

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS GRADUAÇÃO EM ECONOMIA

Edgardo Rodrigues Ladeira

TRÊS ENSAIOS SOBRE ECONOMIA DA EDUCAÇÃO E DO
TRABALHO: Análises Longitudinais sobre desempenho educacional e dinâmicas no
mercado de trabalho.

Juiz de Fora

2026

Edgardo Rodrigues Ladeira

**TRÊS ENSAIOS SOBRE ECONOMIA DA EDUCAÇÃO E DO
TRABALHO:** Análises Longitudinais sobre desempenho educacional e dinâmicas no
mercado de trabalho.

Tese apresentada ao Programa de Pós Graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Economia.
Área de concentração: Economia

Orientador: Doutor Ricardo da Silva Freguglia

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da
UFJF com os dados fornecidos pelo(a) autor(a)

Ladeira, Edgardo Rodrigues.

TRÊS ENSAIOS SOBRE ECONOMIA DA EDUCAÇÃO E DO TRABA-
LHO : Análises Longitudinais sobre desempenho educacional e dinâmicas
no mercado de trabalho. / Edgardo Rodrigues Ladeira. – 2026.

194 f. : il.

Orientador: Ricardo da Silva Freguglia

Tese (Doutorado) – Universidade Federal de Juiz de Fora, Faculdade de
Economia. Programa de Pós Graduação em Economia, 2026.

1. Ocupação dos Pais. 2. Distância Residência-Escola. 3. Desempenho
Escolar. 4. Mercado de Trabalho. 5. Hiperinflação. I. Freguglia, Ricardo
da Silva, orientador

Edgardo Rodrigues Ladeira

TRÊS ENSAIOS SOBRE ECONOMIA DA EDUCAÇÃO E DO TRABALHO: Análises Longitudinais sobre desempenho educacional e dinâmicas no mercado de trabalho.

Tese apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Economia. Área de concentração: Economia.

Aprovada em 16 de julho de 2026.

BANCA EXAMINADORA

Dr. Ricardo da Silva Freguglia - Orientador

Universidade Federal de Juiz de Fora

Dr. Douglas Sad Silveira

Universidade Federal de Juiz de Fora

Dr. Felipe Ferreira de Barros

Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

Dr.^a Jéssica Facioli

Universidade do Estado do Rio de Janeiro

Dr. André Luiz Pereira Mancha

Universidade de São Paulo

Juiz de Fora, 30/06/2026.



Documento assinado eletronicamente por **Ricardo da Silva Freguglia, Professor(a)**, em 17/07/2026, às 08:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jéssica Facioli, Usuário Externo**, em 17/07/2026, às 19:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Felipe Ferreira de Barros, Usuário Externo**, em 18/07/2026, às 14:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **André Luiz Pereira Mancha, Usuário Externo**, em 20/07/2026, às 14:28, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Douglas Sad Silveira, Professor(a)**, em 21/07/2026, às 10:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **3055972** e o código CRC **64AAA040**.

Dedico este trabalho à minha mãe, minha avó (*in memoriam*) e ao meu avô (*in memoriam*), verdadeiras referências e grandes exemplos em minha vida. Com cada gesto, apoio e palavra, ensinaram-me o valor do trabalho honesto, a força da dedicação, a grandeza da humildade e a riqueza que só o conhecimento pode oferecer.

AGRADECIMENTOS

Agradeço a meus professores, em especial meu Orientador, Professor Ricardo Freguglia, pelo apoio e pelo cuidado ao longo deste trabalho. Seu apoio foi extremamente importante para a conclusão deste trabalho, participando em todas as etapas com apoio e dedicação integral. Estendo esse agradecimento ao CNPQ pelo apoio dado ao Orientador, Professor Ricardo Freguglia, apoio este relacionado ao financiamento do CNPQ no âmbito do projeto 409255/2023-8.

Além de ser um período desafiador por conta da pandemia de SARS-Cov-2, Juiz de Fora conviveu com uma de suas maiores tragédias hidrológicas da história ao longo desse período. Assim, agradeço à Universidade Federal de Juiz de Fora por me permitir ser discente desta renomada Universidade, onde sempre sonhei estudar e por onde tenho uma enorme admiração e carinho. Agradeço também ao município de Juiz de Fora, cidade pela qual nutro uma grande admiração e que sempre foi um polo fundamental para todos nós, da zona da mata mineira. Nossa região não seria a mesma sem essa grande, linda e sempre acolhedora cidade.

Agradeço também a meus colegas de trabalho, que muito me auxiliaram ao longo desta jornada de estudos e conhecimento. Esse agradecimento é extensivo ao Banco do Brasil S/A e a todos os líderes que passaram pela minha trajetória ao longo dessa jornada e que sempre me apoiaram.

Agradeço a minha família, em especial minha esposa e minhas enteadas que acompanharam toda essa jornada em minha vida. Agradeço também aos meus tios e minhas tias por terem sempre sido presentes e importantes em minha vida. Acrescento aqui uma tia que apenas conheci pelas doces palavras de minha avó e meu avô, a Consuelo (*in memoriam*), mas a quem presto uma homenagem.

Também agradeço a minha madrinha, Maria Joaquina, a quem tenho como uma segunda mãe, pessoa por quem tenho uma grande admiração e a quem devo minha formação básica, por estar ao meu lado em um momento em que estava me desenvolvendo, ainda criança.

Um agradecimento muito especial à querida cidade de Itamarati de Minas, cidade que tanto amo e por onde tenho uma enorme gratidão. Um agradecimento especial a todos(as) os moradores, amigos, amigas e professoras que foram fundamentais no início da minha formação básica. Itamarati foi a cidade onde me formei como cidadão, a formação mais importante que todos temos em nossas vidas.

E por fim o maior de todos os agradecimentos, a Deus. Esse doutorado contribuiu com uma oportunidade única de estar ainda mais conectado com minhas origens e próximo das pessoas que muito me ajudaram a chegar aqui, inclusive minha vovó, na reta final de sua permanência neste mundo que conhecemos.

« Le droit de l'intolérance est donc absurde & barbare ; c'est le droit des tigres ; & il est bien plus horrible : car les tigres ne déchirent que pour manger, & nous nous sommes exterminés pour des paragraphes. »

(VOLTAIRE, *Traité sur la Tolérance*, 1763, Chapitre VI, p. 41–42)

RESUMO

O Brasil é um país marcado por elevada heterogeneidade social, econômica e territorial. Com dimensões continentais, população superior a 200 milhões de habitantes e cerca de 40 milhões de crianças entre 0 e 14 anos, compreender os mecanismos de formação e reprodução das desigualdades educacionais e laborais constitui uma questão central para a formulação de políticas públicas. A literatura sobre mobilidade intergeracional sugere que características familiares, habilidades dos pais, ambiente doméstico, investimentos parentais e capital humano herdado influenciam as oportunidades futuras dos filhos. Esse processo pode se manifestar tanto pela transmissão de habilidades quanto pela reprodução de contextos socioeconômicos desiguais, refletindo-se, por exemplo, na ocupação de posições profissionais semelhantes entre pais e filhos. Com base nesse debate, esta tese é estruturada em três ensaios empíricos que investigam dimensões complementares da desigualdade educacional e laboral no Brasil. O primeiro artigo analisa, a partir do Estudo Longitudinal da Geração Escolar 2005 (GERES), a associação entre a ocupação dos pais e o desempenho acadêmico dos filhos nos primeiros anos do ensino fundamental. Controlando por escolaridade parental, tipo de escola, características familiares e indicadores de engajamento dos pais no ambiente escolar, os resultados indicam que filhos de pais em ocupações de "colarinho branco" tendem a apresentar desempenho superior em Matemática e Português quando comparados aos alunos cujos pais não estão nessa condição. O estudo também evidencia a relevância da participação familiar na escola como fator associado ao desempenho educacional. O segundo artigo aprofunda a análise das desigualdades educacionais ao investigar a relação entre distância residência–escola, escolha escolar, mobilidade urbana e desempenho acadêmico. Diferentemente da abordagem tradicional, que trata a distância como uma variável exógena de custo de acesso, o estudo argumenta que a distância observada é endógena ao processo de escolha escolar. Para lidar com esse problema, explora-se a existência de políticas municipais de subsídio ao transporte estudantil, conhecidas como passe estudantil, como fonte de variação institucional. O benefício reduz a restrição financeira associada ao deslocamento e pode ampliar o conjunto de escolas factíveis para as famílias. Utilizando dados do GERES e modelos de Mínimos Quadrados em Dois Estágios, os resultados mostram que municípios com benefício apresentam, em média, maior distância percorrida entre residência e escola. No segundo estágio, a parcela da distância associada à política de mobilidade está relacionada a reduções no desempenho acadêmico, sugerindo que trajetos mais longos impõem custos relevantes à aprendizagem. O terceiro artigo desloca o foco para o mercado de trabalho e investiga os efeitos persistentes da exposição à hiperinflação sobre trajetórias laborais. A partir de um estudo de coortes com trabalhadores que ingressaram no mercado formal em 1989 e em 1995, são analisadas diferenças na dinâmica dos rendimentos ao longo do ciclo de vida laboral. O estudo acompanha os trabalhadores por 22 anos, período suficiente

para observar trajetórias de médio e longo prazo no mercado formal, inclusive em relação à janela mínima de contribuições previdenciárias exigida para aposentadoria por idade no Brasil. A análise examina a heterogeneidade dos rendimentos, o risco de quedas salariais, a volatilidade dos ganhos e a permanência no emprego formal ao longo dos estágios inicial, intermediário e avançado da carreira. Os resultados indicam que diferentes contextos macroeconômicos de entrada no mercado de trabalho estão associados a padrões distintos de risco, permanência e dispersão dos rendimentos. Em conjunto, os três ensaios contribuem para o debate sobre mobilidade social, desigualdade educacional e dinâmica do mercado de trabalho no Brasil. A tese mostra que desigualdades se formam e se reproduzem por múltiplos canais, seja no ambiente familiar, nas condições territoriais de acesso à escola e nas trajetórias laborais ao longo do ciclo de vida. Ao articular Economia da Educação e Economia do Trabalho, o estudo evidencia como fatores familiares, institucionais, urbanos e macroeconômicos moldam oportunidades educacionais e laborais em um país profundamente desigual.

Palavras-chave: ocupação dos pais; distância residência–escola; desempenho escolar; hiperinflação; mercado de trabalho formal.

ABSTRACT

Brazil is a country marked by high social, economic, and territorial heterogeneity. With continental dimensions, a population of more than 200 million inhabitants, and around 40 million children between 0 and 14 years old, understanding the mechanisms through which educational and labor inequalities are formed and reproduced is a central issue for public policy design. The literature on intergenerational mobility suggests that family characteristics, parental skills, the home environment, parental investments, and inherited human capital influence children's future opportunities. This process may manifest itself both through the transmission of skills and through the reproduction of unequal socioeconomic contexts, reflected, for example, in the occupation of similar professional positions by parents and children. Based on this debate, this thesis is structured into three empirical essays that investigate complementary dimensions of educational and labor inequality in Brazil. The first article analyzes, using the Longitudinal Study of the 2005 School Generation (GERES), the association between parents' occupation and children's academic performance in the early years of primary education. Controlling for parental education, school type, family characteristics, and indicators of parental engagement in the school environment, the results indicate that children of parents in "white-collar" occupations tend to perform better in Mathematics and Portuguese when compared to students whose parents are not in this condition. The study also highlights the relevance of family participation in school as a factor associated with educational performance. The second article deepens the analysis of educational inequalities by investigating the relationship between home-school distance, school choice, urban mobility, and academic performance. Unlike the traditional approach, which treats distance as an exogenous variable reflecting access costs, the study argues that observed distance is endogenous to the school choice process. To address this issue, the analysis exploits the existence of municipal student transportation subsidy policies, known as student passes, as a source of institutional variation. The benefit reduces the financial constraint associated with commuting and may expand the set of feasible schools available to families. Using GERES data and Two-Stage Least Squares models, the results show that municipalities with the benefit have, on average, longer distances traveled between home and school. In the second stage, the portion of distance associated with the mobility policy is related to reductions in academic performance, suggesting that longer commutes impose relevant costs on learning. The third article shifts the focus to the labor market and investigates the persistent effects of exposure to hyperinflation on labor trajectories. Based on a cohort study of workers who entered the formal labor market in 1989 and in 1995, the article analyzes differences in the dynamics of earnings over the working life cycle. The study follows workers for 22 years, a period long enough to observe medium- and long-term trajectories in the formal labor market, including in relation to the minimum

contribution window required for old-age retirement in Brazil. The analysis examines earnings heterogeneity, the risk of wage declines, earnings volatility, and continued participation in formal employment across the early, intermediate, and advanced stages of the career. The results indicate that different macroeconomic contexts at labor market entry are associated with distinct patterns of risk, permanence, and earnings dispersion. Taken together, the three essays contribute to the debate on social mobility, educational inequality, and labor market dynamics in Brazil. The thesis shows that inequalities are formed and reproduced through multiple channels, whether in the family environment, in the territorial conditions of access to school, or in labor trajectories over the life cycle. By bringing together the Economics of Education and Labor Economics, the study highlights how family, institutional, urban, and macroeconomic factors shape educational and labor opportunities in a deeply unequal country.

Keywords: parental occupation; home–school distance; academic achievement; hyperinflation; formal labor market.

RESUMEN

Brasil es un país marcado por una elevada heterogeneidad social, económica y territorial. Con dimensiones continentales, una población superior a 200 millones de habitantes y alrededor de 40 millones de niños entre 0 y 14 años, comprender los mecanismos de formación y reproducción de las desigualdades educativas y laborales constituye una cuestión central para la formulación de políticas públicas. La literatura sobre movilidad intergeneracional sugiere que las características familiares, las habilidades de los padres, el entorno doméstico, las inversiones parentales y el capital humano heredado influyen en las oportunidades futuras de los hijos. Este proceso puede manifestarse tanto mediante la transmisión de habilidades como mediante la reproducción de contextos socioeconómicos desiguales, reflejándose, por ejemplo, en la ocupación de posiciones profesionales similares entre padres e hijos. A partir de este debate, esta tesis se estructura en tres ensayos empíricos que investigan dimensiones complementarias de la desigualdad educativa y laboral en Brasil. El primer artículo analiza, a partir del Estudio Longitudinal de la Generación Escolar 2005 (GERES), la asociación entre la ocupación de los padres y el desempeño académico de los hijos en los primeros años de la educación primaria. Controlando por escolaridad parental, tipo de escuela, características familiares e indicadores de participación de los padres en el entorno escolar, los resultados indican que los hijos de padres en ocupaciones de "cuello blanco" tienden a presentar un desempeño superior en Matemáticas y Portugués en comparación con los alumnos cuyos padres no se encuentran en esa condición. El estudio también evidencia la relevancia de la participación familiar en la escuela como factor asociado al desempeño educativo. El segundo artículo profundiza el análisis de las desigualdades educativas al investigar la relación entre la distancia residencia-escuela, la elección escolar, la movilidad urbana y el desempeño académico. A diferencia del enfoque tradicional, que trata la distancia como una variable exógena de costo de acceso, el estudio argumenta que la distancia observada es endógena al proceso de elección escolar. Para abordar este problema, se explora la existencia de políticas municipales de subsidio al transporte estudiantil, conocidas como pase estudiantil, como fuente de variación institucional. El beneficio reduce la restricción financiera asociada al desplazamiento y puede ampliar el conjunto de escuelas factibles para las familias. Utilizando datos del GERES y modelos de Mínimos Cuadrados en Dos Etapas, los resultados muestran que los municipios con beneficio presentan, en promedio, una mayor distancia recorrida entre la residencia y la escuela. En la segunda etapa, la parte de la distancia asociada a la política de movilidad se relaciona con reducciones en el desempeño académico, lo que sugiere que los trayectos más largos imponen costos relevantes al aprendizaje. El tercer artículo desplaza el foco hacia el mercado laboral e investiga los efectos persistentes de la exposición a la hiperinflación sobre las trayectorias laborales. A partir de un estudio de cohortes con trabajadores que ingresaron al mercado formal en 1989 y en 1995, se

analizan diferencias en la dinámica de los ingresos a lo largo del ciclo de vida laboral. El estudio sigue a los trabajadores durante 22 años, un período suficiente para observar trayectorias de mediano y largo plazo en el mercado formal, incluso en relación con la ventana mínima de contribuciones previsionales exigida para la jubilación por edad en Brasil. El análisis examina la heterogeneidad de los ingresos, el riesgo de caídas salariales, la volatilidad de las ganancias y la permanencia en el empleo formal a lo largo de las etapas inicial, intermedia y avanzada de la carrera. Los resultados indican que diferentes contextos macroeconómicos de entrada al mercado laboral están asociados con patrones distintos de riesgo, permanencia y dispersión de los ingresos. En conjunto, los tres ensayos contribuyen al debate sobre movilidad social, desigualdad educativa y dinámica del mercado laboral en Brasil. La tesis muestra que las desigualdades se forman y se reproducen por múltiples canales, ya sea en el entorno familiar, en las condiciones territoriales de acceso a la escuela o en las trayectorias laborales a lo largo del ciclo de vida. Al articular la Economía de la Educación y la Economía Laboral, el estudio evidencia cómo los factores familiares, institucionales, urbanos y macroeconómicos moldean las oportunidades educativas y laborales en un país profundamente desigual.

Palabras clave: ocupación de los padres; distancia residencia–escuela; desempeño académico; hiperinflación; mercado laboral formal.

LISTA DE ILUSTRAÇÕES

Mapa de distribuição geográfica dos municípios pesquisados no GERES	30
Boxplot e Densidade de Kernel para matemática e português onda 2	39
Scores em matemática e português considerando os grandes grupos da CBO	40
Pontuação em Matemática por Ocupação da Mãe e Escola para Onda 2	41
Pontuação em Português por Ocupação da Mãe e Tipo de Escola para Onda 2	41
Pontuação em Matemática por Ocupação do Pai e Escola para Onda 2	42
Pontuação em Português por Ocupação do Pai e Tipo de Escola para Onda 2	42
Densidade de Kernel - Matemática e Português	43
Estimativas de Odds Ratios dos Modelos Logit para o Desempenho Escolar	50
Distribuição da distância entre residência e escola, com destaque para as caudas.	102
Distância máxima entre dois pontos urbanos em cinco capitais brasileiras.	103
Densidade de Kernel para matemática e português na onda 2	105
Estimativa de densidade bivariada por núcleo	105
Histórico de inflação no Brasil (1980–2024)	131
Média Salarial 1989 a 2017, para as duas coortes	137
Média Salarial 1995 a 2017, para ambas as coortes.	138
Dispersão relativa dos salários reais de dezembro e tamanho amostral anual, por coorte	139
Evolução das taxas de juros no brasil	140
Varição média dos salários conforme anos de experiência, considerando os anos de experiência desde o nascimento do trabalhador para o mercado formal.	142
Salário real médio de dezembro e número de observações válidas por experiência formal observada, por coorte	143
Percentis do log do salário real de dezembro por ano-calendário e coorte	144
Spreads como medida de desigualdade – desigualdade geral, e desigualdade nos topos – em <i>log</i> do salário real em dezembro	145
Trajetoória de percentis de salários (em logaritmo) por coorte por t_{cal}	146
Trajetoória de percentis de salários (em logaritmo) por coorte, considerando anos de presença no mercado formal, na RAIS.	147
Spreads como medida de desigualdade – desigualdade geral, e desigualdade nos topos – em <i>log</i> do salário real em dezembro para t_{cal} e t_{for}	149
Densidade dos salários em dezembro, anos de entrada (1989 e 1995) <i>vs.</i> último ano observado (2017).	150
Margem extensiva no horizonte de 5 anos, <i>exit_rate</i> por percentil de renda recente (<i>RE_bin</i>) e estágio de carreira ($k = 5$)	167
Dispersão do crescimento residual. $sd(\Delta^5 y^{res})$ por <i>RE_bin</i> e estágio de carreira ($k = 5$)	168
Dispersão robusta do crescimento. $P90 - P10$ de $\Delta^5 y^{res}$ por <i>RE_bin</i> ($k = 5$)	169

Assimetria do crescimento residual. Assimetria (Kelley) de $\Delta^5 y^{res}$ por <i>RE_bin</i> e estágio de carreira ($k = 5$)	170
Eventos extremos no crescimento residual. Curtose (Crow–Siddiqui) de $\Delta^5 y^{res}$ por <i>RE_bin</i> ($k = 5$)	171
Saída do formal ao longo da carreira. <i>exit_rate</i> por fatias de <i>RE</i> ao longo de τ ($k = 5$)	172
Instabilidade do crescimento ao longo da carreira. $sd(\Delta^5 y^{res})$ por fatias de <i>RE</i> ao longo de τ ($k = 5$)	173

LISTA DE TABELAS

Tabela 1 – Número de escolas e alunos por município e onda de aplicação (2005–2008)	31
Tabela 2 – Aplicação dos instrumentos de interesse na pesquisa GERES 2005 . . .	32
Tabela 3 – Conversão das funções da CBO em colarinho branco e colarinho azul .	32
Tabela 4 – Tabela com variáveis utilizadas	33
Tabela 5 – Variáveis transformadas e utilizadas nas regressões	36
Tabela 6 – Estatísticas Descritivas – <i>Cross Section</i> (Onda 2)	38
Tabela 7 – Estatísticas Descritivas – Painel (5 ondas, sem Salvador)	44
Tabela 8 – Resultados das Regressões para a <i>cross section</i> da onda 2	46
Tabela 9 – Estimação Pooled OLS, painel balanceado para as 5 ondas.	48
Tabela 10 – Estimação Logit para o desempenho escolar	49
Tabela 11 – Pooled OLS com notas em nível	58
Tabela 12 – Pooled OLS com notas em desvio-padrão	59
Tabela 13 – Interações entre características invariantes e tempo	62
Tabela 14 – Interações entre características invariantes e tempo (continuação) . . .	63
Tabela 15 – Desempenho em Português e características do pai	66
Tabela 16 – Desempenho em Matemática e características do pai	67
Tabela 17 – Desempenho em Português e características da mãe	68
Tabela 18 – Desempenho em Matemática e características da mãe	69
Tabela 19 – Desempenho em Matemática e características conjuntas dos pais	71
Tabela 20 – Desempenho em Português e características conjuntas dos pais	72
Tabela 21 – Robustez dos resultados de Matemática aos efeitos fixos de escola-série	74
Tabela 22 – Robustez dos resultados de Português aos efeitos fixos de escola-série .	75
Tabela 23 – Painel: desempenho em Português e características do pai	77
Tabela 24 – Painel: desempenho em Matemática e características do pai	78
Tabela 25 – Painel: desempenho em Português e características da mãe	79
Tabela 26 – Painel: desempenho em Matemática e características da mãe	80
Tabela 27 – Painel: desempenho em Matemática e características conjuntas dos pais	82
Tabela 28 – Painel: desempenho em Português e características conjuntas dos pais	83
Tabela 29 – Painel: robustez dos resultados de Matemática	84
Tabela 30 – Painel: robustez dos resultados de Português	85
Tabela 31 – Aplicação dos instrumentos de interesse na pesquisa GERES 2005 . . .	99
Tabela 32 – Distribuição dos estudantes da amostra, considerando a política pública mu- nicipal de subsídio	99
Tabela 33 – Conversão das funções da CBO em Colarinho Branco e Colarinho Azul	100
Tabela 34 – Dicionário com as variáveis utilizadas	101
Tabela 35 – Alunos excluídos após truncamento por município e local de residência	103
Tabela 36 – Estatísticas Descritivas da amostra com e sem truncamento (<i>Percentil 99</i>)	104

Tabela 37 – Forma reduzida: efeito da política municipal de passe estudantil sobre o desempenho em Matemática e Português	107
Tabela 38 – Estimação do 1º estágio: regressão da distância no passe escolar	108
Tabela 39 – Efeito da distância residência–escola sobre o desempenho em Matemática	109
Tabela 40 – Efeito da distância residência–escola sobre o desempenho em Português	111
Tabela 41 – Distância residência–escola e desempenho (em desvios–padrão)	112
Tabela 42 – Heterogeneidade: efeito da distância sobre o desempenho em alunos de escola pública — em <i>Log</i>	121
Tabela 43 – Heterogeneidade: efeito da distância sobre o desempenho entre alunos cujos pais não possuem ensino superior — em <i>Log</i>	122
Tabela 44 – Heterogeneidade: efeito da distância sobre o desempenho em alunos da rede privada — em <i>Log</i>	123
Tabela 45 – Sensibilidade <i>leave-one-city-out</i> — em <i>Log</i>	125
Tabela 46 – Estatísticas descritivas na entrada — Coorte de 1989	134
Tabela 47 – Estatísticas descritivas na entrada — Coorte de 1995	135
Tabela 48 – Estatísticas descritivas do salário real em dezembro por nível de escolaridade	135
Tabela 49 – Estatísticas descritivas no painel completo.	136
Tabela 50 – Frações admissível e RE-elegível por coorte	153
Tabela 51 – Resumo de <i>exit_rate</i> (ausência no formal no lead) no universo RE-elegível	155
Tabela 52 – Momentos calculados nas células da superfície e interpretação econômica	158
Tabela 53 – Guia de interpretação conjunta dos momentos	159
Tabela 54 – Estatísticas descritivas do painel formal por coorte	160
Tabela 55 – Distribuição do salário formal positivo (dezembro) e de $y = \log(1 + Y)$ por coorte	160
Tabela 56 – Cobertura de células (τ, RE_bin) por coorte e horizonte k	163
Tabela 57 – Momentos do crescimento em 5 anos e margem extensiva por estágio de carreira e fatia de <i>Recent Earnings</i> (RE): coortes 1989 e 1995 lado a lado ($k = 5$)	165
Tabela 58 – Heterogeneidade ao longo de <i>RE_bin</i> . Amplitude (máx.–mín.) dos momentos por estágio de carreira ($k = 5$)	166
Tabela 59 – Significância estatística das diferenças entre coortes nas superfícies de momentos	181
Tabela 60 – Dicionário de leitura dos sinais nas diferenças entre coortes	183

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Introdução geral	19
2	TAL PAI, TAL FILHO? UM ENSAIO SOBRE A OCUPAÇÃO DOS PAIS E O DESEMPENHO ACADÊMICO DE SEUS FILHOS. 23	
2.1	Introdução	23
2.2	Referencial Teórico	26
2.2.1	Banco de Dados – O Estudo Longitudinal da Geração Escolar 2005 - GERES.	30
2.3	Metodologia	34
2.4	Estatísticas Descritivas	36
2.4.1	Estatísticas Descritivas para a <i>Cross Section</i> da Onda 2	37
2.4.2	Estatísticas Descritivas para o Painel – 5 ondas.	43
2.5	Resultados	45
2.5.1	Resultados para a Onda 2	45
2.5.2	Resultados obtidos nas regressões para o painel com as 5 ondas	47
2.6	Conclusão	51
	Referências do capítulo	54
	Apêndice ao capítulo 2	57
3	DISTÂNCIA, ESCOLHA ESCOLAR E MOBILIDADE URBANA: EVIDÊNCIAS DOS SUBSÍDIOS AO TRANSPORTE ESTUDANTIL NO BRASIL	87
3.1	Introdução	87
3.2	Distância, Mobilidade e Escolha Escolar	89
3.3	Modelo empírico, metodologia e a distância residência–escola como variável endógena	92
3.3.1	Diálogo com a literatura e hipóteses do estudo	92
3.3.2	Arcabouço conceitual: a escolha escolar e o "custo" de acesso	93
3.3.3	Construção da base de dados e mensuração da distância	94
3.3.4	Estratégia empírica	95
3.3.5	Hipóteses de identificação, interpretação e limitações	96
3.3.6	Heterogeneidade e exercícios de robustez	97
3.4	Base de Dados	98
3.4.1	A base GERES e o recorte analítico	98
3.4.2	Distribuição da política de passe estudantil	99
3.4.3	Construção das variáveis e controles	100
3.5	Estatísticas Descritivas	101
3.6	Resultados	106

3.6.1	Forma reduzida da política municipal	106
3.6.2	Primeiro estágio: política municipal e distância residência–escola	107
3.6.3	Distância residência–escola e desempenho em Matemática	109
3.6.4	Distância residência–escola e desempenho em Português	110
3.6.5	Resultados em desvio-padrão e magnitude dos efeitos	112
3.6.6	Síntese e diálogo dos resultados com a literatura	114
3.7	Conclusão	115
	Referências do capítulo	117
	Apêndice ao capítulo 3	120
4	EXPOSIÇÃO À HIPERINFLAÇÃO E CICLO DE VIDA LABO- RAL: EVIDÊNCIAS DE LONGO PRAZO SOBRE RISCO, REN- DIMENTOS E PERMANÊNCIA NO MERCADO FORMAL.	126
4.1	Introdução	126
4.2	Contexto Econômico e Institucional Brasileiro	129
4.3	Banco de Dados e Estatísticas Descritivas.	132
4.3.1	Banco de Dados	132
4.3.2	Estatísticas Descritivas	134
4.3.3	Distribuição salarial e desigualdade ao longo do ciclo de vida	144
4.4	Metodologia	151
4.4.1	Construção da amostra elegível para o mercado formal brasileiro	151
4.4.2	Momentos da distribuição e leitura econômica das superfícies	156
4.4.3	Estatísticas descritivas do painel construído considerando o mercado formal.	160
4.5	Resultados	161
4.5.1	Composição do painel formal e base efetiva do procedimento	161
4.5.2	Admissibilidade e RE-elegibilidade: tamanho da amostra utilizada . . .	162
4.5.3	Cobertura de células e leitura do número de “grupos”	162
4.5.4	Margem extensiva: ausência no formal no futuro (<i>exit_rate</i>)	163
4.5.5	Evidências sobre heterogeneidade no crescimento formal	166
4.5.6	Superfícies de momentos por <i>RE_bin</i> e evidência visual da heterogeneidade	166
4.6	Conclusão	174
	Referências do capítulo	176
	Apêndice ao capítulo 4	179
5	Conclusão	184
	Referências gerais da tese	188

1 INTRODUÇÃO

1.1 Introdução geral

Esta tese investiga, ao longo de três ensaios, efeitos heterogêneos associados a resultados educacionais e laborais no Brasil. A motivação central parte da complementaridade entre educação e mercado de trabalho, amplamente discutida na literatura econômica. A educação é frequentemente compreendida como um investimento em capital humano, capaz de elevar a produtividade individual e gerar retornos no mercado de trabalho. Nesse sentido, Becker (1964) argumenta que investimentos em educação contribuem para a formação do estoque de capital humano, com efeitos positivos sobre a produtividade. De forma complementar, Mincer (1974a) desenvolve uma formulação empírica que relaciona anos de escolaridade e experiência potencial aos rendimentos auferidos pelos indivíduos. Na mesma tradição, estudos posteriores, como os de David Card, investigaram os retornos salariais da educação adicional, consolidando a escolaridade como uma dimensão central para compreender desigualdades de renda e trajetórias ocupacionais.

Para além de seus efeitos sobre salários individuais, a educação também é considerada um fator relevante para o desenvolvimento econômico. A formação de capital humano amplia as capacidades produtivas, reduz restrições de oportunidade e pode afetar a mobilidade social entre gerações. Em diversos contextos, a decisão entre permanecer estudando ou ingressar precocemente no mercado de trabalho expressa restrições familiares, institucionais e econômicas que condicionam as trajetórias futuras dos indivíduos. Por isso, a literatura sobre mobilidade intergeracional tem tratado a educação como um dos principais mecanismos pelos quais desigualdades de origem podem ser reproduzidas ou mitigadas ao longo do tempo.

A importância dos investimentos educacionais é ainda mais evidente quando se considera o momento em que eles ocorrem. Cunha e Heckman (2007) argumentam que investimentos precoces e realizados em períodos oportunos do ciclo de formação tendem a apresentar retornos mais elevados. Segundo essa abordagem, lacunas formadas nos primeiros anos de vida e de escolarização podem ser difíceis de compensar posteriormente, mesmo com novos investimentos. Assim, compreender quais fatores afetam o desempenho educacional nos anos iniciais é fundamental para orientar políticas públicas voltadas à redução de desigualdades de aprendizagem.

Apesar da centralidade da educação para a formação de capital humano, o Brasil iniciou relativamente tarde a construção de sistemas nacionais de avaliação educacional em larga escala. O Sistema Nacional de Avaliação da Educação Básica (SAEB) começou a ser concebido na segunda metade da década de 1980 e teve sua primeira aplicação em 1990. Em 1995, o SAEB passou por uma reestruturação metodológica que permitiu maior comparabilidade dos resultados ao longo do tempo (INEP, 2023). Posteriormente,

em 2007, foi criado o Índice de Desenvolvimento da Educação Básica (IDEB), indicador que combina o desempenho dos alunos em avaliações padronizadas com medidas de fluxo escolar.¹ O IDEB tornou-se um marco para o monitoramento da educação básica no país, sendo amplamente utilizado em pesquisas, diagnósticos e formulação de políticas.

A literatura nacional também começou a explorar os usos e limites desses indicadores. Sanches (2017), ao analisar 383 teses e dissertações sobre o IDEB defendidas entre 2007 e 2015, identifica diferentes eixos de investigação: o IDEB como indutor de políticas e programas, como resultado de políticas educacionais, como critério de seleção de redes e escolas pesquisadas, como variável explicativa e como instrumento para identificar fatores associados a melhores ou piores desempenhos. Uma limitação importante do IDEB, entretanto, decorre de sua própria construção: por combinar proficiência e fluxo escolar, aumentos nas taxas de aprovação podem compensar desempenhos mais baixos nos testes padronizados. Por essa razão, a avaliação de resultados educacionais requer, muitas vezes, a desagregação dos componentes do indicador e o uso de bases mais detalhadas.

Nesse contexto, estudos longitudinais assumem papel especialmente relevante. Franco, Brooke e Alves (2008) argumentam que avaliações como o SAEB são importantes para revelar tendências gerais da qualidade da educação nacional, mas não têm como objetivo produzir conhecimento detalhado sobre os fatores que influenciam o aprendizado dos estudantes. Foi diante dessa lacuna que surgiu o Estudo Longitudinal da Geração Escolar 2005 (GERES), voltado à produção de evidências sobre qualidade, equidade e desigualdades de oportunidades educacionais. O GERES foi desenvolvido em parceria pela Universidade Federal de Minas Gerais (UFMG), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Universidade Estadual do Mato Grosso do Sul (UEMS), Universidade Federal da Bahia (UFBa), Universidade Estadual de Campinas (UNICAMP) e Universidade Federal de Juiz de Fora (UFJF). A pesquisa acompanhou estudantes em cinco ondas, combinando testes de desempenho com questionários aplicados a alunos, famílias, turmas e escolas. Participaram do estudo os municípios de Belo Horizonte, Campo Grande, Campinas, Rio de Janeiro e Salvador. A riqueza dessa base deu origem ao livro *GERES 2005: Razões e resultados de uma pesquisa longitudinal sobre eficácia escolar* (Brooke; Bonamino, 2011).

Os dois primeiros ensaios desta tese utilizam dados do GERES. A relevância dessa base decorre, em particular, de sua capacidade de combinar informações de desempenho escolar com características familiares e escolares observadas ainda nos primeiros anos de alfabetização. O fato do estudo coletar informações diretamente junto aos pais permite investigar dimensões do ambiente familiar que raramente estão disponíveis em avaliações

¹ O Índice de Desenvolvimento da Educação Básica (IDEB) foi criado em 2007 pelo INEP e combina informações de fluxo escolar, como taxas de aprovação, com o desempenho dos alunos nas avaliações do SAEB, permitindo o acompanhamento da qualidade do ensino e o estabelecimento de metas (Estudos e Pesquisas Educacionais Anísio Teixeira, 2022).

educacionais de larga escala. Essa característica torna o GERES especialmente adequado para analisar como fatores familiares, escolares e territoriais se relacionam ao desempenho das crianças em uma etapa decisiva da formação do aprendizado.

O primeiro ensaio, apresentado no Capítulo 2, investiga a associação entre a ocupação dos pais e o desempenho acadêmico dos filhos. A análise dialoga com a literatura sobre mobilidade intergeracional e transmissão de vantagens entre gerações. Laband e Lentz (1983) argumentam que existe uma tendência de herança ocupacional entre pais e filhos, associada à transmissão de conhecimentos, habilidades e redes de informação. Em vez de observar apenas a educação como determinante dos rendimentos futuros, o ensaio examina uma relação em sentido anterior: como a posição ocupacional dos pais pode estar associada ao desempenho escolar dos filhos ainda nos anos iniciais. Os resultados indicam que a ocupação dos pais, especialmente em ocupações de "colarinho branco", está associada a um melhor desempenho acadêmico. Ao mesmo tempo, os controles de escolaridade parental sugerem maior associação entre a escolaridade materna e os resultados dos filhos, o que pode refletir maior presença ou engajamento das mães no acompanhamento das atividades escolares, em linha com evidências discutidas por Wang *et al.* (2023) e Lee e Borgonovi (2022).

O segundo ensaio, apresentado no Capítulo 3, também utiliza o GERES, mas desloca o foco da família para a dimensão territorial da escolha escolar. O artigo investiga a relação entre a distância residência–escola e o desempenho acadêmico. Diferentemente de abordagens que tratam a distância como uma variável exógena de custo de acesso, como em Card (1993), este ensaio argumenta que, no contexto urbano brasileiro, a distância observada pode ser endógena ao processo de escolha escolar. A existência de políticas municipais de subsídio ao transporte estudantil altera as restrições financeiras enfrentadas pelas famílias e pode ampliar o conjunto de escolas factíveis. Assim, famílias com acesso ao passe estudantil podem escolher escolas mais distantes, o que torna a distância uma variável que combina custo de deslocamento e seleção escolar.

Para lidar com essa questão, o segundo ensaio utiliza a presença da política municipal de subsídio ao transporte estudantil como variável instrumental e estima modelos de Mínimos Quadrados em Dois Estágios. No primeiro estágio, os resultados indicam que municípios com subsídio apresentam, em média, distância residência–escola aproximadamente 1 quilômetro maior. No segundo estágio, a parcela da distância associada à política pública é negativamente relacionada ao desempenho dos alunos. Essa evidência sugere que, embora a distância observada possa refletir escolhas familiares por escolas percebidas como melhores, a distância induzida pelo regime de mobilidade expressa custos efetivos de deslocamento, com potenciais efeitos adversos sobre a aprendizagem.

Após analisar fatores associados ao desempenho educacional, o terceiro ensaio, apresentado no Capítulo 4, desloca a investigação para o mercado de trabalho. Inspirado

em Guvenen *et al.* (2021), o estudo examina a heterogeneidade da dinâmica dos rendimentos formais ao longo do ciclo de vida laboral. O objetivo é investigar se trabalhadores que ingressaram no mercado formal em contextos macroeconômicos distintos apresentam trajetórias diferentes de risco, rendimentos e permanência no emprego formal. Para isso, são comparadas duas coortes: trabalhadores que ingressaram no mercado formal em 1989, período marcado por hiperinflação, e trabalhadores que ingressaram em 1995, após a estabilização macroeconômica associada ao Plano Real.

O terceiro ensaio utiliza microdados da RAIS.² A análise acompanha aproximadamente 3 milhões de trabalhadores ao longo do tempo, de 1989 a 2017 para a primeira coorte e de 1995 a 2017 para a segunda. Para tornar as trajetórias comparáveis, o estudo organiza os dados não apenas pelo ano calendário, mas também por anos de experiência desde a entrada no mercado formal. Dessa forma, o primeiro ano de experiência corresponde a 1989 para a coorte exposta à hiperinflação e a 1995 para a coorte pós-estabilização.

Os resultados do terceiro ensaio indicam que a dinâmica dos rendimentos no mercado formal brasileiro é marcada por forte heterogeneidade ao longo do ciclo de vida. Essa heterogeneidade não se limita ao nível dos rendimentos, mas também se manifesta no risco de saída do mercado formal, na volatilidade dos ganhos, na assimetria das variações salariais e na incidência de eventos extremos. A posição do trabalhador na distribuição de rendimentos recentes e o estágio da carreira condicionam de forma sistemática tanto o crescimento esperado quanto a probabilidade de permanência no setor formal. Além disso, os padrões observados variam entre coortes expostas a diferentes contextos macroeconômicos, sugerindo que condições de entrada no mercado de trabalho podem produzir efeitos persistentes sobre trajetórias laborais.

Em conjunto, os três ensaios desta tese contribuem para o debate sobre desigualdade, mobilidade social, educação e mercado de trabalho no Brasil. Os dois primeiros capítulos mostram que desigualdades educacionais são moldadas por fatores familiares, institucionais e territoriais desde os primeiros anos de escolarização. O terceiro capítulo mostra que desigualdades laborais também se expressam de forma heterogênea ao longo do ciclo de vida, dependendo da posição inicial do trabalhador e do contexto macroeconômico de entrada no mercado formal. Assim, a tese articula Economia da Educação e Economia do Trabalho para compreender como oportunidades são formadas, condicionadas e reproduzidas em diferentes etapas da trajetória individual.

Nas seções seguintes, os três ensaios são apresentados separadamente. Ao final, a conclusão geral sintetiza os principais achados da tese, suas contribuições para a literatura e possíveis desdobramentos para pesquisas futuras.

² A Relação Anual de Informações Sociais (RAIS) é um registro administrativo criado em 1975, com periodicidade anual, cujo objetivo é fornecer informações sobre o mercado de trabalho formal brasileiro, atendendo a necessidades de controle, estatísticas e formulação de políticas públicas.

2 TAL PAI, TAL FILHO? UM ENSAIO SOBRE A OCUPAÇÃO DOS PAIS E O DESEMPENHO ACADÊMICO DE SEUS FILHOS.

Resumo

Este artigo tem como objetivo analisar como a ocupação dos pais pode estar relacionada ao desempenho escolar de seus filhos nos primeiros anos do ensino fundamental. Utilizando os microdados do Estudo Longitudinal da Geração Escolar 2005 (GERES), estimou-se a correlação existente entre a ocupação dos pais e a proficiência de seus filhos na escola, controlando por variáveis relevantes, como a escolaridade dos pais, o engajamento parental nas atividades escolares, que foram medidos a partir dos motivos relatados para visitas à escola e o tipo de instituição frequentada (se pública ou privada). Os resultados indicam que o ambiente familiar exerce influência significativa sobre as habilidades cognitivas desenvolvidas pelas crianças. Em particular, filhos de pais com ocupações classificadas como colarinho branco apresentam desempenho superior em comparação àqueles cujos pais não possuem ocupações com esse perfil. A escolaridade dos pais, seu nível de engajamento escolar e a natureza da escola frequentada também se mostraram determinantes importantes associados ao desempenho acadêmico.

Palavras-chave: Tal pai, tal filho?; Geres 2005; Ocupação dos Pais; Desempenho Escolar; Persistência Intergeracional.

2.1 Introdução

O objetivo deste artigo é analisar como e se a ocupação dos pais explica o desempenho de seus filhos nos primeiros anos do ensino fundamental. Utilizando dados do Estudo Longitudinal da Geração Escolar 2005 (GERES), foi estimada a correlação existente entre a ocupação dos pais e a proficiência de seus filhos na escola, controlando por variáveis relevantes, como a escolaridade dos pais, o engajamento parental nas atividades escolares, engajamento esse medido a partir dos motivos relatados para visitas à escola, além do tipo de instituição frequentada, se pública ou privada.

O desempenho e a permanência dos estudantes na escola são motivos de pesquisa ao longo do tempo e são considerados importantes fatores na formação e desenvolvimento de habilidades e competências, que formam o estoque de capital humano. Card (1993) estimou que ter uma Universidade próxima à moradia ampliava o acesso à Universidade e isso contribuiria para anos a mais de estudo, refletindo em melhor remuneração posterior no mercado de trabalho. Heckman (2006) ao abordar a teoria das taxas de retorno do investimento em capital humano, concluiu que os retornos sobre os investimentos são decrescentes à medida que as crianças evoluem em sua idade escolar.

Uma das principais motivações deste trabalho reside no fato de que o estoque de

capital humano é um importante predecessor do desenvolvimento das sociedades e, ao analisar os fatores que têm relação com o estoque futuro de capital humano, contribui-se para que os formuladores de *políticas* entendam como o presente pode influenciar a persistência de barreiras estruturais à mobilidade intergeracional dos indivíduos nas comunidades.

Foi nesse contexto que Giannelli e Rapallini (2019) desenvolveram, a partir da ocupação dos pais em áreas que exigiam conhecimentos em STEM (ciência, tecnologia, engenharia e matemática), uma variável instrumental de atitude familiar em relação à Matemática e estudaram como essa “variável” poderia influenciar os resultados escolares de seus filhos, demonstrando que crianças cujos pais possuem ocupações em STEM lidam com menos ansiedade em relação à Matemática e conseguem melhores resultados na disciplina. Esse ponto de partida auxiliou no desenvolvimento de parte deste trabalho, embora a abordagem empírica seja mais descritiva neste ensaio quando comparado ao estudo produzido por Giannelli e Rapallini (2019).

Os estudos sobre a ocupação dos pais e o desempenho dos filhos são uma preocupação para pesquisadores e entidades multilaterais, inclusive a Organização para a Cooperação e Desenvolvimento Econômico (OECD), que publicou em sua revista "Pisa in Focus" de 2014 uma matéria que destacou que os alunos cujos pais têm profissões mais qualificadas tiveram um desempenho melhor nas avaliações de Matemática do PISA 2013 em comparação com alunos cujos pais têm ocupações mais elementares. Os estudos sugerem que o ambiente no qual o estudante está inserido fora da escola contribui para seus resultados e que a ocupação dos pais é uma parte importante desse contexto.

