19,009***	-21,515***	-20,596***
	(4,834)	(4,834)	(4,804)	(4,950)	(4,860)
<i>ufPB</i>	1,410	0,635	1,013	-1,786	-0,143
	(2,921)	(3,013)	(2,961)	(3,020)	(2,971)
<i>ufPE</i>	13,341***	13,013***	13,145***	11,739***	12,514***
	(2,651)	(2,678)	(2,684)	(2,689)	(2,721)
<i>ufPI</i>	1,767	1,297	1,485	-0,624	0,513
	(4,097)	(4,212)	(4,116)	(4,272)	(4,098)
<i>ufRJ</i>	-8,187	-8,761	-8,778	-10,331	-9,695
	(6,290)	(6,295)	(6,252)	(6,382)	(6,297)
<i>ufRN</i>	0,966	0,452	0,638	-1,049	-0,084
	(3,111)	(3,215)	(3,163)	(3,240)	(3,176)
<i>ufSE</i>	1,644	1,069	1,477	-1,012	0,146
	(4,623)	(4,477)	(4,588)	(4,322)	(4,483)
<i>ufSP</i>	-14,826*	-15,661*	-15,452*	-15,957*	-15,477*
	(6,410)	(6,456)	(6,384)	(6,454)	(6,348)
<i>gini</i>		-10,348		-66,318***	
		(10,145)		(16,683)	
<i>theil</i>			-3,665		-27,880***
			(4,630)		(7,088)
<i>ypc</i> × <i>gini</i>				0,186***	

				(0,042)	0,073*** (0,016)
<i>ypc</i> × <i>theil</i>					
Num.Obs.	821	821	821	821	821
R2	0,313	0,314	0,313	0,330	0,327
R2 Adj.	0,295	0,295	0,294	0,311	0,307
AIC	6784,3	6784,6	6778,9	6762,7	6762,2
BIC	10444,1	10447,3	10441,8	10411,0	10413,3
Log.Lik.	-5144,876	-5143,126	-5140,392	-5121,611	-5122,771
RMSE	14,56	14,55	14,56	14,38	14,40

p < 0,1, * p < 0,05, ** p < 0,01, *** p < 0,001

Figura A.4 – Resultados Completos – Brasil

	<i>homr</i>				
	(1)	(2)	(3)	(4)	(5)
Intercepto	-11,002 (16,627)	-8,046 (16,745)	-10,305 (16,659)	10,350 (17,303)	-4,346 (16,742)
<i>sld</i>	0,117* (0,052)	0,122* (0,052)	0,120* (0,052)	0,119* (0,052)	0,116* (0,053)
<i>ypc</i>	0,001 (0,004)	0,002 (0,004)	0,002 (0,004)	-0,050*** (0,013)	-0,016* (0,007)
<i>pd</i>	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)
<i>unmp</i>	0,194** (0,075)	0,208** (0,076)	0,201** (0,075)	0,138+ (0,076)	0,146+ (0,076)
<i>be</i>	-0,149* (0,074)	-0,163* (0,074)	-0,159* (0,075)	-0,053 (0,075)	-0,062 (0,079)
<i>ym</i>	1,115** (0,400)	1,054** (0,408)	1,074** (0,403)	1,067** (0,403)	1,103** (0,399)
<i>cps</i>	-1,429*** (0,321)	-1,413*** (0,318)	-1,418*** (0,319)	-1,302*** (0,315)	-1,310*** (0,317)
<i>long</i>	-0,140 (0,208)	-0,160 (0,210)	-0,157 (0,209)	-0,096 (0,210)	-0,113 (0,209)
<i>lat</i>	-0,397+ (0,229)	-0,386+ (0,230)	-0,388+ (0,230)	-0,447+ (0,231)	-0,451+ (0,233)
<i>ufAL</i>	6,075 (8,197)	6,653 (8,268)	6,538 (8,235)	5,181 (8,270)	5,355 (8,266)
<i>ufAM</i>	-2,369 (5,049)	-2,126 (5,064)	-2,242 (5,061)	-1,643 (5,064)	-2,097 (5,102)
<i>ufAP</i>	15,562	15,953+	15,937+	15,277	15,728+