Para este estudo, foram utilizados dados de um painel construído a partir do Estudo Longitudinal da Geração Escolar 2005, conhecido como GERES, pesquisa de campo desenvolvida nas cidades de Campinas - SP, Campo Grande - MT, Salvador - BA, Rio de Janeiro - RJ e Belo Horizonte - MG. Esse projeto contou com a associação de seis centros de pesquisa: a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), a Universidade Federal de Minas Gerais (UFMG), a Universidade Estadual de Campinas (UNICAMP), o IPES, a Universidade Federal da Bahia (UFBA) e a própria Universidade Federal de Juiz de Fora (UFJF).

O GERES foi um estudo longitudinal dividido em 5 ondas^{1,2}, onde se gerou um recorte longitudinal através da aplicação de cinco testes de língua portuguesa (leitura) e de matemática, sendo o primeiro teste aplicado no momento do ingresso do estudante na primeira onda de observação (teste diagnóstico) e os demais aplicados em ondas ao final de cada ano encerrado. Essa base é única, pois existem dados declarados pelos pais dos estudantes sobre suas ocupações, os motivos que os levaram a frequentar a escola ao longo daquele primeiro ano, questionário esse respondido de forma muito próxima à segunda

¹ A 1ª onda é também chamada de onda de diagnóstico.

² A 2ª onda, ou Onda 2, é também chamada de 1ª onda de avaliação.

onda. O fato de serem ocupações declaradas propicia uma rica fonte de informação, pois se pode identificar de forma direta a ocupação dos pais, trazendo dados inclusive de profissionais informais, empreendedores, empresários, profissionais esses que não seria possível localizar em muitos outros bancos de dados.

O estudo acompanhou crianças que estariam ingressando no 2º ano de alfabetização ou, para a maioria das escolas, ingressando no 1º ano do ensino fundamental, à exceção de Salvador, onde as crianças avaliadas estavam ingressando no 2º ano do ensino fundamental, uma vez que os pesquisadores identificaram que não havia, em grande escala, a educação pré-escolar na cidade, o que deixava a responsabilidade pela alfabetização para o 1º ano do ensino fundamental nas escolas da comunidade. Isso traz uma relevância importante para o estudo e, como consequência, sempre que Salvador é incluído no estudo, há uma necessidade de controle por série da dimensão de desempenho escolar.

Os resultados encontraram evidências que sugerem que estudantes filhos de pais que declararam exercer atividades gerenciais e administrativas, também conhecidas como ocupações de colarinho branco³, possuem desempenho superior àqueles estudantes filhos de pais que não declararam ocupação nessa mesma condição. Filhos de pais que visitaram a escola para participar de atividades da comunidade e reunião de pais também tiveram desempenhos superiores àqueles cujos pais declararam que visitaram a escola apenas para reunião com o coordenador ou o professor. Também foi identificado que a escolaridade dos pais tem uma relação de magnitude relevante e direção positiva com o desempenho acadêmico de seus filhos.

Este estudo contribui para um entendimento mais aprofundado dos fatores associados ao desempenho acadêmico das crianças, com especial atenção ao papel do contexto familiar no desenvolvimento de habilidades cognitivas desde a primeira infância. Ao utilizar dados socioeconômicos desagregados, é possível identificar com maior precisão quais dimensões familiares e sociais podem estar associadas à formação do conhecimento e ao desenvolvimento cognitivo infantil. Essa abordagem pode contribuir para o desenho de políticas públicas mais específicas e direcionadas, capazes de mitigar possíveis efeitos negativos sobre o desempenho escolar das crianças.

Este capítulo está dividido em seis seções. Na seção 2.2, são apresentados o referencial teórico que embasa o estudo, a literatura relacionada ao tema e a base de dados utilizada. A seção 2.3 descreve a estratégia empírica e a metodologia adotada; a seção 2.4 apresenta as estatísticas descritivas. Na seção 2.5, discutem-se os resultados e, na seção 2.6, apresentam-se as conclusões e as limitações do estudo.

³ Atividades mais gerenciais, que neste trabalho foram agrupadas pelos 5 primeiros grandes grupos da CBO (0–4).

2.2 Referencial Teórico

Apenas na segunda metade dos anos 80 do século passado é que se começou a pensar no SAEB – Sistema Nacional de Avaliação da Educação Básica, que teve seu primeiro teste aplicado em 1990. Em 1995, o SAEB sofreu uma reestruturação em sua metodologia, possibilitando, então, um acompanhamento do desempenho ao longo dos anos (INEP, 2023). A partir do SAEB, foram pensadas outras metodologias para ampliar o escopo do processo de avaliação, sendo desenvolvido, em 2007, o IDEB – Índice de Desenvolvimento da Educação Básica. Franco, Brooke e Alves (2008) argumentam que as metodologias de avaliação são extremamente relevantes e que o SAEB oferece uma sequência de resultados capazes de revelar a tendência da qualidade da educação nacional. Entretanto, avaliações como o SAEB não oferecem a possibilidade nem possuem como objetivo gerar um conhecimento pormenorizado sobre quais fatores influenciam a qualidade da educação básica. Foi nesse contexto que surgiu o Estudo Longitudinal da Geração Escolar 2005, o GERES, buscando produzir dados sobre qualidade, equidade e desigualdades de oportunidades educacionais (Franco; Brooke; Alves, 2008).

Foi com os resultados do SAEB que Brooke e Bonamino (2011) encontraram evidências de que os alunos chegavam à 4ª série do ensino fundamental com lacunas advindas dos anos iniciais do ensino fundamental, lacunas que são uma base importante para a literatura de mobilidade intergeracional e transmissão de habilidades de natureza (“nature” ou genes) e de educação (“nurture” ou ambiente), conforme relatam diversos autores em estudos focados no tema (Duncan, 1966; DeJong; Milton; Stanley, 1971; Laband; Lentz, 1983; Lentz; Laband, 1989; Black; Devereux, P. J.; Salvanes, 2005, 2009; Black; Devereux, 2010). Neste estudo, busca-se compreender melhor os fatores familiares que poderiam estar associados ao desempenho acadêmico das crianças. Desta forma, seguiu-se a literatura de *nurture* ou ambiente, onde a escola frequentada pode ser representativa como fator de desigualdade, estando relacionada ao tempo e a dedicação escolar dos pais em relação a seus filhos. Cunha e Heckman (2007) demonstram em seu modelo de formação de competências de ciclo de vida que os investimentos na primeira infância podem elevar o nível de capital humano de forma a aumentar a produtividade de investimentos posteriores a essa fase.

Nesta literatura, diversos autores analisaram o efeito da educação sobre rendimentos futuros. Gunderson e Oreopolous (2020) e Patrinos e Psacharopoulos (2020) constataram em diferentes contextos e países que um ano a mais de escolaridade corresponderia a um aumento de 10% nos rendimentos. A literatura Minceriana também destaca os efeitos da educação e da experiência sobre rendimentos (Mincer, 1974b). Também Card (1993, 1994, 1999) apresenta em seus trabalhos o quanto a educação ajuda a explicar os rendimentos futuros.

Em outra linha dessa literatura, Gu *et al.* (2022) estudaram a contribuição dos

país no acesso ao ensino superior na China, onde, desde 1999, esse tipo de ensino cresceu 10 vezes. Descobriram, usando o pressuposto do “liderar pelo exemplo”, que o incremento de 1 ano na educação do pai aumenta em 7,1% a probabilidade de seu filho ingressar na Universidade. Ainda nesse contexto de “herança ambiental”, Solon (1999) afirmou que os rendimentos dos descendentes são fortemente determinados pela renda dos ascendentes.

Considerando a literatura existente, está claro que os rendimentos tendem, em grande parte, a ser obtidos pelas famílias através de suas ocupações, de seus empregos e dos atributos e conhecimentos necessários para executar tais atividades, como a formação acadêmica e as escolhas feitas ao longo da vida. Se isso passa de pai para filho, em um efeito de transmissão, seja pela “liderança pelo exemplo”, seja pela crença, pela dedicação de tempo ou pela decisão de alocação de investimentos financeiros nos filhos, é provável que as ocupações e habilidades também influenciem o desempenho acadêmico dos estudantes.

Assim, ao analisar a literatura tradicional que discute os desempenhos escolares, normalmente se pesquisa dados agregados (como nível socioeconômico), fatores demográficos e geográficos, e questões relacionadas a raça e gênero. Poucos trabalhos discutem como as ocupações dos pais e outros fatores familiares se relacionam com o desempenho da criança na primeira infância. Ao desagregar as variáveis que compõem o *nível socioeconômico* existente, é possível discriminar possíveis mecanismos de transmissão intergeracional de conhecimento e habilidades, contribuindo para a literatura existente e para aqueles que buscam compreender esses possíveis mecanismos de transmissão.

Cunha e Heckman (2007) afirmam que o processo de formação de habilidades é dinâmico e cumulativo, sendo as habilidades adquiridas em um período um fator que influencia diretamente a formação de habilidades futuras, destacando a existência de períodos críticos e sensíveis nesse ciclo de formação, nos quais os investimentos geram maiores retornos, demonstrando que a função de produção de habilidades é crescente e côncava em relação aos investimentos. Utilizando uma tecnologia do tipo CES (Elasticidade de Substituição Constante), generalizam os modelos anteriores ao introduzir o conceito de elasticidade de substituição, denotada por ϕ , que mede o grau de substituíbilidade entre investimentos realizados em diferentes períodos. Quanto mais próximo de zero for ϕ , mais difícil será substituir um investimento inicial por um posterior.

Em uma vertente dessa literatura, mais focada em mobilidade intergeracional, Laband e Lentz (1983) investigam a relação entre a ocupação dos pais e a escolha profissional dos filhos. Concluíram que filhos cujos pais exercem determinadas profissões têm maior probabilidade de seguir carreiras semelhantes, pois recebem de seus genitores conhecimento, habilidades e redes de contato que reduzem as barreiras de entrada nessas ocupações. Além disso, o acesso a informações privilegiadas sobre a profissão dos pais facilita decisões e trajetórias profissionais. Os autores distinguem dois mecanismos principais: um focado na transferência de capital humano e outro na herança de redes de

relacionamento. No primeiro caso, os filhos adquirem desde cedo conhecimentos específicos, o que reduz custos de formação, especialmente em profissões ligadas à agricultura, ao artesanato e a negócios familiares, em que o saber formal tem menor peso. No segundo, os autores introduzem os conceitos de "legado de marca" e "renda monopolista", observando que certas profissões, como dinastias políticas, ocupações sindicalizadas ou empresas familiares, oferecem vantagens associadas ao nome da família, ampliando as chances de inserção profissional por meio de conexões herdadas. Em suas análises, o treinamento formal mostrou-se negativamente relacionado à probabilidade de seguir a ocupação dos pais, sugerindo que a aprendizagem direta no ambiente familiar pode substituir, em parte, a educação formal.

Essa perspectiva é aprofundada por Giannelli e Rapallini (2019), que analisaram como a presença de pais em ocupações relacionadas à matemática (STEM) influencia o desempenho dos filhos nessa disciplina, com base nos dados do Pisa de 2012. As autoras mensuraram a atitude matemática dos alunos por meio de fatores tangíveis, como o auxílio dos pais nos deveres de casa, e intangíveis, como crenças parentais, e ansiedade⁴ em relação à matemática por parte dos alunos. Para lidar com a endogeneidade, utilizaram a presença de um familiar em carreira matemática como variável instrumental. As pesquisadoras identificam quatro mecanismos principais de transmissão intergeracional: (i) conversas e orientações que moldam crenças e valores sobre a importância da matemática; (ii) reforço comunicativo que influencia a motivação instrumental dos filhos ao associar esforço na disciplina ao sucesso futuro; (iii) redução da ansiedade matemática por meio do exemplo dos pais; e (iv) apoio direto nas tarefas escolares, criando um canal tangível de transmissão de conhecimento. Esses mecanismos atuam de forma combinada, gerando um efeito multiplicador no aprendizado. Concluem que a presença de pais em carreiras matemáticas não apenas melhora a motivação e percepção dos filhos, como também reduz barreiras emocionais e cognitivas, resultando em melhor desempenho na disciplina.

Seguindo essa linha de investigações centradas nos resultados do PISA, Wang *et al.* (2023) realizaram uma revisão sistemática da literatura com o objetivo de compreender os fatores que afetam o desempenho dos alunos em matemática nos exames do programa. O estudo analisou pesquisas publicadas entre 2011 e 2021, abrangendo dados de 85 países.

⁴ Giannelli e Rapallini (2019) tratam a ansiedade em matemática como uma série de mecanismos que interferem na forma como o aluno lida com a disciplina. Para medir a ansiedade dos alunos, as autoras utilizaram a variável *PISA ANXMAT*, onde uma pergunta em escala é realizada e tratada como sinal negativo em relação a expectativa de desempenho; ou seja, quanto maior a ansiedade, menor a expectativa de desempenho. A pergunta utilizada pelas autoras é apresentada a seguir: "Pensando em estudar matemática: até que ponto você concorda com as seguintes afirmações?". 1) "Frequentemente me preocupo que será difícil para mim nas aulas de matemática"; 2) "Fico muito tenso quando preciso fazer lição de matemática"; 3) "Fico muito nervoso fazendo problemas de matemática"; 4) "Eu me sinto impotente ao resolver um problema de matemática"; e 5) "Me preocupo em tirar notas ruins em matemática".

Em relação à ocupação dos pais, os autores observaram 22 estudos, dos quais dezessete apontaram uma associação positiva entre o status ocupacional dos pais e o desempenho em matemática.

De forma complementar, Plasman *et al.* (2021) também conduziram uma revisão sistemática com foco na influência da ocupação dos pais sobre o sucesso dos filhos em disciplinas STEM, considerando ainda recortes de gênero e etnia. Os autores identificaram que filhos de pais em ocupações profissionais (de alto status) apresentaram, em média, uma diferença de 55 pontos no desempenho em matemática no PISA quando comparados a filhos de pais em ocupações de menor prestígio. Concluem que a ocupação parental influencia diretamente o desempenho e a persistência dos filhos em ciências e matemática, e destacam a importância de políticas educacionais que considerem o papel da família e do ambiente social para ampliar o acesso e promover um desempenho mais equitativo nessas áreas.

Já Black, Paul J. Devereux e Salvanes (2005) investigaram a relação entre a escolaridade dos pais e a dos filhos utilizando uma mudança na lei de escolaridade obrigatória na Noruega como experimento natural. Embora dados observacionais (OLS) mostrem uma forte correlação entre o nível educacional dos pais e o dos filhos, os autores encontraram pouca evidência de causalidade nessa relação; ou seja, o aumento da escolaridade dos pais nem sempre leva diretamente a mais educação para os filhos. Já Black, Paul G. Devereux e Salvanes (2004) investigaram a relação entre o tamanho da família, a ordem de nascimento e os níveis educacionais dos filhos. O estudo buscou verificar se famílias maiores influenciam negativamente a escolaridade dos filhos e constatou que, de fato, um aumento no número de filhos reduz levemente o nível educacional médio. No entanto, esse efeito é consideravelmente atenuado quando se controlam características familiares fixas, sugerindo que fatores não observáveis afetam a relação. Em relação à ordem de nascimento, os autores concluíram que filhos mais velhos tendem a alcançar níveis educacionais mais elevados, o que pode estar associado à maior atenção e aos recursos concentrados nos primogênitos ao longo do tempo.

Os trabalhos apresentados neste capítulo reforçam a relevância do ambiente familiar para a compreensão da mobilidade intergeracional, indicando que os pais são vetores centrais na transmissão de habilidades, valores e aspirações. O envolvimento parental, seja por meio do apoio direto nas atividades escolares, seja pelo uso dos recursos disponíveis para prover e investir na formação dos filhos, tem um papel essencial na construção do desempenho acadêmico e, por conseguinte, na mobilidade social.

Entretanto, vale destacar que grande parte desses estudos baseou-se em dados do Programa Internacional de Avaliação de Estudantes (PISA), que avalia o desempenho em leitura, matemática e ciências de estudantes com cerca de 15 anos, idade próxima à conclusão da educação básica na maioria dos países. Segundo o INEP (2024), o PISA

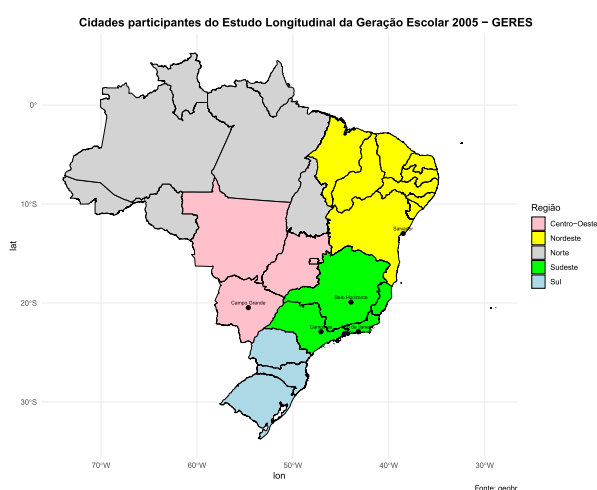
é conduzido com o objetivo de mensurar competências essenciais para a vida adulta e, portanto, fornece uma base robusta para estudos sobre desigualdades educacionais. E aqui se encontra a importância e o diferencial deste estudo, pois se avaliará o desempenho dos estudantes no início da formação básica.

Assim, na seção seguinte, será apresentado o Estudo Longitudinal da Geração Escolar 2005, conhecido como GERES, que complementa essa literatura e de onde partem os microdados utilizados neste estudo.

2.2.1 Banco de Dados – O Estudo Longitudinal da Geração Escolar 2005 - GERES.

Utilizou-se um painel de dados obtidos por meio do Estudo Longitudinal da Geração Escolar 2005 – GERES. O GERES foi desenvolvido em parceria pela Universidade Federal de Minas Gerais (UFMG), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Universidade Estadual do Mato Grosso do Sul (UEMS), Universidade Federal da Bahia (UFBa), Universidade Estadual de Campinas (UNICAMP) e a própria Universidade Federal de Juiz de Fora (UFJF). Foi construído em cinco ondas, sendo a primeira delas de diagnóstico antes do ingresso do estudante no segundo ano de alfabetização, e as demais ondas de avaliação foram aplicadas ano a ano, em todos os quatro primeiros anos iniciais do ensino fundamental, à exceção de Salvador⁵, que iniciou sua participação já no 2º ano do ensino (Brooke; Bonamino, 2011). Foram cinco municípios em diferentes estados do país. As cidades de Belo Horizonte em Minas Gerais, Campinas em São Paulo, Campo Grande em Mato Grosso do Sul, Rio de Janeiro e Salvador na Bahia participaram do processo, conforme se apresenta na figura 1.

Figura 1 – Mapa de distribuição geográfica dos municípios pesquisados no GERES



Fonte: Elaboração própria a partir de dados do GERES (2005-2008).

⁵ Salvador ingressou no 2º ano do ensino fundamental porque, segundo os autores, o acesso ao pré escolar não era tão amplo, motivo pelo qual a alfabetização naquele município se daria no segundo ano, diferentemente dos demais. Por isso, Salvador não participa da 5ª onda.

Foram 303 escolas (ondas 1, 2 e 3), totalizando 312 escolas⁶ acompanhadas durante as ondas 4 e 5. A tabela 1 apresenta a composição de escolas e número de alunos por onda, por município e tipo de rede escolar. Os pesquisadores utilizaram alguns critérios estatísticos e qualitativos⁷ para definição final da população pesquisada, conforme se pode observar na tabela 1.

Tabela 1 – Número de escolas e alunos por município e onda de aplicação (2005–2008)

Município	Rede	Número de Escolas por Onda					Número de Alunos por Onda				
		1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a
Belo Horizonte	Especial	1	1	1	1	1	88	85	87	89	92
	Estadual	20	20	20	20	20	1.521	1.554	1.649	1.670	1.654
	Municipal	20	20	20	20	20	1.840	1.888	2.096	2.044	2.000
	Privada	19	19	19	19	19	641	646	655	653	653
Campinas	Estadual	20	20	20	20	16	1.939	2.017	2.095	2.201	1.860
	Municipal	21	21	21	21	21	1.691	1.761	1.886	2.084	2.203
	Privada	20	20	20	19	18	768	781	748	766	738
Campo Grande	Estadual	19	19	19	18	18	734	825	864	875	833
	Municipal	20	20	20	20	20	2.171	2.247	2.197	2.311	2.115
	Privada	20	20	20	18	14	315	330	260	211	188
Rio de Janeiro	Especial	8	8	8	8	8	730	720	740	764	799
	Municipal	30	30	30	36	35	2.271	2.230	2.224	3.575	3.802
	Privada	30	30	30	30	28	971	949	934	935	778
Salvador	Estadual	11	11	11	10	–	571	575	757	845	–
	Municipal	26	26	26	26	–	2.278	2.332	2.629	2.657	–
	Privada	18	18	18	17	–	495	488	512	477	–
Total		303	303	303	238	–	19.024	19.428	20.333	22.157	17.715

Fonte: Brooke e Bonamino (2011).

Nota: 1^a onda (diagnóstico) e 2^a onda (avaliação) ocorreram em 2005.

Além do desempenho em português e matemática auferido na 1^a onda, os pesquisadores encaminharam um questionário de entrevista aos pais dos estudantes próximo à 2^a onda de avaliação. A tabela 2 apresenta o cronograma de aplicação dos instrumentos utilizados neste trabalho. A fonte das variáveis explicativas foi o questionário respondido pelos pais, apresentado na tabela 2, que foi respondido quase de forma concomitante à 2^a onda de avaliação e, por isso, não é possível acompanhar sua variação no tempo.

⁶ Esse aumento no número de escolas se deu por conta da decisão da Secretaria Municipal de Educação do Rio de Janeiro de migrar cerca de 600 alunos entre escolas no período que antecedeu a 4^a onda. Foi feita uma amostra probabilística de alunos, turmas e escolas com base no Censo Escolar de 2003 (Brooke; Bonamino, 2011),

⁷ Foram excluídas escolas com turmas multisseriadas na 1^a série, privadas com menos de 10 alunos no 2^o ano, públicas com menos de 20 alunos, localizadas na zona rural ou com alunos da 1^a série no horário noturno. Escolas privadas com até 3 turmas no 2^o ano foram incluídas. Escolas estaduais não municipalizadas no Rio de Janeiro também foram excluídas, exceto uma escola técnica. Nas escolas selecionadas, foram aplicados testes cognitivos e questionários a alunos, pais, professores e diretores (Brooke; Bonamino, 2011).

Tabela 2 – Aplicação dos instrumentos de interesse na pesquisa GERES 2005

Instrumento	1ª onda	2ª onda	3ª onda	4ª onda	5ª onda
Teste de Português	X	X	X	X	X
Teste de Matemática	X	X	X	X	X
Questionário da Escola	X				
Questionário da Turma	X				
Questionário de Pais	X				

Fonte: Brooke e Bonamino (2011) .

Os pesquisadores, após recolherem os questionários respondidos pelos pais, analisaram e tabularam os dados de ocupação, utilizando os códigos da CBO - Classificação Brasileira de Ocupações⁸. Para todos os valores vazios, mal esclarecidos, desocupados ou situações diversas que envolviam a especificação correta para a classificação das ocupações, os pesquisadores adotaram o código 0000, código inexistente na CBO, que foi classificado como *X*. Para fins de definição dos grandes grupos de atividades, a conversão das ocupações em colarinho branco, colarinho azul e não especificadas seguiu os grandes grupos da CBO 2002, apresentados na tabela 3. Silva (2019) em seu estudo utilizou-se da mesma analogia. A tabela 3 apresenta a construção das variáveis de colarinho.

Tabela 3 – Conversão das funções da CBO em colarinho branco e colarinho azul

Inicial da CBO	Título da Função	Tipo de Função
0	Forças Armadas, Policiais e Bombeiros Militares	Colarinho Branco
1	Membros Superiores do Poder Público, Dirigentes de Organizações de Interesse Público e de Empresas e Gerentes	Colarinho Branco
2	Profissionais das Ciências e das Artes	Colarinho Branco
3	Técnicos de Nível Médio	Colarinho Branco
4	Trabalhadores de Serviços Administrativos	Colarinho Branco
5	Trabalhadores dos serviços, vendedores dos comércios em lojas e mercados	Colarinho Azul
6	Trabalhadores da agropecuária, florestais e da pesca	Colarinho Azul
7	Trabalhadores que conhecem e trabalham em todas as fases de produção	Colarinho Azul
8	Operadores de máquinas	Colarinho Azul
9	Trabalhadores de manutenção e reparação – ocupações mais elementares	Colarinho Azul
Z	Ocupações mal definidas ou sem emprego atual	Mal especificado ou Desocupado

Fonte: Elaboração própria a partir da Classificação Brasileira de Ocupações (CBO 2002).

A partir da classificação das ocupações, foi possível discriminar o tipo principal de atividade e separar as ocupações em mais artesanais (colarinho azul) ou mais profis-

⁸ A Classificação Brasileira de Ocupações - CBO, instituída por portaria ministerial nº. 397, de 9 de outubro de 2002, tem por finalidade a identificação das ocupações no mercado de trabalho, para fins classificatórios junto aos registros administrativos e domiciliares.

sionais e gerenciais (colarinho branco). Os termos colarinho branco e colarinho azul são conhecidos há muitas décadas e, desde o século passado, são utilizados para distinguir o trabalho artesão e manual das atividades técnicas e gerenciais⁹.

Após a construção da variável de colarinho, na tabela 4 serão apresentadas as variáveis utilizadas neste trabalho. Como as variáveis explicativas são invariantes no tempo, dado o fluxo da coleta de informações apresentadas na tabela 2, será feita uma *cross section* específica para a onda 2, momento em que foram colhidas as variáveis explicativas. A utilização da variável visita à escola pode ser uma boa *proxy* de participação dos pais na vida em comunidade e vetor de avaliação da participação voluntária dos pais (engajamento) com as atividades escolares de seus filhos.

Tabela 4 – Tabela com variáveis utilizadas

Relação das Variáveis Utilizadas				
Variável GERES	Variável Transformada	Tipo de Variável	Descrição da Variável	Período
Variáveis Dependentes				
Proficiência em matemática	—	Contínua	Nota do Aluno <i>i</i> na disciplina de Matemática	2005–2008
Proficiência em português	—	Contínua	Nota do Aluno <i>i</i> na disciplina de Português	2005–2008
Variáveis explicativas relacionadas ao estudante e sua família (invariantes no tempo)				
Ocupação do Pai	Colarinho Branco Pai	Dummy	Agrupadas como colarinho branco, azul ou sem ocupação	2005
Ocupação da Mãe	Colarinho Branco Mãe	Dummy	Agrupadas como colarinho branco, azul ou sem ocupação	2005
Visita à escola	Reunião de Pais	Dummy	Se os pais compareceram à escola em 2005 para reunião	2005
Visita à escola	Conversa com Docente	Dummy	Se os pais compareceram à escola em 2005 para conversar com docente	2005
Visita à escola	Atividade Comunidade	Dummy	Se os pais participaram de atividade em comunidade na escola em 2005	2005
Escolaridade da Mãe	Grau de Escolaridade	Dummy	Grau de escolaridade da mãe	2005
Escolaridade do Pai	Grau de Escolaridade	Dummy	Grau de escolaridade do pai	2005
Variáveis de escola				
Município da Escola	—	Categórica	Município da escola frequentada pelo aluno	2005–2008
Rede	—	Dummy	Se a escola é pública ou privada	2005–2008

Fonte: Elaboração própria usando os dados do GERES (2005–2008).

O GERES é uma fonte relevante de conhecimento acerca de fatores que afetam a educação na primeira infância. A diferença entre esse trabalho e os trabalhos que avaliam a associação das ocupações dos pais no desempenho escolar dos filhos, como os conduzidos pela OECD (2014) e por Giannelli e Rapallini (2019), é que, com o GERES, é possível avaliar situações análogas em um momento inicial da formação escolar do estudante. Dessa forma, o presente estudo contribui para que os responsáveis por políticas públicas possam

⁹ Em 1945, o contra-almirante FG Crisp, em um depoimento junto ao Congresso Americano, disse: “...Os trabalhadores de colarinho azul, em geral, são aqueles que produzem com as mãos. Eles incluem mecânicos, operários, ajudantes, soldadores, eletricitas, armadores... Os trabalhadores de colarinho branco incluem não apenas datilógrafos, estenógrafos e escrivães de arquivo, mas também executivos, engenheiros, cientistas, desenhistas, guardas...” (Statement of Hear Adm F. G. Crisp, United States Navy, Director of Shore Establishments, 1945, página 86).

atuar maximizando o retorno sobre os investimentos, com intervenções já no início do ciclo de formação de habilidades, conforme Cunha e Heckman (2007). Na próxima seção, será detalhada a metodologia utilizada e, na sequência, serão apresentadas as estatísticas descritivas.

2.3 Metodologia

Buscou-se investigar possíveis fatores associados aos pais que possam se relacionar com o desempenho acadêmico de seus filhos. Uma vez que as variáveis explicativas foram coletadas no ano de 2005, de forma próxima ao momento de aplicação da primeira onda de avaliação, segunda onda do GERES, avaliar os resultados nesse período pode auxiliar a compreender de forma mais objetiva e direcionada o quanto essas variáveis estariam relacionadas aos escores de matemática e português obtidos pelos estudantes. A equação 2.1 apresenta a primeira regressão linear proposta, que será estimada pelo método dos mínimos quadrados ordinários (OLS), onde:

$$\log(Y_{id}) = \alpha + \beta \cdot \text{Colarinho}_i + \delta X_i + \varepsilon_i \quad (2.1)$$

Y_{id} é o desempenho do aluno i na disciplina d (matemática ou português) na onda 2 em logaritmo¹⁰. O termo α é a constante da regressão, que indica o desempenho esperado quando todas as variáveis independentes são iguais a zero. A variável Colarinho_i é uma *dummy* que assume valor 1 caso o responsável pela criança exerça ocupação classificada como colarinho branco (conforme Tabela 3) e 0 caso contrário. O vetor X_i inclui características observáveis da família e da escola. O termo de erro ε_i capta variações não observadas.

Após apresentar os modelos utilizados para avaliar a *Cross Section* para a onda 2, serão avaliados esses achados para os anos iniciais do ensino fundamental, ampliando o escopo de uma coorte para o painel longitudinal, acompanhando os estudantes nas 5 ondas. Por questões inerentes à pesquisa realizada, será excluído Salvador do painel, pois o município não participou das 5 ondas. Sob o prisma metodológico, foi necessário lidar com a limitação relacionada aos dados, já que as variáveis explicativas são invariantes no tempo e os modelos com efeitos fixos não são viáveis. Assim, a equação 2.2 representa o modelo Pooled OLS, nossa estimação para os dados em painel.

$$\log(Y_{it}) = \alpha + \beta_1 \cdot \text{DB}_i + \beta' \mathbf{X}_i + \varepsilon_{it} \quad (2.2)$$

¹⁰ Optou-se por usar o logaritmo da nota nas especificações principais, porque essa transformação permite interpretar os coeficientes em termos proporcionais e reduz a dependência da escala original do teste.

A interpretação dos termos está descrita a seguir, onde Y_{it} é a nota do aluno i no tempo t (português ou matemática) em logaritmo¹¹, α é o intercepto, DB_i é uma variável *dummy* que indica se os pais possuem ocupação de colarinho branco e $\mathbf{X}_i$ é o vetor de variáveis de controle, que inclui *dummies* de tipo de escola (se pública ou privada), escolaridade da mãe e do pai, visita dos pais à escola com motivos além das ondas da pesquisa (tempo). ε_{it} é o erro idiossincrático.

Foram encontrados 635 alunos no painel que participaram de duas ondas em uma mesma série, o que representa cerca de 9% do total de alunos do painel estudado (excluindo salvador). Pelo desenho da pesquisa, apenas na primeira série isso aconteceria, que seria o ano em que ocorreria, no início, uma onda diagnóstica e, depois, a primeira onda de avaliação. Criou-se uma variável *dummy* que representa essa repetição de série. Também foram transformadas as variáveis contínuas de notas em variáveis *dummies* (decil e quintil), e, através de um modelo logit, avaliou-se a probabilidade de um estudante filho de pais de colarinho branco ser reprovado ou estar em algum dos extremos em notas em comparação a filhos de pais que não possuem a mesma condição declarada. Dadas as variáveis dependentes criadas, a equação 2.3 apresenta o modelo logit utilizado.

$$\Pr(Y_{it} = 1 \mid X_{it}) = \frac{e^{\alpha + \beta X_{it}}}{1 + e^{\alpha + \beta X_{it}}} \quad (2.3)$$

A especificação do modelo segue a forma funcional da equação 2.3, onde X_{it} representa um vetor de variáveis explicativas, β é o vetor de coeficientes a serem estimados e α é o intercepto. A função logística expressa a probabilidade condicional de ocorrência do evento.

Na Tabela 5, apresentam-se as variáveis utilizadas na regressão por Mínimos Quadrados Ordinários. A variável explicativa de interesse, se os pais possuíam ou não colarinho branco, foi estimada considerando a CBO, conforme se apresenta na tabela 3. Além dela, foram construídas as variáveis *dummies* para a educação da mãe e do pai, o tipo de escola e o motivo da ida a escola que os pais declaram ao longo das respostas.

Foi considerada uma boa *proxy* para representar o nível de engajamento dos pais em ir à escola para participar de reuniões e atividades acadêmicas, atributos que normalmente envolvem todos os pais em um único evento ou em atividades voluntárias. Já ir à

¹¹ Como exercício adicional de robustez, também foram estimadas especificações alternativas dos modelos *pooled OLS*, variando a forma de mensuração da variável dependente. Além da especificação principal, em que as notas são utilizadas em logaritmo, foram realizadas estimações com as notas em nível e com as notas padronizadas em desvios-padrão. Em todas as especificações, manteve-se a mesma estrutura de controles adotada nos modelos principais, incluindo características ocupacionais dos pais, tipo de escola, escolaridade materna e paterna, participação dos responsáveis na vida escolar e efeitos fixos de onda. Esse procedimento permite avaliar se os resultados são sensíveis à escala da variável de desempenho escolar. Os resultados desses exercícios são apresentados no Apêndice e servem como teste de robustez das conclusões obtidas no corpo principal do trabalho.

escola para conversar com um docente¹² pode ser uma resposta a outras situações, como problemas de desempenho, mau comportamento, ou seja, uma *proxy* para ir a escola por um motivo não voluntário.

Tabela 5 – Variáveis transformadas e utilizadas nas regressões

Variáveis	Descrição	Valores da Variável
Variável Dependente		
Desempenho em Matemática	Desempenho dos alunos nos exames de Matemática.	Score de matemática: mínimo 15,31, máximo 227,91.
Desempenho em Português	Desempenho dos alunos nos exames de Português.	Score de português: mínimo 67,15, máximo 183,93.
Variável Independente		
Colarinho Branco (mãe)	Colarinho Branco representa ocupações onde se utiliza de habilidades técnicas e gerenciais ou socialmente tratadas como tal, tais como bombeiros, policiais e seguranças (militares e civis).	<i>Dummy</i> igual a 1 se atende esse critério e 0 se não.
Colarinho Branco (pai)		
Ambos os pais colarinho branco		
Nenhum dos pais colarinho branco		
Colarinho Branco (apenas a mãe)		
Colarinho Branco (apenas o pai)		
Outros controles de aluno/família/escola		
Não concluiu a 4 ^a série – Pai	Nunca estudou ou não concluiu	Variáveis omitidas
Não concluiu a 4 ^a série – Mãe	Nunca estudou ou não concluiu	
Pai concluiu a 4 ^a série	Concluiu o respectivo ano	<i>Dummy</i> = 1 se sim
Pai concluiu o ensino fundamental	Concluiu o ens. fundamental	
Pai concluiu o ensino médio	Concluiu o ensino médio	
Pai concluiu o ensino superior	Concluiu o ens. superior	
Mãe concluiu a 4 ^a série	Concluiu o respectivo ano	<i>Dummy</i> = 1 se sim
Mãe concluiu o ensino fundamental	Concluiu o ens. fundamental	
Mãe concluiu o ensino médio	Concluiu o ensino médio	
Mãe concluiu o ensino superior	Concluiu o ensino superior	
Escola Pública	Se escola é pública	<i>Dummy</i> = 1 se sim
Reunião	Ida a escola para reunião de pais	<i>Dummy</i> = 1 se sim
Conversa Docente	Ida a escola para falar com docente	
Atividade Comunidade	Participação em atividades na escola	

Fonte: Elaboração própria com base nos dados do GERES (2005–2008).

Tendo em vista a importância de se acompanhar os resultados para o momento mais próximo da coleta de informações por parte dos pesquisadores, serão apresentadas na próxima seção as estatísticas descritivas para a Onda 2, a primeira de avaliação, incorporando todos os municípios a partir dessa *cross section*. Na sequência, serão apresentadas as estatísticas descritivas para todas as ondas.

2.4 Estatísticas Descritivas

Esta seção está estruturada em duas etapas. A primeira apresenta uma análise mais aprofundada da Onda 2, com o objetivo de compreender o desempenho inicial dos estudantes, tomando como referência os resultados observados ao final do primeiro ciclo

¹² Esse grupo de alunos merece uma atenção especial e pode ser objeto de futuros estudos. Infelizmente, são poucas as informações de contexto para analisar com maior profundidade a heterogeneidade dos comportamentos que poderiam ter resultado nesse motivo de visita à escola.

de avaliação. Esse recorte permite verificar se, já após o primeiro ano do processo de alfabetização, existem diferenças de rendimento entre os alunos. Além disso, a escolha desse período justifica-se por sua maior proximidade temporal com a coleta das variáveis explicativas utilizadas neste estudo.

2.4.1 Estatísticas Descritivas para a *Cross Section* da Onda 2

As estatísticas descritivas do painel utilizado para a *Cross Section* da onda 2 serão apresentadas na Tabela 6. Como explicitado anteriormente, Salvador (com baixo número de pré-escolas) participou com seus alunos na 2^a série do ensino fundamental (2^o ano de alfabetização), e os demais municípios com alunos terminando a 1^a série do ensino fundamental (2^o ano de alfabetização devido ao acesso pré-escolar anterior). Para esse recorte, considerando a manutenção de Salvador, será extremamente importante controlar o modelo para avaliar o quanto do desempenho estaria relacionado à série de ingresso e, conseqüentemente, pela idade, uma vez que os estudantes de Salvador, estando uma série adiante, tendem a ter um ano a mais de idade em comparação aos alunos dos demais municípios. Ao se apresentarem as estatísticas descritivas para a onda 2 na Tabela 6, observam-se 14.290 estudantes no painel já sem valores ausentes.

As médias dos escores em Português e Matemática foram, respectivamente, 127 e 138 pontos, com desvio padrão de 22,62 e 32,27 pontos, indicando maior dispersão no desempenho em Matemática. A maioria dos estudantes estava matriculada na primeira série e frequentava escolas públicas. Verifica-se ainda elevada frequência de visitas dos pais à escola (95%), sendo predominantes os encontros em reuniões ou conversas com docentes.

Em termos de perfil ocupacional, observa-se que cerca de 27% dos pais e 23% das mães estavam inseridos em ocupações de colarinho branco, e 41% das mães não possuem algum tipo de ocupação padronizada com algum código aceitável da CBO. Já em relação a escolaridade parental, 62% das mães e 64% dos pais estudaram até, no máximo, os anos finais do ensino fundamental.

Tabela 6 – Estatísticas Descritivas – *Cross Section* (Onda 2)

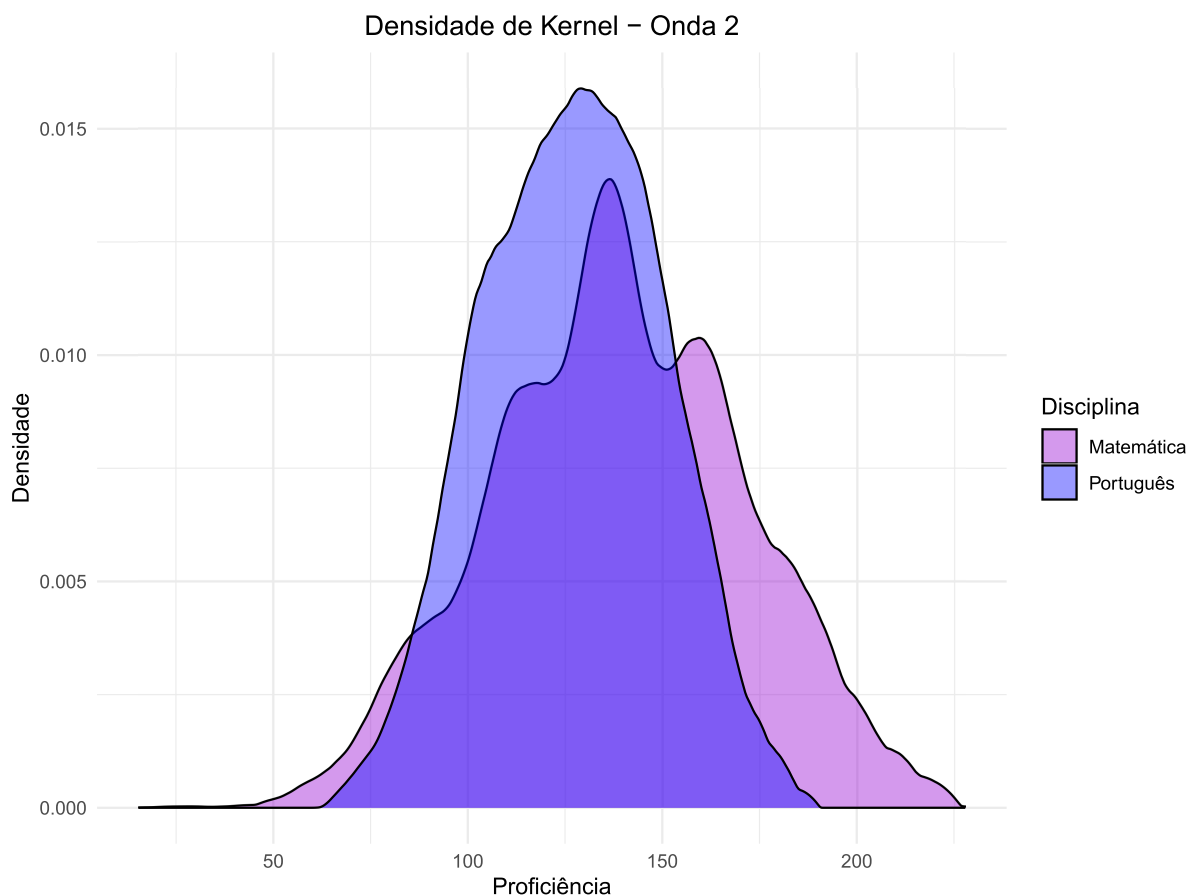
Variável	Média	Desv. Padrão	Mín.	Máx.	Observ.
<i>Variáveis Dependentes</i>					
Score em português (Y_{is})	127,00	22,62	67,15	183,93	14.290
Score em matemática (Y_{is})	138,85	32,27	15,31	227,91	14.290
<i>Variáveis Explicativas</i>					
Série	1,16	0,37	1,00	2,00	14.290
Ida à Escola	0,95	0,21	0,00	1,00	14.290
Reunião	0,81	0,39	0,00	1,00	14.290
Conversa Docente	0,55	0,50	0,00	1,00	14.290
Atividade Comunidade	0,17	0,38	0,00	1,00	14.290
Outra Razão	0,21	0,41	0,00	1,00	14.290
Escola Pública	0,81	0,39	0,00	1,00	14.290
Colarinho Azul Mãe	0,36	0,48	0,00	1,00	14.290
Colarinho Branco Mãe	0,23	0,42	0,00	1,00	14.290
Sem Ocupação Mãe	0,41	0,49	0,00	1,00	14.290
Colarinho Azul Pai	0,49	0,50	0,00	1,00	14.290
Colarinho Branco Pai	0,27	0,45	0,00	1,00	14.290
Sem Ocupação Pai	0,23	0,42	0,00	1,00	14.290
Colarinho Branco ambos	0,11	0,32	0,00	1,00	14.290
Colarinho Branco apenas mãe	0,11	0,32	0,00	1,00	14.290
Colarinho Branco apenas pai	0,16	0,37	0,00	1,00	14.290
Colarinho Branco nenhum	0,61	0,49	0,00	1,00	14.290
<i>Escolaridade da Mãe</i>					
Nunca estudou ou não terminou a 4ª série	0,14	0,34	0,00	1,00	14.290
Terminou a 4ª série	0,26	0,44	0,00	1,00	14.290
Terminou a 8ª série	0,22	0,42	0,00	1,00	14.290
Terminou o ensino médio	0,27	0,44	0,00	1,00	14.290
Terminou a faculdade	0,11	0,31	0,00	1,00	14.290
<i>Escolaridade do Pai</i>					
Nunca estudou ou não terminou a 4ª série	0,14	0,34	0,00	1,00	14.290
Terminou a 4ª série	0,27	0,44	0,00	1,00	14.290
Terminou a 8ª série	0,23	0,42	0,00	1,00	14.290
Terminou o ensino médio	0,26	0,44	0,00	1,00	14.290
Terminou a faculdade	0,10	0,31	0,00	1,00	14.290

Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

De forma descritiva, a Figura 2 apresenta a dispersão das notas em Matemática e Leitura por meio de estimativas de densidade kernel. Esse recurso permite observar visualmente a distribuição do desempenho dos alunos nas duas áreas. Nota-se que a distribuição das notas em Leitura apresenta formato mais simétrico e unimodal, o que sugere menor variabilidade no desempenho dos estudantes em comparação com Matemática.

Em contraste, a distribuição das notas em Matemática apresenta um formato multimodal e mais irregular, indicando maior heterogeneidade no desempenho dos estudantes. Embora ambas as variáveis apresentem distribuições aproximadamente normais, observa-se que elas diferem quanto ao grau de dispersão, conforme corroboram as estatísticas descritivas apresentadas na Tabela 6.

Figura 2 – Boxplot e Densidade de Kernel para matemática e português onda 2



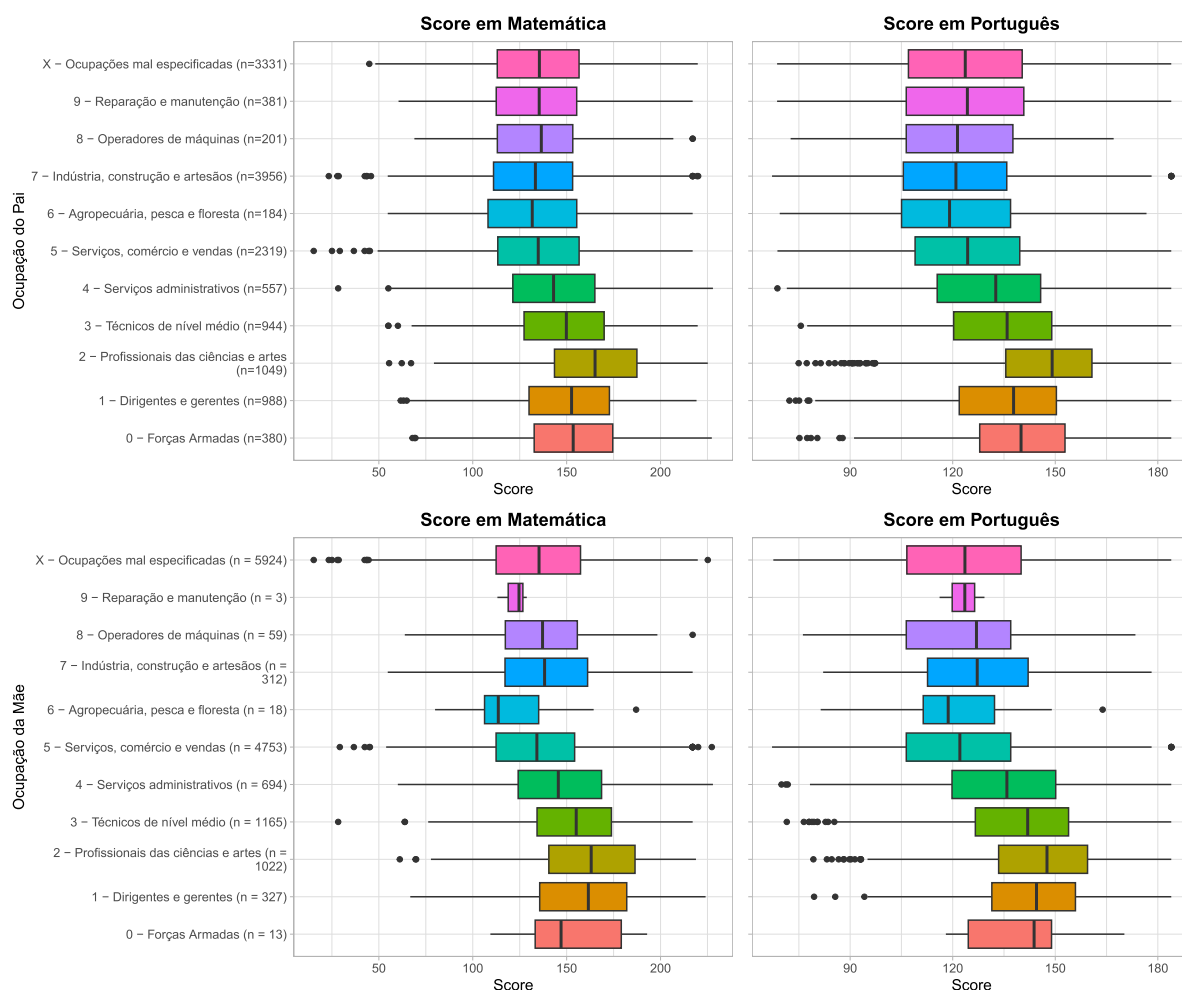
Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

Para aprofundar a análise da dispersão, avaliou-se o comportamento das variáveis dependentes segundo os grandes grupos ocupacionais da CBO, dos quais derivam as variáveis explicativas, em especial a classificação dos estudantes como filhos de trabalhadores de colarinho branco ou não.

A Figura 3 apresenta os escores em Matemática e Língua Portuguesa segundo a ocupação declarada pelos pais e pelas mães no questionário aplicado a essa coorte. Observa-se que os grandes grupos de 0 a 4, definidos neste estudo como correspondentes a ocupações de colarinho branco, apresentam desempenho superior em comparação aos grandes grupos de 5 a 9, bem como ao grupo de ocupações mal especificadas ou sem ocupação.

Ao analisar as distribuições das notas considerando as ocupações, verifica-se que as ocupações posicionadas nos grandes grupos que vão de 0 a 4 possuem desempenho mais alinhado à direita, tanto em matemática quanto em português, sinalizando um desempenho melhor em ambas as disciplinas, com destaque especial no desempenho entre os grandes grupos desse eixo, especialmente aqueles que concentram os profissionais das ciências e das artes, o grande grupo 2.

Figura 3 – Scores em matemática e português considerando os grandes grupos da CBO



Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

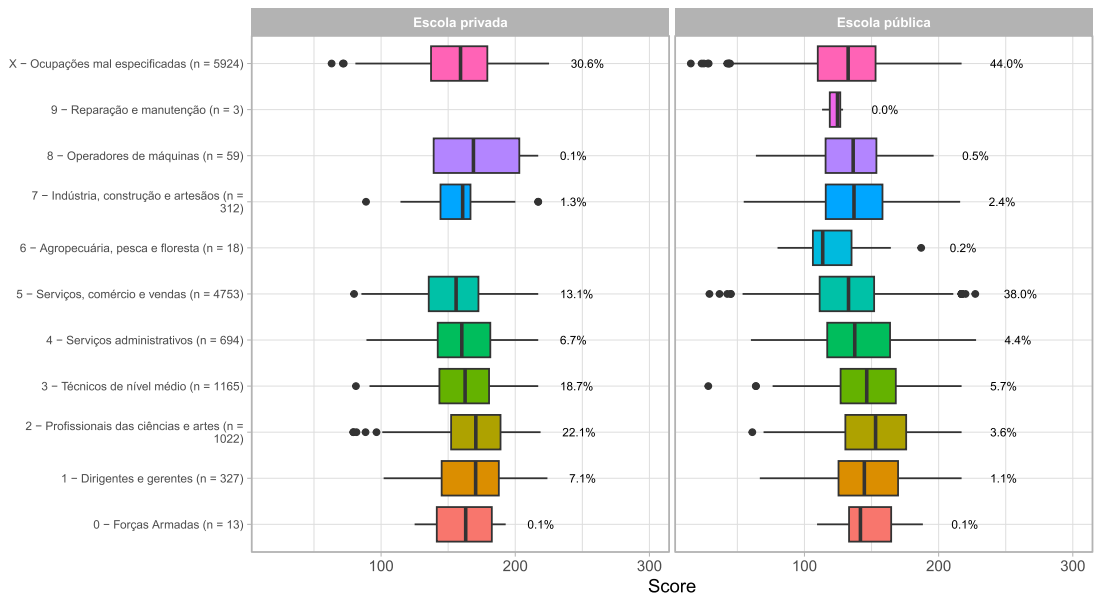
Além de apresentar os resultados para o desempenho escolar, a figura 3 apresenta o número de pais posicionados em cada grande grupo, possibilitando assim avaliar a distribuição dos pais dos alunos nos grandes grupos ocupacionais da CBO. As mães são a maioria no grupo 5, que inclui trabalhadores que produzem serviços pessoais, bem como aqueles que trabalham na intermediação de vendas de bens e serviços (MTE, 2002). Os pais, além de ter uma composição de destaque nesse mesmo grande grupo, são mais representativos no grande grupo 7, que consiste nos profissionais da indústria que desempenham trabalhos "artesaniais", podendo estes utilizar equipamentos rudimentares ou sofisticados para realizar suas atividades. A diferença entre o grande grupo 7 e o 8 é que o 8 está reservado para os operadores de máquinas.

Merecem atenção os filhos de pais que se declararam em ocupações no grande grupo ocupacional 6, reservado a trabalhadores da agropecuária, pesca e florestas. Embora em número menor, estes merecem atenção pelo posicionamento mais à esquerda do desempenho tanto em matemática quanto em português, quando comparados aos demais

grupos ocupacionais da CBO.

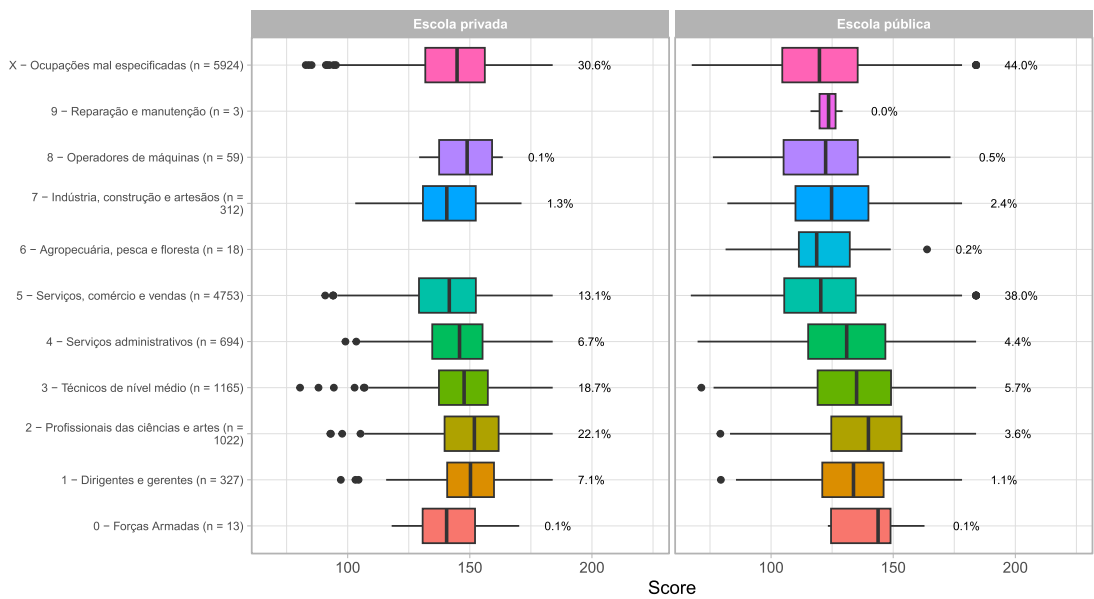
As figuras 4 e 5 apresentam a mesma análise, mas agora colocando lado a lado a disciplina e o tipo de escola para a ocupação da mãe. Essa é uma segregação importante para verificar se, mesmo em escolas públicas e privadas, o comportamento acaba por se repetir.

Figura 4 – Pontuação em Matemática por Ocupação da Mãe e Escola para Onda 2



Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

Figura 5 – Pontuação em Português por Ocupação da Mãe e Tipo de Escola para Onda 2

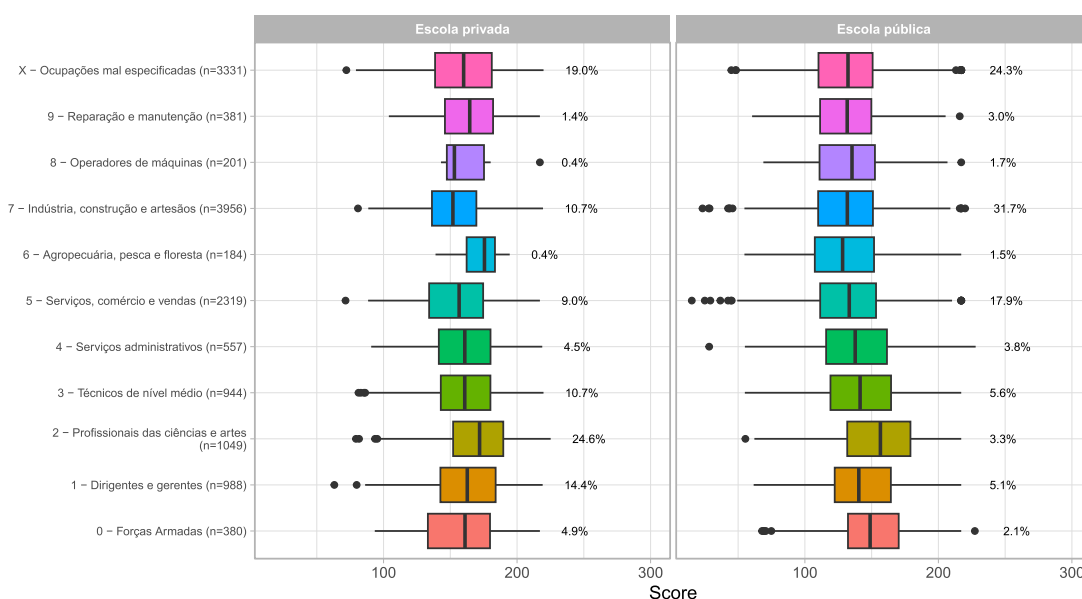


Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

Como se pode perceber nas figuras 4 e 5, existem ocupações para as quais sequer

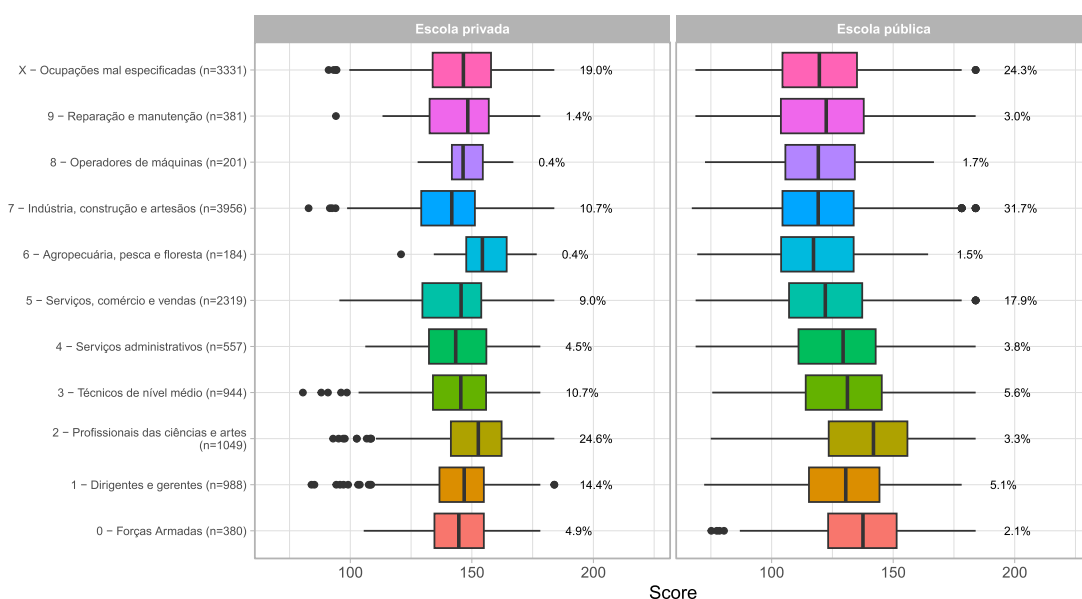
há estudantes matriculados em escolas privadas, como é o caso dos grandes grupos 6 e 9. Já em relação aos demais, percentualmente, o número de estudantes também é bastante baixo. As figuras 6 7 apresentam a pontuação dos estudantes para as disciplinas de matemática e português, mas agora considerando a ocupação dos pais. É perceptível o incremento de estudantes agora distribuídos entre as diversas ocupações.

Figura 6 – Pontuação em Matemática por Ocupação do Pai e Escola para Onda 2



Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

Figura 7 – Pontuação em Português por Ocupação do Pai e Tipo de Escola para Onda 2



Fonte: Elaboração própria a partir de dados do GERES (2005–2008), para a onda 2.

Um aspecto relevante é que, nas escolas privadas, a magnitude das diferenças

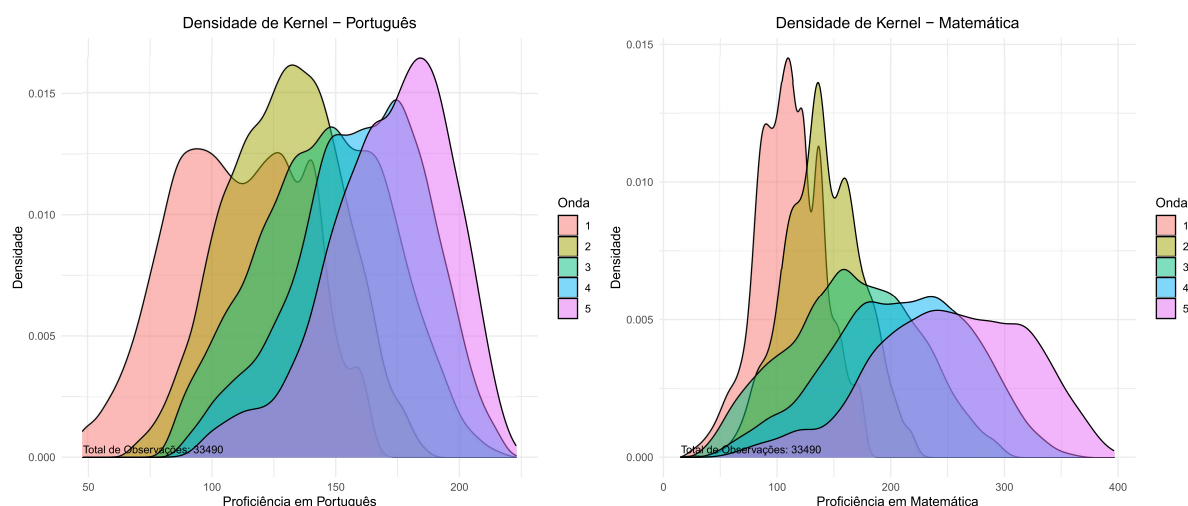
observadas é menos intensa. Esse padrão pode ser identificado em ambas as análises, tanto para Língua Portuguesa quanto para Matemática.

Se encerra esta seção de estatísticas descritivas da Onda 2 destacando que as ocupações declaradas pelos pais tendem a se concentrar em categorias mais bem definidas e mais alinhadas aos grandes grupos da CBO, quando comparadas às ocupações declaradas pelas mães. Entre estas, observa-se maior presença de situações de vulnerabilidade ocupacional, com registros fora dos grandes grupos especificados ou associados a atividades como “do lar”, “cuidadora da casa”, entre outras denominações semelhantes. Em muitos casos, essas categorias indicam ausência de vínculo formal ou de atividade ocupacional remunerada.

2.4.2 Estatísticas Descritivas para o Painel – 5 ondas.

Conforme descrito na seção de metodologia, o objetivo desta etapa é apresentar as estatísticas descritivas referentes ao painel completo, composto pela onda de diagnóstico e pelas quatro ondas de avaliação. Como Salvador participou de apenas quatro ondas, essa localidade foi excluída do painel, a fim de preservar a comparabilidade entre as observações e evitar possíveis distorções nos resultados analisados. Na figura 8, se apresenta a dispersão das notas para os alunos em Matemática e Português.

Figura 8 – Densidade de Kernel - Matemática e Português



Fonte: Elaboração própria a partir de dados do GERES (2005–2008).

Embora se observe uma evolução nas notas ao longo do tempo, verifica-se também um aumento da distância entre os extremos da distribuição. Esse padrão sugere que, com o passar dos anos, os alunos tendem a se concentrar de forma mais acentuada tanto nos decis superiores quanto nos decis inferiores de desempenho. Em outras palavras, as caudas da distribuição parecem se alongar, conferindo-lhe contornos mais próximos

de uma distribuição normal, ao mesmo tempo em que indicam maior dispersão entre os estudantes com melhores e piores resultados.

Esse movimento aponta para um possível aumento das desigualdades de desempenho ao longo das ondas avaliadas, uma vez que parte dos alunos avança para níveis mais elevados de proficiência, enquanto outra parcela permanece relativamente mais atrás. A Tabela 7 apresenta as estatísticas descritivas do painel balanceado.

Tabela 7 – Estatísticas Descritivas – Painel (5 ondas, sem Salvador)

Variável	Média	Desv. Padrão	Mín.	Máx.	Observ.
<i>Variáveis Dependentes</i>					
Score em português (Y_{is})	143,18	33,45	47,80	222,96	33.490
Score em matemática (Y_{is})	175,66	71,55	15,31	396,55	33.490
<i>Variáveis Explicativas</i>					
Série	2,17	1,15	1,00	4,00	33.490
Ida à Escola	0,97	0,17	0,00	1,00	33.490
Reunião	0,85	0,36	0,00	1,00	33.490
Conversa Docente	0,51	0,50	0,00	1,00	33.490
Atividade Comunidade	0,20	0,40	0,00	1,00	33.490
Escola Pública	0,81	0,39	0,00	1,00	33.490
Colarinho Branco Mãe	0,26	0,44	0,00	1,00	33.490
Colarinho Branco Pai	0,30	0,46	0,00	1,00	33.490
Colarinho Branco ambos	0,14	0,34	0,00	1,00	33.490
Colarinho Branco apenas mãe	0,12	0,32	0,00	1,00	33.490
Colarinho Branco apenas pai	0,16	0,37	0,00	1,00	33.490
Colarinho Branco nenhum	0,58	0,49	0,00	1,00	33.490
<i>Escolaridade da mãe</i>					
Nunca estudou ou não terminou a 4 ^a série	0,11	0,31	0,00	1,00	33.490
Terminou a 4 ^a série	0,25	0,43	0,00	1,00	33.490
Terminou a 8 ^a série	0,21	0,41	0,00	1,00	33.490
Terminou o ensino médio	0,29	0,45	0,00	1,00	33.490
Terminou a faculdade	0,14	0,34	0,00	1,00	33.490
<i>Escolaridade do pai</i>					
Nunca estudou ou não terminou a 4 ^a série	0,11	0,32	0,00	1,00	33.490
Terminou a 4 ^a série	0,26	0,44	0,00	1,00	33.490
Terminou a 8 ^a série	0,23	0,42	0,00	1,00	33.490
Terminou o ensino médio	0,27	0,44	0,00	1,00	33.490
Terminou a faculdade	0,13	0,33	0,00	1,00	33.490
<i>Desempenho Relativo</i>					
Repetiu o ano	0,09	0,29	0,00	1,00	33.490
Decil Superior Matemática	0,10	0,30	0,00	1,00	33.490
Decil Inferior Matemática	0,10	0,30	0,00	1,00	33.490
Quintil Superior em Matemática	0,20	0,40	0,00	1,00	33.490
Quintil Inferior Matemática	0,20	0,40	0,00	1,00	33.490
Decil Superior Português	0,10	0,30	0,00	1,00	33.490
Decil Inferior Português	0,10	0,30	0,00	1,00	33.490
Quintil Superior em Português	0,20	0,40	0,00	1,00	33.490
Quintil Inferior Português	0,20	0,40	0,00	1,00	33.490

Fonte: Elaboração própria com base nos dados do GERES (2005–2008).

Após a apresentação das estatísticas descritivas, a próxima seção discute os resultados das estimações. Inicialmente, são apresentados os resultados referentes à Onda 2, seguidos das análises realizadas para o painel longitudinal completo e balanceado.

2.5 Resultados

Esta seção apresenta os resultados das estimações por Mínimos Quadrados Ordinários (MQO), com a inclusão progressiva de controles. As especificações retomam a equação 2.1 e incorporam controles para a série de realização do exame, escolaridade do pai e da mãe, tipo de escola e engajamento familiar com a escola.

Além disso, as variáveis de ocupação de colarinho branco foram combinadas entre pai e mãe, permitindo distinguir alunos com ambos os pais em ocupações de colarinho branco, apenas o pai, apenas a mãe ou nenhum dos pais nessa condição.

Os resultados são apresentados em duas etapas. Primeiro, analisam-se as estimativas para a Onda 2, que reúne todos os municípios avaliados e coincide com o período de coleta das informações ocupacionais dos pais. Em seguida, são discutidos os resultados para o painel longitudinal balanceado.

2.5.1 Resultados para a Onda 2

Os resultados apresentados na Tabela 8 indicam coeficientes estatisticamente significativos na maior parte das especificações estimadas. A exceção ocorre no Modelo 6, referente ao desempenho em Matemática dos estudantes cujas mães, exclusivamente, declaram exercer ocupações de colarinho branco, cujo coeficiente não apresentou significância estatística.

Quando ambos os pais estavam inseridos em ocupações classificadas como de colarinho branco, os filhos apresentaram desempenho médio superior em Matemática e em Língua Portuguesa, com coeficientes de 2,1% e 2,8%, respectivamente, em relação aos demais estudantes. Por outro lado, nos casos em que nenhum dos pais exercia ocupação de colarinho branco, observou-se desempenho inferior de 3,7% em Matemática e 3,1% em Língua Portuguesa.

No caso em que apenas a mãe exercia ocupação de colarinho branco, o coeficiente foi positivo e estatisticamente significativo apenas para a Língua Portuguesa, com ganho médio de 1,1%, e não sendo significativo para o desempenho em Matemática. Já quando apenas o pai ocupava uma posição de colarinho branco, os coeficientes permaneceram estatisticamente significativos e apresentaram magnitude próxima à observada entre estudantes com ambos os pais nessa condição.

Esses resultados sugerem que a ocupação paterna apresenta uma associação mais consistente com o desempenho escolar dos filhos, em linha com as evidências discutidas por Wang *et al.* (2023) e Lee e Borgonovi (2022). De forma complementar, os achados também dialogam com Caplan (2019), ao sugerirem que diferenças no ambiente familiar e no capital humano acumulado podem estar associadas às desigualdades observadas no desempenho educacional.

Tabela 8 – Resultados das Regressões para a *cross section* da onda 2

Variável dependente	Português				Matemática				
	Score em:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Se ambos os pais possuíam atividade de colarinho branco	0.021*** (0.005)					0.028*** (0.007)			
Apenas a Mãe possuía atividade de colarinho branco		0.011** (0.004)					0.007 (0.006)		
Apenas o Pai possuía atividade de colarinho branco			0.02*** (0.004)					0.026*** (0.005)	
Nenhum dos Pais possuíam atividade de colarinho branco				-0.031*** (0.003)					-0.037*** (0.005)
Se aluno de escola pública	-0.088*** (0.004)	-0.089*** (0.004)	-0.088*** (0.004)	-0.085*** (0.004)	-0.099*** (0.006)	-0.101*** (0.006)	-0.1*** (0.006)	-0.097*** (0.006)	
Mãe (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Mãe terminou a 4ª Série	0.036*** (0.005)	0.036*** (0.005)	0.036*** (0.005)	0.036*** (0.005)	0.049*** (0.007)	0.049*** (0.007)	0.049*** (0.007)	0.048*** (0.007)	
Mãe concluiu o ens. fundamental	0.065*** (0.005)	0.065*** (0.005)	0.065*** (0.005)	0.062*** (0.005)	0.083*** (0.007)	0.082*** (0.007)	0.082*** (0.007)	0.079*** (0.007)	
Mãe concluiu o ens. Médio	0.11*** (0.005)	0.109*** (0.005)	0.112*** (0.005)	0.102*** (0.005)	0.125*** (0.007)	0.126*** (0.008)	0.127*** (0.007)	0.116*** (0.008)	
Mãe concluiu o ens. superior	0.148*** (0.007)	0.152*** (0.007)	0.16*** (0.007)	0.14*** (0.007)	0.183*** (0.01)	0.191*** (0.01)	0.199*** (0.01)	0.175*** (0.01)	
Pai (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Pai terminou a 4ª Série	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	0.022*** (0.005)	0.025*** (0.007)	0.025*** (0.007)	0.024*** (0.007)	0.024*** (0.007)	
Pai concluiu o ens. fundamental	0.049*** (0.005)	0.049*** (0.005)	0.048*** (0.005)	0.047*** (0.005)	0.047*** (0.007)	0.047*** (0.007)	0.045*** (0.007)	0.044*** (0.007)	
Pai concluiu o ens. médio	0.076*** (0.005)	0.078*** (0.005)	0.074*** (0.005)	0.07*** (0.005)	0.085*** (0.007)	0.087*** (0.007)	0.082*** (0.007)	0.078*** (0.007)	
Pai concluiu o ens. superior	0.104*** (0.007)	0.111*** (0.007)	0.103*** (0.007)	0.098*** (0.007)	0.122*** (0.01)	0.13*** (0.01)	0.121*** (0.01)	0.116*** (0.01)	
Ida a Escola (Reunião)	0.026*** (0.003)	0.027*** (0.003)	0.026*** (0.003)	0.026*** (0.003)	0.032*** (0.005)	0.033*** (0.005)	0.032*** (0.005)	0.032*** (0.005)	
Ida a Escola (Conversa Docente)	-0.025*** (0.003)	-0.025*** (0.003)	-0.025*** (0.003)	-0.025*** (0.003)	-0.035*** (0.004)	-0.036*** (0.004)	-0.035*** (0.004)	-0.035*** (0.004)	
Ida a Escola (Atividade Comunidade)	0.033*** (0.004)	0.033*** (0.004)	0.033*** (0.004)	0.033*** (0.004)	0.041*** (0.005)	0.042*** (0.005)	0.042*** (0.005)	0.041*** (0.005)	
Série em que o aluno está matriculado.	0.05*** (0.004)	0.05*** (0.004)	0.049*** (0.004)	0.05*** (0.004)	0.117*** (0.005)	0.117*** (0.005)	0.117*** (0.005)	0.118*** (0.005)	
Constante	4.757*** (0.007)	4.758*** (0.007)	4.756*** (0.007)	4.783*** (0.007)	4.811*** (0.009)	4.813*** (0.009)	4.81*** (0.009)	4.842*** (0.01)	
Número de Observações	14290	14290	14290	14290	14290	14290	14290	14290	
R² Ajustado	0.281	0.28	0.281	0.284	0.211	0.21	0.211	0.213	

Fonte: Elaboração própria a partir de dados do GERES (2005-2008).

Nota 1 -> Significância: `\*\*\*` $p < 0.01$; `\*\*` $p < 0.05$; `\*` $p < 0.1$; espaço vazio: $p \geq 0.1$.

Nota 2 -> Desvio padrão entre parênteses;

Nota 3 -> Nos modelos 1 e 5 tem-se a relação entre a variável explicativa de interesse, se ambos os pais exerciam atividade de colarinho branco e a variável dependente Score em Matemática e Português. Nos modelos 2,3, 6 e 7 se avaliou quando apenas um dos pais exercia atividade de colarinho branco. Nos modelos 4 e 8 avaliou-se a situação quando nenhum dos pais exerciam atividades de colarinho branco.

Entre os controles incluídos nesse recorte, destaca-se a variável referente à série do aluno. Estar matriculado na segunda série do ensino fundamental esteve associado a um incremento de aproximadamente 5% na nota de Língua Portuguesa e de quase 12% na nota de Matemática. Embora todos os alunos tenham vivenciado dois anos de alfabetização, aqueles matriculados na segunda série tendem, pela organização do sistema educacional brasileiro, a ser mais velhos. Esse resultado sugere que a permanência no

ensino fundamental pode estar associada ao desenvolvimento de habilidades capturadas pelos testes aplicados. Além disso, o achado levanta uma possível crítica à efetividade da etapa pré-escolar, que parece não produzir ganhos equivalentes aos observados no primeiro ano de alfabetização no ensino fundamental.

Os resultados dialogam com Giannelli e Rapallini (2019) e com a publicação da OECD (2014), que identificaram melhor desempenho acadêmico em Matemática entre filhos de pais inseridos em ocupações STEM ou em ocupações profissionais quando comparados aos demais estudantes. Este estudo avança ao incorporar controles adicionais, como escolaridade dos pais, tipo de escola e engajamento familiar com as atividades escolares dos filhos.

Em relação ao tipo de escola, os resultados indicam que alunos matriculados em escolas públicas apresentam desempenho inferior ao observado entre alunos de escolas privadas. Quanto à escolaridade dos pais, verifica-se uma associação positiva entre maior escolaridade parental e desempenho escolar dos filhos. No entanto, a magnitude dos coeficientes é maior para a escolaridade das mães, quando comparada à escolaridade dos pais. Esse resultado pode ajudar a explicar a ausência de significância estatística da variável de ocupação de colarinho branco da mãe no Modelo 6, uma vez que parte da associação entre características maternas e desempenho escolar pode estar sendo capturada pela escolaridade.

No que se refere ao engajamento familiar, a ida dos pais à escola para participação em reuniões ou atividades com a comunidade escolar apresenta associação positiva com o desempenho dos filhos. Em contrapartida, quando o comparecimento à escola ocorre para conversas com a coordenação ou com docentes, a relação estimada torna-se negativa. Esse resultado sugere que diferentes motivos de comparecimento à escola capturam dimensões distintas da relação família-escola, uma vez que, enquanto a participação em atividades regulares pode refletir maior engajamento preventivo, as visitas motivadas por conversas com a equipe escolar podem estar associadas a dificuldades acadêmicas ou comportamentais já existentes.

Encerrada a análise dos resultados para a Onda 2, passa-se à apresentação das estimativas para o painel longitudinal, composto pelas cinco ondas avaliadas. Nesta etapa, serão apresentados os resultados obtidos nos modelos Pooled OLS, apresentados na Tabela 9. A estimação ocorreu com os dados em painel balanceado, reduzindo o risco de distorções associadas à ausência sistemática de observações ao longo do tempo.

2.5.2 Resultados obtidos nas regressões para o painel com as 5 ondas

Se inicia esta seção de resultados apresentando na tabela 9 os resultados das primeiras estimações, realizadas através de um pooled OLS, apresentado na equação 2.2.

Tabela 9 – Estimação Pooled OLS, painel balanceado para as 5 ondas.

Variável dependente em:	Score	Português				Matemática		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Se ambos os pais possuíam atividade de colarinho branco	0.017*** (0.003)				0.025*** (0.005)			
Apenas a Mãe possuía atividade de colarinho branco		0.01*** (0.003)				0.016*** (0.005)		
Apenas o Pai possuía atividade de colarinho branco			0.03*** (0.003)				0.047*** (0.004)	
Nenhum dos Pais possuíam atividade de colarinho branco				-0.039*** (0.002)				-0.06*** (0.004)
Se aluno de escola pública	-0.078*** (0.003)	-0.08*** (0.003)	-0.078*** (0.003)	-0.074*** (0.003)	-0.116*** (0.005)	-0.118*** (0.005)	-0.115*** (0.005)	-0.11*** (0.005)
Mãe (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Mãe terminou a 4ª Série	0.033*** (0.003)	0.033*** (0.003)	0.033*** (0.003)	0.032*** (0.003)	0.069*** (0.006)	0.069*** (0.006)	0.068*** (0.006)	0.067*** (0.006)
Mãe concluiu o ens. fundamental	0.063*** (0.004)	0.062*** (0.004)	0.061*** (0.004)	0.058*** (0.004)	0.117*** (0.006)	0.116*** (0.006)	0.114*** (0.006)	0.109*** (0.006)
Mãe concluiu o ens. Médio	0.104*** (0.004)	0.103*** (0.004)	0.105*** (0.004)	0.092*** (0.004)	0.173*** (0.006)	0.171*** (0.006)	0.175*** (0.006)	0.155*** (0.006)
Mãe concluiu o ens. superior	0.135*** (0.005)	0.137*** (0.005)	0.147*** (0.005)	0.122*** (0.005)	0.23*** (0.008)	0.234*** (0.008)	0.25*** (0.008)	0.21*** (0.008)
Pai (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Pai terminou a 4ª Série	0.022*** (0.003)	0.022*** (0.003)	0.021*** (0.003)	0.021*** (0.003)	0.033*** (0.006)	0.033*** (0.006)	0.031*** (0.006)	0.031*** (0.006)
Pai concluiu o ens. fundamental	0.058*** (0.004)	0.058*** (0.004)	0.056*** (0.004)	0.055*** (0.004)	0.081*** (0.006)	0.081*** (0.006)	0.078*** (0.006)	0.076*** (0.006)
Pai concluiu o ens. médio	0.085*** (0.004)	0.086*** (0.004)	0.08*** (0.004)	0.077*** (0.004)	0.126*** (0.006)	0.128*** (0.006)	0.119*** (0.006)	0.113*** (0.006)
Pai concluiu o ens. superior	0.115*** (0.005)	0.121*** (0.005)	0.111*** (0.005)	0.106*** (0.005)	0.181*** (0.008)	0.191*** (0.008)	0.174*** (0.008)	0.167*** (0.008)
Ida a Escola (Reunião)	0.03*** (0.003)	0.03*** (0.003)	0.029*** (0.003)	0.03*** (0.003)	0.056*** (0.004)	0.057*** (0.004)	0.055*** (0.004)	0.056*** (0.004)
Ida a Escola (Conversa Docente)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.038*** (0.003)	-0.038*** (0.003)	-0.038*** (0.003)	-0.037*** (0.003)
Ida a Escola (Atividade Comunidade)	0.032*** (0.002)	0.033*** (0.002)	0.032*** (0.002)	0.031*** (0.002)	0.043*** (0.004)	0.043*** (0.004)	0.043*** (0.004)	0.041*** (0.004)
Factor (Onda = 1)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Factor (Onda = 2)	0.174*** (0.003)	0.174*** (0.003)	0.174*** (0.003)	0.174*** (0.003)	0.24*** (0.005)	0.24*** (0.005)	0.24*** (0.005)	0.24*** (0.005)
Factor (Onda = 3)	0.293*** (0.003)	0.293*** (0.003)	0.293*** (0.003)	0.293*** (0.003)	0.372*** (0.005)	0.372*** (0.005)	0.372*** (0.005)	0.372*** (0.005)
Factor (Onda = 4)	0.391*** (0.003)	0.391*** (0.003)	0.391*** (0.003)	0.391*** (0.003)	0.615*** (0.005)	0.615*** (0.005)	0.615*** (0.005)	0.615*** (0.005)
Factor (Onda = 5)	0.463*** (0.003)	0.463*** (0.003)	0.463*** (0.003)	0.463*** (0.003)	0.819*** (0.005)	0.819*** (0.005)	0.819*** (0.005)	0.819*** (0.005)
Constante	4.584*** (0.005)	4.585*** (0.005)	4.582*** (0.005)	4.617*** (0.006)	4.519*** (0.009)	4.52*** (0.009)	4.516*** (0.009)	4.569*** (0.009)
R² Ajustado	0.569	0.569	0.571	0.573	0.579	0.579	0.58	0.582
Número de Observações	33490	33490	33490	33490	33490	33490	33490	33490

Fonte: Elaboração própria a partir de dados do GERES (2005-2008).

Nota 1 -> Significância: **** $p < 0.01$; *** $p < 0.05$; ** $p < 0.1$; espaço vazio: $p \geq 0.1$.

Nota 2 -> Desvio padrão entre parênteses.

Nota 3 -> Nos modelos 1 e 5 foi estimada a relação entre a variável explicativa se ambos os pais exerciam atividade de colarinho branco e as variáveis dependentes score em matemática e português de seus filhos. Nos modelos 2, 3, 6 e 7 apenas um dos pais exercia as mesmas atividades e nos modelos 4 e 8 o desempenho dos estudantes quando nenhum dos pais exerciam atividades de colarinho branco.

Observa-se que ser filho de pais em ocupações de colarinho branco apresenta associação positiva e estatisticamente significativa com o desempenho escolar. Quando ambos os pais exerciam ocupações dessa natureza, o desempenho dos alunos foi superior em 1,7% em Língua Portuguesa e 2,5% em Matemática. Quando apenas a mãe estava nessa condição, os coeficientes indicaram desempenho superior de 1% em Língua Portuguesa e 1,6% em Matemática. Já nos casos em que apenas o pai exercia ocupação de colarinho branco, os ganhos estimados foram maiores, alcançando 3,0% em Língua Portuguesa e 4,7% em Matemática. Por outro lado, quando nenhum dos pais exercia ocupação de colarinho branco, a direção dos coeficientes se inverteu: o desempenho foi inferior em 3,9%

em Língua Portuguesa e 6,0% em Matemática.

Após a análise do modelo pooled OLS, apresentam-se os resultados para os desfechos relacionados à reprovação e ao posicionamento dos alunos nos quintis e decis da distribuição de desempenho. Para isso, utiliza-se o modelo logit, com o objetivo de investigar a probabilidade de alunos filhos de pais em ocupações de colarinho branco serem reprovados ou figurarem entre os quintis e decis superiores e inferiores de desempenho, em comparação com os estudantes cujos pais não declararam ocupações dessa natureza. A Tabela 10 apresenta os resultados das estimações realizadas por meio do modelo logit especificado na equação 2.3.

Tabela 10 – Estimação Logit para o desempenho escolar

Variável Dependente	Probabilidade de figurar entre os melhores				Probabilidade de figurar entre os piores desempenho				Repetir Série
	Decil Superior (1)	(2)	Quintil Superior (3)	(4)	Decil Inferior (5)	(6)	Quintil Inferior (7)	(8)	
Nenhum dos Pais possuíam atividade de colarinho branco	0.654*** (0.054)	0.687*** (0.052)	0.684*** (0.038)	0.692*** (0.038)	1.474*** (0.053)	1.667*** (0.056)	1.515*** (0.039)	1.682*** (0.04)	1.263*** (0.052)
Se aluno de escola pública	0.577*** (0.047)	0.574*** (0.047)	0.524*** (0.038)	0.507*** (0.038)	4.712*** (0.142)	7.639*** (0.188)	4.35*** (0.089)	5.813*** (0.106)	1.999*** (0.093)
Mãe (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Mãe terminou a 4ª Série	1.274 (0.159)	1.546*** (0.156)	1.417*** (0.092)	1.405*** (0.09)	0.658*** (0.052)	0.707*** (0.052)	0.738*** (0.044)	0.774*** (0.044)	0.784*** (0.056)
Mãe concluiu o ens. fundamental	1.771*** (0.156)	2.112*** (0.154)	1.799*** (0.092)	1.782*** (0.091)	0.46*** (0.062)	0.514*** (0.061)	0.543*** (0.049)	0.579*** (0.049)	0.572*** (0.065)
Mãe concluiu o ens. Médio	3.14*** (0.152)	3.3*** (0.151)	2.75*** (0.091)	2.623*** (0.09)	0.388*** (0.07)	0.32*** (0.074)	0.438*** (0.054)	0.423*** (0.055)	0.425*** (0.073)
Mãe concluiu o ens. superior	5.272*** (0.158)	4.739*** (0.157)	4.64*** (0.099)	3.89*** (0.098)	0.215*** (0.173)	0.187*** (0.195)	0.263*** (0.113)	0.22*** (0.127)	0.442*** (0.126)
Pai (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Pai terminou a 4ª Série	1.29 (0.156)	1.249 (0.148)	1.213** (0.088)	1.192** (0.087)	0.828*** (0.053)	0.809*** (0.052)	0.838*** (0.044)	0.775*** (0.044)	0.923 (0.056)
Pai concluiu o ens. fundamental	1.816*** (0.153)	1.905*** (0.145)	1.585*** (0.088)	1.633*** (0.087)	0.577*** (0.062)	0.519*** (0.062)	0.637*** (0.049)	0.554*** (0.049)	0.579*** (0.066)
Pai concluiu o ens. médio	3.122*** (0.151)	3.109*** (0.143)	2.406*** (0.087)	2.523*** (0.086)	0.533*** (0.071)	0.531*** (0.072)	0.532*** (0.055)	0.483*** (0.055)	0.656*** (0.072)
Pai concluiu o ens. superior	5.05*** (0.156)	4.788*** (0.149)	4.037*** (0.095)	3.816*** (0.094)	0.274*** (0.172)	0.29*** (0.186)	0.331*** (0.11)	0.284*** (0.121)	0.43*** (0.131)
Ida a Escola (Reunião)	1.336*** (0.059)	1.38*** (0.059)	1.394*** (0.046)	1.396*** (0.045)	0.683*** (0.048)	0.695*** (0.049)	0.724*** (0.039)	0.698*** (0.039)	0.699*** (0.048)
Ida a Escola (Conversa Docente)	0.802*** (0.04)	0.759*** (0.039)	0.737*** (0.031)	0.743*** (0.031)	1.205*** (0.038)	1.262*** (0.038)	1.284*** (0.029)	1.239*** (0.029)	1.389*** (0.039)
Ida a Escola (Atividade Comunidade)	1.256*** (0.044)	1.462*** (0.043)	1.276*** (0.036)	1.465*** (0.035)	0.666*** (0.059)	0.722*** (0.059)	0.68*** (0.043)	0.688*** (0.044)	0.579*** (0.061)
Constante	0.023*** (0.187)	0.021*** (0.184)	0.086*** (0.113)	0.088*** (0.112)	0.072*** (0.161)	0.039*** (0.203)	0.15*** (0.108)	0.116*** (0.122)	0.127*** (0.121)
Pseudo-R ²	0.201	0.178	0.193	0.183	0.113	0.131	0.126	0.146	0.073
Número de Observações	33490	33490	33490	33490	33490	33490	33490	33490	33490

Fonte: Elaboração própria a partir de dados do GERES (2005-2008).