	(9,576)	(9,603)	(9,597)	(9,350)	(9,417)
<i>ufBA</i>	-4,767	-4,344	-4,356	-6,222	-5,626
	(7,570)	(7,617)	(7,606)	(7,628)	(7,644)
<i>ufCE</i>	4,451	4,907	4,896	3,194	3,756
	(7,806)	(7,854)	(7,833)	(7,852)	(7,860)
<i>ufDF</i>	26,890***	27,070***	27,088***	25,654***	26,560***
	(6,998)	(7,009)	(7,016)	(6,967)	(6,986)
<i>ufES</i>	13,662+	14,069+	14,067+	11,734	12,419
	(8,274)	(8,320)	(8,321)	(8,339)	(8,370)
<i>ufGO</i>	9,112	9,206	9,257	8,106	8,641
	(6,436)	(6,446)	(6,454)	(6,460)	(6,494)
<i>ufMA</i>	-2,164	-1,807	-1,772	-3,421	-2,718
	(6,905)	(6,937)	(6,925)	(6,937)	(6,954)
<i>ufMG</i>	-1,559	-1,360	-1,329	-2,950	-2,426
	(7,404)	(7,425)	(7,431)	(7,454)	(7,483)
<i>ufMS</i>	11,748+	11,913+	11,972+	10,730	11,310+
	(6,586)	(6,602)	(6,619)	(6,635)	(6,671)
<i>ufMT</i>	24,537***	24,658***	24,659***	24,032***	24,508***
	(6,060)	(6,069)	(6,080)	(6,098)	(6,129)
<i>ufPA</i>	3,050	3,376	3,346	2,856	3,274
	(6,389)	(6,416)	(6,403)	(6,419)	(6,442)
<i>ufPB</i>	7,536	7,796	7,858	5,401	6,409
	(8,446)	(8,472)	(8,463)	(8,474)	(8,491)
<i>ufPE</i>	21,450**	21,922**	21,869**	20,078*	20,597*
	(8,278)	(8,335)	(8,309)	(8,332)	(8,338)
<i>ufPI</i>	2,812	3,171	3,219	1,351	2,138
	(7,338)	(7,375)	(7,369)	(7,377)	(7,395)
<i>ufPR</i>	5,171	5,337	5,379	3,466	3,916
	(6,995)	(7,011)	(7,026)	(7,043)	(7,085)
<i>ufRJ</i>	15,793*	15,951*	15,981*	14,393+	14,829+
	(8,027)	(8,033)	(8,042)	(8,055)	(8,088)
<i>ufRN</i>	6,917	7,317	7,268	5,181	5,802
	(8,306)	(8,346)	(8,327)	(8,347)	(8,354)
<i>ufRO</i>	17,316**	17,340**	17,348**	17,592**	17,846**
	(6,546)	(6,549)	(6,552)	(6,577)	(6,606)
<i>ufRR</i>	32,501***	32,641***	32,423***	33,667***	32,944***
	(7,816)	(7,797)	(7,808)	(7,835)	(7,814)
<i>ufRS</i>	1,874	1,812	1,927	0,943	1,022
	(7,480)	(7,476)	(7,493)	(7,490)	(7,538)
<i>ufSC</i>	1,309	1,322	1,399	0,418	0,541
	(7,442)	(7,443)	(7,456)	(7,459)	(7,503)
<i>ufSE</i>	10,114	10,459	10,481	8,129	8,904

	(8,347)	(8,392)	(8,384)	(8,386)	(8,413)
<i>ufSP</i>	7,085	7,020	7,074	6,662	6,670
	(6,979)	(6,975)	(6,985)	(6,982)	(7,020)
<i>ufTO</i>	12,968+	13,388*	13,331*	12,205+	12,539+
	(6,714)	(6,759)	(6,746)	(6,768)	(6,786)
<i>gini</i>		-5,762		-30,697***	
		(5,858)		(8,503)	
<i>theil</i>			-2,214		-8,919**
			(2,676)		(3,243)
<i>ypc</i> × <i>gini</i>				0,075***	
				(0,018)	
<i>ypc</i> × <i>theil</i>					0,018***
					(0,005)
Num.Obs.	2699	2699	2699	2699	2699
R2	0,243	0,243	0,243	0,249	0,247
R2 Adj.	0,233	0,233	0,233	0,239	0,236
AIC	22568,5	22569,7	22569,9	22542,4	22551,3
BIC	34922,2	34926,6	34927,4	34885,3	34903,8
Log.Lik.	-17314,934	-17313,191	-17313,598	-17288,578	-17297,854
RMSE	15,65	15,65	15,65	15,58	15,61

p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001