Nota 1 -> Significância: **** $p < 0.01$; *** $p < 0.05$; ** $p < 0.1$; espaço vazio: $p \geq 0.1$.

Nota 2 -> Desvio padrão entre parênteses.

Nota 3 -> Nos modelos 1 e 3 foram estimadas as probabilidades para os alunos figurarem nos melhores desempenhos em matemática e nos modelos 2 e 4 para português. Nos modelos 5 e 7 foram estimadas as probabilidades para os alunos figurarem entre os piores desempenhos em matemática e nos modelos 6 e 8 para português. Em ambos se considerou como variável explicativa se ambos os pais não possuíam atividades de colarinho branco. No último modelo estimamos a probabilidade de repetir a mesma série.

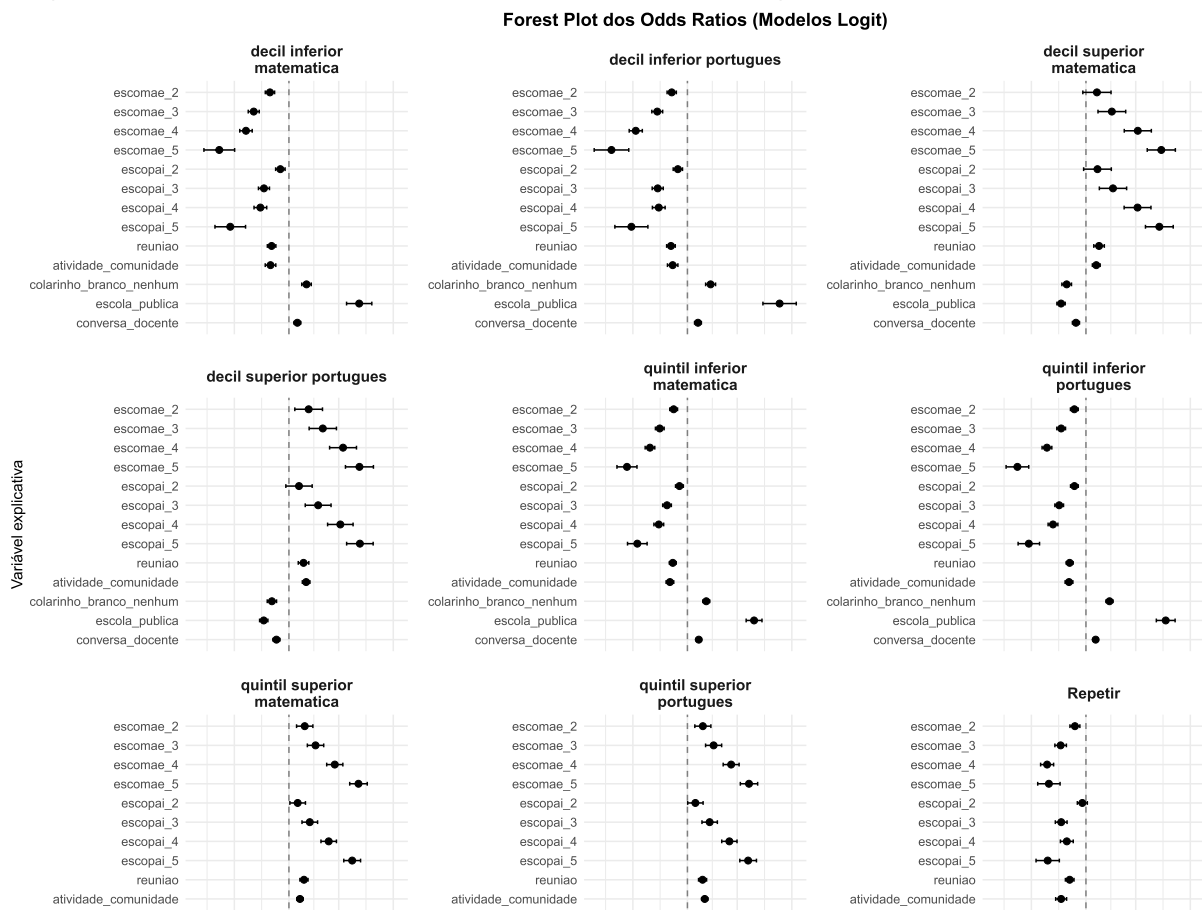
Os coeficientes estimados pelos modelos logit foram convertidos em razões de chances (odds ratios) por meio da exponenciação dos coeficientes log-odds ($OR = e^{\beta}$), o que permite uma interpretação mais direta dos efeitos sobre a probabilidade relativa dos desfechos. O aumento percentual nas chances foi calculado com base na fórmula: $(OR-1) \times 100$, representando a variação proporcional da chance associada a cada variável explicativa.

Assim, ao se analisar o efeito das ocupações parentais por meio dos modelos logit e convertendo os coeficientes em razões de chances, foram identificados, através dos resultados das estimações, que não possuir nenhum dos pais em atividades de colarinho branco

representa uma clara desvantagem educacional para os alunos, em comparação àqueles com pais em ocupações mais qualificadas. Essa condição está associada a menores chances de os estudantes figurarem entre os melhores desempenhos (decil ou quintil superior) em matemática e língua portuguesa, e a maiores chances de estarem entre os desempenhos mais baixos.

Como exemplo, a chance de um aluno sem pais em ocupações de colarinho branco estar no quintil inferior de desempenho é aproximadamente 51,5% maior em matemática e 68,2% maior em português. Resultados semelhantes foram encontrados para o decil inferior. Além disso, a probabilidade de repetência¹³ escolar também é superior nesse grupo, com um aumento de cerca de 26,3%, reforçando o padrão de desvantagem já observado. A figura 9 apresenta as estimativas odds ratios.

Figura 9 – Estimativas de Odds Ratios dos Modelos Logit para o Desempenho Escolar



Fonte: Elaboração Própria com dados do GERES 2005 – 2008.

Os resultados encontrados dialogam com a literatura que aborda o desempenho dos estudantes, analisando características familiares. Quando se utilizaram controles por

¹³ Não se apresenta resultados para evasão escolar, pois, apesar do esforço dos pesquisadores em acompanhar os alunos mesmo em caso de mudança de domicílio escolar, a ausência em uma onda poderia ter ocorrido pelo motivo de transferência entre escolas e estados. Por isso, optou-se por balancear o painel para todas as ondas e analisar apenas alunos que participaram do estudo de forma completa.

escolaridade materna, os resultados se mostraram em linha com Wang *et al.* (2023) e Lee e Borgonovi (2022), ao sugerir uma maior presença ou engajamento das mães no acompanhamento das atividades escolares de seus filhos, mecanismo que corresponde ao fato de que a escolaridade da mãe é uma variável hierárquica de direção igual mas de magnitude maior que a escolaridade dos pais. Quando se olha para a variável explicativa principal dos estudos, a ocupação dos pais, Giannelli e Rapallini (2019) encontraram como resultados que filhos de pais em ocupações STEM lidam melhor com a ansiedade e, com isso, apresentam desempenhos melhores em Matemática quando comparados a filhos de pais que não possuem as mesmas características, resultados também encontrados neste estudo ao se analisar filhos de pais em ocupações menos artesanais. Esses resultados seguem de forma similar aos encontrados por Plasman *et al.* (2021), que, ao conduzir uma revisão sistemática com foco na influência da ocupação dos pais sobre o sucesso dos filhos nos testes PISA, identificaram que filhos de pais em ocupações profissionais (de alto status) apresentaram, em média, uma diferença de 55 pontos no desempenho em matemática no PISA.

Considerando os resultados apresentados, as políticas educacionais precisam considerar o papel da família e do ambiente social como forma de ampliar o acesso e promover um desempenho escolar mais equitativo. Na próxima seção, serão apresentadas as principais conclusões, as limitações e as sugestões para estudos futuros.

2.6 Conclusão

As estimações iniciaram-se com a análise da *cross-section* correspondente à Onda 2 do GERES. Nessa etapa, observou-se que estudantes com ambos os pais inseridos em ocupações de colarinho branco apresentaram desempenho superior em Matemática e Português. Um padrão semelhante foi encontrado na análise longitudinal das cinco ondas. Além da ocupação, a escolaridade dos pais mostrou-se uma dimensão hierárquica relevante: de maneira geral, quanto maior o nível de escolaridade materna ou paterna, maior o desempenho escolar dos filhos.

Os exercícios adicionais apresentados no Apêndice C, reforçam esse resultado. Ao incluir efeitos fixos de escola, os coeficientes foram estimados a partir da comparação entre estudantes da mesma instituição; nas especificações mais restritivas, a comparação foi realizada entre alunos da mesma escola e da mesma série. Mesmo após a inclusão progressiva da escolaridade, da ocupação do outro responsável e dos indicadores de participação familiar, permaneceram gradientes positivos associados às características educacionais e ocupacionais dos pais. A estabilidade das estimativas nas especificações com efeitos fixos de escola-série indica que os resultados não decorrem apenas de diferenças na distribuição dos estudantes entre escolas ou séries. Essas evidências, contudo, representam associações condicionais e não permitem identificar efeitos causais da escolaridade ou da ocupação pa-

rental.

Os resultados são compatíveis com a atuação de mecanismos complementares. A escolaridade dos pais pode refletir maior capital humano, maior capacidade de acompanhar as atividades escolares, transmitir expectativas educacionais e auxiliar na organização dos estudos. A ocupação, por sua vez, pode captar diferenças de renda, estabilidade profissional, condições de trabalho, acesso à informação, redes sociais e recursos disponíveis para o investimento na educação dos filhos. A redução dos coeficientes ocupacionais após a inclusão da escolaridade sugere que essas dimensões estão parcialmente relacionadas, mas a permanência de associações ocupacionais positivas nas especificações completas indica que a ocupação contém informação adicional sobre o ambiente familiar.

Em relação à mãe, os coeficientes associados à escolaridade foram, em geral, superiores aos observados para a escolaridade paterna. Esse resultado é compatível com uma participação mais intensa das mães nas rotinas educacionais e no acompanhamento cotidiano das crianças, embora tal mecanismo não seja diretamente identificado pelas estimações. Da mesma forma, a perda de magnitude ou de significância da variável referente à ocupação materna, especialmente quando apenas a mãe se encontra em ocupação de colarinho branco, pode estar associada a restrições de tempo, à condição de principal provedora do domicílio ou a outras características familiares não observadas. Essas interpretações devem, portanto, ser tomadas como hipóteses compatíveis com os resultados, e não como mecanismos causais comprovados.

Outro resultado relevante foi a associação entre a participação familiar na escola e o desempenho dos estudantes. A presença em reuniões e atividades escolares apresentou coeficientes positivos, sugerindo que o engajamento parental pode constituir uma dimensão importante do ambiente educacional da criança. Por outro lado, o coeficiente negativo associado à conversa com docentes deve ser interpretado com cautela, pois o contato com professores pode ocorrer com maior frequência justamente quando o estudante apresenta dificuldades, configurando possível seleção ou causalidade reversa.

A escola também se mostrou uma dimensão relevante para o desempenho nas disciplinas avaliadas. Nas especificações sem efeitos fixos de escola, estudantes da rede privada apresentaram desempenho superior. Entretanto, essa associação não deve ser interpretada isoladamente como efeito causal do tipo de escola, uma vez que pode refletir diferenças de composição socioeconômica, seleção dos estudantes, infraestrutura, localização e outras características institucionais. Os modelos com efeitos fixos de escola complementam essa análise ao concentrar a identificação nas diferenças entre alunos da mesma instituição.

Em síntese, estudantes cujos pais exerciam atividades gerenciais ou profissionais, agrupadas neste estudo como ocupações de colarinho branco, apresentaram desempenho superior em Matemática e Português quando comparados àqueles cujos pais não declararam ocupações nessa condição. Também foram encontradas evidências de que não possuir

país em ocupações de colarinho branco está associado a uma maior probabilidade de o estudante figurar nos grupos de menor desempenho e a uma menor probabilidade de alcançar os decis e quintis superiores. Esses resultados convergem com o gradiente observado para a escolaridade parental e reforçam a importância do contexto familiar para a compreensão das desigualdades educacionais.

Entre as limitações, destaca-se a abrangência geográfica do GERES. Embora a seleção dos polos tenha seguido critérios técnicos definidos pelos responsáveis pela pesquisa, o estudo contemplou apenas cinco municípios brasileiros, sendo um no Nordeste, um no Centro-Oeste e três no Sudeste, sem representação das regiões Norte e Sul. Essa composição restringe a generalização dos resultados para o conjunto do país, especialmente diante da heterogeneidade econômica, social, institucional e educacional existente entre municípios e comunidades escolares. Estudos futuros deveriam replicar a análise em um número maior de municípios, redes de ensino e comunidades, incorporando diferentes regiões e contextos socioeconômicos.

Também é importante reproduzir o estudo com dados mais recentes. As relações entre ocupação, escolaridade parental, organização familiar, participação na escola e desempenho acadêmico podem ter sido modificadas por transformações no mercado de trabalho, pela expansão educacional, por mudanças tecnológicas e por alterações nas formas de interação entre famílias e escolas. A utilização de bases contemporâneas e de maior cobertura territorial contribuiria para ampliar a validade externa dos resultados e verificar se os gradientes identificados permanecem presentes nas atuais gerações de estudantes.

Outra limitação decorre da ausência de atualização de determinados questionários em todas as ondas, o que dificulta acompanhar algumas características familiares durante todo o ciclo e incluí-las em modelos com efeitos fixos de aluno e de tempo. Além disso, mesmo com os controles observados e os exercícios de robustez, podem permanecer características familiares não mensuradas, como renda, condições do domicílio, expectativas educacionais e disponibilidade de materiais de aprendizagem. Por essa razão, os resultados não devem ser interpretados como identificação causal.

A principal contribuição deste capítulo consiste em mostrar que desigualdades associadas à escolaridade e à inserção ocupacional dos pais já podem ser observadas nas etapas iniciais do ensino fundamental e permanecem ao longo do ciclo de aprendizagem analisado. Ao destacar a atuação conjunta de recursos familiares, capital educacional, condições ocupacionais e engajamento parental, os resultados podem contribuir para o desenho de políticas voltadas à redução das lacunas iniciais de aprendizagem e ao fortalecimento da relação entre famílias, escolas e comunidades.

Referências do capítulo

BLACK, Sandra E.; DEVEREUX, Paul G.; SALVANES, Kjell G. **The more the merrier? The effect of family composition on children's education.** [*S. l.*], 2004. (NBER Working Paper Series, 10720). DOI: 10.3386/w10720. Disponível em: <http://www.nber.org/papers/w10720>.

BLACK, Sandra E.; DEVEREUX, Paul J. **Recent Developments in Intergenerational Mobility.** Massachusetts, 2010. DOI: 10.3386/w15889. Disponível em: <https://doi.org/10.3386/w15889>.

BLACK, Sandra E.; DEVEREUX, Paul J.; SALVANES, Kjell G. Like father, like son? A note on the intergenerational transmission of IQ scores. **Economics Letters**, v. 105, n. 1, p. 138–140, 2009. ISSN 0165-1765. DOI: 10.1016/j.econlet.2009.06.022. Disponível em: <https://doi.org/10.1016/j.econlet.2009.06.022>.

BLACK, Sandra E.; DEVEREUX, Paul J.; SALVANES, Kjell G. Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. **The American Economic Review**, American Economic Association, v. 95, n. 1, p. 437–449, 2005. ISSN 00028282. Disponível em: <http://www.jstor.org/stable/4132690>. Acesso em: 3 ago. 2025.

BROOKE, N.; BONAMINO, A. **Geres 2005: Razões e Resultado de uma pesquisa longitudinal sobre eficácia escolar.** Rio de Janeiro: WalPrint, 2011. Disponível em: https://www.laedpuc.net.br/_files/ugd/fc35a5_a0f50bf2665448af8c06be6b3c089460.pdf.

CARD, David. **Earnings, Schooling, and Ability Revisited.** [*S. l.*], 1994. Acessado em 21 de outubro de 2025. DOI: 10.3386/w4832. Disponível em: <https://doi.org/10.3386/w4832>.

CARD, David. The Causal Effect of Education on Earnings. *In*: ASHENFELTER, Orley; CARD, David (ed.). **Handbook of Labor Economics.** [*S. l.*]: Elsevier, 1999. 3A. p. 1801–1863. DOI: 10.1016/S1573-4463(99)03011-4.

CARD, David. **Using geographic variation in college proximity to estimate the return to schooling.** [*S. l.*], 1993. DOI: 10.3386/w4483. Disponível em: <https://doi.org/10.3386/w4483>.

CUNHA, Flavio; HECKMAN, James. The Technology of Skill Formation. **American Economic Review**, v. 97, n. 2, p. 31–47, 2007. DOI: 10.1257/aer.97.2.31.

DEJONG, Peter Y.; MILTON, Brawer J.; STANLEY, Robin S. Patterns of Female Intergenerational Occupational Mobility: A Comparison with Male Patterns of Intergenerational Occupational Mobility. **American Sociological Review**, v. 36, n. 6, p. 1033–1042, 1971. DOI: 10.2307/2093763. Disponível em: <https://doi.org/10.2307/2093763>.

DUNCAN, Otis D. Occupation Trends and Patterns of Net Mobility in the United States. **Demography**, v. 3, n. 1, p. 1–18, 1966. DOI: 10.2307/2060060. Disponível em: <https://doi.org/10.2307/2060060>.

FRANCO, C.; BROOKE, N.; ALVES, F. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 16, n. 61, p. 625–637, 2008.

GIANNELLI, Gianna Claudia; RAPALLINI, Chiara. Parental occupation and children's school outcomes in math. **Research in Economics**, v. 73, n. 4, p. 293–303, 2019. ISSN 1090-9443. DOI: 10.1016/j.rie.2019.08.003. Disponível em: <https://doi.org/10.1016/j.rie.2019.08.003>.

GU, Xiang *et al.* Like father, like son? Parental input, access to higher education, and social mobility in China. **China Economic Review**, Elsevier, v. 72, p. 101761, 2022.

GUNDERSON, Morley; OREOPOLOUS, Philip. Chapter 3 - Returns to education in developed countries. *In*: BRADLEY, Steve; GREEN, Colin (ed.). **The Economics of Education (Second Edition)**. Second Edition. [S. l.]: Academic Press, 2020. p. 39–51. ISBN 978-0-12-815391-8. DOI: <https://doi.org/10.1016/B978-0-12-815391-8.00003-3>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780128153918000033>.

HECKMAN, James J. Skill Formation and the Economics of Investing in Disadvantaged Children. **Science**, v. 312, p. 1900–1902, 2006.

INEP. **Histórico do Saeb (Sistema de Avaliação da Educação Básica)**. [S. l.: s. n.], 2023. Publicado em 31 ago. 2020. Atualizado em 22 set. 2023. Acesso em: 3 ago. 2025. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/historico>.

INEP. **Programa Internacional de Avaliação de Estudantes (Pisa)**. [S. l.: s. n.], 2024. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa>.

LABAND, David N.; LENTZ, Bernard F. Like Father, like Son: Toward an Economic Theory of Occupational Following. **Southern Economic Journal**, v. 50, n. 2, p. 474–493, 1983. DOI: 10.2307/1058220. Disponível em: <https://doi.org/10.2307/1058220>.

LEE, Jihyun; BORGONOV, Francesca. Relationships between family socioeconomic status and mathematics achievement in OECD and non-OECD countries. **Comparative Education Review**, The University of Chicago Press Chicago, IL, v. 66, n. 2, p. 199–227, 2022.

LENTZ, Bernard F.; LABAND, David N. Why So Many Children of Doctors Become Doctors: Nepotism vs. Human Capital Transfers. **The Journal of Human Resources**, v. 24, n. 3, p. 396–413, 1989. DOI: 10.2307/145820. Disponível em: <https://doi.org/10.2307/145820>.

MINCER, Jacob A. **Schooling, Experience, and Earnings**. New York: Columbia University Press for NBER, 1974. Disponível em: <https://www.nber.org/books-and-chapters/schooling-experience-and-earnings>.

OECD. Do Parents' Occupations Have an Impact on Student Performance? **PISA in Focus**, OECD Publishing, Paris, n. 36, 2014. DOI: 10.1787/5jz8mr7kp026-en. Disponível em: <https://doi.org/10.1787/5jz8mr7kp026-en>.

PATRINOS, Harry Anthony; PSACHAROPOULOS, George. Returns to education in developing countries. *In: THE economics of education*. [S. l.]: Elsevier, 2020. p. 53–64.

PLASMAN, Jay *et al.* Parents' Occupations and Students' Success in STEM Fields: A Systematic Review and Narrative Synthesis. **Adolescent Research Review**, v. 6, n. 1, p. 33–44, 2021. ISSN 2363-8354. DOI: 10.1007/s40894-020-00136-z. Disponível em: <https://doi.org/10.1007/s40894-020-00136-z>.

SILVA, A. T. F. da. Aplicação de Regressão Logística na Associação entre Ocupação e Saúde Cardiovascular - Pesquisa Nacional de Saúde 2019. Projeto (Bacharelado em Estatística) – Departamento de Estatística, Universidade de Brasília. Brasília, 2019.

SOLON, Gary. Intergenerational Mobility in the Labor Market. *In: HANDBOOK of Labor Economics*. [S. l.]: Elsevier Science B.V., 1999. v. 3.A. p. 1761–1800. DOI: 10.1016/S1573-4463(99)03010-22. Disponível em: [https://doi.org/10.1016/S1573-4463\(99\)03010-22](https://doi.org/10.1016/S1573-4463(99)03010-22).

WANG, Xiaofang Sarah *et al.* Factors predicting mathematics achievement in PISA: a systematic review. **Large-scale Assessments in Education**, v. 11, n. 1, p. 24, 2023. ISSN 2196-0739. DOI: 10.1186/s40536-023-00174-8. Disponível em: <https://doi.org/10.1186/s40536-023-00174-8>.

APÊNDICE

Apêndice A - Especificações alternativas para a variável de desempenho

Este apêndice apresenta exercícios adicionais de robustez para as estimações reportadas no texto principal. Com o objetivo de avaliar se os resultados dependem da forma de mensuração da variável de desempenho escolar, optou-se por reestimar os modelos *pooled OLS*, utilizando duas escalas alternativas para as notas, inicialmente em nível e depois em desvios-padrão. Em todas as especificações, foi mantida a mesma estrutura de controles empregada no corpo do trabalho, incluindo as variáveis de ocupação dos pais, o indicador de escola pública, os controles de escolaridade materna e paterna, as medidas de participação dos responsáveis na vida escolar e os efeitos fixos de onda. Assim, a única alteração em relação às estimações principais consiste na escala da variável dependente. No caso das notas padronizadas, a transformação foi realizada por meio da expressão usual do *score-z*, dada por:

$$Z_{it} = \frac{Y_{it} - \bar{Y}}{s_Y}, \quad (2.4)$$

em que Y_{it} representa a nota observada do aluno i na onda t , $\bar{Y}$ corresponde à média da variável de desempenho na amostra considerada e s_Y denota o respectivo desvio-padrão. Desse modo, a variável padronizada passa a ter média zero e desvio-padrão igual a um, permitindo interpretar os coeficientes estimados em frações de desvio-padrão. Essa parametrização é útil para facilitar a comparação entre magnitudes estimadas e para verificar se os resultados permanecem estáveis quando a variável dependente é expressa em uma escala adimensional.

De forma geral, as especificações estimadas no apêndice podem ser representadas pela seguinte equação:

$$\begin{aligned} Y_{it}^m = & \alpha + \beta_1 CB_{it} + \beta_2 EP_{it} + \beta_3 EM_{it} \\ & + \beta_4 EPa_{it} + \beta_5 PF_{it} + \delta_t + \varepsilon_{it}, \end{aligned} \quad (2.5)$$

em que Y_{it}^m representa o desempenho do aluno i na onda t , medido alternativamente em nível ou em desvios-padrão; CB_{it} corresponde ao conjunto de variáveis indicadoras para a condição ocupacional dos pais (colarinho branco); EP_{it} identifica os estudantes matriculados em escola pública; EM_{it} e EPa_{it} representam, respectivamente, os vetores de controles para a escolaridade da mãe e do pai; PF_{it} reúne as variáveis de participação dos responsáveis na vida escolar; δ_t denota os efeitos fixos de onda; e ε_{it} é o termo de erro idiossincrático. Assim, a estrutura do modelo permanece idêntica àquela utilizada no texto principal, alterando-se apenas a forma de mensuração da variável dependente. A Tabela 11 apresenta os resultados com as notas em nível, preservando a escala original das provas.

Tabela 11 – Pooled OLS com notas em nível

Variável dependente	Português				Matemática			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score em:								
Se ambos os pais possuíam atividade de colarinho branco	2.498*** (0.423)				4.692*** (0.871)			
Apenas a Mãe possuía atividade de colarinho branco		1.321*** (0.383)				2.094*** (0.788)		
Apenas o Pai possuía atividade de colarinho branco			3.883*** (0.327)				7.628*** (0.673)	
Nenhum dos Pais possuíam atividade de colarinho branco				-5.222*** (0.296)				-9.852*** (0.610)
Se aluno de escola pública	-10.514*** (0.370)	-10.686*** (0.369)	-10.458*** (0.369)	-10.009*** (0.369)	-19.696*** (0.761)	-20.020*** (0.759)	-19.571*** (0.758)	-18.743*** (0.760)
Mãe (nunca estudou ou não terminou a 4ª série)	Omitida 4.144*** (0.442)	Omitida 4.133*** (0.443)	Omitida 4.054*** (0.442)	Omitida 3.957*** (0.441)	Omitida 8.603*** (0.910)	Omitida 8.586*** (0.911)	Omitida 8.424*** (0.909)	Omitida 8.249*** (0.908)
Mãe terminou a 4ª Série	7.874*** (0.470)	7.802*** (0.471)	7.658*** (0.470)	7.220*** (0.469)	15.564*** (0.967)	15.453*** (0.968)	15.138*** (0.966)	14.330*** (0.967)
Mãe concluiu o ens. fundamental	13.295*** (0.482)	13.201*** (0.489)	13.528*** (0.479)	11.787*** (0.488)	25.341*** (0.991)	25.258*** (1.006)	25.779*** (0.986)	22.493*** (1.005)
Mãe concluiu o ens. Médio	17.974*** (0.640)	18.444*** (0.634)	19.757*** (0.623)	16.343*** (0.634)	36.549*** (1.316)	37.564*** (1.305)	39.973*** (1.283)	33.464*** (1.306)
Mãe concluiu o ens. superior	Omitida 2.648*** (0.437)	Omitida 2.643*** (0.437)	Omitida 2.521*** (0.437)	Omitida 2.473*** (0.436)	Omitida 4.599*** (0.900)	Omitida 4.591*** (0.900)	Omitida 4.350*** (0.899)	Omitida 4.269*** (0.897)
Pai (nunca estudou ou não terminou a 4ª série)	7.332*** (0.464)	7.311*** (0.464)	7.063*** (0.464)	6.903*** (0.463)	11.589*** (0.955)	11.554*** (0.955)	11.060*** (0.955)	10.778*** (0.953)
Pai terminou a 4ª Série	11.277*** (0.484)	11.441*** (0.484)	10.727*** (0.487)	10.201*** (0.487)	20.071*** (0.997)	20.377*** (0.996)	18.978*** (1.001)	18.040*** (1.002)
Pai concluiu o ens. fundamental	15.874*** (0.628)	16.752*** (0.623)	15.369*** (0.625)	14.678*** (0.625)	31.055*** (1.292)	32.631*** (1.283)	30.011*** (1.287)	28.794*** (1.287)
Pai concluiu o ens. médio	3.976*** (0.326)	4.048*** (0.327)	3.932*** (0.326)	3.997*** (0.325)	8.505*** (0.671)	8.629*** (0.672)	8.416*** (0.671)	8.544*** (0.669)
Pai concluiu o ens. superior	-3.227*** (0.234)	-3.238*** (0.234)	-3.195*** (0.234)	-3.135*** (0.233)	-6.478*** (0.481)	-6.500*** (0.481)	-6.414*** (0.481)	-6.304*** (0.480)
Ida a Escola (Reunião)	4.467*** (0.298)	4.488*** (0.298)	4.454*** (0.298)	4.324*** (0.297)	6.887*** (0.613)	6.930*** (0.613)	6.857*** (0.612)	6.617*** (0.612)
Ida a Escola (Conversa Docente)	Omitida 19.039*** (0.369)	Omitida 19.039*** (0.369)	Omitida 19.039*** (0.369)	Omitida 19.039*** (0.368)	Omitida 29.154*** (0.759)	Omitida 29.154*** (0.760)	Omitida 29.154*** (0.758)	Omitida 29.154*** (0.757)
Factor (Onda = 1)	35.519*** (0.369)	35.519*** (0.369)	35.519*** (0.369)	35.519*** (0.368)	54.805*** (0.759)	54.805*** (0.760)	54.805*** (0.758)	54.805*** (0.757)
Factor (Onda = 2)	49.979*** (0.369)	49.979*** (0.369)	49.979*** (0.369)	49.979*** (0.368)	97.265*** (0.759)	97.265*** (0.760)	97.265*** (0.758)	97.264*** (0.757)
Factor (Onda = 3)	61.536*** (0.369)	61.536*** (0.369)	61.536*** (0.368)	61.536*** (0.368)	141.697*** (0.759)	141.697*** (0.760)	141.697*** (0.758)	141.697*** (0.757)
Factor (Onda = 4)	99.041*** (0.667)	99.134*** (0.667)	98.800*** (0.666)	103.416*** (0.706)	90.030*** (1.372)	90.218*** (1.372)	89.545*** (1.371)	98.284*** (1.453)
Factor (Onda = 5)								
Constante	0.592	0.592	0.594	0.596	0.623	0.622	0.624	0.625
R ² Ajustado								
Número de Observações	33490	33490	33490	33490	33490	33490	33490	33490

Fonte: Elaboração própria a partir de dados do GERES (2005-2008). Nota 1: Significância: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$; espaço vazio: $p \geq 0.1$.

Nota 2: Desvio-padrão entre parênteses, em linha inferior ao coeficiente. Nota 3: Nos modelos 1 e 5 foi estimada a relação entre a variável explicativa se ambos os pais exerciam atividade de colarinho branco e as variáveis dependentes score em matemática e português de seus filhos. Nos modelos 2, 3, 6 e 7 apenas quando um dos pais exercia as mesmas atividades e nos modelos 4 e 8 o desempenho dos estudantes quando nenhum dos pais exerciam atividades de colarinho branco.

Essa parametrização permite interpretar os coeficientes diretamente em pontos da proficiência, o que facilita a leitura das magnitudes estimadas. Em comparação com a especificação principal em *log*, os resultados permanecem qualitativamente estáveis, com as variáveis associadas a ocupações de colarinho branco dos pais estando relacionadas a um melhor desempenho de seus filhos, o indicador de escola pública permanece negativamente associado às notas e as *dummies* de escolaridade materna e paterna mantêm o gradiente positivo observado no texto principal. Em particular, as categorias de ocupação de colarinho branco dos pais continuam associadas a melhor desempenho relativo quando comparadas às categorias de referência, ao passo que a categoria em que nenhum dos pais exerce esse tipo de ocupação permanece negativamente associada ao desempenho escolar. Esse padrão é observado tanto em Português quanto em Matemática.

Já a Tabela 12 reporta as mesmas especificações com as notas expressas em desvios-padrão.

Tabela 12 – Pooled OLS com notas em desvio-padrão

Variável dependente	Português				Matemática			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score em:								
Se ambos os pais possuíam atividade de colarinho branco	0.075*** (0.013)				0.066*** (0.012)			
Apenas a Mãe possuía atividade de colarinho branco		0.039*** (0.011)				0.029*** (0.011)		
Apenas o Pai possuía atividade de colarinho branco			0.116*** (0.010)				0.107*** (0.009)	
Nenhum dos Pais possuíam atividade de colarinho branco				-0.156*** (0.009)				-0.138*** (0.009)
Se aluno de escola pública	-0.314*** (0.011)	-0.319*** (0.011)	-0.313*** (0.011)	-0.299*** (0.011)	-0.275*** (0.011)	-0.280*** (0.011)	-0.274*** (0.011)	-0.262*** (0.011)
Mãe (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Mãe terminou a 4ª Série	0.124*** (0.013)	0.124*** (0.013)	0.121*** (0.013)	0.118*** (0.013)	0.120*** (0.013)	0.120*** (0.013)	0.118*** (0.013)	0.115*** (0.013)
Mãe concluiu o ens. fundamental	0.235*** (0.014)	0.233*** (0.014)	0.229*** (0.014)	0.216*** (0.014)	0.218*** (0.014)	0.216*** (0.014)	0.212*** (0.014)	0.200*** (0.014)
Mãe concluiu o ens. Médio	0.397*** (0.014)	0.395*** (0.015)	0.404*** (0.014)	0.352*** (0.015)	0.354*** (0.014)	0.353*** (0.014)	0.360*** (0.014)	0.314*** (0.014)
Mãe concluiu o ens. superior	0.537*** (0.019)	0.551*** (0.019)	0.591*** (0.019)	0.489*** (0.019)	0.511*** (0.018)	0.525*** (0.018)	0.559*** (0.018)	0.468*** (0.018)
Pai (nunca estudou ou não terminou a 4ª série)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Pai terminou a 4ª Série	0.079*** (0.013)	0.079*** (0.013)	0.075*** (0.013)	0.074*** (0.013)	0.064*** (0.013)	0.064*** (0.013)	0.061*** (0.013)	0.060*** (0.013)
Pai concluiu o ens. fundamental	0.219*** (0.014)	0.219*** (0.014)	0.211*** (0.014)	0.206*** (0.014)	0.162*** (0.013)	0.161*** (0.013)	0.155*** (0.013)	0.151*** (0.013)
Pai concluiu o ens. médio	0.337*** (0.014)	0.342*** (0.014)	0.321*** (0.015)	0.305*** (0.015)	0.281*** (0.014)	0.285*** (0.014)	0.265*** (0.014)	0.252*** (0.014)
Pai concluiu o ens. superior	0.475*** (0.019)	0.501*** (0.019)	0.459*** (0.019)	0.439*** (0.019)	0.434*** (0.018)	0.456*** (0.018)	0.419*** (0.018)	0.402*** (0.018)
Ida a Escola (Reunião)	0.119*** (0.010)	0.121*** (0.010)	0.118*** (0.010)	0.119*** (0.010)	0.119*** (0.009)	0.121*** (0.009)	0.118*** (0.009)	0.119*** (0.009)
Ida a Escola (Conversa Docente)	-0.096*** (0.007)	-0.097*** (0.007)	-0.096*** (0.007)	-0.094*** (0.007)	-0.091*** (0.007)	-0.091*** (0.007)	-0.090*** (0.007)	-0.088*** (0.007)
Ida a Escola (Atividade Comunidade)	0.134*** (0.009)	0.134*** (0.009)	0.133*** (0.009)	0.129*** (0.009)	0.096*** (0.009)	0.097*** (0.009)	0.096*** (0.009)	0.092*** (0.009)
Factor (Onda = 1)	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida	Omitida
Factor (Onda = 2)	0.569*** (0.011)	0.569*** (0.011)	0.569*** (0.011)	0.569*** (0.011)	0.407*** (0.011)	0.407*** (0.011)	0.407*** (0.011)	0.407*** (0.011)
Factor (Onda = 3)	1.062*** (0.011)	1.062*** (0.011)	1.062*** (0.011)	1.062*** (0.011)	0.766*** (0.011)	0.766*** (0.011)	0.766*** (0.011)	0.766*** (0.011)
Factor (Onda = 4)	1.494*** (0.011)	1.494*** (0.011)	1.494*** (0.011)	1.494*** (0.011)	1.359*** (0.011)	1.359*** (0.011)	1.359*** (0.011)	1.359*** (0.011)
Factor (Onda = 5)	1.840*** (0.011)	1.840*** (0.011)	1.840*** (0.011)	1.840*** (0.011)	1.980*** (0.011)	1.980*** (0.011)	1.980*** (0.011)	1.980*** (0.011)
Constante	-1.320*** (0.020)	-1.317*** (0.020)	-1.327*** (0.020)	-1.189*** (0.021)	-1.197*** (0.019)	-1.194*** (0.019)	-1.203*** (0.019)	-1.081*** (0.020)
R ² Ajustado	0.592	0.592	0.594	0.596	0.623	0.622	0.624	0.625
Número de Observações	33490	33490	33490	33490	33490	33490	33490	33490

Fonte: Elaboração própria a partir de dados do GERES (2005-2008).

Nota 1 -> Significância: *** p < 0.01; ** p < 0.05; * p < 0.1; espaço vazio: p ≥ 0.1.

Nota 2 -> Desvio-padrão entre parênteses, em linha inferior ao coeficiente.

Nota 3 -> Nos modelos 1 e 5 foi estimada a relação entre a variável explicativa se ambos os pais exerciam atividade de colarinho branco e as variáveis dependentes score em matemática e português de seus filhos. Nos modelos 2, 3, 6 e 7 apenas quando um dos pais exercia as mesmas atividades e nos modelos 4 e 8 o desempenho dos estudantes quando nenhum dos pais exerciam atividades de colarinho branco.

Nessa escala, os coeficientes passam a ser interpretados em frações de desvio-padrão da distribuição da nota, o que facilita a comparação entre magnitudes estimadas e entre disciplinas. Novamente, os resultados permanecem alinhados aos da especificação em log apresentada no texto principal, onde os sinais dos coeficientes centrais se mantêm, bem como a significância estatística das variáveis mais importantes. O indicador de escola pública continua apresentando associação negativa com o desempenho, as escolaridades materna e paterna preservam um gradiente positivo, e as variáveis de participação escolar dos responsáveis mantêm o mesmo padrão qualitativo já observado anteriormente. Assim, a padronização da variável dependente altera a unidade de interpretação dos coeficientes, mas não modifica as conclusões.

Em conjunto, ambas as Tabelas 11 e 12 reforçam a robustez dos resultados principais, onde, mesmo sujeitando a variável dependente a escalas diferentes, o logaritmo apresentado no trabalho principal, nível e desvios-padrão apresentados no apêndice, as

conclusões não dependem da forma de mensuração da variável de desempenho escolar. A especificação em *log* é mantida como principal no corpo do texto por oferecer uma interpretação proporcional mais direta dos coeficientes, enquanto as versões em nível e em desvios-padrão cumprem o papel de demonstrar que os resultados permanecem estáveis quando a variável dependente é expressa em escalas alternativas.

Apêndice B – Robustez com efeitos fixos de aluno e de onda e interações entre características invariantes e tempo

Este apêndice apresenta um exercício adicional de robustez baseado em uma especificação com efeitos fixos de aluno e de onda. O objetivo é verificar se as desigualdades associadas ao contexto familiar e escolar permanecem relevantes mesmo após controlar pela heterogeneidade invariável ao nível do estudante. A especificação estimada pode ser representada por:

$$Y_{it} = \alpha_i + \lambda_t + \beta_2(G_i \times O2_t) + \beta_3(G_i \times O3_t) + \beta_4(G_i \times O4_t) + \beta_5(G_i \times O5_t) + u_{it}$$

Nessa expressão, Y_{it} representa a proficiência padronizada do aluno i na onda t , estimada separadamente para Português e Matemática. O termo α_i representa o efeito fixo do aluno, enquanto λ_t representa o efeito fixo da onda. A variável G_i indica o pertencimento do aluno a determinado grupo familiar ou escolar, como a ocupação dos pais, o tipo de escola ou a escolaridade dos pais. As variáveis $O2_t$, $O3_t$, $O4_t$ e $O5_t$ indicam, respectivamente, as ondas 2, 3, 4 e 5. A onda 1 (diagnóstica) é omitida e funciona como categoria de referência.

Como as características familiares e escolares consideradas neste exercício não variam dentro do aluno ao longo do painel, seus efeitos de nível são absorvidos pelos efeitos fixos individuais. Portanto, os coeficientes das interações não devem ser interpretados como diferenças médias permanentes entre grupos, mas sim, como indicadores da trajetória de desempenho de determinado grupo, sinalizando se essa trajetória se alterou relativamente ao grupo de referência ao longo das ondas, sempre em comparação com a onda 1, de diagnóstico.

A Tabela 13 reúne os resultados para a ocupação dos pais, tipo de escola e parte das interações com a escolaridade materna. Em Português, os diferenciais associados à ocupação dos pais são geralmente modestos e pouco persistentes. Em Matemática, por outro lado, os coeficientes tornam-se positivos e estatisticamente significativos a partir das ondas intermediárias, especialmente quando ambos os pais ou apenas o pai estão em ocupações de colarinho branco. Esse resultado sugere que estudantes oriundos de famílias com inserção ocupacional de colarinho branco apresentam uma trajetória relativamente mais positiva em Matemática ao longo do painel.

As interações entre onda e escola pública apresentam sinais opostos entre as disciplinas. Em Português, os coeficientes são positivos e crescentes, indicando uma evolução relativa mais favorável dos estudantes de escolas públicas em relação à onda inicial. Em Matemática, os coeficientes são negativos e estatisticamente significativos, apontando para uma deterioração relativa da trajetória desses estudantes. Esse contraste sugere que a dinâmica das desigualdades escolares varia conforme a dimensão de aprendizagem analisada.

Ainda na Tabela 13, os resultados para escolaridade materna indicam um gradiente mais forte em Matemática. Para níveis mais elevados de escolaridade da mãe, os coeficientes em Matemática são positivos, crescentes e estatisticamente significativos, o que dialoga com o trabalho de Giannelli e Rapallini (2019). Em Português, embora alguns coeficientes também sejam positivos, as magnitudes são menores e a significância estatística é menos persistente.

A Tabela 14 complementa essa análise ao apresentar a continuação dos resultados para escolaridade materna e as interações entre tempo e escolaridade paterna. Novamente, os resultados mais fortes aparecem em Matemática. Tanto para escolaridade materna quanto para escolaridade paterna, os coeficientes tornam-se positivos e estatisticamente significativos, sobretudo a partir da terceira onda, indicando uma evolução relativamente mais favorável dos estudantes com pais mais escolarizados. Em Português, os coeficientes associados à escolaridade dos pais são menores e menos sistemáticos, embora alguns efeitos positivos apareçam nas ondas posteriores. Isso sugere que a diferenciação temporal em Português existe, mas é menos intensa do que em Matemática.

Os resultados das Tabelas 13 e 14 reforçam que mesmo em uma especificação mais restritiva, na qual toda heterogeneidade invariável ao nível do aluno é absorvida pelos efeitos fixos individuais, permanecem evidências de que características familiares e escolares estão associadas à evolução diferencial da proficiência ao longo do tempo, especialmente em Matemática, disciplina na qual os diferenciais associados à escolaridade dos pais, à ocupação familiar e ao tipo de escola são mais pronunciados.

Tabela 13 – Interações entre características invariantes e tempo

	Variável dependente	
	Port. (z)	Mat. (z)
<i>Painel A. Ocupação dos pais</i>		
O2 × ambos pais: col. branco	-0.018 (0.023)	0.008 (0.016)
O3 × ambos pais: col. branco	0.041 (0.028)	0.084*** (0.026)
O4 × ambos pais: col. branco	0.034 (0.030)	0.100*** (0.029)
O5 × ambos pais: col. branco	0.010 (0.030)	0.137*** (0.032)
O2 × mãe: col. branco	-0.010 (0.022)	-0.012 (0.014)
O3 × mãe: col. branco	0.024 (0.026)	0.068*** (0.023)
O4 × mãe: col. branco	0.013 (0.028)	0.073*** (0.026)
O5 × mãe: col. branco	0.017 (0.027)	0.092*** (0.029)
O2 × pai: col. branco	-0.032* (0.018)	0.015 (0.012)
O3 × pai: col. branco	0.022 (0.022)	0.103*** (0.019)
O4 × pai: col. branco	0.016 (0.023)	0.113*** (0.023)
O5 × pai: col. branco	0.005 (0.023)	0.167*** (0.025)
<i>Painel B. Escola</i>		
O2 × esc. pública	0.066*** (0.017)	-0.053*** (0.012)
O3 × esc. pública	0.091*** (0.021)	-0.181*** (0.019)
O4 × esc. pública	0.148*** (0.022)	-0.136*** (0.022)
O5 × esc. pública	0.192*** (0.021)	-0.156*** (0.023)
<i>Painel C. Esc. da mãe</i>		
O2 × mãe: 4 ^a s.	0.028 (0.024)	-0.002 (0.015)
O3 × mãe: 4 ^a s.	0.049* (0.028)	0.072*** (0.025)
O4 × mãe: 4 ^a s.	0.064** (0.031)	0.094*** (0.029)
O5 × mãe: 4 ^a s.	0.074** (0.032)	0.096*** (0.033)
O2 × mãe: EF	0.044* (0.026)	0.038** (0.016)
O3 × mãe: EF	0.078*** (0.030)	0.140*** (0.027)
O4 × mãe: EF	0.115*** (0.033)	0.179*** (0.031)
O5 × mãe: EF	0.137*** (0.034)	0.218*** (0.035)

Fonte: Elaboração Própria a partir de dados do GERES 2005 – 2008. Notas: Robustez com efeitos fixos de aluno e onda. Erros-padrão agrupados por ID do aluno entre parênteses. Referência temporal: Onda 1. Abreviações: col. branco = colarinho branco; EF = ensino fundamental. Significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Tabela 14 – Interações entre características invariantes e tempo (continuação)

	Variável dependente	
	Port. (z)	Mat. (z)
<i>Painel C. Esc. da mãe (cont.)</i>		
O2 × mãe: EM	-0.003 (0.027)	0.029* (0.016)
O3 × mãe: EM	0.039 (0.031)	0.220*** (0.028)
O4 × mãe: EM	0.051 (0.034)	0.266*** (0.033)
O5 × mãe: EM	0.060* (0.035)	0.315*** (0.037)
O2 × mãe: ES	-0.032 (0.034)	0.023 (0.022)
O3 × mãe: ES	0.025 (0.040)	0.271*** (0.037)
O4 × mãe: ES	0.022 (0.044)	0.341*** (0.042)
O5 × mãe: ES	0.009 (0.044)	0.359*** (0.047)
<i>Painel D. Esc. do pai</i>		
O2 × pai: 4 ^a s.	0.017 (0.024)	-0.003 (0.015)
O3 × pai: 4 ^a s.	0.035 (0.029)	0.065** (0.026)
O4 × pai: 4 ^a s.	0.035 (0.031)	0.067** (0.029)
O5 × pai: 4 ^a s.	0.026 (0.033)	0.033 (0.033)
O2 × pai: EF	0.022 (0.025)	-0.003 (0.016)
O3 × pai: EF	0.067** (0.030)	0.118*** (0.027)
O4 × pai: EF	0.056* (0.033)	0.127*** (0.031)
O5 × pai: EF	0.053 (0.034)	0.111*** (0.035)
O2 × pai: EM	0.021 (0.026)	0.007 (0.017)
O3 × pai: EM	0.114*** (0.031)	0.181*** (0.029)
O4 × pai: EM	0.088*** (0.034)	0.218*** (0.032)
O5 × pai: EM	0.071** (0.035)	0.236*** (0.037)
O2 × pai: ES	0.022 (0.033)	0.014 (0.022)
O3 × pai: ES	0.095** (0.039)	0.259*** (0.036)
O4 × pai: ES	0.084** (0.042)	0.304*** (0.041)
O5 × pai: ES	0.056 (0.042)	0.332*** (0.045)
FE aluno	Sim	Sim
FE onda	Sim	Sim
Obs.	33.490	33.490
R^2	0.8899	0.8857
Within R^2	0.0123	0.1110

Fonte: Elaboração Própria a partir de dados do GERES 2005 – 2008. Notas: Robustez com efeitos fixos de aluno e onda. Erros-padrão agrupados por ID do aluno entre parênteses. Referência temporal: Onda 1. Abreviações: col. branco = colarinho branco; EF = ensino fundamental. Significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Apêndice C – Robustez com efeitos fixos de aluno e de escola, para a *Cross Section* da Segunda Onda

Este apêndice investiga se a escolaridade e a inserção ocupacional dos pais estão associadas ao desempenho dos estudantes em Matemática e Português, comparando alunos que frequentam a mesma escola. A especificação principal é:

$$\log(Y_{isg}) = \alpha_s + \lambda_g + \beta F_{isg} + \gamma X_{isg} + u_{isg}$$

em que Y_{isg} representa a nota do aluno i , matriculado na escola s e na série g ; F_{isg} representa as características educacionais e ocupacionais dos pais; X_{isg} contém os controles familiares; α_s representa o efeito fixo da escola; λ_g representa o efeito fixo da série; e u_{isg} é o termo de erro.

Os efeitos fixos de escola fazem com que os coeficientes sejam estimados a partir da comparação entre alunos da mesma instituição. Dessa forma, características constantes no nível da escola, como localização, dependência administrativa e infraestrutura, são absorvidas pelo modelo.

Os erros-padrão são clusterizados no nível da escola, permitindo a correlação entre os resíduos dos estudantes pertencentes à mesma instituição.

Apêndice C.1 – Progressão das especificações

As especificações são apresentadas de forma progressiva, mantendo a mesma amostra dentro de cada bloco de resultados.

Nos modelos referentes ao pai, as colunas incluem:

- (1) ocupação do pai;
- (2) ocupação e escolaridade do pai;
- (3) características do pai e participação familiar;
- (4) características do pai, da mãe e participação familiar.

Nos modelos referentes à mãe, as colunas incluem:

- (1) ocupação da mãe;
- (2) ocupação e escolaridade da mãe;
- (3) características da mãe e participação familiar;
- (4) características da mãe, do pai e participação familiar.

Nos modelos conjuntos, são incluídas progressivamente as categorias ocupacionais dos pais, a escolaridade paterna, a escolaridade materna e os indicadores de participação familiar.

Apêndice C.2 – Resultados para o pai e para a mãe (separados)

As Tabelas 16 e 15 apresentam os resultados para as características paternas, considerando o logaritmo das notas de Matemática e Português como variáveis dependentes enquanto as tabelas 17 e 18 apresentam os resultados para as características maternas.

A Tabela 15 indica que a escolaridade paterna está positivamente associada ao desempenho em Português. Na especificação completa, os coeficientes das categorias 2 a 5 são 0,019, 0,035, 0,044 e 0,049. O coeficiente da ocupação de colarinho branco do pai diminui de 0,025 para 0,012 com a inclusão dos controles, permanecendo significativo. A redução das estimativas após a inclusão das características maternas mostra que parte da associação paterna está relacionada a características compartilhadas pelo domicílio.

Na Tabela 16, observa-se um padrão semelhante. Os coeficientes das categorias de escolaridade paterna aumentam de 0,018 para 0,052 na especificação completa, indicando um desempenho progressivamente maior nas categorias educacionais mais elevadas. O coeficiente da ocupação de colarinho branco do pai diminui de 0,036 para 0,020, mas permanece estatisticamente significativo após a inclusão da escolaridade materna e dos demais controles familiares.

Já a Tabela 17 mostra uma associação positiva e crescente entre a escolaridade materna e o desempenho em Português. Na especificação completa, os coeficientes das categorias 2 a 5 são 0,028, 0,049, 0,070 e 0,076, respectivamente, todos significativos ao nível de 1%. A ocupação de colarinho branco da mãe, por sua vez, diminui de 0,031 na especificação inicial para 0,008 após a inclusão dos demais controles, sugerindo que parte da associação ocupacional inicial é explicada pela escolaridade e pelas características do pai.

Na Tabela 18, a escolaridade materna também apresenta um gradiente positivo. Na especificação completa, os coeficientes variam de 0,040 para a categoria 2 a 0,094 para a categoria 5. O coeficiente da ocupação de colarinho branco da mãe diminui de 0,035 para 0,009 após a inclusão progressiva dos controles. Embora a magnitude dos coeficientes educacionais seja reduzida quando as características do pai são adicionadas, a associação permanece positiva e estatisticamente significativa.

Em conjunto, os resultados revelam gradientes educacionais e ocupacionais consistentes no desempenho dos estudantes. A redução dos coeficientes associados às ocupações de colarinho branco, após a inclusão da escolaridade dos pais, indica que essas dimensões estão parcialmente relacionadas, uma vez que níveis educacionais mais elevados aumentam a probabilidade de inserção em ocupações de maior qualificação. Ainda assim, a permanência de coeficientes ocupacionais positivos e estatisticamente significativos nas especificações completas sugere que a ocupação dos pais contém informação adicional sobre o ambiente familiar, possivelmente relacionada à renda, à estabilidade profissional, às condições de trabalho, ao acesso a informações e aos recursos disponíveis para a educação

Tabela 15 – Desempenho em Português e características do pai

	Log da nota de Português			
	(1)	(2)	(3)	(4)
colarinho_branco_pai1	0.025*** (0.003)	0.014*** (0.003)	0.014*** (0.003)	0.012*** (0.003)
Escolaridade do pai: categoria 2		0.032*** (0.004)	0.032*** (0.004)	0.019*** (0.004)
Escolaridade do pai: categoria 3		0.058*** (0.004)	0.058*** (0.004)	0.035*** (0.004)
Escolaridade do pai: categoria 4		0.074*** (0.005)	0.074*** (0.005)	0.044*** (0.005)
Escolaridade do pai: categoria 5		0.084*** (0.006)	0.083*** (0.006)	0.049*** (0.006)
reuniao1			0.022*** (0.004)	0.021*** (0.004)
conversa_docente1			-0.021*** (0.003)	-0.022*** (0.003)
atividade_comunidade1			0.021*** (0.004)	0.018*** (0.004)
colarinho_branco_mae1				0.008** (0.003)
Escolaridade da mãe: categoria 2				0.028*** (0.005)
Escolaridade da mãe: categoria 3				0.049*** (0.005)
Escolaridade da mãe: categoria 4				0.070*** (0.005)
Escolaridade da mãe: categoria 5				0.076*** (0.006)
Características da mãe				
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.414	0.427	0.434	0.445
Within R ²	0.005	0.028	0.039	0.057
Within Adjusted R ²	0.005	0.027	0.039	0.057
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria 1 da escolaridade do pai é a categoria de referência. A coluna (1) inclui somente a ocupação do pai; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características da mãe. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 16 – Desempenho em Matemática e características do pai

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
colarinho_branco_pai1	0.036*** (0.004)	0.023*** (0.005)	0.022*** (0.004)	0.020*** (0.004)
Escolaridade do pai: categoria 2		0.036*** (0.007)	0.036*** (0.006)	0.018*** (0.007)
Escolaridade do pai: categoria 3		0.058*** (0.006)	0.058*** (0.006)	0.030*** (0.007)
Escolaridade do pai: categoria 4		0.083*** (0.007)	0.084*** (0.007)	0.048*** (0.008)
Escolaridade do pai: categoria 5		0.093*** (0.009)	0.092*** (0.009)	0.052*** (0.009)
reuniao1			0.024*** (0.006)	0.024*** (0.006)
conversa_docente1			-0.027*** (0.005)	-0.029*** (0.004)
atividade_comunidade1			0.027*** (0.005)	0.025*** (0.005)
colarinho_branco_mae1				0.009* (0.005)
Escolaridade da mãe: categoria 2				0.040*** (0.007)
Escolaridade da mãe: categoria 3				0.066*** (0.008)
Escolaridade da mãe: categoria 4				0.081*** (0.009)
Escolaridade da mãe: categoria 5				0.094*** (0.010)
Características da mãe				
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.320	0.329	0.335	0.343
Within R ²	0.005	0.017	0.026	0.037
Within Adjusted R ²	0.004	0.017	0.025	0.036
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria 1 da escolaridade do pai é a categoria de referência. A coluna (1) inclui somente a ocupação do pai; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características da mãe. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 17 – Desempenho em Português e características da mãe

	Log da nota de Português			
	(1)	(2)	(3)	(4)
colarinho_branco_mae1	0.031*** (0.004)	0.010*** (0.004)	0.009*** (0.003)	0.008** (0.003)
Escolaridade da mãe: categoria 2		0.036*** (0.005)	0.036*** (0.005)	0.028*** (0.005)
Escolaridade da mãe: categoria 3		0.064*** (0.005)	0.064*** (0.005)	0.049*** (0.005)
Escolaridade da mãe: categoria 4		0.090*** (0.005)	0.090*** (0.005)	0.070*** (0.005)
Escolaridade da mãe: categoria 5		0.101*** (0.006)	0.101*** (0.006)	0.076*** (0.006)
reuniao1			0.021*** (0.004)	0.021*** (0.004)
conversa_docente1			-0.022*** (0.003)	-0.022*** (0.003)
atividade_comunidade1			0.019*** (0.004)	0.018*** (0.004)
colarinho_branco_pai1				0.012*** (0.003)
Escolaridade do pai: categoria 2				0.019*** (0.004)
Escolaridade do pai: categoria 3				0.035*** (0.004)
Escolaridade do pai: categoria 4				0.044*** (0.005)
Escolaridade do pai: categoria 5				0.049*** (0.006)
Características do pai				
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.414	0.433	0.440	0.445
Within R ²	0.006	0.037	0.049	0.057
Within Adjusted R ²	0.006	0.037	0.048	0.057
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria 1 da escolaridade da mãe é a categoria de referência. A coluna (1) inclui somente a ocupação da mãe; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características do pai. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** p < 0,01; ** p < 0,05; * p < 0,10.

Tabela 18 – Desempenho em Matemática e características da mãe

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
colarinho_branco_mae1	0.035*** (0.005)	0.011** (0.005)	0.010** (0.005)	0.009* (0.005)
Escolaridade da mãe: categoria 2		0.047*** (0.007)	0.048*** (0.007)	0.040*** (0.007)
Escolaridade da mãe: categoria 3		0.080*** (0.008)	0.081*** (0.008)	0.066*** (0.008)
Escolaridade da mãe: categoria 4		0.102*** (0.008)	0.103*** (0.008)	0.081*** (0.009)
Escolaridade da mãe: categoria 5		0.122*** (0.010)	0.121*** (0.010)	0.094*** (0.010)
reuniao1			0.024*** (0.006)	0.024*** (0.006)
conversa_docente1			-0.029*** (0.004)	-0.029*** (0.004)
atividade_comunidade1			0.026*** (0.006)	0.025*** (0.005)
colarinho_branco_pai1				0.020*** (0.004)
Escolaridade do pai: categoria 2				0.018*** (0.007)
Escolaridade do pai: categoria 3				0.030*** (0.007)
Escolaridade do pai: categoria 4				0.048*** (0.008)
Escolaridade do pai: categoria 5				0.052*** (0.009)
Características do pai				
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.320	0.333	0.339	0.343
Within R ²	0.004	0.023	0.031	0.037
Within Adjusted R ²	0.004	0.022	0.031	0.036
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria 1 da escolaridade da mãe é a categoria de referência. A coluna (1) inclui somente a ocupação da mãe; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características do pai. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

dos filhos.

O gradiente de escolaridade, por sua vez, pode refletir mecanismos associados ao capital humano familiar, como maior capacidade de acompanhamento das atividades escolares, auxílio nos estudos, transmissão de expectativas educacionais e valorização da continuidade escolar. As associações mais elevadas observadas para a escolaridade materna podem estar relacionadas à maior participação das mães nas rotinas educacionais, embora esse mecanismo não seja diretamente identificado. Assim, os resultados sugerem que escolaridade, ocupação e participação familiar representam canais complementares na associação entre o contexto familiar e o desempenho escolar. Como esses mecanismos não são isolados causalmente pelas estimações, os coeficientes devem ser interpretados como associações condicionais.

Apêndice C.2 – Resultados conjuntos e testes de robustez

As Tabelas 19 e 20 incluem simultaneamente as características educacionais e ocupacionais do pai e da mãe.

Os resultados conjuntos para o Português, apresentados na Tabela 20, mostram que as escolaridades materna e paterna permanecem positivamente associadas ao desempenho quando incluídas simultaneamente. Na categoria mais elevada, os coeficientes são 0,077 para a mãe e 0,049 para o pai. Os coeficientes das categorias ocupacionais diminuem com a inclusão das escolaridades, mas permanecem positivos, sendo maiores quando ambos os pais estão em ocupações de colarinho branco. O R^2 *within* aumenta de 0,011 para 0,058 ao longo das especificações.

Na Tabela 19, a escolaridade materna também apresenta coeficientes superiores aos da escolaridade paterna. Na especificação completa, a categoria mais elevada corresponde a 0,094 para a mãe e 0,052 para o pai. O coeficiente associado a ambos os pais em ocupações de colarinho branco diminui de 0,062 para 0,027 após a inclusão das escolaridades e dos indicadores de participação familiar. O R^2 *within* aumenta de 0,008 para 0,037, indicando maior poder explicativo dentro das escolas à medida que os controles são adicionados.

Como exercício de robustez, são estimados modelos com efeitos fixos para cada combinação entre escola e série:

$$\log(Y_{isg}) = \alpha_{sg} + \beta F_{isg} + \gamma X_{isg} + u_{isg}$$

em que α_{sg} representa o efeito fixo específico de cada combinação entre escola e série. Nesta especificação, os coeficientes são estimados com base na comparação entre estudantes matriculados simultaneamente na mesma escola e na mesma série.

Tabela 19 – Desempenho em Matemática e características conjuntas dos pais

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
colarinho_branco_ambos1	0.062*** (0.007)	0.043*** (0.007)	0.029*** (0.007)	0.027*** (0.007)
colarinho_branco_apenas_mae1	0.038*** (0.006)	0.028*** (0.006)	0.011* (0.006)	0.011* (0.006)
colarinho_branco_apenas_pai1	0.038*** (0.005)	0.025*** (0.005)	0.022*** (0.005)	0.021*** (0.005)
Escolaridade do pai: categoria 2		0.035*** (0.006)	0.018*** (0.007)	0.018*** (0.007)
Escolaridade do pai: categoria 3		0.056*** (0.006)	0.030*** (0.007)	0.030*** (0.007)
Escolaridade do pai: categoria 4		0.079*** (0.007)	0.047*** (0.008)	0.048*** (0.008)
Escolaridade do pai: categoria 5		0.088*** (0.009)	0.053*** (0.009)	0.052*** (0.009)
Escolaridade da mãe: categoria 2			0.039*** (0.007)	0.040*** (0.007)
Escolaridade da mãe: categoria 3			0.065*** (0.008)	0.066*** (0.008)
Escolaridade da mãe: categoria 4			0.080*** (0.009)	0.081*** (0.009)
Escolaridade da mãe: categoria 5			0.095*** (0.010)	0.094*** (0.010)
reuniaol				0.024*** (0.006)
conversa_docente1				-0.029*** (0.004)
atividade_comunidade1				0.025*** (0.005)
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.323	0.330	0.337	0.343
Within R ²	0.008	0.019	0.029	0.037
Within Adjusted R ²	0.008	0.018	0.028	0.036
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria ocupacional de referência corresponde aos estudantes para os quais nenhum dos pais está em ocupação de colarinho branco. As categorias 1 de escolaridade do pai e da mãe são as categorias de referência. A coluna (1) inclui as categorias ocupacionais; a coluna (2) acrescenta a escolaridade do pai; a coluna (3) acrescenta a escolaridade da mãe; e a coluna (4) acrescenta participação familiar. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 20 – Desempenho em Português e características conjuntas dos pais

	Log da nota de Português			
	(1)	(2)	(3)	(4)
colarinho_branco_ambos1	0.049*** (0.005)	0.033*** (0.005)	0.020*** (0.005)	0.019*** (0.005)
colarinho_branco_apenas_mae1	0.035*** (0.004)	0.026*** (0.004)	0.010** (0.004)	0.010** (0.004)
colarinho_branco_apenas_pai1	0.028*** (0.003)	0.017*** (0.003)	0.014*** (0.003)	0.013*** (0.003)
Escolaridade do pai: categoria 2		0.031*** (0.004)	0.019*** (0.004)	0.019*** (0.004)
Escolaridade do pai: categoria 3		0.056*** (0.004)	0.035*** (0.004)	0.035*** (0.004)
Escolaridade do pai: categoria 4		0.070*** (0.005)	0.043*** (0.005)	0.044*** (0.005)
Escolaridade do pai: categoria 5		0.079*** (0.006)	0.050*** (0.006)	0.049*** (0.006)
Escolaridade da mãe: categoria 2			0.028*** (0.005)	0.028*** (0.005)
Escolaridade da mãe: categoria 3			0.048*** (0.005)	0.049*** (0.005)
Escolaridade da mãe: categoria 4			0.069*** (0.005)	0.070*** (0.005)
Escolaridade da mãe: categoria 5			0.077*** (0.006)	0.077*** (0.006)
reuniaol				0.021*** (0.004)
conversa_docente1				-0.022*** (0.003)
atividade_comunidade1				0.018*** (0.004)
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.417	0.429	0.438	0.445
Within R ²	0.011	0.031	0.046	0.058
Within Adjusted R ²	0.010	0.030	0.045	0.056
Escola E.F.	Sim	Sim	Sim	Sim
Série E.F.	Sim	Sim	Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. A categoria ocupacional de referência corresponde aos estudantes para os quais nenhum dos pais está em ocupação de colarinho branco. As categorias 1 de escolaridade do pai e da mãe são as categorias de referência. A coluna (1) inclui as categorias ocupacionais; a coluna (2) acrescenta a escolaridade do pai; a coluna (3) acrescenta a escolaridade da mãe; e a coluna (4) acrescenta participação familiar. Todos os modelos incluem efeitos fixos de escola e série. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Os resultados para as estimações de robustez são apresentados nas tabelas 21 e 22 e mostram que os resultados permanecem praticamente inalterados quando os efeitos fixos separados de escola e série são substituídos por efeitos fixos de escola-série. A estabilidade dos coeficientes indica que as associações não são explicadas por diferenças na distribuição dos estudantes entre séries dentro de uma mesma escola.

Em conjunto, os resultados combinados e os testes de robustez indicam que as características educacionais e ocupacionais dos pais permanecem associadas ao desempenho escolar, mesmo quando consideradas simultaneamente e quando a comparação é restrita a estudantes da mesma escola e da mesma série. Esse padrão é compatível com a atuação de mecanismos complementares: a escolaridade pode refletir maior capital humano, capacidade de acompanhamento das atividades escolares e expectativas educacionais mais elevadas, enquanto a ocupação pode captar diferenças de renda, estabilidade profissional, condições de trabalho, acesso à informação e recursos disponíveis para o investimento na educação dos filhos.

A estabilidade dos coeficientes nas especificações com efeitos fixos escola-série reforça que essas associações não decorrem apenas da composição das turmas ou de diferenças entre séries dentro das escolas. Ainda assim, os resultados devem ser interpretados como evidência de gradientes familiares condicionais e não como efeitos causais isolados da escolaridade ou da ocupação dos pais.

Apêndice C.4 – Estratégia e progressão das especificações

Complementando os resultados da análise transversal apresentados no Apêndice C.1, C.2 e C.3, este apêndice segue, agora utilizando as cinco ondas do painel GERES. As estimações são realizadas por meio de modelos agrupados, com efeitos fixos de escola e de onda e erros-padrão agrupados simultaneamente nos níveis do aluno e da escola. A especificação geral é representada por:

$$\log(Y_{ist}) = \alpha_s + \lambda_t + \beta F_{ist} + \gamma X_{ist} + u_{ist}$$

em que Y_{ist} representa a nota do estudante i , matriculado na escola s e observado na onda t ; α_s corresponde aos efeitos fixos de escola; λ_t representa os efeitos fixos de onda; F_{ist} reúne as características educacionais e ocupacionais dos pais; e X_{ist} contém os demais controles familiares. Como não são incluídos os efeitos fixos de aluno, os coeficientes são identificados principalmente pela comparação entre estudantes da mesma escola, acompanhados ao longo das ondas, e não exclusivamente pela variação das características de um mesmo aluno no tempo.

Tabela 21 – Robustez dos resultados de Matemática aos efeitos fixos de escola-série

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
colarinho_branco_ambos1	0.029*** (0.007)	0.027*** (0.007)	0.029*** (0.007)	0.028*** (0.007)
colarinho_branco_apenas_mae1	0.011* (0.006)	0.011* (0.006)	0.011* (0.006)	0.011* (0.006)
colarinho_branco_apenas_pai1	0.022*** (0.005)	0.021*** (0.005)	0.022*** (0.005)	0.021*** (0.005)
Escolaridade do pai: categoria 2	0.018*** (0.007)	0.018*** (0.007)	0.019*** (0.007)	0.018*** (0.007)
Escolaridade do pai: categoria 3	0.030*** (0.007)	0.030*** (0.007)	0.030*** (0.007)	0.030*** (0.007)
Escolaridade do pai: categoria 4	0.047*** (0.008)	0.048*** (0.008)	0.047*** (0.008)	0.048*** (0.007)
Escolaridade do pai: categoria 5	0.053*** (0.009)	0.052*** (0.009)	0.053*** (0.009)	0.052*** (0.009)
Escolaridade da mãe: categoria 2	0.039*** (0.007)	0.040*** (0.007)	0.039*** (0.007)	0.040*** (0.007)
Escolaridade da mãe: categoria 3	0.065*** (0.008)	0.066*** (0.008)	0.064*** (0.008)	0.065*** (0.008)
Escolaridade da mãe: categoria 4	0.080*** (0.009)	0.081*** (0.009)	0.080*** (0.009)	0.081*** (0.009)
Escolaridade da mãe: categoria 5	0.095*** (0.010)	0.094*** (0.010)	0.094*** (0.010)	0.094*** (0.010)
reuniao1		0.024*** (0.006)		0.024*** (0.006)
conversa_docente1		-0.029*** (0.004)		-0.029*** (0.004)
atividade_comunidade1		0.025*** (0.005)		0.025*** (0.005)
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.337	0.343	0.338	0.344
Within R ²	0.029	0.037	0.029	0.037
Within Adjusted R ²	0.028	0.036	0.028	0.036
Escola E.F.	Sim	Sim		
Série E.F.	Sim	Sim		
Escola-série E.F.			Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. As colunas (1) e (2) incluem separadamente efeitos fixos de escola e de série. As colunas (3) e (4) incluem efeitos fixos para cada combinação escola-série. As colunas (2) e (4) acrescentam os controles de participação familiar. Erros-padrão clusterizados por escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 22 – Robustez dos resultados de Português aos efeitos fixos de escola-série

	Log da nota de Português			
	(1)	(2)	(3)	(4)
colarinho_branco_ambos1	0.020*** (0.005)	0.019*** (0.005)	0.020*** (0.005)	0.019*** (0.004)
colarinho_branco_apenas_mae1	0.010** (0.004)	0.010** (0.004)	0.010** (0.004)	0.010** (0.004)
colarinho_branco_apenas_pai1	0.014*** (0.003)	0.013*** (0.003)	0.014*** (0.003)	0.013*** (0.003)
Escolaridade do pai: categoria 2	0.019*** (0.004)	0.019*** (0.004)	0.019*** (0.004)	0.019*** (0.004)
Escolaridade do pai: categoria 3	0.035*** (0.004)	0.035*** (0.004)	0.035*** (0.004)	0.035*** (0.004)
Escolaridade do pai: categoria 4	0.043*** (0.005)	0.044*** (0.005)	0.043*** (0.005)	0.043*** (0.005)
Escolaridade do pai: categoria 5	0.050*** (0.006)	0.049*** (0.006)	0.050*** (0.006)	0.049*** (0.006)
Escolaridade da mãe: categoria 2	0.028*** (0.005)	0.028*** (0.005)	0.027*** (0.005)	0.028*** (0.005)
Escolaridade da mãe: categoria 3	0.048*** (0.005)	0.049*** (0.005)	0.048*** (0.005)	0.049*** (0.005)
Escolaridade da mãe: categoria 4	0.069*** (0.005)	0.070*** (0.005)	0.069*** (0.005)	0.070*** (0.005)
Escolaridade da mãe: categoria 5	0.077*** (0.006)	0.077*** (0.006)	0.077*** (0.006)	0.077*** (0.006)
reuniao1		0.021*** (0.004)		0.021*** (0.004)
conversa_docente1		-0.022*** (0.003)		-0.022*** (0.003)
atividade_comunidade1		0.018*** (0.004)		0.018*** (0.004)
Participação familiar				
Observations	14,290	14,290	14,290	14,290
R ²	0.438	0.445	0.438	0.445
Within R ²	0.046	0.058	0.046	0.058
Within Adjusted R ²	0.045	0.056	0.045	0.057
Escola E.F.	Sim	Sim		
Série E.F.	Sim	Sim		
Escola-série E.F.			Sim	Sim

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. As colunas (1) e (2) incluem separadamente efeitos fixos de escola e de série. As colunas (3) e (4) incluem efeitos fixos para cada combinação escola-série. As colunas (2) e (4) acrescentam os controles de participação familiar. Erros-padrão clusterizados por escola entre parênteses. *** p < 0,01; ** p < 0,05; * p < 0,10.

Apêndice C.5 – Resultados para o pai e para a mãe (painel)

As Tabelas 24 e 23 apresentam os resultados para as características paternas, enquanto as Tabelas 26 e 25 apresentam as estimações referentes às características maternas. Todas as especificações utilizam 33.478 observações, correspondentes a 6.696 estudantes, distribuídos em 233 escolas e acompanhados durante cinco ondas.

Na Tabela 23, a ocupação de colarinho branco do pai apresenta coeficiente de 0,033 na especificação inicial e de 0,019 após a inclusão dos demais controles, permanecendo estatisticamente significativa. A escolaridade paterna também apresenta um gradiente positivo: na especificação completa, os coeficientes das categorias 2 a 5 são 0,013, 0,043, 0,049 e 0,056. Na Tabela 24, observa-se padrão semelhante, com o coeficiente ocupacional diminuindo de 0,054 para 0,031 e os coeficientes das categorias educacionais mais elevadas alcançando 0,055, 0,070 e 0,084. Apenas a categoria 2 perde significância estatística após a inclusão das características maternas.

Os resultados maternos também evidenciam um gradiente educacional positivo. Na Tabela 25, o coeficiente da ocupação de colarinho branco da mãe diminui de 0,029 para 0,007 e deixa de ser estatisticamente significativo na especificação completa. Em contrapartida, os coeficientes das categorias 2 a 5 de escolaridade materna permanecem positivos e significativos, assumindo valores de 0,022, 0,041, 0,060 e 0,062. Na Tabela 26, o coeficiente ocupacional diminui de 0,045 para 0,011, também perdendo significância após a inclusão das características paternas. Os coeficientes da escolaridade materna permanecem positivos e crescentes, variando de 0,052, na categoria 2, a 0,112, na categoria 5.

Em conjunto, os resultados do painel confirmam os gradientes educacionais e ocupacionais encontrados na análise transversal. A redução dos coeficientes ocupacionais após a inclusão da escolaridade indica que essas dimensões captam componentes parcialmente relacionados do ambiente familiar. Ainda assim, a ocupação paterna permanece positivamente associada ao desempenho nas especificações completas, enquanto a ocupação materna perde significância quando as características do pai são consideradas. A escolaridade de ambos os responsáveis permanece relevante, especialmente nos níveis educacionais mais elevados, sendo as magnitudes associadas à escolaridade materna geralmente superiores em Matemática. Esses padrões são compatíveis com mecanismos relacionados ao capital humano familiar, aos recursos econômicos, às condições de trabalho, às expectativas educacionais e à capacidade de acompanhamento das atividades escolares, embora tais mecanismos não sejam diretamente identificados pelas estimações.

Apêndice C.6 – Resultados conjuntos e testes de robustez

As Tabelas 27 e 28 incluem simultaneamente as características educacionais e ocupacionais do pai e da mãe. Na especificação completa para Português, apresentada na

Tabela 23 – Painel: desempenho em Português e características do pai

	Log da nota de Português			
	(1)	(2)	(3)	(4)
Pai em ocupação de colarinho branco	0.033*** (0.003)	0.022*** (0.003)	0.020*** (0.003)	0.019*** (0.003)
Escolaridade do pai: categoria 2		0.024*** (0.007)	0.024*** (0.007)	0.013* (0.007)
Escolaridade do pai: categoria 3		0.061*** (0.007)	0.062*** (0.006)	0.043*** (0.006)
Escolaridade do pai: categoria 4		0.075*** (0.008)	0.074*** (0.008)	0.049*** (0.007)
Escolaridade do pai: categoria 5		0.085*** (0.008)	0.084*** (0.008)	0.056*** (0.008)
Participação em reunião escolar			0.026*** (0.005)	0.025*** (0.005)
Conversa com o docente			-0.023*** (0.003)	-0.024*** (0.003)
Participação em atividade comunitária			0.021*** (0.005)	0.019*** (0.005)
Mãe em ocupação de colarinho branco				0.007 (0.004)
Escolaridade da mãe: categoria 2				0.022*** (0.007)
Escolaridade da mãe: categoria 3				0.041*** (0.007)
Escolaridade da mãe: categoria 4				0.060*** (0.007)
Escolaridade da mãe: categoria 5				0.062*** (0.008)
Escolaridade do pai		Sim	Sim	Sim
Participação familiar			Sim	Sim
Características da mãe				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.632	0.640	0.644	0.648
R ² within	0.007	0.028	0.040	0.051
R ² within ajustado	0.007	0.028	0.040	0.050

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria 1 da escolaridade do pai é a referência. A coluna (1) inclui a ocupação do pai; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características da mãe. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 24 – Painel: desempenho em Matemática e características do pai

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
Pai em ocupação de colarinho branco	0.054*** (0.006)	0.036*** (0.006)	0.034*** (0.006)	0.031*** (0.006)
Escolaridade do pai: categoria 2		0.042*** (0.013)	0.042*** (0.013)	0.020 (0.013)
Escolaridade do pai: categoria 3		0.089*** (0.012)	0.090*** (0.012)	0.055*** (0.012)
Escolaridade do pai: categoria 4		0.115*** (0.014)	0.114*** (0.014)	0.070*** (0.014)
Escolaridade do pai: categoria 5		0.134*** (0.015)	0.134*** (0.015)	0.084*** (0.015)
Participação em reunião escolar			0.049*** (0.009)	0.048*** (0.009)
Conversa com o docente			-0.035*** (0.005)	-0.036*** (0.005)
Participação em atividade comunitária			0.030*** (0.008)	0.027*** (0.008)
Mãe em ocupação de colarinho branco				0.011 (0.007)
Escolaridade da mãe: categoria 2				0.052*** (0.014)
Escolaridade da mãe: categoria 3				0.080*** (0.013)
Escolaridade da mãe: categoria 4				0.105*** (0.014)
Escolaridade da mãe: categoria 5				0.112*** (0.016)
Escolaridade do pai		Sim	Sim	Sim
Participação familiar			Sim	Sim
Características da mãe				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.625	0.631	0.635	0.639
R ² within	0.007	0.023	0.033	0.044
R ² within ajustado	0.007	0.022	0.033	0.043

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria 1 da escolaridade do pai é a referência. A coluna (1) inclui a ocupação do pai; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características da mãe. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 25 – Painel: desempenho em Português e características da mãe

	Log da nota de Português			
	(1)	(2)	(3)	(4)
Mãe em ocupação de colarinho branco	0.029*** (0.004)	0.008** (0.004)	0.008* (0.004)	0.007 (0.004)
Escolaridade da mãe: categoria 2		0.030*** (0.008)	0.030*** (0.008)	0.022*** (0.007)
Escolaridade da mãe: categoria 3		0.060*** (0.007)	0.060*** (0.007)	0.041*** (0.007)
Escolaridade da mãe: categoria 4		0.085*** (0.008)	0.084*** (0.008)	0.060*** (0.007)
Escolaridade da mãe: categoria 5		0.093*** (0.009)	0.091*** (0.009)	0.062*** (0.008)
Participação em reunião escolar			0.025*** (0.005)	0.025*** (0.005)
Conversa com o docente			-0.023*** (0.003)	-0.024*** (0.003)
Participação em atividade comunitária			0.020*** (0.005)	0.019*** (0.005)
Pai em ocupação de colarinho branco				0.019*** (0.003)
Escolaridade do pai: categoria 2				0.013* (0.007)
Escolaridade do pai: categoria 3				0.043*** (0.006)
Escolaridade do pai: categoria 4				0.049*** (0.007)
Escolaridade do pai: categoria 5				0.056*** (0.008)
Escolaridade da mãe		Sim	Sim	Sim
Participação familiar			Sim	Sim
Características do pai				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.631	0.639	0.643	0.648
R ² within	0.005	0.027	0.038	0.051
R ² within ajustado	0.005	0.026	0.038	0.050

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria 1 da escolaridade da mãe é a referência. A coluna (1) inclui a ocupação da mãe; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características do pai. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 26 – Painel: desempenho em Matemática e características da mãe

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
Mãe em ocupação de colarinho branco	0.045*** (0.007)	0.013* (0.007)	0.013* (0.007)	0.011 (0.007)
Escolaridade da mãe: categoria 2		0.063*** (0.014)	0.064*** (0.014)	0.052*** (0.014)
Escolaridade da mãe: categoria 3		0.106*** (0.013)	0.106*** (0.013)	0.080*** (0.013)
Escolaridade da mãe: categoria 4		0.140*** (0.014)	0.139*** (0.013)	0.105*** (0.014)
Escolaridade da mãe: categoria 5		0.158*** (0.016)	0.155*** (0.015)	0.112*** (0.016)
Participação em reunião escolar			0.047*** (0.009)	0.048*** (0.009)
Conversa com o docente			-0.036*** (0.005)	-0.036*** (0.005)
Participação em atividade comunitária			0.029*** (0.008)	0.027*** (0.008)
Pai em ocupação de colarinho branco				0.031*** (0.006)
Escolaridade do pai: categoria 2				0.020 (0.013)
Escolaridade do pai: categoria 3				0.055*** (0.012)
Escolaridade do pai: categoria 4				0.070*** (0.014)
Escolaridade do pai: categoria 5				0.084*** (0.015)
Escolaridade da mãe		Sim	Sim	Sim
Participação familiar			Sim	Sim
Características do pai				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.623	0.631	0.635	0.639
R ² within	0.004	0.024	0.035	0.044
R ² within ajustado	0.004	0.024	0.034	0.043

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria 1 da escolaridade da mãe é a referência. A coluna (1) inclui a ocupação da mãe; a coluna (2) acrescenta sua escolaridade; a coluna (3) acrescenta participação familiar; e a coluna (4) acrescenta as características do pai. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 28, os coeficientes são iguais a 0,022 quando ambos os pais estão em ocupações de colarinho branco, 0,012 quando somente a mãe está nessa condição e 0,023 quando somente o pai está em ocupação de colarinho branco. Todos permanecem positivos e estatisticamente significativos. Na categoria mais elevada de escolaridade, os coeficientes são 0,057 para o pai e 0,062 para a mãe. O R^2 *within* aumenta de 0,012 para 0,051 ao longo das especificações.

Na Tabela 27, o coeficiente associado a ambos os pais em ocupações de colarinho branco diminui de 0,082 para 0,035 após a inclusão das escolaridades e dos controles de participação familiar. Os coeficientes são iguais a 0,021 quando somente a mãe está em ocupação de colarinho branco e a 0,039 quando somente o pai está nessa condição. Na categoria educacional mais elevada, os coeficientes alcançam 0,086 para o pai e 0,113 para a mãe. O R^2 *within* aumenta de 0,011 para 0,044, indicando maior capacidade de explicar as diferenças de desempenho entre estudantes da mesma escola à medida que as características familiares são incorporadas.

Como exercício de robustez, os efeitos fixos separados de escola e onda são substituídos por efeitos fixos específicos para cada combinação entre escola e onda:

$$\log(Y_{ist}) = \alpha_{st} + \beta F_{ist} + \gamma X_{ist} + u_{ist}$$

em que α_{st} representa o efeito fixo de cada combinação escola-onda. Essa especificação absorve fatores comuns aos estudantes de uma mesma escola em cada momento da avaliação, como mudanças na gestão, no corpo docente, na infraestrutura, na composição escolar ou em outras condições institucionais que possam variar ao longo do painel.

Os resultados apresentados nas Tabelas 29 e 30 mostram elevada estabilidade entre as especificações com efeitos fixos separados de escola e onda e aquelas com efeitos fixos escola-onda. Em Matemática, por exemplo, o coeficiente associado a ambos os pais em ocupações de colarinho branco permanece em 0,035 na especificação completa, enquanto os coeficientes das categorias mais elevadas de escolaridade paterna e materna permanecem próximos de 0,085 e 0,113. Em Português, os respectivos coeficientes permanecem próximos de 0,022, 0,057 e 0,063. Os indicadores de participação familiar também praticamente não se alteram entre as especificações.

Tabela 27 – Painel: desempenho em Matemática e características conjuntas dos pais

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
Ambos em ocupação de colarinho branco	0.082*** (0.009)	0.055*** (0.009)	0.038*** (0.009)	0.035*** (0.009)
Somente a mãe em ocupação de colarinho branco	0.057*** (0.009)	0.042*** (0.009)	0.021** (0.009)	0.021** (0.009)
Somente o pai em ocupação de colarinho branco	0.064*** (0.008)	0.046*** (0.008)	0.041*** (0.008)	0.039*** (0.008)
Escolaridade do pai: categoria 2		0.040*** (0.013)	0.020 (0.013)	0.020 (0.013)
Escolaridade do pai: categoria 3		0.086*** (0.012)	0.053*** (0.012)	0.054*** (0.012)
Escolaridade do pai: categoria 4		0.108*** (0.014)	0.069*** (0.014)	0.069*** (0.014)
Escolaridade do pai: categoria 5		0.128*** (0.015)	0.085*** (0.015)	0.086*** (0.015)
Escolaridade da mãe: categoria 2			0.051*** (0.014)	0.051*** (0.014)
Escolaridade da mãe: categoria 3			0.080*** (0.013)	0.080*** (0.013)
Escolaridade da mãe: categoria 4			0.105*** (0.014)	0.103*** (0.013)
Escolaridade da mãe: categoria 5			0.115*** (0.016)	0.113*** (0.016)
Participação em reunião escolar				0.048*** (0.009)
Conversa com o docente				-0.036*** (0.005)
Participação em atividade comunitária				0.026*** (0.008)
Escolaridade do pai		Sim	Sim	Sim
Escolaridade da mãe			Sim	Sim
Participação familiar				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.626	0.631	0.635	0.639
R ² within	0.011	0.025	0.034	0.044
R ² within ajustado	0.011	0.025	0.033	0.044

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria ocupacional de referência corresponde aos estudantes para os quais nenhum dos pais está em ocupação de colarinho branco. As categorias 1 das escolaridades paterna e materna são as referências. A coluna (1) inclui as categorias ocupacionais; a coluna (2) acrescenta a escolaridade do pai; a coluna (3) acrescenta a escolaridade da mãe; e a coluna (4) acrescenta participação familiar. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 28 – Painel: desempenho em Português e características conjuntas dos pais

	Log da nota de Português			
	(1)	(2)	(3)	(4)
Ambos em ocupação de colarinho branco	0.051*** (0.005)	0.035*** (0.005)	0.023*** (0.005)	0.022*** (0.005)
Somente a mãe em ocupação de colarinho branco	0.036*** (0.006)	0.027*** (0.006)	0.012** (0.006)	0.012** (0.006)
Somente o pai em ocupação de colarinho branco	0.039*** (0.004)	0.028*** (0.004)	0.025*** (0.004)	0.023*** (0.004)
Escolaridade do pai: categoria 2		0.023*** (0.007)	0.013* (0.007)	0.013* (0.007)
Escolaridade do pai: categoria 3		0.059*** (0.007)	0.041*** (0.006)	0.042*** (0.006)
Escolaridade do pai: categoria 4		0.071*** (0.008)	0.048*** (0.008)	0.048*** (0.007)
Escolaridade do pai: categoria 5		0.081*** (0.009)	0.056*** (0.008)	0.057*** (0.008)
Escolaridade da mãe: categoria 2			0.021*** (0.008)	0.022*** (0.007)
Escolaridade da mãe: categoria 3			0.041*** (0.007)	0.040*** (0.007)
Escolaridade da mãe: categoria 4			0.060*** (0.007)	0.059*** (0.007)
Escolaridade da mãe: categoria 5			0.064*** (0.008)	0.062*** (0.008)
Participação em reunião escolar				0.026*** (0.005)
Conversa com o docente				-0.023*** (0.003)
Participação em atividade comunitária				0.019*** (0.005)
Escolaridade do pai		Sim	Sim	Sim
Escolaridade da mãe			Sim	Sim
Participação familiar				Sim
Efeitos fixos de escola	Sim	Sim	Sim	Sim
Efeitos fixos de onda	Sim	Sim	Sim	Sim
Efeitos fixos escola-onda				
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.634	0.641	0.644	0.648
R ² within	0.012	0.031	0.040	0.051
R ² within ajustado	0.012	0.030	0.039	0.051

Notas: A variável dependente é o logaritmo da nota. Todas as colunas de cada tabela utilizam a mesma amostra. A categoria ocupacional de referência corresponde aos estudantes para os quais nenhum dos pais está em ocupação de colarinho branco. As categorias 1 das escolaridades paterna e materna são as referências. A coluna (1) inclui as categorias ocupacionais; a coluna (2) acrescenta a escolaridade do pai; a coluna (3) acrescenta a escolaridade da mãe; e a coluna (4) acrescenta participação familiar. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 29 – Painel: robustez dos resultados de Matemática

	Log da nota de Matemática			
	(1)	(2)	(3)	(4)
Ambos em ocupação de colarinho branco	0.038*** (0.009)	0.035*** (0.009)	0.038*** (0.009)	0.035*** (0.009)
Somente a mãe em ocupação de colarinho branco	0.021** (0.009)	0.021** (0.009)	0.020** (0.009)	0.021** (0.009)
Somente o pai em ocupação de colarinho branco	0.041*** (0.008)	0.039*** (0.008)	0.041*** (0.008)	0.039*** (0.008)
Escolaridade do pai: categoria 2	0.020 (0.013)	0.020 (0.013)	0.019 (0.013)	0.019 (0.013)
Escolaridade do pai: categoria 3	0.053*** (0.012)	0.054*** (0.012)	0.052*** (0.012)	0.054*** (0.012)
Escolaridade do pai: categoria 4	0.069*** (0.014)	0.069*** (0.014)	0.069*** (0.014)	0.069*** (0.014)
Escolaridade do pai: categoria 5	0.085*** (0.015)	0.086*** (0.015)	0.084*** (0.015)	0.085*** (0.015)
Escolaridade da mãe: categoria 2	0.051*** (0.014)	0.051*** (0.014)	0.051*** (0.014)	0.052*** (0.014)
Escolaridade da mãe: categoria 3	0.080*** (0.013)	0.080*** (0.013)	0.080*** (0.013)	0.080*** (0.013)
Escolaridade da mãe: categoria 4	0.105*** (0.014)	0.103*** (0.013)	0.105*** (0.014)	0.104*** (0.013)
Escolaridade da mãe: categoria 5	0.115*** (0.016)	0.113*** (0.016)	0.116*** (0.016)	0.113*** (0.016)
Participação em reunião escolar		0.048*** (0.009)		0.048*** (0.009)
Conversa com o docente		-0.036*** (0.005)		-0.036*** (0.005)
Participação em atividade comunitária		0.026*** (0.008)		0.027*** (0.008)
Participação familiar		Sim		Sim
Efeitos fixos de escola	Sim	Sim		
Efeitos fixos de onda	Sim	Sim		
Efeitos fixos escola-onda			Sim	Sim
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.635	0.639	0.650	0.654
R ² within	0.034	0.044	0.035	0.046
R ² within ajustado	0.033	0.044	0.035	0.046

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. As colunas (1) e (2) incluem efeitos fixos separados de escola e onda. As colunas (3) e (4) incluem efeitos fixos para cada combinação escola-onda. As colunas (2) e (4) acrescentam os controles de participação familiar. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Tabela 30 – Pannel: robustez dos resultados de Português

	Log da nota de Português			
	(1)	(2)	(3)	(4)
Ambos em ocupação de colarinho branco	0.023*** (0.005)	0.022*** (0.005)	0.023*** (0.005)	0.022*** (0.005)
Somente a mãe em ocupação de colarinho branco	0.012** (0.006)	0.012** (0.006)	0.012** (0.006)	0.012** (0.006)
Somente o pai em ocupação de colarinho branco	0.025*** (0.004)	0.023*** (0.004)	0.025*** (0.004)	0.024*** (0.004)
Escolaridade do pai: categoria 2	0.013* (0.007)	0.013* (0.007)	0.013* (0.007)	0.013* (0.007)
Escolaridade do pai: categoria 3	0.041*** (0.006)	0.042*** (0.006)	0.041*** (0.006)	0.042*** (0.006)
Escolaridade do pai: categoria 4	0.048*** (0.008)	0.048*** (0.007)	0.048*** (0.007)	0.048*** (0.007)
Escolaridade do pai: categoria 5	0.056*** (0.008)	0.057*** (0.008)	0.056*** (0.008)	0.057*** (0.008)
Escolaridade da mãe: categoria 2	0.021*** (0.008)	0.022*** (0.007)	0.021*** (0.008)	0.022*** (0.007)
Escolaridade da mãe: categoria 3	0.041*** (0.007)	0.040*** (0.007)	0.041*** (0.007)	0.041*** (0.007)
Escolaridade da mãe: categoria 4	0.060*** (0.007)	0.059*** (0.007)	0.060*** (0.007)	0.059*** (0.007)
Escolaridade da mãe: categoria 5	0.064*** (0.008)	0.062*** (0.008)	0.064*** (0.008)	0.063*** (0.008)
Participação em reunião escolar		0.026*** (0.005)		0.026*** (0.005)
Conversa com o docente		-0.023*** (0.003)		-0.023*** (0.003)
Participação em atividade comunitária		0.019*** (0.005)		0.019*** (0.005)
Participação familiar		Sim		Sim
Efeitos fixos de escola	Sim	Sim		
Efeitos fixos de onda	Sim	Sim		
Efeitos fixos escola-onda			Sim	Sim
Erros-padrão agrupados	Aluno e escola	Aluno e escola	Aluno e escola	Aluno e escola
Observações	33,478	33,478	33,478	33,478
Alunos	6,696	6,696	6,696	6,696
Escolas	233	233	233	233
Ondas	5	5	5	5
R ²	0.644	0.648	0.673	0.677
R ² within	0.040	0.051	0.043	0.055
R ² within ajustado	0.039	0.051	0.043	0.055

Notas: A variável dependente é o logaritmo da nota. Todas as colunas utilizam a mesma amostra. As colunas (1) e (2) incluem efeitos fixos separados de escola e onda. As colunas (3) e (4) incluem efeitos fixos para cada combinação escola-onda. As colunas (2) e (4) acrescentam os controles de participação familiar. Erros-padrão agrupados por aluno e escola entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Em conjunto, os resultados do painel e os testes de robustez reforçam as evidências obtidas na análise transversal. As características educacionais e ocupacionais dos pais permanecem associadas ao desempenho dos estudantes quando são consideradas simultaneamente, quando se controla por diferenças permanentes entre escolas e quando são absorvidos fatores específicos de cada combinação escola-onda. A estabilidade das estimativas sugere que os gradientes observados não decorrem exclusivamente de mudanças gerais entre as ondas ou de alterações na composição e nas condições das escolas ao longo do período.

Os resultados são compatíveis com a atuação de mecanismos complementares. A escolaridade parental pode refletir maior capital humano, maior capacidade de auxílio nas atividades escolares, expectativas educacionais mais elevadas e maior familiaridade com o sistema de ensino. A ocupação pode captar recursos econômicos, estabilidade laboral, condições de trabalho, redes de informação e possibilidades de investimento na educação dos filhos. Os coeficientes positivos associados à participação em reuniões e atividades comunitárias também sugerem a relevância do envolvimento familiar, enquanto o coeficiente negativo da conversa com docentes pode refletir que esse contato ocorre com maior frequência quando o estudante já apresenta dificuldades. Apesar da consistência dos resultados, as estimações identificam associações condicionais e não efeitos causais isolados da escolaridade, da ocupação ou da participação dos pais.

3 DISTÂNCIA, ESCOLHA ESCOLAR E MOBILIDADE URBANA: EVIDÊNCIAS DOS SUBSÍDIOS AO TRANSPORTE ESTUDANTIL NO BRASIL

Resumo

Neste estudo, investiga-se a relação entre a distância residência-escola e o desempenho acadêmico no primeiro ano de alfabetização. Em contraste com a abordagem usual, que trata a distância como uma medida exógena de custo de acesso, argumenta-se que ela pode ser endógena ao processo de escolha escolar, pois reflete simultaneamente restrições de mobilidade e a seleção das famílias em direção a escolas percebidas como mais atrativas, considerando, nesse contexto, a política municipal de passe estudantil como fonte de variação institucional na distância observada. O subsídio ao transporte reduz o custo monetário do deslocamento e pode ampliar o conjunto de escolas factíveis para as famílias. Com dados da segunda onda do Estudo Longitudinal da Geração Escolar 2005 (GERES), estimam-se modelos de mínimos quadrados ordinários, forma reduzida e variáveis instrumentais. Os resultados mostram um contraste entre as estimativas, pois, enquanto a distância observada aparece positivamente associada ao desempenho nos modelos observacionais, a parcela da distância associada à política de mobilidade apresenta efeito negativo e estatisticamente significativo sobre a proficiência em Matemática e Português. Embora a interpretação causal exija cautela, os achados sugerem que a distância residência-escola não deve ser entendida apenas como uma barreira geográfica, mas como uma medida cujo significado depende das condições locais de mobilidade e do processo de escolha escolar, contribuindo, desta forma, para uma avaliação mais ampla sobre a variável distância e seus potenciais custos no contexto educacional brasileiro.

Palavras-chave: distância residência-escola; mobilidade escolar; subsídio ao transporte estudantil; escolha escolar; desempenho acadêmico.

3.1 Introdução

A escola frequentada por uma criança nem sempre é a escola mais próxima de sua residência. Essa diferença é particularmente relevante nos anos iniciais do ensino fundamental, quando o deslocamento diário depende da rotina familiar, da disponibilidade de transporte e da capacidade dos responsáveis de acomodar o tempo e o custo da viagem. Em contextos urbanos, estudar mais longe pode significar uma restrição de acesso, mas também pode refletir uma escolha, onde as famílias podem aceitar trajetos maiores para matricular os filhos em escolas que consideram mais adequadas, seguras ou bem avaliadas.

Essa ambiguidade torna a distância residência-escola uma variável de interpretação menos direta do que parece à primeira vista. No caso da escola efetivamente frequentada, a distância observada já incorpora decisões familiares, restrições de mobilidade e carac-

terísticas da oferta escolar local. Assim, tratá-la simplesmente como custo exógeno pode levar a interpretações incompletas sobre sua relação com o desempenho acadêmico.

Este artigo investiga como a distância entre residência e escola se relaciona com o desempenho em Matemática e Português nos anos iniciais do ensino fundamental. Busca-se responder se alunos que estudam mais longe apresentam desempenho diferente e, sobretudo, se essa relação muda quando se leva em conta que a distância observada pode ser endógena ao processo de escolha escolar.

O tratamento de interesse é a distância residência–escola, medida em quilômetros a partir da geocodificação dos endereços dos estudantes e das escolas. Utilizam-se dados da onda 2 do Estudo Longitudinal da Geração Escolar 2005 (GERES), em uma amostra de 12.157 alunos dos municípios de Belo Horizonte, Campinas, Campo Grande, Rio de Janeiro e Salvador, após o truncamento da distribuição da distância no percentil 99. A estratégia empírica combina modelos de mínimos quadrados ordinários, forma reduzida, primeiro estágio e variáveis instrumentais. O instrumento utilizado é a política municipal de passe estudantil, interpretada como exposição ao regime municipal de mobilidade escolar, já que não se observa o uso individual do benefício.

Os resultados mostram que a política municipal de passe estudantil está positivamente associada à distância entre residência e escola. Na especificação principal, a existência de subsídio aumenta a distância observada em aproximadamente 0,91 km. Nos modelos observacionais, cada quilômetro adicional está associado a um aumento de cerca de 0,2% na proficiência em Matemática e Português. Quando a distância é instrumentada pela política de passe estudantil, o sinal se inverte: um quilômetro adicional na distância instrumentada está associado a reduções de aproximadamente 3,9% em Matemática e 3,3% em Português. A mudança de sinal sugere que a associação positiva observada por MQO reflete, ao menos em parte, seleção familiar em direção a escolas percebidas como melhores.

Os exercícios complementares reforçam essa interpretação, mas também delimitam seu alcance. As estimativas padronizadas indicam magnitudes economicamente relevantes, e as análises de heterogeneidade sugerem efeitos mais intensos entre alunos da rede pública e estudantes cujos pais não possuem ensino superior. Esses são grupos para os quais o subsídio ao transporte tende a ser mais relevante como relaxamento de restrição orçamentária. Por outro lado, o instrumento se mostra fraco na subamostra da rede privada e nas especificações que distinguem o subsídio parcial e integral¹.

Este artigo busca investigar uma lacuna na literatura econômica, propondo uma avaliação relacionada à forma como se utiliza a distância ou a proximidade geográfica entre a residência e a escola do estudante como uma medida do custo de acesso à educação.

¹ Apresentados apenas como exercícios de robustez, dado que o teste de Sargan rejeita a validade conjunta dos instrumentos.

Enquanto trabalhos como Card (1993) e Duflo (2001) exploram a proximidade a instituições ou a expansão da oferta escolar como fontes de variação no acesso educacional, este artigo busca demonstrar que, em contextos urbanos, essa distância não é apenas uma medida de acesso físico, mas um reflexo de escolhas familiares, disponibilidade de escolas e condições locais de mobilidade.

A principal contribuição é mostrar que essas duas dimensões da distância produzem sinais distintos. A distância observada nos modelos de MQO está positivamente associada ao desempenho, resultado compatível com seleção positiva. Já a parcela da distância associada à política municipal de transporte estudantil está negativamente relacionada à proficiência. Essa diferença ajuda a reconciliar resultados aparentemente contraditórios e sugere que a distância residência–escola deve ser interpretada com cautela em estudos sobre mobilidade e desempenho educacional.

Do ponto de vista da política pública, os resultados indicam que subsídios ao transporte estudantil podem ampliar o conjunto de escolas acessíveis às famílias, mas não eliminam todos os custos associados a trajetos mais longos. Embora este artigo não observe diretamente o tempo de viagem, o cansaço ou a confiabilidade do transporte, esses canais são plausíveis e devem ser considerados em políticas que articulem a mobilidade escolar, o planejamento territorial da oferta educacional e as desigualdades socioeconômicas.

3.2 Distância, Mobilidade e Escolha Escolar

A distância entre a residência e a escola costuma ser interpretada como uma medida do custo de acesso à educação. Sob essa perspectiva, deslocamentos mais longos aumentam o tempo gasto no trajeto, tornam a rotina escolar mais custosa e podem reduzir o tempo disponível para estudo, descanso e acompanhamento familiar. A distância, portanto, não importa apenas por representar uma separação geográfica entre casa e escola, mas porque sintetiza custos monetários e não monetários que incidem sobre a rotina das famílias e das crianças.

Essa interpretação, contudo, é incompleta quando a escola observada resulta de um processo de escolha familiar. Em contextos urbanos, a família não necessariamente matricula a criança na escola mais próxima da residência. A decisão pode envolver atributos percebidos da escola, como reputação, segurança, rede de ensino, disponibilidade de vagas e compatibilidade com a rotina dos responsáveis. Assim, estudar mais longe pode refletir uma restrição imposta pela localização da oferta escolar, mas também pode indicar que a família aceitou um deslocamento maior para acessar uma escola considerada mais adequada.

Essa ambiguidade diferencia o problema analisado neste artigo de parte da literatura que utiliza a proximidade geográfica como medida de acesso educacional. Em Card

(1993), a proximidade a uma faculdade durante a juventude é utilizada como fonte de variação na escolaridade, enquanto Duflo (2001) analisa a expansão da oferta de escolas primárias na Indonésia como política de ampliação do acesso educacional.² Nesses trabalhos, a proximidade está associada à redução de barreiras de acesso. No caso da distância residência–escola observada, entretanto, a variável pode ser consequência da própria escolha escolar, o que torna mais frágil sua interpretação como medida exógena de custo.

A literatura sobre mobilidade urbana e escolha escolar aproxima-se mais diretamente do mecanismo analisado neste artigo. Dustan e Ngo (2018) estudam a expansão do transporte público na Cidade do México e mostraram que a introdução do Trem Suburbano alterou escolhas escolares, especialmente entre estudantes de maior desempenho e com pais mais escolarizados. Segundo os autores, esses estudantes passaram a escolher escolas de elite com maior probabilidade e a percorrer maiores distâncias até a escola escolhida. De forma complementar, Moreno-Monroy, Lovelace e Ramos (2018) analisam a acessibilidade ao transporte público e as desigualdades educacionais na Região Metropolitana de São Paulo, mostrando que a distribuição espacial das escolas e os tempos de deslocamento podem afetar de forma desigual estudantes de diferentes grupos socioeconômicos. Esses estudos sugerem que mudanças nas condições de mobilidade podem alterar não apenas o custo de acesso à escola, mas também o conjunto de escolas efetivamente consideradas pelas famílias.

Essa relação entre transporte e escolha escolar exige distinguir contextos distintos de política pública. Em áreas rurais, o transporte escolar costuma ter como finalidade principal garantir o acesso físico à escola, especialmente quando a residência se localiza em comunidades afastadas. Nesse caso, a distância aparece predominantemente como barreira direta de acesso. Em áreas urbanas, por outro lado, o passe estudantil opera sobre a rede de transporte coletivo municipal e pode alterar o custo relativo de frequentar escolas mais distantes. O subsídio não muda a geografia entre casa e escola, mas pode modificar o preço econômico dessa distância, ampliando o raio de escolha escolar das famílias.³

² Card (1993) utiliza a proximidade a uma faculdade como variável instrumental para a escolaridade ao estimar os retornos da educação. Duflo (2001) explora o programa de construção de escolas primárias na Indonésia, implementado entre 1973 e 1978, encontrando efeitos positivos sobre a escolaridade e os rendimentos futuros. Esses trabalhos são importantes para a interpretação da proximidade como redução do custo de acesso, mas diferem deste artigo porque não analisam a distância até a escola efetivamente escolhida pela família.

³ No Brasil, parte importante da literatura sobre transporte escolar concentra-se em áreas rurais. Martins (2010) analisa o transporte escolar rural em Anápolis, Goiás, enquanto Goulart, Morais e Vieira Júnior (2019) examina deslocamentos em uma escola rural de Caracaraí, Roraima. Esses trabalhos dialogam com a agenda de acesso e permanência, mas tratam de contextos distintos daqueles observados em municípios urbanos com transporte coletivo subsidiado.

Entre os estudos brasileiros mais próximos deste artigo, Pepe (2017) avalia o Passe Livre Estudantil no município de São Paulo e conclui que a política possui potencial para ampliar a liberdade de escolha educacional e reduzir desigualdades de acesso, embora não encontre efeitos claros sobre o absenteísmo. Barbosa (2024), por sua vez, analisa a relação entre a distância casa–escola, o transporte escolar e o desempenho em Maracanaú, no Ceará, encontrando uma relação negativa entre a distância e o desempenho em Matemática, mas não encontra os mesmos efeitos em leitura. Esses trabalhos aproximam transporte estudantil, mobilidade urbana e resultados educacionais no Brasil. Ainda assim, permanece pouco explorada a distinção entre a distância como barreira de acesso e a distância como resultado de seleção familiar em direção a escolas percebidas como melhores.

A evidência internacional sobre deslocamento escolar reforça os canais e os mecanismos pelos quais trajetos mais longos podem prejudicar a experiência educacional. Estudos em diferentes contextos associam maiores distâncias ou tempos de deslocamento a pior desempenho, maior absenteísmo, atrasos, fadiga, menor bem-estar ou maior evasão escolar.⁴ Esses estudos ajudam a caracterizar os mecanismos de custos associados ao deslocamento, mas nem sempre enfrentam diretamente o problema de seleção envolvido na escolha da escola frequentada.

Se busca demonstrar ao longo deste trabalho que a distância até a escola frequentada não pode ser interpretada da mesma forma que a distância até a escola mais próxima ou até uma oferta educacional recém-criada. No primeiro caso, a distância observada já incorpora uma decisão familiar. Ela pode revelar dificuldade de acesso, mas também pode indicar que a família abriu mão da escola mais próxima para buscar outra unidade, por razões associadas à qualidade percebida, segurança, rede de relações, disponibilidade de vagas ou conveniência da rotina familiar. Por isso, a associação entre distância e desempenho não é direta; ou seja, alunos que estudam mais longe podem estar sujeitos a maiores custos de deslocamento, mas também podem ser justamente aqueles cujas famílias conseguiram acessar escolas consideradas mais desejáveis.

A implicação econométrica é que a distância observada pode estar correlacionada com determinantes não observados do desempenho, como informação familiar, preferências por qualidade escolar, capacidade de mobilidade e atributos não observados das escolas. Dessa forma, a associação estimada por mínimos quadrados ordinários pode combinar seleção positiva e custos reais de deslocamento. Políticas de subsídio ao transporte estu-

⁴ Baliyan e Khama (2020) identificam, em Botsuana, que a distância até a escola e a falta de tempo de estudo em casa comprometeram o desempenho em Matemática. Bammou, El Bouhali e El Alaoui (2024), em uma escola rural no Marrocos, associam longos deslocamentos a atrasos, absenteísmo e menor desempenho. Peteros *et al.* (2022), nas Filipinas, encontram relação negativa entre a distância residência–escola e o desempenho em Matemática. Ding e Feng (2022), na China, relacionam maiores tempos de deslocamento a pior bem-estar psicológico e desempenho acadêmico. Zeragaber, Teame e Tsighe (2024), na Eritreia, encontram relação positiva entre distância e evasão escolar.

dantil são relevantes nesse contexto porque alteram o peso econômico da distância e, com isso, podem mudar o conjunto de escolas que entram no horizonte de escolha das famílias. Neste artigo, a distância residência–escola é tratada justamente como uma variável que reúne esses dois componentes: de um lado, o custo do trajeto e, do outro, a possibilidade de frequentar uma escola considerada melhor, ainda que mais distante.

3.3 Modelo empírico, metodologia e a distância residência–escola como variável endógena

Esta seção apresenta as hipóteses levantadas, a estratégia empírica e o diálogo com a literatura que sustentam este estudo. O ponto de partida é a hipótese de que a distância entre residência e escola não constitui uma medida puramente exógena do custo de acesso à educação. Ao analisar os contextos urbanos no Brasil, em especial aqueles de maior dimensão geográfica e populacional, a escola frequentada resulta de um processo de escolha em que as famílias ponderam atributos da escola, restrições orçamentárias e custos de deslocamento. Explorando esse contexto, busca-se analisar as políticas públicas de subsídio ao transporte estudantil e como estas podem alterar o custo monetário do trajeto e, com isso, modificar o conjunto de escolas efetivamente viáveis para determinados alunos.

A distância observada entre residência e escola incorpora simultaneamente duas dimensões, sendo a primeira delas a margem de oportunidade, que está associada à possibilidade de acessar escolas percebidas como superiores, com características como melhor desempenho em ensino e avaliações, melhores qualidades dos docentes, dentre outros aspectos que as famílias podem observar no momento da escolha. A outra dimensão, que se denomina margem de custo, está associada ao esforço e às fricções do deslocamento diário. Essa dupla natureza torna a variável distância potencialmente endógena nas regressões de desempenho escolar e, para investigar esse problema, propomos explorar a política municipal de passe estudantil como fonte de variação institucional na distância observada, interpretando seus resultados com cautela, dada a natureza agregada do instrumento e as limitações que serão apresentadas ao longo deste trabalho.

3.3.1 Diálogo com a literatura e hipóteses do estudo

A literatura de economia da educação frequentemente interpreta a distância ou a proximidade como uma *proxy* de custo de acesso à escolarização. Em trabalhos clássicos, a proximidade geográfica aparece como um fator que reduz o custo de estudar e afeta decisões educacionais na margem, como em Card (1993). Em termos econômicos, a distância não importa apenas porque mede quilômetros, mas porque resume custos de tempo, dinheiro, esforço e restrição de escolhas.

Uma segunda vertente da literatura mostra que políticas públicas podem alterar o acesso à educação por meio de choques institucionais na oferta ou no custo de acesso.

Duflo (2001), por exemplo, explora a construção de escolas na Indonésia como uma política que alterou o acesso educacional de forma diferencial entre regiões e coortes. Embora o desenho de identificação desse trabalho seja distinto do adotado aqui, sua contribuição é importante para o presente estudo porque reforça a ideia de que políticas públicas podem deslocar, de forma exógena, os custos efetivos de acesso à educação.

Mais próximo do mecanismo que se avalia neste estudo está a literatura sobre transporte, acessibilidade e escolha escolar. Estudos recentes mostram que melhorias no transporte público ou subsídios ao deslocamento podem alterar o conjunto de escolas viáveis para as famílias, levando alunos a acessar escolas mais distantes e, em alguns casos, de melhor qualidade (Dustan; Ngo, 2018; Ramírez-Hassan *et al.*, 2023; Moreno-Monroy; Lovelace; Ramos, 2018). Essa literatura sugere que, em grandes centros urbanos, a distância física não deve ser interpretada isoladamente, uma vez que seu significado econômico depende das condições locais de mobilidade.

À luz desse debate, a hipótese central deste artigo é que a distância observada entre residência e escola é endógena à própria escolha escolar. Famílias podem aceitar deslocamentos maiores para acessar escolas percebidas como superiores, especialmente quando o custo do deslocamento é subsidiado. Assim, a associação observada entre distância e desempenho pode confundir dois mecanismos distintos: a seleção para escolas mais atraídas e o custo efetivo de trajetos mais longos. O papel da política de passe estudantil, neste estudo, é fornecer uma variação institucional informativa sobre essa endogeneidade.

3.3.2 Arcabouço conceitual: a escolha escolar e o "custo" de acesso

A decisão da família sobre a escola frequentada pelo aluno pode ser entendida como um problema de maximização de utilidade, no espírito dos modelos de utilidade aleatória de McFadden (1974). Seja U_{ijm} a utilidade que a família i associa à escola j , localizada no município m :

$$U_{ijm} = V_{ijm} + \varepsilon_{ijm}, \quad (3.1)$$

em que V_{ijm} representa a parte determinística da utilidade, composta por atributos observáveis da escola e do aluno, e ε_{ijm} o termo de erro que capta fatores idiossincráticos não observados. Assumindo que a utilidade determinística depende de atributos da escola, características familiares e do custo de acesso à escola. Assim, se denota:

$$V_{ijm} = X_j\beta + Z_i\gamma - \lambda C_{ijm}, \quad (3.2)$$

em que X_j representa atributos observáveis da escola, Z_i reúne características observáveis da família e $\lambda > 0$ mede o peso do custo de acesso na utilidade. Se representa esse custo de acesso como:

$$C_{ijm} = d_{ij}(1 - \alpha s_m), \quad (3.3)$$

em que d_{ij} é a distância entre a residência da família i e a escola j , s_m representa o regime de subsídio ao transporte estudantil no município m , e α mede em que grau esse subsídio reduz o peso econômico da distância. Substituindo C_{ijm} em V_{ijm} , tem-se:

$$V_{ijm} = X_j\beta + Z_i\gamma - \lambda d_{ij}(1 - \alpha s_m). \quad (3.4)$$

Essa formulação deixa claro que a política pública não altera a geografia física entre domicílio e escola, mas reduz o custo efetivo do deslocamento e, por isso, modifica o conjunto de escolas factíveis para a família.

3.3.3 Construção da base de dados e mensuração da distância

Se avalia a medida da distância residência-escola, construída a partir da geolocalização dos logradouros contidos no banco de dados do GERES (2005). Como os endereços apresentavam inconsistências de preenchimento, erros de grafia e ausência de padronização, realizaram-se inicialmente procedimentos de limpeza textual e normalização dos logradouros, obtiveram-se suas coordenadas e calculou-se a distância entre os logradouros da residência e da escola.

A distribuição da variável distância mostrou forte assimetria à direita e presença de valores extremos. Diante disso, optou-se por restringir a amostra a alunos residentes no município da escola ou em municípios limítrofes e truncar a distribuição no percentil 99 da distância. Essa decisão visa reduzir a influência de observações extremas potencialmente associadas a erros de geocodificação, registros inconsistentes ou deslocamentos pouco compatíveis com a escala urbana dos municípios analisados⁵.

Foram agregadas a medida de distância aos demais dados do Estudo Longitudinal da Geração Escolar 2005 (GERES), que acompanhou estudantes ao longo dos anos iniciais do ensino fundamental em cinco municípios brasileiros. Embora o GERES possua estrutura longitudinal, o presente estudo adota uma *cross section* referente à onda 2, pois é nesse momento que se concentram, de forma mais próxima no tempo, as informações utilizadas para construir a variável distância, os controles familiares e a variável institucional de interesse. Dessa forma, as variáveis dependentes são as proficiências em Matemática e Português, e as variáveis de controle incluem escolaridade e ocupação dos pais, engajamento parental com a escola, tipo de rede escolar e série do aluno. Como a base não contém uma medida direta de renda familiar, utilizou-se a escolaridade e a ocupação dos responsáveis como *proxies* para habilidades e contexto econômico familiar.

⁵ Ao longo da próxima seção, onde se apresentam as estatísticas descritivas e a distribuição geográfica dos municípios, sustentam essa escolha.

3.3.4 Estratégia empírica

A estratégia empírica articula quatro relações complementares, iniciadas pela associação observacional entre a distância residência-escola e o desempenho, o efeito da política municipal de passe estudantil sobre a distância observada, o efeito reduzido da política institucional de subsídios sobre o desempenho; e, por fim, a estimação por variáveis instrumentais, que utiliza a política de passe como fonte de variação institucional na distância observada.

Modelo observacional estimado por Mínimos Quadrados Ordinários

Como referência inicial, estima-se um modelo observacional da distância sobre o desempenho a partir da seguinte equação, utilizando um estimador de Mínimos Quadrados Ordinários:

$$\log(\text{desempenho}_{id}) = \alpha + \beta \text{distancia}_i + \mathbf{X}_i \boldsymbol{\gamma} + \varepsilon_i, \quad (3.5)$$

em que $\log(\text{desempenho}_{id})$ representa o logaritmo da proficiência do aluno i na disciplina $d \in \{\text{Português, Matemática}\}$, distancia_i é a distância residência-escola em quilômetros, e $\mathbf{X}_i$ é o vetor de controles observáveis. Essa equação fornece a associação bruta entre distância e desempenho, condicionada aos controles. No entanto, β não deve ser interpretado como um parâmetro causal, pois a distância observada pode refletir autoseleção familiar, qualidade não observada da escola escolhida e diferenças de mobilidade entre municípios.

Forma reduzida da política municipal

Como a elegibilidade para o passe estudantil é definida em nível municipal, também se estima a relação reduzida entre a política de subsídio e o desempenho escolar:

$$\log(\text{desempenho}_{id}) = \alpha + \tau Z_m + \mathbf{X}_i \boldsymbol{\eta} + \nu_i, \quad (3.6)$$

em que Z_m é um indicador da existência de política municipal de passe estudantil no município m . Essa equação mede o efeito da exposição à política em nível municipal sobre o desempenho e pode ser lida como um efeito de intenção de tratar (*intent-to-treat*) ao nível do município. Trata-se de uma peça importante da análise, pois permite verificar se a política afeta o desfecho mesmo antes de impor a interpretação estrutural do estimador de variáveis instrumentais.

Primeiro estágio

A etapa seguinte consiste em verificar se a política de passe estudantil está associada à distância residência-escola. Assim, se estima:

$$\text{distancia}_i = \pi_0 + \pi_1 Z_m + \mathbf{X}_i \boldsymbol{\delta} + u_i, \quad (3.7)$$

em que Z_m representa a política municipal de subsídio ao transporte estudantil. Na especificação principal, utiliza-se uma *dummy* que assume o valor 1 para municípios com algum subsídio e 0 caso contrário. Em exercícios de robustez, apresenta-se também uma codificação com três categorias: ausência de subsídio, subsídio parcial e subsídio integral.

O objetivo do primeiro estágio é verificar a relevância empírica do instrumento. Um coeficiente $\pi_1 > 0$ indica que, em média, a política amplia a distância observada entre residência e escola, resultado compatível com a hipótese de que o subsídio reduz o custo monetário do deslocamento e expande o conjunto de escolas factíveis para as famílias.

Segundo estágio

Por fim, estima-se o modelo por Mínimos Quadrados em Dois Estágios:

$$\log(\text{desempenho}_{id}) = \beta_0 + \beta_1 \widehat{\text{distancia}}_i + \mathbf{X}_i \boldsymbol{\gamma} + \epsilon_i, \quad (3.8)$$

em que $\widehat{\text{distancia}}_i$ corresponde ao componente da distância previsto pela política municipal de passe estudantil e pelas covariáveis observadas.

É importante ressaltar que não se interpreta $\widehat{\text{distancia}}_i$ como uma medida direta de “acessibilidade”. A acessibilidade escolar é entendida, neste trabalho, como um conceito mais amplo e latente, relacionado ao custo efetivo de acesso à escola. A variável instrumentada deve ser interpretada como a parcela da distância observada associada ao regime institucional de subsídio ao transporte.

3.3.5 Hipóteses de identificação, interpretação e limitações

A interpretação do estimador por variáveis instrumentais requer, além da relevância do instrumento, hipóteses fortes de exogeneidade e exclusão. A primeira exige que a política municipal de passe estudantil não esteja correlacionada, condicionalmente aos controles, com determinantes não observados do desempenho escolar. A segunda exige que a política afete a proficiência apenas por meio da distância residência-escola.

Essas hipóteses são muito exigentes no contexto deste estudo. Em primeiro lugar, a política varia em nível municipal, de modo que a identificação decorre de diferenças entre municípios com e sem subsídio, e não de variações individuais dentro de um mesmo município. Em segundo lugar, municípios que oferecem passe estudantil podem diferir sistematicamente dos demais em dimensões não completamente observadas, como capacidade fiscal, qualidade da rede de transporte, estrutura da oferta escolar e composição socioeconômica. Em terceiro lugar, o passe escolar pode funcionar também como um

alívio orçamentário para famílias elegíveis, afetando o desempenho por canais distintos da distância, o que constitui uma ameaça direta à restrição de exclusão.

Essa limitação é particularmente relevante no presente caso, pois a codificação binária da política coincide em larga medida com a comparação entre Belo Horizonte e os demais municípios da amostra. Assim, embora o exercício de IV seja informativo sobre a endogeneidade da distância observada e sobre a relevância institucional do regime de mobilidade, seus resultados devem ser interpretados com cautela. Ao longo da seção de resultados, enfatiza-se tanto as estatísticas de primeiro estágio quanto o contraste entre as estimativas por Mínimos Quadrados Ordinários, forma reduzida e Variáveis Instrumentais, evitando tratar o estimador IV como evidência causal irrefutável.

Entretanto, cabe ainda observar que, sob hipóteses adicionais de monotonicidade e exclusão, o estimador de Wald pode receber uma interpretação local (Imbens; Angrist, 1994). No entanto, dada a natureza municipal do instrumento, o número reduzido de contextos comparados e as possíveis violações da restrição de exclusão, optou-se por interpretar essa estimativa de forma prudente, como evidência sugestiva do efeito da distância induzida pela política e não como parâmetro estrutural definitivo para toda a população.

3.3.6 Heterogeneidade e exercícios de robustez

Além das especificações principais, foram realizados exercícios complementares para avaliar a robustez e a plausibilidade econômica dos resultados. Inicialmente, estimou-se uma versão alternativa do instrumento, distinguindo entre subsídio parcial e integral. Embora essa codificação acrescente flexibilidade, ela também impõe hipóteses adicionais de exogeneidade, razão pela qual seus resultados são interpretados apenas como robustez. Após, serão reapresentados os resultados em desvios-padrão, a fim de facilitar a comparação de magnitudes entre disciplinas.

Na sequência, será explorada a heterogeneidade em grupos para os quais o subsídio ao transporte é economicamente mais plausível como relaxamento de restrição orçamentária, em especial alunos da rede pública e estudantes cujos pais não possuem ensino superior, usando a escolaridade dos responsáveis como *proxy* de vulnerabilidade socioeconômica. Essa análise é importante porque, se o mecanismo do passe opera principalmente via relaxamento de restrição financeira, os efeitos esperados devem ser mais fortes nesses grupos.

Por fim, se complementar a análise com exercícios de sensibilidade que excluam um município por vez, dada a forte dependência da identificação em variação entre poucos contextos urbanos. Esse procedimento permite avaliar em que medida os resultados são influenciados por diferenças estruturais específicas entre municípios, em particular pela comparação implícita entre Belo Horizonte e os demais casos da amostra.

Por fim, dado que a política varia no nível municipal, a inferência estatística deve

ser interpretada com cautela. A presença de poucos municípios limita a precisão de erros-padrão convencionais, de modo que os resultados são apresentados como evidência sugestiva, buscando provocar o início de um debate de maior profundidade sobre os efeitos de políticas públicas no desempenho escolar, mas não como um resultado ou uma interpretação definitiva. Desta forma, apresentam-se de forma explícita sugestões para a continuidade do presente estudo ao final da seção conclusão.

3.4 Base de Dados

Esta seção descreve a base de dados utilizada, o recorte amostral adotado e o comportamento das principais variáveis do estudo. O objetivo é contextualizar empiricamente a estratégia apresentada na seção anterior e documentar os padrões descritivos que motivam as estimações subsequentes.

3.4.1 A base GERES e o recorte analítico

Foram utilizados dados do Estudo Longitudinal da Geração Escolar 2005 (GERES), desenvolvido em parceria com diferentes instituições brasileiras⁶ voltado ao acompanhamento de estudantes ao longo dos anos iniciais do ensino fundamental. O GERES acompanhou os alunos em cinco ondas de avaliação (2005–2008), incluindo testes padronizados de Matemática e Português e questionários aplicados a escolas, turmas e famílias (Brooke; Bonamino, 2011).

Embora a base tenha natureza longitudinal, este artigo adota uma *cross section* referente à onda 2. A escolha decorre do fato de que as variáveis centrais da análise, como a variável explicativa distância residência–escola e os controles familiares, foram observadas nesse momento da pesquisa. Além disso, essa escolha também se deve ao fato de que Salvador não participou de todas as ondas, em razão do desenho específico do acompanhamento naquele município, o que acaba por prejudicar a comparabilidade temporal das variáveis de interesse. Como existe uma diferença entre as séries, já que Salvador ingressa no 2º ano do ensino fundamental e os demais municípios ingressam no 1º ano (nesta pesquisa), este é um importante controle para o painel da *cross-section* observada.

⁶ O GERES foi desenvolvido em parceria pela Universidade Federal de Minas Gerais (UFMG), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Universidade Estadual do Mato Grosso do Sul (UEMS), Universidade Federal da Bahia (UFBA), Universidade Estadual de Campinas (UNICAMP) e a própria Universidade Federal de Juiz de Fora (UFJF). O estudo foi construído em cinco ondas, sendo a primeira delas de avaliação (diagnóstico) antes do ingresso do estudante na primeira série do ensino fundamental, e as demais ondas de avaliação aplicadas ano a ano, em todos os quatro primeiros anos iniciais do ensino fundamental, à exceção de Salvador,

A tabela 31 resume os instrumentos do GERES utilizados no artigo e seus períodos de aplicação.

Tabela 31 – Aplicação dos instrumentos de interesse na pesquisa GERES 2005

Instrumento	1 ^a onda	2 ^a onda	3 ^a onda	4 ^a onda	5 ^a onda
Teste de Português	X	X	X	X	X
Teste de Matemática	X	X	X	X	X
Questionário da Escola	X				
Questionário da Turma	X				
Questionário de Pais	X				

Fonte: Brooke e Bonamino (2011).

3.4.2 Distribuição da política de passe estudantil

A política de passe estudantil, variável institucional central deste estudo, é observada em nível municipal. A Tabela 32 apresenta a distribuição dos estudantes de acordo com a existência e a intensidade do subsídio⁷ ao transporte escolar em seus respectivos municípios.

Tabela 32 – Distribuição dos estudantes da amostra, considerando a política pública municipal de subsídio

Passe Estudantil	Frequência	Percentual (%)
Nenhum benefício no município	3.107	25,3
Benefício de subsídio parcial da tarifa	4.905	39,9
Benefício de subsídio integral da tarifa	4.268	34,8

Fonte: Elaboração própria com base nos dados do GERES (Ano – 2005).

Nota: O benefício é definido em nível municipal. Não se observa diretamente o uso individual do passe, apenas a existência da política no município.

Como se observa, aproximadamente três quartos da amostra residem em municípios com algum tipo de subsídio ao transporte estudantil. Essa distribuição já antecipa uma limitação importante da estratégia empírica, já que a variação institucional ocorre

⁷ A variável de benefício no transporte público foi construída com base na legislação municipal e estadual vigente em 2005 nas localidades analisadas. Belo Horizonte foi classificada como município sem benefício identificado no período, embora a Lei Orgânica do Município previsse sua instituição em situações específicas; essa só foi regulamentada muito mais tarde e inclui apenas estudantes do EJA (educação de jovens e adultos). Campinas foi classificada como um caso de benefício parcial, considerando-se o desconto de 60% no transporte municipal e de 50% no transporte metropolitano, conforme a Lei Municipal nº 9.788, de 2 de julho de 1998, e o Decreto Estadual nº 12.976, de 19 de outubro de 1998. Campo Grande foi classificada como um caso de gratuidade integral no âmbito municipal, nos termos da Lei Municipal nº 3.026, de 27 de dezembro de 1993. O Rio de Janeiro foi classificado como um caso de gratuidade integral, em razão da previsão de benefício total em âmbito municipal e estadual, com base no Decreto nº 4.472, de 2 de março de 1984, e na Lei Estadual nº 4.510, de 13 de janeiro de 2005. Salvador foi classificada como um caso de benefício parcial, com desconto de 50% no transporte municipal, nos termos da Lei nº 6.907, de 10 de janeiro de 1992, e da Lei Municipal nº 5.699/2000.

entre poucos municípios, razão pela qual os resultados posteriores devem ser interpretados com cautela.

3.4.3 Construção das variáveis e controles

Além da variável de política, utilizam-se controles familiares e escolares observados na base. A Tabela 33 resume a classificação das ocupações segundo a CBO.

Tabela 33 – Conversão das funções da CBO em Colarinho Branco e Colarinho Azul

Inicial da CBO	Título da Função	Tipo de Função
0	Forças Armadas, Policiais e Bombeiros Militares	Colarinho Branco
1	Membros Superiores do Poder Público, Dirigentes de Organizações de Interesse Público e de Empresas e Gerentes	Colarinho Branco
2	Profissionais das Ciências e das Artes	Colarinho Branco
3	Técnicos de Nível Médio	Colarinho Branco
4	Trabalhadores de Serviços Administrativos	Colarinho Branco
5	Trabalhadores dos serviços, vendedores dos comércios em lojas e mercados	Colarinho Azul
6	Trabalhadores da agropecuária, florestais e da pesca	Colarinho Azul
7	Trabalhadores que conhecem e trabalham em todas as fases de produção	Colarinho Azul
8	Operadores de máquinas	Colarinho Azul
9	Trabalhadores de manutenção e reparação – ocupações mais elementares	Colarinho Azul
Z	Ocupações mal definidas ou sem emprego atual	Mal especificado

Fonte: Elaboração própria a partir da Classificação Brasileira de Ocupações (CBO 2002).

A Tabela 34 apresenta o dicionário das variáveis utilizadas na análise da *cross section* da onda 2, que foram utilizadas no presente estudo.

Tabela 34 – Dicionário com as variáveis utilizadas

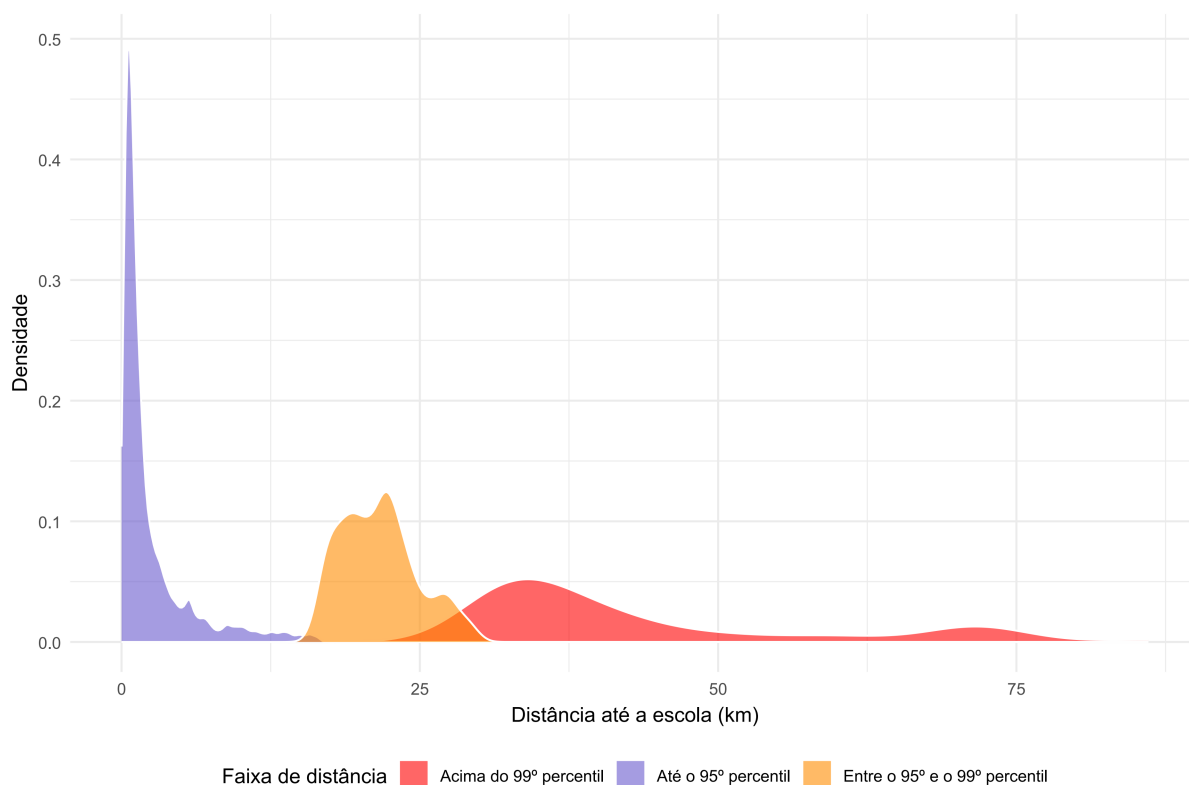
Relação das Variáveis Utilizadas				
Variável GERES	Variável Transformada	Tipo de Variável	Descrição da Variável	Período
Variáveis Dependentes				
Proficiência em matemática	—	Contínua	Nota do Aluno i em Matemática	2005
Proficiência em português	—	Contínua	Nota do Aluno i em Português	2005
Variáveis explicativas relacionadas ao estudante e sua família (invariáveis no tempo)				
Ocupação do Pai	Colarinho Branco Pai	Dummy	Agrupadas como colarinho branco, azul ou sem ocupação	2005
Ocupação da Mãe	Colarinho Branco Mãe	Dummy	Agrupadas como colarinho branco, azul ou sem ocupação	2005
Visita à escola	Reunião de Pais	Dummy	Se os pais compareceram à escola em 2005 para reunião	2005
Visita à escola	Conversa com Docente	Dummy	Se os pais compareceram à escola em 2005 para conversar com docente	2005
Visita à escola	Atividade Comunidade	Dummy	Se os pais participaram de atividade em comunidade na escola em 2005	2005
Escolaridade da Mãe	Grau de Escolaridade	Dummy	Grau de escolaridade da mãe	2005
Escolaridade do Pai	Grau de Escolaridade	Dummy	Grau de escolaridade do pai	2005
Distância Residência-Escola	—	Contínua	Apurada com base na distância entre logradouro da família e da escola	2005
Subsídio Passe Estudantil	—	Dummy	Medida que avalia se em nível municipal existe subsídio, sendo 0 (não) e 1 (sim)	2005
Variáveis de escola				
Município da Escola	—	Catégorica	Município da escola frequentada	2005
Rede	—	Dummy	Se a escola é pública ou privada	2005

Fonte: Elaboração própria usando os dados do GERES (Ano – 2005) e legislações municipais.

3.5 Estatísticas Descritivas

Esta seção apresenta as principais estatísticas descritivas para o painel utilizado. Inicialmente, apresentam-se as características centrais da amostra no que tange à distância entre os logradouros de residência e escola. A Figura 10 mostra que a distribuição da distância entre os logradouros de residência e escola é fortemente assimétrica à direita, com caudas longas e poucos valores extremos. Esses casos podem refletir tanto deslocamentos efetivamente atípicos quanto problemas de geocodificação e inconsistências cadastrais.

Figura 10 – Distribuição da distância entre residência e escola, com destaque para as caudas.



Fonte: Elaboração própria a partir do GERES 2005.

Para avaliar a plausibilidade dos valores extremos, se inspeciona a escala espacial dos municípios analisados. A Figura 11 apresenta a distância máxima entre pontos urbanos nas cidades onde se localizavam as escolas. Comparando a distância limite entre os pontos extremos das cidades pesquisadas durante o projeto GERES e a distância apresentada na figura 10 para o grupo de 1% com maior distância entre os logradouros residência-escola, optou-se por testar e avaliar a composição da amostra truncada no 99º percentil, excluindo esse 1% que representa uma amostra muito pequena, mas com dimensões de distância muito elevadas.

Figura 11 – Distância máxima entre dois pontos urbanos em cinco capitais brasileiras.



Fonte: Elaboração própria com base em dados do IBGE e SRTM.

A Tabela 35 resume o número de observações que são excluídas após o procedimento de truncamento mencionado, distinguindo casos de residência no mesmo município e casos em que o logradouro da residência se localiza em município distinto do logradouro da escola.

Tabela 35 – Alunos excluídos após truncamento por município e local de residência

Município	Alunos	Alunos (P99)	Excluídos (mesma cidade)	Excluídos (outra cidade)	Total excluídos
RIO DE JANEIRO	2.423	2.353	56	14	70
CAMPINAS	3.110	3.077	22	11	33
SALVADOR	1.795	1.781	13	1	14
BELO HORIZONTE	3.107	3.101	2	4	6
CAMPO GRANDE	1.845	1.845	0	0	0

Fonte: Elaboração própria com base nos dados do GERES (Ano – 2005).

Como observado, foram poucos alunos excluídos na amostra (123 alunos), mas com distâncias efetivamente muito grandes quando consideradas as distâncias extremas em nível dos municípios estudados (figura 11), o que reforça o tratamento anterior. Assim, se apresentam na Tabela 36 as estatísticas descritivas da amostra completa e da amostra com truncamento no percentil 99 da distância. Observa-se que o truncamento reduz substancialmente a dispersão da variável distância, sem alterar de forma relevante as distribuições das demais variáveis observadas.

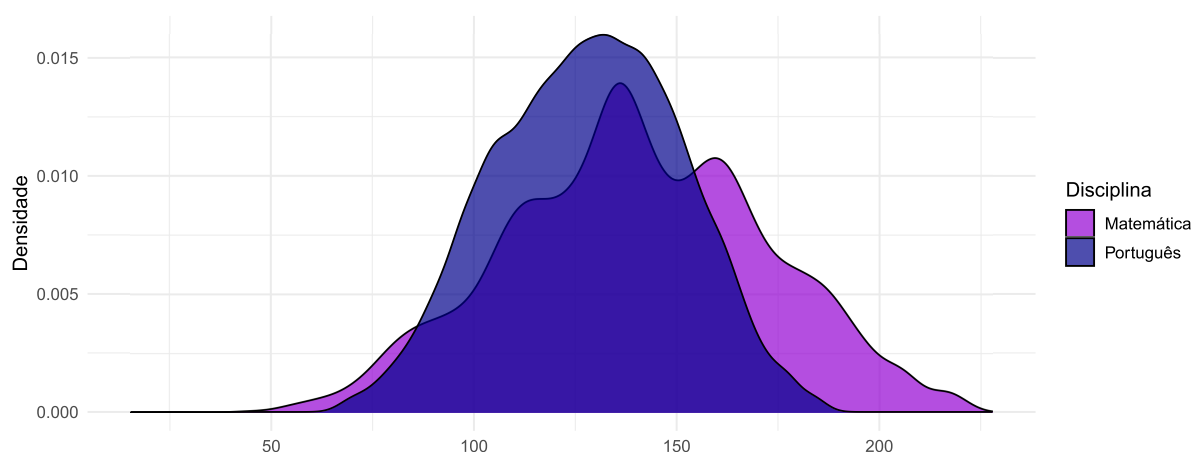
Tabela 36 – Estatísticas Descritivas da amostra com e sem truncamento (*Percentil 99*)

Variável	Amostra completa (com <i>outliers</i>)					Amostra com truncamento (P99)				
	Média	DP	Mín.	Máx.	Obs.	Média	DP	Mín.	Máx.	Obs.
Variáveis Dependentes										
Score em Português (Yis)	128.19	22.67	67.15	183.93	12,280	128.14	22.67	67.15	183.93	12,157
Score em Matemática (Yis)	140.33	32.22	15.31	227.91	12,280	140.26	32.21	15.31	227.91	12,157
Série	1.15	0.36	1.00	2.00	12,280	1.15	0.36	1.00	2.00	12,157
Reunião	0.82	0.39	0.00	1.00	12,280	0.82	0.39	0.00	1.00	12,157
Conversa Docente	0.56	0.50	0.00	1.00	12,280	0.56	0.50	0.00	1.00	12,157
Atividade Comunidade	0.18	0.39	0.00	1.00	12,280	0.18	0.39	0.00	1.00	12,157
Escola Pública	0.79	0.41	0.00	1.00	12,280	0.79	0.41	0.00	1.00	12,157
Colarinho Branco Pais (ambos)	0.13	0.33	0.00	1.00	12,280	0.13	0.33	0.00	1.00	12,157
Colarinho Branco Pais (nenhum)	0.59	0.49	0.00	1.00	12,280	0.59	0.49	0.00	1.00	12,157
Escolaridade da Mãe										
Nunca estudou ou não terminou a 4ª série	0.13	0.33	0.00	1.00	12,280	0.13	0.33	0.00	1.00	12,157
Terminou a 4ª série	0.24	0.43	0.00	1.00	12,280	0.25	0.43	0.00	1.00	12,157
Terminou a 8ª série	0.22	0.42	0.00	1.00	12,280	0.22	0.42	0.00	1.00	12,157
Terminou o ensino médio	0.29	0.45	0.00	1.00	12,280	0.29	0.45	0.00	1.00	12,157
Terminou a Faculdade	0.12	0.33	0.00	1.00	12,280	0.12	0.32	0.00	1.00	12,157
Escolaridade do Pai										
Nunca estudou ou não terminou a 4ª série	0.13	0.33	0.00	1.00	12,280	0.13	0.33	0.00	1.00	12,157
Terminou a 4ª série	0.26	0.44	0.00	1.00	12,280	0.26	0.44	0.00	1.00	12,157
Terminou a 8ª série	0.23	0.42	0.00	1.00	12,280	0.23	0.42	0.00	1.00	12,157
Terminou o ensino médio	0.27	0.44	0.00	1.00	12,280	0.27	0.44	0.00	1.00	12,157
Terminou a Faculdade	0.12	0.32	0.00	1.00	12,280	0.12	0.32	0.00	1.00	12,157
Distância Residência–Escola	3.59	6.48	0.00	86.07	12,280	1.28	1.05	0.00	3.00	12,157
Subsídio Passe Estudantil	0.75	0.43	0.00	1.00	12,280	0.74	0.44	0.00	1.00	12,157

Fonte: Elaboração própria com base nos dados do GERES (Ano – 2005).

Uma vez encerradas as questões do tratamento dado à relação distância residência-escola e apresentadas as estatísticas descritivas para a amostra adotada ao longo do trabalho, a Figura 12 apresenta a distribuição, usando a densidade de kernel para os escores de Matemática e Português. As duas distribuições exibem formato aproximadamente unimodal, embora Matemática apresente maior dispersão e maior heterogeneidade relativa que Português.

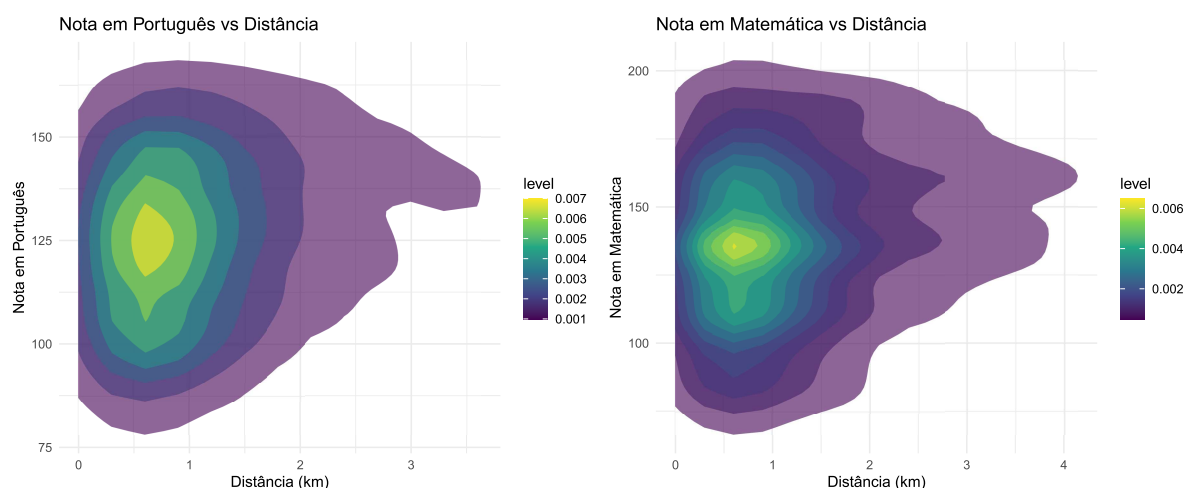
Figura 12 – Densidade de Kernel para matemática e português na onda 2



Fonte: Elaboração própria a partir de dados do GERES (Ano – 2005).

A Figura 13 apresenta a distribuição conjunta entre a distância residência-escola e o desempenho dos estudantes, a partir de uma estimativa de densidade bivariada por núcleo. Em ambas as disciplinas, a maior concentração de alunos com desempenho intermediário e alto ocorre em curtas distâncias, especialmente entre 0 e 2,5 km, sugerindo que trajetos mais curtos estão associados a melhores resultados acadêmicos.

Figura 13 – Estimativa de densidade bivariada por núcleo



Fonte: Elaboração própria com dados do GERES (Ano – 2005). Nota e distância.

Esse padrão deve ser interpretado com cautela, pois a associação bivariada entre distância e desempenho pode refletir simultaneamente o custo de deslocamento e a autoseleção em direção a escolas mais atraentes. É precisamente essa possibilidade que motiva as estimações econométricas para as quais serão reportados os resultados na seção seguinte.

3.6 Resultados

Os resultados apresentados estão divididos em cinco etapas. Primeiro, apresenta-se a forma reduzida da política municipal de passe estudantil sobre o desempenho escolar. Em seguida, apresenta-se o primeiro estágio, isto é, a relação entre a política de subsídio e a distância residência–escola. Depois, apresentam-se as estimações por Mínimos Quadrados Ordinários e por variáveis instrumentais para Matemática e Português e, na sequência, discute-se as magnitudes em desvio-padrão. Finaliza-se a seção sintetizando os resultados à luz da literatura e das principais limitações da estratégia empírica.

Em toda a seção, busca-se adotar uma interpretação cautelosa. Como a política varia no nível do município, a identificação decorre de diferenças entre poucos contextos urbanos e não de variação individual dentro de um mesmo município. Por esse motivo, os resultados por variáveis instrumentais buscam evidenciar a endogeneidade da distância, mas não como prova definitiva de um efeito causal plenamente isolado.

3.6.1 Forma reduzida da política municipal

Antes da estimativa por variáveis instrumentais, é útil examinar a forma reduzida da política de passe estudantil sobre o desempenho escolar. Como a política é definida em nível municipal, essa regressão fornece a evidência mais direta da associação entre a exposição ao ambiente institucional de mobilidade e a proficiência escolar. Na tabela 37 apresentam-se os resultados das estimações da equação 3.6.

Como se pode observar, os resultados da Tabela 37 mostram que a política municipal de passe estudantil está negativamente associada ao desempenho escolar, tanto em Matemática quanto em Português. Em Matemática, o coeficiente estimado é de aproximadamente $-0,035$, enquanto em Português é de cerca de $-0,030$, ambos estatisticamente significativos ao nível de 1%.

Esse resultado não deve ser lido como evidência de que a política de subsídio “reduz” diretamente a aprendizagem. Em vez disso, ele sugere que os municípios com passe estudantil diferem sistematicamente dos municípios sem passe em dimensões também relevantes para o desempenho escolar. Em outras palavras, a forma reduzida já antecipa que o instrumento pode estar captando não apenas a variação institucional no custo do deslocamento, mas também as diferenças estruturais entre contextos urbanos.

Tabela 37 – Forma reduzida: efeito da política municipal de passe estudantil sobre o desempenho em Matemática e Português

Variáveis	Matemática	Português
Política pública municipal de subsídio escolar (<i>dummy</i>)	-0.035*** (0.005)	-0.030*** (0.003)
Nenhum dos pais em ocupação de colarinho branco	-0.033*** (0.005)	-0.028*** (0.004)
Aluno de escola pública	-0.090*** (0.006)	-0.083*** (0.004)
Mãe: nunca estudou / não concluiu 4 ^a série	Categoria de referência	
Mãe: concluiu 4 ^a série	0.051*** (0.007)	0.038*** (0.005)
Mãe: concluiu ensino fundamental	0.080*** (0.008)	0.065*** (0.005)
Mãe: concluiu ensino médio	0.119*** (0.008)	0.106*** (0.006)
Mãe: concluiu ensino superior	0.179*** (0.011)	0.145*** (0.008)
Pai: nunca estudou / não concluiu 4 ^a série	Categoria de referência	
Pai: concluiu 4 ^a série	0.023** (0.007)	0.021*** (0.005)
Pai: concluiu ensino fundamental	0.044*** (0.008)	0.046*** (0.005)
Pai: concluiu ensino médio	0.084*** (0.008)	0.073*** (0.006)
Pai: concluiu ensino superior	0.120*** (0.011)	0.101*** (0.007)
Ida à escola: conversa com docente	-0.035*** (0.004)	-0.025*** (0.003)
Ida à escola: reunião	0.033*** (0.005)	0.029*** (0.004)
Ida à escola: atividade com a comunidade	0.038*** (0.005)	0.030*** (0.004)
Série do aluno	0.128*** (0.006)	0.062*** (0.004)
Constante	4.731*** (0.013)	4.735*** (0.009)
Número de observações	12157	12157
R ² ajustado	0.220	0.295

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota: Variável dependente em log. Coeficientes com erros-padrão entre parênteses. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

3.6.2 Primeiro estágio: política municipal e distância residência–escola

No primeiro estágio, o objetivo é verificar se a política municipal de passe estudantil está associada à distância observada entre a residência e a escola. Essa é a condição de relevância do instrumento. A Tabela 38 apresenta os resultados para a estimação da equação 3.7 e mostra que a existência de passe estudantil está positivamente associada à distância residência–escola. Na especificação principal (*dummy* binária), a política está associada a um aumento médio de quase 0,91 km na distância observada. Na especificação alternativa (subsídio parcial e integral), os coeficientes permanecem positivos e

significativos.

Tabela 38 – Estimação do 1º estágio: regressão da distância no passe escolar

	Modelo 1: <i>dummy</i>	Modelo 2: categorias
Política pública municipal de subsídio escolar (<i>dummy</i>)	0.906*** (0.105)	
Política pública municipal de subsídio escolar (parcial)		1.157*** (0.126)
Política pública municipal de subsídio escolar (integral)		0.730*** (0.116)
Nenhum dos pais em ocupação de colarinho branco	-0.104 (0.110)	-0.109 (0.110)
Aluno de escola pública	-0.390*** (0.133)	-0.340** (0.134)
Ida à escola: conversa com docente	-0.064 (0.089)	-0.036 (0.089)
Ida à escola: reunião	0.187 (0.114)	0.176 (0.114)
Ida à escola: atividade com a comunidade	-0.029 (0.117)	-0.025 (0.117)
Mãe: nunca estudou / não concluiu 4ª série	Categoria de referência	
Mãe: concluiu 4ª série	-0.244 (0.160)	-0.243 (0.160)
Mãe: concluiu ensino fundamental	-0.407** (0.170)	-0.401** (0.170)
Mãe: concluiu ensino médio	0.121 (0.179)	0.123 (0.179)
Mãe: concluiu ensino superior	0.980*** (0.239)	0.986*** (0.239)
Pai: nunca estudou / não concluiu 4ª série	Categoria de referência	
Pai: concluiu 4ª série	-0.146 (0.158)	-0.146 (0.158)
Pai: concluiu ensino fundamental	-0.325* (0.169)	-0.321* (0.169)
Pai: concluiu ensino médio	-0.182 (0.178)	-0.174 (0.178)
Pai: concluiu ensino superior	0.096 (0.235)	0.116 (0.235)
Série do aluno	0.710*** (0.131)	0.454*** (0.149)
Constante	2.098*** (0.287)	2.302*** (0.292)
Número de observações	12157	12157
R ² ajustado	0.024	0.025

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota: Variável dependente: distância residência-escola em quilômetros. Coeficientes com erros-padrão entre parênteses. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Os resultados demonstram que o passe estudantil não altera a geografia física entre casa e escola, mas reduz o custo monetário do deslocamento, ampliando o conjunto de escolas factíveis para as famílias. Em consequência, alunos expostos à política podem frequentar escolas mais distantes. Ainda assim, o primeiro estágio não deve ser confundido com a validação completa do instrumento, pois ele demonstra relevância empírica, mas não resolve, por si só, as preocupações de exogeneidade e restrição de exclusão.

3.6.3 Distância residência–escola e desempenho em Matemática

A Tabela 39 apresenta as estimativas para a proficiência em Matemática, cuja especificação foi apresentada na equação 3.8.

Tabela 39 – Efeito da distância residência–escola sobre o desempenho em Matemática

Variáveis	OLS principal	OLS + passe	IV – <i>dummy</i>	IV – categorias
Distância residência–escola (km)	0.002*** (0.0004)	0.002*** (0.0004)	-0.039*** (0.0071)	-0.062*** (0.0084)
Política pública municipal de subsídio escolar (<i>dummy</i>)		-0.037*** (0.0048)		
Nenhum dos pais em ocupação de colarinho branco	-0.033*** (0.0050)	-0.033*** (0.0050)	-0.038*** (0.0068)	-0.040*** (0.0088)
Aluno de escola pública	-0.087*** (0.0060)	-0.089*** (0.0060)	-0.105*** (0.0087)	-0.115*** (0.0111)
Mãe: nunca estudou / não concluiu 4ª série		Categoria de referência		
Mãe: concluiu 4ª série	0.053*** (0.0072)	0.052*** (0.0072)	0.042*** (0.0100)	0.036*** (0.0128)
Mãe: concluiu ensino fundamental	0.080*** (0.0077)	0.081*** (0.0077)	0.064*** (0.0108)	0.055*** (0.0138)
Mãe: concluiu ensino médio	0.117*** (0.0081)	0.118*** (0.0081)	0.123*** (0.0111)	0.127*** (0.0142)
Mãe: concluiu ensino superior	0.175*** (0.0108)	0.177*** (0.0108)	0.218*** (0.0164)	0.242*** (0.0208)
Pai: nunca estudou / não concluiu 4ª série		Categoria de referência		
Pai: concluiu 4ª série	0.024*** (0.0072)	0.023*** (0.0071)	0.017* (0.0097)	0.013 (0.0125)
Pai: concluiu ensino fundamental	0.045*** (0.0077)	0.045*** (0.0077)	0.031*** (0.0107)	0.023* (0.0136)
Pai: concluiu ensino médio	0.083*** (0.0081)	0.084*** (0.0081)	0.077*** (0.0110)	0.073*** (0.0141)
Pai: concluiu ensino superior	0.120*** (0.0106)	0.120*** (0.0106)	0.124*** (0.0144)	0.127*** (0.0185)
Ida à escola: conversa com docente	-0.035*** (0.0040)	-0.034*** (0.0040)	-0.037*** (0.0055)	-0.038*** (0.0070)
Ida à escola: reunião	0.033*** (0.0052)	0.033*** (0.0052)	0.041*** (0.0071)	0.045*** (0.0091)
Ida à escola: atividade com a comunidade	0.042*** (0.0053)	0.038*** (0.0053)	0.037*** (0.0072)	0.033*** (0.0092)
Série do aluno	0.115*** (0.0058)	0.127*** (0.0059)	0.156*** (0.0105)	0.179*** (0.0131)
Constante	4.711*** (0.013)	4.726*** (0.013)	4.813*** (0.025)	4.870*** (0.031)
Número de observações	12157	12157	12157	12157
R ² ajustado	0.218	0.222	-0.428	-1.364
Weak instruments (F)			74.00***	43.69***
Wu–Hausman			61.49***	179.41***
Sargan (sobreidentificação)			–	42.07***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota: Variável dependente em log da proficiência em Matemática. Coeficientes com erros-padrão entre parênteses. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

A coluna de Mínimos Quadrados Ordinários (MQO) indica associação positiva entre distância e desempenho, com coeficiente de aproximadamente 0,002 no log da nota por quilômetro. Esse resultado é compatível com a seleção positiva na distância observada, onde famílias mais informadas ou mais capazes de contornar as fricções de deslocamento

podem aceitar trajetos mais longos para matricular seus filhos em escolas percebidas como superiores. Ao se adicionar a variável *dummy* de passe como controle no MQO, o coeficiente da distância permanece positivo e o coeficiente da política aparece negativo. Essa especificação é apenas um exercício de sensibilidade descritiva e não deve ser confundida com a estratégia de identificação por variáveis instrumentais. Esse resultado dialoga com o resultado apresentado na tabela 37.

O resultado central aparece nas colunas quando são adicionadas as Variáveis Instrumentais. Quando a distância é instrumentada pela *dummy* de passe estudantil, o sinal se inverte e a magnitude aumenta substancialmente: o coeficiente da distância passa a $-0,039$, estatisticamente significativo a 1%. Esse contraste entre o estimador de MQO e o de IV é um dos principais achados do trabalho, pois sugere que a distância observada combina uma margem de oportunidade e uma margem de custo. No dado bruto, trajetos mais longos podem estar associados ao acesso a escolas mais atraentes. Quando se isola a parcela da distância associada ao regime institucional de mobilidade, emerge o componente de custo do deslocamento, que aparece negativamente relacionado à aprendizagem.

Os testes diagnósticos reforçam a leitura anterior. A estatística de instrumento fraco no modelo principal é elevada ($F = 74$), indicando a relevância do instrumento, e o teste de Wu–Hausman rejeita a hipótese de exogeneidade da distância. Já a especificação com categorias produz um coeficiente ainda mais negativo, mas o teste de Sargan rejeita a validade conjunta dos instrumentos, razão pela qual é interpretada apenas como robustez nesta seção.

3.6.4 Distância residência–escola e desempenho em Português

A Tabela 40 apresenta os resultados para a proficiência em Português, também seguindo a especificação apresentada na equação 3.8. O padrão é muito semelhante ao encontrado em Matemática. No modelo estimado por MQO, a distância aparece positivamente associada ao desempenho, com coeficiente de aproximadamente 0,002. Mais uma vez, essa associação é consistente com um processo de *sorting* escolar no qual famílias com maior capacidade de escolha aceitam trajetos mais longos para acessar escolas melhores.

Quando a distância é instrumentada pela política municipal de passe estudantil, o sinal se inverte. Na especificação principal de Variáveis Instrumentais, o coeficiente passa a $-0,033$, estatisticamente significativo. Isso indica que, ao isolar a parcela da distância associada ao regime municipal de subsídio ao transporte, trajetos mais longos passam a se relacionar a um pior desempenho em Português.

Tabela 40 – Efeito da distância residência–escola sobre o desempenho em Português

Variáveis	OLS principal	OLS + passe	IV – <i>dummy</i>	IV – categorias
Distância residência–escola (km)	0.002*** (0.0003)	0.002*** (0.0003)	-0.033*** (0.0055)	-0.056*** (0.0071)
Política pública municipal de subsídio escolar (<i>dummy</i>)		-0.032*** (0.0034)		
Nenhum dos pais em ocupação de colarinho branco	-0.028*** (0.0035)	-0.028*** (0.0035)	-0.032*** (0.0052)	-0.034*** (0.0073)
Aluno de escola pública	-0.080*** (0.0043)	-0.082*** (0.0042)	-0.095*** (0.0067)	-0.106*** (0.0093)
Mãe: nunca estudou / não concluiu 4ª série		Categoria de referência		
Mãe: concluiu 4ª série	0.039*** (0.0051)	0.038*** (0.0051)	0.030*** (0.0077)	0.024** (0.0107)
Mãe: concluiu ensino fundamental	0.065*** (0.0054)	0.066*** (0.0054)	0.052*** (0.0083)	0.043*** (0.0116)
Mãe: concluiu ensino médio	0.105*** (0.0057)	0.106*** (0.0057)	0.110*** (0.0085)	0.113*** (0.0119)
Mãe: concluiu ensino superior	0.141*** (0.0076)	0.143*** (0.0076)	0.177*** (0.0126)	0.201*** (0.0174)
Pai: nunca estudou / não concluiu 4ª série		Categoria de referência		
Pai: concluiu 4ª série	0.022*** (0.0050)	0.022*** (0.0050)	0.016** (0.0075)	0.013 (0.0105)
Pai: concluiu ensino fundamental	0.047*** (0.0054)	0.046*** (0.0054)	0.035*** (0.0082)	0.027** (0.0114)
Pai: concluiu ensino médio	0.073*** (0.0057)	0.074*** (0.0057)	0.067*** (0.0085)	0.064*** (0.0118)
Pai: concluiu ensino superior	0.100*** (0.0075)	0.101*** (0.0075)	0.104*** (0.0111)	0.107*** (0.0155)
Ida à escola: conversa com docente	-0.025*** (0.0028)	-0.025*** (0.0028)	-0.027*** (0.0042)	-0.028*** (0.0059)
Ida à escola: reunião	0.029*** (0.0037)	0.029*** (0.0036)	0.035*** (0.0055)	0.040*** (0.0077)
Ida à escola: atividade com a comunidade	0.034*** (0.0037)	0.030*** (0.0037)	0.029*** (0.0055)	0.026*** (0.0078)
Série do aluno	0.051*** (0.0041)	0.060*** (0.0042)	0.085*** (0.0081)	0.108*** (0.0110)
Constante	4.718*** (0.009)	4.731*** (0.009)	4.804*** (0.019)	4.861*** (0.026)
Número de observações	12157	12157	12157	12157
R ² ajustado	0.293	0.298	-0.541	-2.025
Weak instruments (F)			74.00***	43.69***
Wu–Hausman			87.83***	293.16***
Sargan (sobreidentificação)			–	59.29***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota: Variável dependente em log da proficiência em Português. Coeficientes com erros-padrão entre parênteses. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Assim como em Matemática, a especificação categórica produz um efeito ainda mais negativo, mas a rejeição do teste de Sargan a impede de ser tratada como especificação principal. Em ambos os casos, o contraste entre MQO e IV reforça a tese central do artigo, onde se busca associar a distância residência-escola como uma medida que não pode ser tratada como puramente exógena em relação ao custo de acesso, pois incorpora simultaneamente fricções de deslocamento e a escolha de escolas mais atrativas.

3.6.5 Resultados em desvio-padrão e magnitude dos efeitos

Na literatura sobre desempenho educacional, é comum estimar os *resultados* de desempenho em desvios padrão. Assim, a tabela 41 reapresenta os resultados principais em escala padronizada, de modo a tornar comparáveis as magnitudes estimadas por MQO e por variáveis instrumentais. Nessa especificação, tanto as proficiências em Matemática e Português quanto a distância residência-escola são expressas em desvios-padrão.

Tabela 41 – Distância residência-escola e desempenho (em desvios-padrão)

	OLS Matemática	OLS Português	IV Matemática	IV Português
Distância residência-escola (sd-score)	0,044*** (0,008)	0,051*** (0,008)	-0,840*** (0,145)	-0,917*** (0,149)
Nenhum dos pais ocupação de colarinho branco	-0,135*** (0,020)	-0,156*** (0,019)	-0,154*** (0,028)	-0,175*** (0,029)
Aluno de escola pública	-0,365*** (0,024)	-0,447*** (0,023)	-0,444*** (0,036)	-0,533*** (0,037)
Ida à escola (conversa com docente)	-0,143*** (0,016)	-0,137*** (0,015)	-0,153*** (0,023)	-0,149*** (0,023)
Ida à escola (reunião)	0,129*** (0,021)	0,153*** (0,020)	0,162*** (0,030)	0,190*** (0,030)
Ida à escola (atividade da comunidade)	0,178*** (0,021)	0,192*** (0,020)	0,154*** (0,030)	0,165*** (0,031)
Mãe terminou a 4ª série	0,183*** (0,029)	0,197*** (0,027)	0,136*** (0,042)	0,145*** (0,043)
Mãe concluiu o ensino fundamental	0,291*** (0,031)	0,332*** (0,029)	0,223*** (0,045)	0,257*** (0,046)
Mãe concluiu o ensino médio	0,446*** (0,032)	0,552*** (0,031)	0,472*** (0,046)	0,581*** (0,047)
Mãe concluiu o ensino superior	0,725*** (0,043)	0,779*** (0,041)	0,913*** (0,068)	0,985*** (0,070)
Pai terminou a 4ª série	0,098*** (0,028)	0,108*** (0,027)	0,068* (0,041)	0,074* (0,042)
Pai concluiu o ensino fundamental	0,180*** (0,030)	0,244*** (0,029)	0,120** (0,044)	0,178*** (0,046)
Pai concluiu o ensino médio	0,347*** (0,032)	0,389*** (0,031)	0,319*** (0,046)	0,358*** (0,047)
Pai concluiu o ensino superior	0,533*** (0,042)	0,571*** (0,040)	0,553*** (0,060)	0,593*** (0,062)
Série	0,460*** (0,023)	0,269*** (0,022)	0,641*** (0,044)	0,467*** (0,045)
Constante	-0,766*** (0,051)	-0,577*** (0,049)	-0,895*** (0,075)	-0,718*** (0,077)
Número de observações	12.157	12.157	12.157	12.157
R^2 ajustado	0,241	0,310	-0,528	-0,611
Weak instruments	—	—	74,00***	74,00***
Wu-Hausman	—	—	75,41***	99,62***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota 1: Coeficientes com erros-padrão entre parênteses. As variáveis dependentes e a distância foram padronizadas em z-score. Assim, o coeficiente da distância indica a variação, em desvios-padrão do desempenho, associada a um aumento de 1 desvio-padrão na distância residência-escola. Nas colunas IV, a distância foi instrumentada pela *dummy* de política municipal de passe estudantil. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Os resultados por MQO padronizado mantêm o mesmo padrão já observado nas

regressões em logaritmo das notas, onde a distância aparece associada positivamente ao desempenho, embora com magnitude pequena. Em Matemática, um aumento de 1 desvio-padrão na distância está associado a um aumento de aproximadamente 0,044 desvio-padrão na proficiência. Em Português, a magnitude correspondente é de cerca de 0,051 desvio-padrão. Esses coeficientes positivos são consistentes com a interpretação de seleção positiva, onde alunos que estudam mais longe podem estar acessando escolas percebidas como mais atrativas, de modo que a distância observada incorpora não apenas custo de deslocamento, mas também uma margem de oportunidade.

Quando se evolui para a estimação por variáveis instrumentais, o padrão se altera de forma substantiva. No modelo IV com a *dummy* de política municipal de passe estudantil como instrumento, um aumento de 1 desvio-padrão na distância está associado a uma redução de aproximadamente 0,840 desvio-padrão no desempenho em Matemática e de cerca de 0,917 desvio-padrão no desempenho em Português. Em outras palavras, quando se isola a parcela da distância associada ao regime institucional de mobilidade escolar, a distância deixa de refletir uma margem de acesso a escolas potencialmente melhores e passa a revelar, com maior nitidez, seu componente de custo.

Esse contraste entre MQO e IV reforça a interpretação principal do artigo, combinando a distância observada nos dados brutos a duas margens distintas: de um lado, uma margem de escolha e oportunidade, pela qual algumas famílias aceitam trajetos maiores para acessar escolas mais desejáveis; e, por outro lado, uma margem de custo, associada ao esforço, ao tempo e às fricções do deslocamento diário. O estimador por variáveis instrumentais sugere que, uma vez isolada a variação institucional da distância, prevalece o componente de custo, negativamente relacionado à aprendizagem.

Em ambos os casos, a estatística de instrumentos fracos permanece elevada ($F = 74$), rejeitando a hipótese de instrumento fraco. O teste de Wu–Hausman também rejeita a exogeneidade da distância, tanto em Matemática quanto em Português, indicando que o estimador por MQO está enviesado e reforçando a relevância da estratégia de variáveis instrumentais.

Ainda assim, essas magnitudes devem ser interpretadas com cautela. Os coeficientes padronizados estimados por IV são elevados quando comparados aos efeitos usualmente relatados para intervenções educacionais, o que sugere que o estimador pode estar captando não apenas a variação da distância induzida pela política de passe, mas também diferenças municipais mais amplas relacionadas à estrutura urbana, à oferta escolar e às condições locais de mobilidade. Por essa razão, a evidência deve ser lida como um fato importante relacionado à distância induzida pela política de subsídio e não como um efeito médio universal aplicável a toda a população.

3.6.6 Síntese e diálogo dos resultados com a literatura

Os resultados apontam para quatro conclusões principais. A primeira delas é que a política municipal de passe estudantil está positivamente associada à distância residência-escola, o que é compatível com a hipótese de que o subsídio amplia o raio de escolha das famílias. A segunda delas é que a forma reduzida da política sobre o desempenho é negativa, sugerindo que municípios com e sem passe diferem sistematicamente em dimensões relevantes para a proficiência escolar. Além disso, a distância observada apresenta associação positiva com o desempenho em MQO, o que é consistente com a seleção para escolas mais atrativas. Por fim, quando se isola a parcela da distância associada ao regime institucional de mobilidade, o sinal se torna negativo em ambas as disciplinas.

Esses resultados reforçam que a distância residência-escola não deve ser tratada como uma barreira geográfica exógena e homogênea, deixando claro que, em contextos urbanos, ela resulta da interação entre escolha escolar, restrições familiares e condições locais de mobilidade. Ao mesmo tempo, a evidência não deve ser tratada como uma relação causal irrestrita, já que a identificação depende de poucos municípios e a restrição de exclusão permanece exigente. Por isso, a principal contribuição deste artigo não é afirmar um efeito causal definitivo da distância, mas mostrar que sua interpretação depende do regime local de mobilidade e que a política de transporte escolar constitui uma margem institucional relevante para compreender a geografia das oportunidades educacionais.

Dialogando com a literatura, os resultados se aproximam de duas tradições complementares. A primeira é a literatura que interpreta a proximidade ou acessibilidade como determinantes relevantes das trajetórias educacionais. Em Card (1993), a proximidade a instituições de ensino funciona como uma fonte de variação exógena no acesso à educação, reforçando a ideia de que a geografia importa não apenas como localização física, mas como restrição econômica ao investimento educacional. Duflo (2001) mostra que políticas públicas que alteram a oferta ou o acesso à escola podem produzir mudanças substantivas na escolarização e nos retornos futuros, tendo em vista que o incremento de instalações de escolas próximas amplia em 12% a probabilidade de uma criança afetada concluir o ensino primário, além de aumentar a permanência no ensino. Embora o desenho empírico do presente artigo seja distinto, esses trabalhos oferecem o enquadramento mais geral para nossa hipótese, comprovando que mudanças institucionais no custo de acesso à escola podem alterar decisões educacionais observadas.

Em um segundo arcabouço, avalia-se a literatura que está mais diretamente relacionada ao mecanismo avaliado neste trabalho, que é a literatura sobre transporte, acessibilidade e escolha escolar. Dustan e Ngo (2018) destacam, ao estudarem a cidade do México, que a expansão do transporte público pode aumentar o acesso a escolas de alta qualidade e mais distantes para determinados grupos de estudantes (estudantes de alto desempenho com pais altamente educados), mas não é suficiente para beneficiar todos

os alunos igualmente. Na mesma direção, Ramírez-Hassan *et al.* (2023) encontram que subsídios ao transporte escolar permitem que estudantes de baixa renda acessem escolas de melhor qualidade, justamente porque reduzem o custo monetário do deslocamento e ampliam o espectro de escolha. Já Moreno-Monroy, Lovelace e Ramos (2018), estudando a Região Metropolitana de São Paulo, enfatizam que a desigualdade educacional urbana está profundamente associada à desigualdade de acessibilidade por transporte público, reforçando a ideia de que a distância física, isoladamente, é uma medida incompleta do custo efetivo de acesso à escola.

Os exercícios complementares reportados no Apêndice ajudam a qualificar essa interpretação. Nas subamostras de alunos da rede pública e de estudantes cujos pais não possuem ensino superior, os coeficientes de IV permanecem negativos e estatisticamente significativos. Entre alunos da rede pública, um quilômetro adicional na distância instrumentada está associado a reduções aproximadas de 3,7% em Matemática e 3,1% em Português. Entre alunos cujos pais não possuem ensino superior, as magnitudes são ainda maiores, cerca de 4,5% em Matemática e 3,8% em Português. Esses resultados são consistentes com a hipótese de que o custo de deslocamento pesa mais para grupos mais expostos a restrições de mobilidade. Na rede privada, por outro lado, o instrumento é fraco e os coeficientes de IV não são estatisticamente significativos, o que impede uma interpretação desse recorte.

O exercício de sensibilidade *leave-one-city-out*, também apresentado no Apêndice, mostra que as estimativas dependem da composição municipal da amostra. A exclusão de Belo Horizonte elimina a variação do instrumento, pois todos os municípios remanescentes passam a apresentar algum tipo de subsídio, enquanto a exclusão de Campinas reduz drasticamente o efeito estimado, e a exclusão do Rio de Janeiro torna os coeficientes substancialmente mais negativos. Esse exercício não deve ser lido como evidência de efeitos municipais específicos, mas como indicação de que a estimativa principal sintetiza contextos institucionais distintos e depende da comparação entre poucos municípios.

3.7 Conclusão

Este artigo investigou a relação entre a distância residência-escola e o desempenho acadêmico nos anos iniciais do ensino fundamental. Buscou-se fomentar a discussão de que a distância até a escola frequentada não deve ser interpretada automaticamente como uma medida exógena de custo de acesso, pois, em contextos urbanos, ela pode refletir tanto restrições de mobilidade quanto escolhas familiares por escolas percebidas como mais atrativas.

Nas estimações por Mínimos Quadrados Ordinários, a distância residência-escola aparece positivamente associada ao desempenho em Matemática e Português. Essa associação, contudo, não deve ser lida como evidência de que estudar mais distante melhora

o aprendizado, mas sim que a distância observada incorpora seleção familiar, onde alunos que percorrem trajetos maiores podem pertencer a famílias com maior informação, maior capacidade de mobilidade ou maior disposição para acessar escolas consideradas melhores.

Quando a distância é instrumentada pela política municipal de passe estudantil, o sinal se inverte. O primeiro estágio indica que a existência de subsídio ao transporte está associada a maiores distâncias entre a residência e a escola. No segundo estágio, um quilômetro adicional na distância instrumentada pela política institucional em nível municipal está associado a reduções de aproximadamente 3,9% na proficiência em Matemática e 3,3% em Português. O contraste entre MQO e IV sugere que a distância observada combina dois componentes: a seleção positiva na escolha escolar e os custos reais de deslocamento.

Os exercícios complementares reforçam essa interpretação, mas também delimitam seu alcance. Os resultados são mais claros entre alunos da rede pública e entre estudantes cujos pais não possuem ensino superior, grupos para os quais o custo monetário do transporte tende a ser mais relevante. Na rede privada, o instrumento é fraco, indicando que o passe estudantil explica pouco da variação da distância nesse grupo, que tende a ser socialmente menos vulnerável e menos exposto aos benefícios e restrições causadas por essa política institucional. Além disso, os exercícios de sensibilidade mostram que as estimativas dependem da composição municipal da amostra, em particular da comparação entre Belo Horizonte, que foi o único município pesquisado sem benefício identificado para o público analisado, e os demais municípios.

A contribuição do artigo está em preencher uma lacuna na literatura econômica que trata a distância como uma variável instrumental potencialmente exógena. Ao mostrar que a distância até a escola frequentada pode assumir sinais distintos conforme a fonte de variação explorada, o artigo aproxima-se da literatura sobre mobilidade urbana, escolha escolar e desigualdade educacional, demonstrando que transporte e oferta escolar não operam separadamente. A política de mobilidade pode ampliar o conjunto de escolas acessíveis às famílias, mas não elimina todos os custos associados a trajetos mais longos.

Do ponto de vista de política pública, os resultados sugerem que subsídios ao transporte estudantil devem ser pensados em conjunto com a distribuição territorial das escolas e com a qualidade da oferta educacional. Ampliar o raio de escolha das famílias pode ser desejável, mas trajetos mais longos continuam envolvendo custos de tempo, esforço, previsibilidade e organização familiar. Políticas de transporte escolar, portanto, não substituem o planejamento territorial da rede de ensino, conforme apresentado por Duflo (2001).

Limitações importantes devem ser consideradas. Embora sejam resultados importantes no contexto da literatura de mobilidade escolar, os resultados devem ser interpretados com cautela. A política de passe estudantil é observada no nível municipal e não

no nível individual; por isso, não se sabe quais alunos utilizaram efetivamente o benefício, nem se o subsídio alterou diretamente a escola escolhida. A identificação também se apoia em poucos municípios, e a restrição de exclusão é exigente, pois o passe pode afetar o desempenho por canais distintos da distância, como alívio orçamentário, frequência escolar ou estabilidade do deslocamento. Além disso, este estudo se restringiu a uma pesquisa realizada há mais de 20 anos, o que pode não refletir os padrões atuais de mobilidade. Essa composição restringe a generalização dos resultados para o conjunto do país, especialmente diante da heterogeneidade econômica, social, institucional e educacional existente entre municípios e comunidades escolares. Estudos futuros deveriam replicar a análise em um número maior de municípios, redes de ensino e comunidades, incorporando diferentes regiões e contextos socioeconômicos.

Assim, os resultados devem ser interpretados de forma cautelosa e como evidência sobre a endogeneidade da distância residência-escola e não como efeito causal homogêneo de uma única política. Para uma agenda futura, seria importante combinar bases com informações individuais sobre elegibilidade e uso efetivo do passe, um maior número de municípios e medidas mais completas de acessibilidade.

Referências do capítulo

BALIYAN, Som Pal; KHAMA, Dira. How Distance to School and Study Hours after School Influence Students' Performance in Mathematics and English: A Comparative Analysis. **Journal of Education and e-Learning Research**, Asian Online Journal Publishing Group, v. 7, n. 2, p. 209–217, 2020. Received: 29 April 2020; Revised: 5 June 2020; Accepted: 7 July 2020; Published: 20 July 2020. DOI: 10.20448/journal.509.2020.72.209.217.

BAMMOU, Hamid; EL BOUHALI, Mohammed; EL ALAOUI, Aicha. Impact of the Distance Travelled to School on Students' Performance, Tardiness, and Absenteeism in Public Schools in Morocco. **European Scientific Journal, ESJ**, European Scientific Institute, v. 20, n. 34, p. 117, 2024. Submitted: 16 Sep 2024; Accepted: 17 Dec 2024; Published: 31 Dec 2024. DOI: 10.19044/esj.2024.v20n34p117. Disponível em: <https://doi.org/10.19044/esj.2024.v20n34p117>.

BARBOSA, Felipe Alves Saraiva. **Impacto da distância casa-escola e do transporte escolar no desempenho escolar dos alunos da rede pública de Maracanaú**. 2024. f. 31. Dissertação de Mestrado em Economia do Setor Público – Faculdade de Economia, Administração, Atuária e Contabilidade (FEAAC), Universidade Federal do Ceará (UFC), Fortaleza, CE.

BROOKE, N.; BONAMINO, A. **Geres 2005: Razões e Resultado de uma pesquisa longitudinal sobre eficácia escolar**. Rio de Janeiro: WalPrint, 2011. Disponível em: https://www.laedpuc.net.br/_files/ugd/fc35a5_a0f50bf2665448af8c06be6b3c089460.pdf.

CARD, David. **Using geographic variation in college proximity to estimate the return to schooling**. [S. l.], 1993. DOI: 10.3386/w4483. Disponível em: <https://doi.org/10.3386/w4483>.

DING, Peng; FENG, Shuaizhang. How School Travel Affects Children's Psychological Well-Being and Academic Achievement in China. **International Journal of Environmental Research and Public Health**, MDPI, v. 19, n. 21, p. 13881, 2022. DOI: 10.3390/ijerph192113881. Disponível em: <https://doi.org/10.3390/ijerph192113881>.

DUFLO, Esther. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. **American Economic Review**, v. 91, n. 4, p. 795–813, 2001. DOI: 10.1257/aer.91.4.795. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.91.4.795>.

DUSTAN, Andrew; NGO, Diana K. L. Commuting to Educational Opportunity? School Choice Effects of Mass Transit Expansion in Mexico City. **Economics of Education Review**, v. 63, p. 116–133, 2018. ISSN 0272-7757. DOI: 10.1016/j.econedurev.2018.01.006. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0272775717303369>.

GERES. **Banco de Dados do GERES, 2005: GERES – Estudo Longitudinal da Geração Escolar 2005**. [S. l.: s. n.], 2005. Banco de dados, notas técnicas e informações complementares. Acesso em: Wed, 08 Feb 2023 12:53:03 GMT. Disponível em: <https://www.laedpuc.net.br/projetogeres>. Acesso em: 8 fev. 2023.

GOULART, Livia Mara Lima; MORAIS, Elaine Andrade de; VIEIRA JÚNIOR, Niltom. Tempo de permanência no transporte escolar sobre o desempenho estudantil. **Revista Interterritórios**, Instituto Federal de Roraima, v. 5, n. 9, 2019. DOI: 10.33052/inter.v5i9.243594. Disponível em: <https://doi.org/10.33052/inter.v5i9.243594>.

IMBENS, Guido W.; ANGRIST, Joshua D. Identification and Estimation of Local Average Treatment Effects. **Econometrica**, Wiley e Econometric Society, v. 62, n. 2, p. 467–475, 1994. ISSN 0012-9682. Disponível em: <https://www.jstor.org/stable/2951620>. Acesso em: 10 fev. 2026.

MARTINS, Ana Paula Antunes. **Análise dos impactos das condições do transporte escolar rural no rendimento escolar dos alunos**. 2010. xi, 118 f. Dissertação de Mestrado – Universidade de Brasília, Brasília. Orientador: José Matsuo Shimoishi. Disponível em: <https://repositorio.unb.br/handle/10482/8791>.

MCFADDEN, Daniel. Conditional Logit Analysis of Qualitative Choice Behavior. In: ZAREMBKA, Paul (ed.). **Frontiers in Econometrics**. New York: Academic Press, 1974. p. 105–142.

MORENO-MONROY, Ana I.; LOVELACE, Robin; RAMOS, Frederico R. Public transport and school location impacts on educational inequalities: Insights from São Paulo. **Journal of Transport Geography**, v. 67, p. 110–118, 2018. ISSN 0966-6923. DOI: <https://doi.org/10.1016/j.jtrangeo.2017.08.012>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0966692316303453>.

PEPE, Marco Antonio Monteiro Villela. **Passe livre estudantil e frequência escolar: uma avaliação dos efeitos do Passe Livre Estudantil na frequência escolar dos educandos da rede pública municipal de São Paulo**. 2017. Dissertação de Mestrado – Escola de Administração de Empresas de São Paulo (EAESP), São Paulo. Orientador: Ciro Biderman. Membros da banca: Ricardo Paes de Barros, André Portela. Disponível em: <https://hdl.handle.net/10438/18192>.

PETEROS, Emerson D. *et al.* Effects of School Proximity on Students' Performance in Mathematics. **Open Journal of Social Sciences**, v. 10, p. 365–376, 2022. ISSN 2327-5960. DOI: 10.4236/jss.2022.107027. Disponível em: <https://www.scirp.org/journal/jss>.

RAMÍREZ-HASSAN, Andrés *et al.* What kind of schools parents choose when they have more options? Effects of school transport subsidies. **Socio-Economic Planning Sciences**, v. 87, p. 101509, 2023. ISSN 0038-0121. DOI: <https://doi.org/10.1016/j.seps.2023.101509>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012123000022>.

ZERAGABER, Tsinat Yemane; TEAME, Ghirmai Tesfamariam; TSIGHE, Zemenfes. Assessing the effect of home-to-school distance on student dropout rate in Adi-Keyih sub-zone, Eritrea. **International Journal of Educational Research Open**, v. 7, p. 100340, 2024. ISSN 2666-3740. DOI: <https://doi.org/10.1016/j.ijedro.2024.100340>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666374024000220>.

APÊNDICE

Apêndice A - Heterogeneidade

As Tabelas 42 e 43 apresentam exercícios de heterogeneidade em dois grupos nos quais o mecanismo econômico do passe estudantil tende a possuir maior sensibilidade: alunos da rede pública e alunos cujos pais não possuem ensino superior. Realizam-se as estimações por variáveis instrumentais em três subamostras: *(i)* alunos matriculados em escolas públicas, *(ii)* alunos cujos pais não possuem ensino superior e *(iii)* alunos matriculados em escolas privadas.

O objetivo é avaliar se o efeito da distância é mais intenso em grupos para os quais o custo de deslocamento e as restrições orçamentárias são potencialmente mais relevantes e comparar com um grupo onde tais restrições financeiras não sejam tão relevantes. Nos dois primeiros casos, os coeficientes associados à distância instrumentada permanecem negativos e estatisticamente significativos, tanto para Matemática quanto para Português, reforçando a interpretação de que a parcela exógena da distância associada ao regime de mobilidade constitui uma barreira relevante ao desempenho escolar. Já para a terceira subamostra, os resultados não foram estatisticamente significativos.

Na subamostra de alunos da rede pública, apresentada na tabela 42, a comparação entre MQO e IV reforça a endogeneidade da distância residência–escola. No modelo observacional, a distância aparece positivamente associada ao desempenho, com coeficientes de 0,003 em Matemática e 0,002 em Português, ambos estatisticamente significativos, o que sugeriria, em uma primeira abordagem, que trajetos mais longos estariam associados a notas ligeiramente mais altas.

Entretanto, quando a distância é instrumentada pela política municipal de passe estudantil, o sinal se inverte e a magnitude aumenta substancialmente: o coeficiente passa para -0,037 em Matemática e -0,031 em Português, ambos significativos a 1%, indicando que um quilômetro adicional na parcela exógena da distância está associado, aproximadamente, a reduções de 3,7% e 3,1% nas proficiências em matemática e português, respectivamente.

Tabela 42 – Heterogeneidade: efeito da distância sobre o desempenho em alunos de escola pública — em *Log*

	Matemática		Português	
	OLS	IV	OLS	IV
Distância residência–escola (km)	0.003*** (0.000)	-0.037*** (0.008)	0.002*** (0.000)	-0.031*** (0.006)
Nenhum dos pais em ocupação de colarinho branco	-0.037*** (0.006)	-0.038*** (0.008)	-0.033*** (0.004)	-0.033*** (0.006)
Mãe: nunca estudou / não concluiu 4ª série	Categoria de referência			
Mãe: concluiu 4ª série	0.054*** (0.008)	0.047*** (0.010)	0.039*** (0.005)	0.034*** (0.008)
Mãe: concluiu ensino fundamental	0.078*** (0.008)	0.062*** (0.011)	0.062*** (0.006)	0.049*** (0.009)
Mãe: concluiu ensino médio	0.111*** (0.009)	0.118*** (0.012)	0.099*** (0.006)	0.105*** (0.009)
Mãe: concluiu ensino superior	0.197*** (0.015)	0.280*** (0.025)	0.164*** (0.011)	0.234*** (0.020)
Pai: nunca estudou / não concluiu 4ª série	Categoria de referência			
Pai: concluiu 4ª série	0.024*** (0.008)	0.017* (0.010)	0.022*** (0.005)	0.015** (0.008)
Pai: concluiu ensino fundamental	0.043*** (0.008)	0.028** (0.011)	0.043*** (0.006)	0.030*** (0.009)
Pai: concluiu ensino médio	0.082*** (0.009)	0.073*** (0.012)	0.072*** (0.006)	0.064*** (0.009)
Pai: concluiu ensino superior	0.134*** (0.015)	0.178*** (0.021)	0.114*** (0.010)	0.150*** (0.016)
Ida à escola: conversa com docente	-0.032*** (0.005)	-0.032*** (0.006)	-0.024*** (0.003)	-0.024*** (0.005)
Ida à escola: reunião	0.035*** (0.006)	0.044*** (0.008)	0.032*** (0.004)	0.039*** (0.006)
Ida à escola: atividade com a comunidade	0.052*** (0.007)	0.043*** (0.009)	0.043*** (0.005)	0.035*** (0.007)
Série do aluno	0.121*** (0.006)	0.165*** (0.012)	0.056*** (0.004)	0.094*** (0.009)
Constante	4.614*** (0.013)	4.680*** (0.022)	4.630*** (0.009)	4.686*** (0.017)
Número de observações	9622	9622	9622	9622
R ² ajustado	0.140	-0.459	0.180	-0.628
Weak instruments	—	61.68***	—	61.68***
Wu–Hausman	—	43.16***	—	61.10***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Notas: Coeficientes com erros-padrão entre parênteses. A variável “escola pública” não aparece nesta subamostra por não haver variação dentro do grupo analisado. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Já na tabela 43 se apresentam os resultados para o efeito da distância sobre o desempenho entre alunos cujos pais não possuem ensino superior. Na subamostra de alunos cujos pais não possuem ensino superior, apresentada na tabela 43 o contraste entre MQO e IV é ainda mais nítido e sugere que a endogeneidade da distância é particularmente relevante entre famílias com maior restrição socioeconômica. Nos modelos de MQO, a distância residência-escola aparece positivamente associada ao desempenho, com coeficientes de 0,002 tanto em Matemática quanto em Português, ambos estatisticamente significativos, o que indicaria, em uma leitura puramente observacional, que alunos que es-

tudam mais longe tendem a apresentar notas ligeiramente mais altas. No entanto, quando a distância é instrumentada pela política municipal de passe estudantil, o sinal se inverte e a magnitude cresce substancialmente, com comportamento similar ao apresentado na primeira especificação.

Tabela 43 – Heterogeneidade: efeito da distância sobre o desempenho entre alunos cujos pais não possuem ensino superior — em *Log*

	Matemática		Português	
	OLS	IV	OLS	IV
Distância residência–escola (km)	0.002*** (0.000)	-0.045*** (0.009)	0.002*** (0.000)	-0.038*** (0.007)
Nenhum dos pais em ocupação de colarinho branco	-0.035*** (0.005)	-0.035*** (0.008)	-0.029*** (0.004)	-0.029*** (0.006)
Aluno de escola pública	-0.102*** (0.008)	-0.157*** (0.015)	-0.096*** (0.005)	-0.143*** (0.012)
Mãe: nunca estudou / não concluiu 4 ^a série		Categoria de referência		
Mãe: concluiu 4 ^a série	0.051*** (0.008)	0.039*** (0.011)	0.038*** (0.005)	0.028*** (0.008)
Mãe: concluiu ensino fundamental	0.077*** (0.008)	0.058*** (0.012)	0.063*** (0.006)	0.047*** (0.009)
Mãe: concluiu ensino médio	0.110*** (0.009)	0.113*** (0.012)	0.098*** (0.006)	0.101*** (0.009)
Pai: nunca estudou / não concluiu 4 ^a série		Categoria de referência		
Pai: concluiu 4 ^a série	0.025*** (0.007)	0.017 (0.011)	0.023*** (0.005)	0.017** (0.008)
Pai: concluiu ensino fundamental	0.045*** (0.008)	0.029** (0.012)	0.047*** (0.006)	0.033*** (0.009)
Pai: concluiu ensino médio	0.081*** (0.008)	0.071*** (0.012)	0.070*** (0.006)	0.061*** (0.009)
Ida à escola: conversa com docente	-0.033*** (0.005)	-0.033*** (0.006)	-0.025*** (0.003)	-0.025*** (0.005)
Ida à escola: reunião	0.036*** (0.006)	0.045*** (0.009)	0.032*** (0.004)	0.040*** (0.007)
Ida à escola: atividade com a comunidade	0.049*** (0.006)	0.044*** (0.009)	0.039*** (0.004)	0.034*** (0.007)
Série do aluno	0.116*** (0.006)	0.167*** (0.013)	0.051*** (0.004)	0.094*** (0.010)
Constante	4.723*** (0.014)	4.863*** (0.033)	4.733*** (0.010)	4.850*** (0.026)
Número de observações	10157	10157	10157	10157
R ² ajustado	0.138	-0.740	0.192	-0.974
Weak instruments	—	55.04***	—	55.04***
Wu–Hausman	—	56.38***	—	80.08***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Notas: Coeficientes com erros-padrão entre parênteses. As categorias “Mãe: concluiu ensino superior” e “Pai: concluiu ensino superior” não aparecem nesta subamostra por não haver variação dentro do grupo analisado. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Analisando os padrões para alunos matriculados na rede privada, passa-se agora a analisar como a distância afeta especificamente essa subamostra, que tende a ser uma possível *proxy* para rendimentos. Na tabela 44 apresentam-se os resultados para os alunos da rede privada.

Tabela 44 – Heterogeneidade: efeito da distância sobre o desempenho em alunos da rede privada — em *Log*

	Matemática		Português	
	OLS	IV	OLS	IV
Distância residência–escola (km)	-0.002** (0.001)	-0.163 (0.111)	-0.002*** (0.000)	-0.145 (0.097)
Nenhum dos pais em ocupação de colarinho branco	-0.005 (0.010)	-0.014 (0.043)	0.005 (0.006)	-0.003 (0.037)
Mãe: nunca estudou / não concluiu 4ª série		Categoria de referência		
Mãe: concluiu 4ª série	0.021 (0.033)	-0.468 (0.364)	0.023 (0.021)	-0.409 (0.318)
Mãe: concluiu ensino fundamental	0.077** (0.032)	-0.308 (0.298)	0.068*** (0.020)	-0.272 (0.260)
Mãe: concluiu ensino médio	0.111*** (0.031)	-0.338 (0.336)	0.096*** (0.020)	-0.301 (0.294)
Mãe: concluiu ensino superior	0.157*** (0.032)	-0.223 (0.295)	0.123*** (0.020)	-0.213 (0.258)
Pai: nunca estudou / não concluiu 4ª série		Categoria de referência		
Pai: concluiu 4ª série	0.009 (0.031)	0.023 (0.137)	0.034* (0.020)	0.046 (0.120)
Pai: concluiu ensino fundamental	0.031 (0.031)	-0.005 (0.137)	0.066*** (0.020)	0.035 (0.120)
Pai: concluiu ensino médio	0.049 (0.030)	-0.028 (0.142)	0.061*** (0.019)	-0.006 (0.124)
Pai: concluiu ensino superior	0.079*** (0.030)	-0.045 (0.159)	0.086*** (0.020)	-0.023 (0.139)
Ida à escola: conversa com docente	-0.041*** (0.007)	-0.073* (0.038)	-0.025*** (0.005)	-0.054 (0.033)
Ida à escola: reunião	0.030*** (0.008)	0.045 (0.039)	0.022*** (0.005)	0.036 (0.034)
Ida à escola: atividade com a comunidade	0.019** (0.008)	0.040 (0.037)	0.013** (0.005)	0.031 (0.033)
Série do aluno	-0.047** (0.024)	-0.391 (0.258)	-0.105*** (0.015)	-0.409* (0.225)
Constante	4.943*** (0.049)	6.422*** (1.035)	4.915*** (0.032)	6.223*** (0.904)
Número de observações	2535	2535	2535	2535
R ² ajustado	0.119	-16.120	0.175	-28.660
Weak instruments	—	2.25	—	2.25
Wu–Hausman	—	42.24***	—	81.31***

Fonte: Elaboração própria com base nos dados do GERES (2005).

Notas: Coeficientes com erros-padrão entre parênteses. A variável “escola pública” não aparece nesta subamostra por não haver variação dentro do grupo analisado. Significância: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Na subamostra de alunos da rede privada, os resultados diferem daqueles obtidos na amostra principal. Nos modelos de MQO, a distância residência-escola apresenta coeficientes negativos e estatisticamente significativos, mas de pequena magnitude, sendo de $-0,0016$ para Matemática e $-0,0018$ para Português por quilômetro adicional. Nos modelos de variáveis instrumentais, os coeficientes tornam-se mais negativos ($-0,163$ em Matemática e $-0,145$ em Português); porém, deixam de ser estatisticamente significativos. Esse resultado deve ser interpretado com cautela, pois o instrumento é fraco nessa subamostra. A estatística de primeiro estágio é de apenas 2,25 em ambas as especifica-

ções, com $p = 0,133$, indicando que o passe estudantil não gera variação suficientemente forte da distância entre alunos da rede privada.

Dessa forma, nesta subamostra, os resultados de IV não oferecem uma base sólida para inferência causal, padrão consistente com a hipótese de que, entre famílias de maior renda, o custo de deslocamento é menos decisivo para a escolha escolar. Nesses casos, o benefício do transporte tende a ter menor capacidade de alterar a escola frequentada, o que enfraquece o mecanismo central investigado neste artigo, ou seja, a redução do custo de acesso à escola. Esse achado reforça a interpretação de que os efeitos do transporte escolar são mais relevantes quando a restrição financeira é efetivamente importante para a decisão das famílias.

Assim, como principais achados de robustez por heterogeneidade, os resultados sugerem que esse mecanismo é particularmente relevante em grupos mais vulneráveis. Entre alunos cujos pais não possuem ensino superior, os coeficientes IV permanecem negativos e, em alguns casos, são mais intensos do que na amostra total, o que é compatível com a hipótese de que o subsídio ao transporte relaxa restrições de mobilidade de forma mais acentuada para famílias com menor capacidade de absorver custos de deslocamento. Esses resultados também são encontrados para alunos da rede pública, mas devem ser interpretados com cautela.

Apêndice B - Sensibilidade à exclusão de municípios

Dado que a identificação do modelo principal decorre de variação em nível municipal e se apoia em um número reduzido de contextos urbanos, será realizado um exercício complementar de sensibilidade do tipo *leave-one-city-out*, no qual se reestimarão a especificação principal de variáveis instrumentais, excluindo alternadamente um município da amostra. O objetivo desse exercício não é transformar o artigo em uma análise de efeitos diferenciais por município, o que exigiria uma base mais ampla e maior variação institucional, mas verificar em que medida os resultados principais dependem da composição municipal da amostra e, sobretudo, explicitar que o instrumento agregado utilizado no artigo combina contextos de mobilidade bastante distintos.

A Tabela 45 apresenta, para cada reestimação, o coeficiente associado à variável distância residência-escola no segundo estágio do modelo IV, tanto para Matemática quanto para Português. Como o desfecho está em logaritmo e a distância é medida em quilômetros, esses coeficientes podem ser lidos como semi-elasticidades. Assim, um coeficiente de $-0,041$ indica que um quilômetro adicional na parcela instrumentada da distância está associado, aproximadamente, a uma redução de 4,1% na proficiência. A coluna “Obs.” informa o número de observações em cada reestimação, enquanto a coluna “Nota” indica se o modelo permaneceu identificado.

Tabela 45 – Sensibilidade *leave-one-city-out* — em *Log*

Município excluído	Matemática	Português	Obs.	Nota
Salvador	-0.0411	-0.0351	10.376	ok
Campo Grande	-0.0269	-0.0261	10.312	ok
Rio de Janeiro	-0.1187	-0.0967	9.804	ok
Campinas	-0.0054	0.0006	9.080	ok
Belo Horizonte	—	—	9.056	sem variação no instrumento

Fonte: Elaboração própria com base nos dados do GERES (2005).

Nota: Cada linha reapresenta o coeficiente da distância residência–escola no modelo principal de variáveis instrumentais após excluir um município da amostra. Os coeficientes de Matemática e Português correspondem ao parâmetro estimado para a variável distância no segundo estágio do IV. Ao excluir Belo Horizonte, o instrumento perde variação, pois todos os municípios remanescentes passam a apresentar algum tipo de passe estudantil.

Os resultados mostram que a sensibilidade das estimativas não decorre apenas do pequeno número de municípios, mas também da heterogeneidade do próprio desenho institucional do instrumento. Nossa especificação utiliza uma *dummy* de existência de passe estudantil. A exclusão de Belo Horizonte inviabiliza completamente a identificação, pois elimina o único município sem subsídio e, com isso, suprime toda a variação do instrumento. Já a exclusão de Campinas praticamente elimina o efeito estimado, com coeficientes de $-0,005$ em Matemática e $0,001$ em Português. Em contraste, a exclusão do Rio de Janeiro torna os coeficientes substancialmente mais negativos, elevando-os para $-0,119$ em Matemática e $-0,097$ em Português. Isso indica que a presença do Rio de Janeiro atenua a magnitude média do efeito estimado na amostra completa. Salvador e Campo Grande, por sua vez, preservam o sinal negativo das estimativas, embora também modifiquem sua magnitude, o que é consistente com a ideia de que cada município contribui de modo diferente para a variação efetivamente explorada pelo estimador IV.

4 EXPOSIÇÃO À HIPERINFLAÇÃO E CICLO DE VIDA LABORAL: EVIDÊNCIAS DE LONGO PRAZO SOBRE RISCO, RENDIMENTOS E PERMANÊNCIA NO MERCADO FORMAL.

Resumo

Este ensaio tem como objetivo investigar os impactos duradouros da hiperinflação sobre uma geração exposta a choques econômicos severos no início da vida laboral. Foi utilizado um painel longitudinal com microdados administrativos oriundos da RAIS, um banco de dados com registros dos trabalhadores no mercado formal brasileiro, onde foram acompanhados mais de 3 milhões de trabalhadores durante 22 anos de carreira ou até 28 anos de trajetória, e comparam-se as duas coortes: uma que ingressou no emprego formal em 1989, durante um surto de hiperinflação, e uma que ingressou em 1995, pós estabilização do plano real. Construiu-se estatísticas de superfície para mapear a evolução conjunta de rendimentos e permanência no setor formal, mostrando que momentos de crescimento revelam heterogeneidade que se perderia em médias agregadas. Os resultados indicam forte variação por estágio de carreira e pela composição dos rendimentos recentes, afetando o risco e as caudas da distribuição ao longo do ciclo de vida. Diferenças persistentes entre as coortes sugerem que ambientes macroeconômicos distintos no momento do ingresso no mercado de trabalho estão associados a mudanças localizadas na dinâmica de riscos associados aos rendimentos e na probabilidade de permanecer no setor formal do mercado de trabalho.

Palavras-chave: Hiperinflação; mercado de trabalho; heterogeneidade nos rendimentos futuros; ciclo de vida do trabalhador; desigualdade.

4.1 Introdução

Estudar os efeitos da hiperinflação sobre o mercado de trabalho é importante porque episódios dessa natureza podem deixar marcas que ultrapassam o período imediato da crise. A exposição a um choque macroeconômico severo no início da vida laboral pode influenciar decisões ocupacionais, trajetórias de rendimentos e padrões de permanência no emprego formal.

Neste trabalho, investiga-se o ciclo de vida dos trabalhadores no mercado formal brasileiro, acompanhando longitudinalmente duas coortes: 1989 e 1995, definidas pelo ano de ingresso no setor formal. Inspirados em Guvenen *et al.* (2021), foram reindexadas as trajetórias pelo tempo de carreira, τ , de modo a comparar trabalhadores em estágios equivalentes desde sua entrada no mercado formal. Com isso, a análise deixa de se apoiar apenas no tempo calendário e passa a considerar o tempo econômico da carreira. Esse procedimento permite comparar trajetórias sob uma métrica comum, ainda que os tra-

balhadores tenham ingressado no mercado formal em anos e contextos macroeconômicos bastante distintos.

O diferencial deste trabalho está no fato de que nossa análise compara indivíduos que iniciaram suas trajetórias profissionais em um ambiente de forte instabilidade macroeconômica com trabalhadores que ingressaram no mercado formal em um contexto mais estável, utilizando para isso os microdados da RAIS¹.

O painel longitudinal acompanha essas duas coortes de trabalhadores entre 1989 e 2017, sendo a primeira formada pelos indivíduos que ingressaram no mercado formal em 1989, ano em que o Brasil viveu um surto inflacionário, com inflação anual de 1.972%. A segunda reúne os trabalhadores que ingressaram em 1995, após a implementação do Plano Real, quando a inflação anual foi de 22%, a menor taxa registrada desde 1980. Ao longo do artigo, os trabalhadores que ingressaram em 1989 são tratados como grupo exposto, também denominado grupo tratado, enquanto aqueles que ingressaram em 1995 compõem o que será denominado grupo de controle. Não se trata de um estudo de eventos, mas se adotará essa denominação para fins de tangibilizar a comparação entre os diferentes grupos.

A literatura já documentou diferentes canais e mecanismos pelos quais surtos inflacionários se relacionam com ajustes na renda e no emprego. Dvorkin e Marks (2024), por exemplo, analisaram reajustes salariais nominais de trabalhadores que permaneceram no mesmo emprego por pelo menos um ano. Os autores mostram que, após a pandemia, os salários passaram a variar com maior frequência e intensidade, embora esse comportamento tenha perdido força a partir de 2023, em um contexto de maior estabilização econômica.

Para a América Latina, Gambetti e Messina (2018) analisaram a evolução dos salários reais e do emprego no México, na Colômbia, no Chile e no Brasil. Entre seus resultados, chama atenção o comportamento assimétrico dos salários reais ao longo do ciclo econômico. Em períodos recessivos, os salários tendem a cair rapidamente; com a redução da inflação, porém, tornam-se menos sensíveis a ajustes negativos, revelando maior rigidez nominal para baixo. Nas expansões, por outro lado, os salários crescem de forma acelerada, em ritmo até superior ao observado nas quedas durante as recessões. Os autores também mostram que, na Colômbia, o emprego se tornou mais pró-cíclico em um contexto de maior rigidez salarial. No Brasil, o ajuste ocorreu principalmente por meio das horas trabalhadas e de mudanças setoriais.

Em outra linha, Mello, Waisman e Zilberman (2014) avaliaram os efeitos da exposição à hiperinflação sobre a escolha pelo trabalho autônomo, considerando imigrantes

¹ A RAIS (Relação Anual de Informações Sociais) é um registro administrativo criado em 1975, com periodicidade anual, cujo objetivo é fornecer informações sobre o mercado de trabalho formal brasileiro, atendendo a demandas de controle, estatísticas e formulação de políticas públicas por parte de entidades governamentais (MTE, 2025).

residentes nos Estados Unidos entre 1910 e 2000. Os resultados indicam que indivíduos expostos à hiperinflação apresentaram maior probabilidade de optar pelo trabalho autônomo em vez do emprego assalariado.

Outra parte da literatura utiliza versões estendidas da função de rendimentos de Mincer para examinar os determinantes da renda dos trabalhadores sob diferentes abordagens teóricas e empíricas. Entre os trabalhos mais influentes, Card (1994, 1999) destacam questões econométricas ligadas à identificação da relação causal entre escolaridade e salários. Heckman, Lochner e Todd (2003) incorporam a incerteza quanto aos rendimentos futuros e propõem refinamentos importantes ao modelo tradicional de mensuração dos rendimentos. Já Lemieux (2006) investiga em que medida a formulação clássica da função de Mincer continua adequada para dados mais recentes, discutindo suas limitações e implicações contemporâneas. Em conjunto, esses estudos consolidaram uma base teórica e empírica importante para compreender a determinação dos salários, os retornos à educação e a dinâmica dos rendimentos ao longo da vida laboral.

No caso brasileiro, diversos estudos com dados da RAIS têm investigado a mobilidade dos trabalhadores entre firmas e seus efeitos sobre os salários. Freguglia e Procópio (2013), por exemplo, analisaram os diferenciais salariais no mercado formal, distinguindo o impacto da mudança de emprego, associado ao efeito firma, do impacto da migração intermunicipal ou interestadual. O estudo utiliza dados da RAISmiga para o período de 1995 a 2006. Os resultados indicam que ambos os movimentos afetam os salários, embora por canais distintos. A troca de firma está associada a uma queda imediata de aproximadamente 3,5% no salário real, enquanto a migração geográfica exerce efeito ligeiramente maior. Esses achados sugerem que a mobilidade laboral pode impor uma penalidade salarial de curto prazo ao trabalhador, ainda que também possa refletir processos de realocação produtiva no longo prazo.

Neste artigo, construiu-se, para cada estágio da trajetória laboral, uma medida de rendimentos recentes (*Recent Earnings*), definida como a média dos rendimentos observados nos cinco anos anteriores da carreira. A partir desta variável, foram classificados os trabalhadores em percentis de rendimentos recentes dentro de cada estágio de carreira. Em seguida, foram estimados, para horizontes em anos $k \in \{1, 5\}$, diferentes momentos da distribuição do crescimento dos rendimentos formais. Foram analisados a dispersão por meio do desvio-padrão e dos diferenciais $P90 - P10$ e $P90 - P50$, a assimetria por meio da *Kelley skewness* e o comportamento das caudas por meio da curtose de *Crow-Siddiqui*.

Também foi estruturada uma medida de margem extensiva², que mensura a pro-

² Distingui-se margem intensiva e margem extensiva, à luz da literatura em economia do trabalho, para separar dois tipos de ajuste no mercado de trabalho. A margem intensiva refere-se a variações *condicionais à permanência no estado observado*, isto é, mudanças no *nível* e no *crescimento* dos rendimentos formais (Y e $\Delta^k y$) para indivíduos que permanecem com vínculo registrado. A margem extensiva, por sua vez, captura variações no *status de participação*, isto

babilidade de o indivíduo não aparecer no registro formal em períodos futuros, variável essa que é interpretada como uma medida de risco de saída do mercado formal. Em conjunto, essas estatísticas descrevem a dinâmica dos rendimentos formais ao longo da carreira por dois ângulos, ou seja, a permanência no setor formal e o risco distributivo entre os trabalhadores que continuam observados nesse mercado.

Os resultados mostram que, tanto na probabilidade de o trabalhador deixar de aparecer no registro formal no futuro, na dispersão do crescimento dos rendimentos, na assimetria da distribuição e na ocorrência de eventos extremos, existe uma heterogeneidade observada relevante, demonstrando que trabalhadores em posições distintas da distribuição de renda recente enfrentam salários diferentes, mas também riscos distintos, trajetórias distintas e formatos distintos de evolução dos rendimentos e dos riscos associados a permanência no formal ao longo da carreira.

Isto posto, este trabalho contribui com a literatura ao complementar a análise sobre possíveis efeitos persistentes da hiperinflação, avaliando como esses padrões, tanto na margem intensiva quanto na margem extensiva, diferem entre as coortes de 1989 e 1995, e avaliando se as diferenças nas condições de entrada no mercado formal estão associadas a trajetórias persistentemente distintas de permanência, volatilidade e exposição a eventos extremos. Assim, o objetivo não é apenas descrever a evolução dos rendimentos formais, mas caracterizar a distribuição dos riscos que acompanha essas trajetórias ao longo do ciclo de vida laboral.

A comparação entre coortes precisa ser lida à luz do ambiente histórico em que cada uma delas ingressou no mercado formal. Por essa razão, a próxima seção descreve o contexto macroeconômico e institucional brasileiro em torno dos anos de entrada das coortes de 1989 e 1995, destacando os elementos que motivam sua comparação ao longo do artigo.

4.2 Contexto Econômico e Institucional Brasileiro

Esta seção discute o contexto histórico que orienta a escolha das coortes de entrada analisadas no artigo. A comparação entre trabalhadores que ingressaram no mercado formal em 1989 e em 1995 é relevante porque esses dois grupos iniciaram sua trajetória laboral sob regimes macroeconômicos profundamente distintos. De um lado, a coorte de 1989 entrou no mercado em um ambiente marcado por inflação elevada e persistente, forte

é, entrada e saída do emprego formal. Neste trabalho, essa margem é medida pela probabilidade de o trabalhador *não aparecer na RAIS* em período futuro (*exit_rate*), interpretada como saída do registro formal. Essa medida não implica ausência de renda total, uma vez que o setor informal não é observado. Em síntese, a margem intensiva descreve *quanto* os rendimentos mudam entre os trabalhadores observados, enquanto a margem extensiva descreve *quem permanece* e *quem sai* do universo formal observado.

incerteza e sucessivas tentativas de estabilização malsucedidas. Além disso, foi um ano de eleições, as primeiras diretas desde a redemocratização.

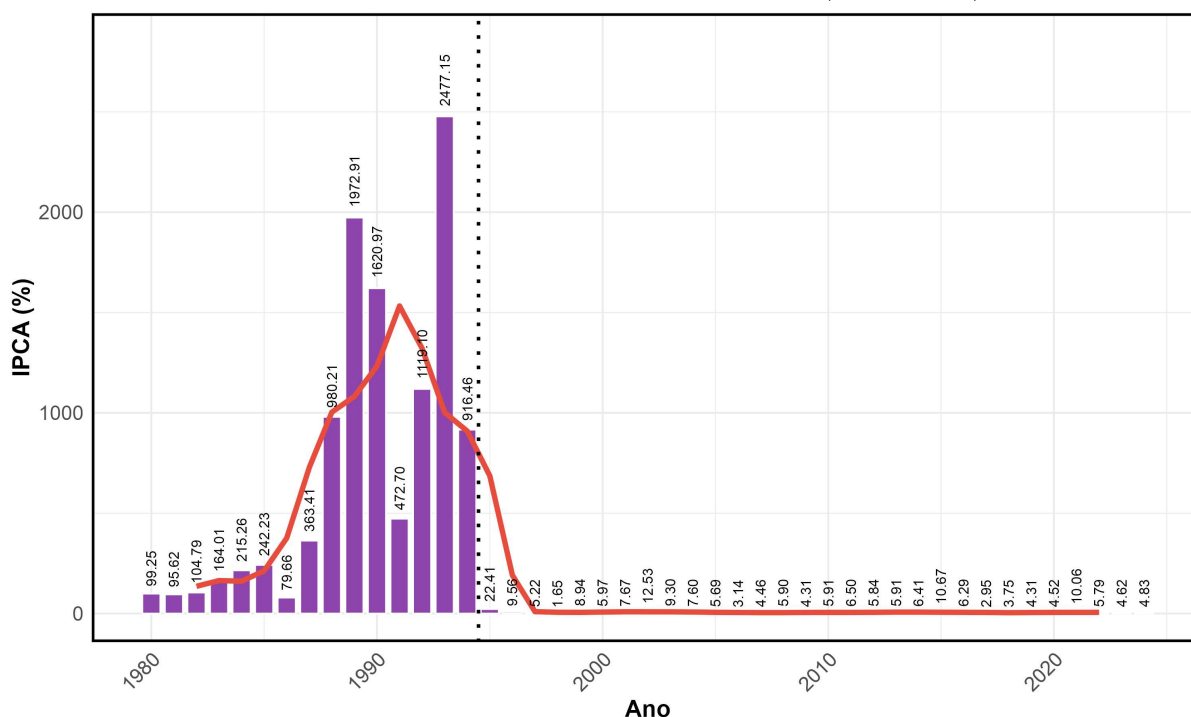
De outro lado, a coorte de 1995 iniciou sua trajetória após a ruptura representada pelo Plano Real, em um cenário de estabilização monetária e reorganização do ambiente econômico. Reconstruir esse pano de fundo é importante porque ele fornece a base para interpretar, ao longo do artigo, de que maneira as condições de entrada no mercado formal podem estar associadas a diferenças persistentes nas trajetórias de rendimentos e de permanência no emprego formal.

Esse contraste entre as coortes remete a uma transformação macroeconômica central da economia brasileira entre o final da década de 1980 e a primeira metade da década de 1990. Ao longo desse período, o país conviveu com inflação elevada e persistente IBGE (2025). O Plano Cruzado, lançado em 1986 durante o governo José Sarney, procurou conter a inflação por meio do congelamento de preços e salários e da introdução de uma nova moeda, o Cruzado. A intenção era interromper a indexação automática da economia (BACEN, 2024a). Após o fracasso dessa iniciativa, seguiram-se o Plano Bresser, em 1987, e o Plano Verão, em 1989, ambos também apoiados em congelamentos de preços e mudanças monetárias (BACEN, 2024a).

Já os Planos Collor I, de 1990, e Collor II, de 1991, implementados no governo Fernando Collor de Mello, combinaram bloqueio de liquidez, congelamento de preços e reformas fiscais como tentativa de estabilização (BACEN, 2024a). Diferentemente dos planos anteriores, o Plano Collor ficou marcado pelo bloqueio de ativos financeiros, restringindo por um período considerável o acesso dos poupadores aos seus próprios recursos. Essa medida ocorreu em um ambiente de desvalorização acelerada dos ativos monetários e ampliou a percepção de incerteza entre famílias, empresas e trabalhadores. As iniciativas de estabilização contribuíram para reduzir a popularidade do governo, que já contava com uma base política frágil no Congresso e passou a enfrentar denúncias de corrupção. Esse conjunto de fatores culminou no primeiro processo de impeachment de um presidente eleito por voto direto no Brasil após a redemocratização.

Após a saída de Fernando Collor, em 1992, o governo Itamar Franco assumiu a estabilização monetária como prioridade e introduziu, em 1993, a Unidade Real de Valor (URV) (BACEN, 2024b). Em 1994, entrou em vigor o Plano Real, acompanhado da criação da nova moeda, o Real. A partir desse momento, a inflação foi substancialmente reduzida, inaugurando um novo regime monetário e um ambiente macroeconômico distinto daquele observado nos anos anteriores (BACEN, 2024b). A Figura 14 resume essa trajetória e evidencia tanto a persistência inflacionária do período anterior quanto a ruptura associada à estabilização de 1994.

Figura 14 – Histórico de inflação no Brasil (1980–2024)



Fonte: Elaboração própria com dados do IPCA do IBGE (2025)

A escolha da coorte de 1989 decorre precisamente desse contexto. Os trabalhadores que ingressaram no mercado formal naquele ano iniciaram sua trajetória em um momento particularmente adverso, caracterizado por inflação em níveis extremos e elevada volatilidade nominal. Além de coincidir com o surto hiperinflacionário do período, esse recorte captura os maiores níveis de inflação observados na série e concentra, nos anos iniciais de carreira, um ambiente de forte instabilidade macroeconômica, que inclui a transição pelos Planos Collor. Nesse sentido, a coorte de 1989 oferece um ponto de contraste especialmente informativo em relação à coorte de 1995, cuja entrada ocorre já sob o regime de estabilidade introduzido pelo Plano Real.

A literatura recente mostra que a relação entre hiperinflação e mercado de trabalho tem sido estudada, em grande medida, a partir de uma perspectiva macroeconômica. Parte desses estudos examina como choques inflacionários afetam os níveis de emprego, os rendimentos dos trabalhadores e a própria dinâmica de ajuste das economias atingidas. Outra parte se concentra nos efeitos das políticas de combate à inflação, como a elevação das taxas de juros, restrição de liquidez e mudanças institucionais, sobre o funcionamento do mercado de trabalho. Há ainda uma vertente que inverte essa relação e investiga de que forma a estrutura do mercado de trabalho, os níveis de emprego e a dinâmica dos salários podem influenciar os próprios indicadores macroeconômicos. Assim, a literatura ora trata o mercado de trabalho como consequência de choques e políticas macroeconômicas, ora como um dos elementos que ajudam a explicar esses mesmos desequilíbrios.

Este trabalho se afasta parcialmente desse enquadramento agregado. O contexto

macroeconômico é utilizado para definir o momento de entrada dos trabalhadores no mercado formal e interpretar as diferenças entre coortes. A análise, inspirada em Guvenen *et al.* (2021), desloca o foco para o ciclo de vida laboral dos indivíduos, explorando a evolução dos rendimentos, da permanência no setor formal e das heterogeneidades de risco ao longo das trajetórias profissionais. O ponto de largada é a entrada no mercado formal e como os rendimentos recentes desse trabalhador influenciam o futuro.

A seguir, a seção 4.3 apresenta de forma detalhada a construção do banco de dados e as estatísticas descritivas das variáveis utilizadas ao longo do trabalho. Na sequência, a seção 4.4 descreve a metodologia adotada, inspirada em Guvenen *et al.* (2021). A seção 4.5 apresenta os resultados obtidos. Por fim, a seção 4.6 reúne as principais conclusões, além das limitações do estudo e das sugestões para pesquisas futuras.

4.3 Banco de Dados e Estatísticas Descritivas.

A partir do contexto institucional apresentado na seção anterior, foi realizada a seleção da amostra a ser utilizada e estruturado o painel contendo as duas coortes, de 1989 a 2017 e de 1995 a 2017. Será aberta uma subseção para explicar a seleção e construção do painel, com uma descrição detalhada do processo de tratamento dos microdados da RAIS.

4.3.1 Banco de Dados

Utilizou-se o painel longitudinal da RAIS (Relação Anual de Informações Sociais), base administrativa que reúne informações sobre empregados e empregadores do mercado formal brasileiro. Estudos recentes, como o de Engbom e Moser (2022), destacam a importância dos microdados da RAIS para a análise de trajetórias salariais, mobilidade laboral e desigualdade. A partir desses registros, construiu-se um painel longitudinal, de 1989 a 2017, composto por duas coortes de trabalhadores que ingressaram no mercado formal em momentos distintos. A primeira reúne os indivíduos que aparecem pela primeira vez na RAIS em 1989; a segunda, aqueles que aparecem pela primeira vez em 1995. Por definição, nenhum desses trabalhadores poderia constar na base nos anos anteriores ao ingresso. Como os dados disponíveis para esta pesquisa abrangem o período de 1985 a 2017, os trabalhadores da coorte de 1989 não poderiam aparecer entre 1985 e 1988, enquanto os da coorte de 1995 não poderiam constar na RAIS entre 1985 e 1994.

A primeira coorte, referente a 1989³ contempla trabalhadores que iniciaram sua

³ Conforme se observa na Figura 14, foi composta por trabalhadores que ingressaram no mercado formal em um dos momentos mais críticos da década em termos inflacionários, após sucessivos planos econômicos e tentativas de estabilização malsucedidas. Nos anos seguintes, esses trabalhadores ainda enfrentariam condições macroeconômicas adversas, incluindo a maior média móvel do indicador de inflação e o maior pico observado no período, em 1993. Por esse motivo, optou-se por selecionar os profissionais que ingressaram no mercado formal

trajetória formal em um ano de hiperinflação no Brasil, quando a taxa anual de inflação chegou a 1.972% ao ano (BACEN, 2019). Para integrar essa coorte, o trabalhador deveria aparecer na RAIS em 1989, não constar na base nos anos anteriores e ter entre 18 e 24 anos no ano de ingresso.

Após a definição da coorte exposta à hiperinflação, que será denominada como coorte tratada⁴. A coorte de controle foi composta por indivíduos que ingressaram no mercado formal em 1995, o primeiro ano completo após a implementação do Plano Real.⁵ A resposta à nova política econômica foi rápida, com a inflação desabando para 22,41% ao ano em 1995, valor muito inferior aos 916% ao ano registrados em 1994, ano em que ocorreu o choque de estabilização no início do segundo semestre. A regra de inclusão aplicada ao grupo de controle seguiu o mesmo critério utilizado para o grupo tratado, onde trabalhadores que apareceram na RAIS em 1995, não constavam em nenhum dos anos anteriores e tinham entre 18 e 24 anos no ano de ingresso.

Foram mantidos apenas os vínculos ativos em 31 de dezembro de cada ano analisado. Também foram excluídos registros com remuneração nula ou ausente nas variáveis de salário, exceto no caso da remuneração real em dezembro, registros com tempo de emprego igual a zero ou sem informação; e vínculos com carga horária semanal contratada inferior a 20 horas. Em relação às variáveis salariais, foram excluídos trabalhadores que, em dezembro de cada ano, receberam salário-hora inferior ao mínimo legal vigente no período. Também foram retiradas observações com salário real em dezembro superior a R\$ 50.000,00, com o objetivo de reduzir a influência de valores extremos ou *outliers*.

Os rendimentos informados na RAIS foram deflacionados pelo INPC, tomando 2017 como ano-base.⁶ A escolha do INPC como deflator se justifica por sua abrangência e por sua maior aderência ao perfil da amostra utilizada neste trabalho, composta por trabalhadores formais em início de trajetória laboral.

Ainda no tratamento da base, excluíram-se servidores públicos e trabalhadores vinculados a sociedades de economia mista. Essa decisão decorre do fato de que a dinâmica ocupacional e salarial desses trabalhadores tende a diferir daquela observada no setor

naquele ano.

⁴ Importante destacar que não se trata de um estudo de eventos, nem se tem a pretensão de gerar neste ensaio algo similar ou interpretações como tal. Apenas se denomina de coorte tratada pelo fato de ter sido exposta ao ambiente hiperinflacionário

⁵ O Plano Real foi implementado em 1º de julho de 1994, como parte de um conjunto de reformas econômicas adotadas no Brasil com o objetivo de estabilizar a economia e controlar a hiperinflação. O plano introduziu uma nova moeda, o real, em substituição ao cruzeiro real, e teve papel central na estabilização dos preços e na recuperação da confiança na economia brasileira (BACEN, 2024b).

⁶ O INPC busca corrigir o poder de compra dos salários, medindo as variações de preços da cesta de consumo das famílias assalariadas de menor renda. Essa faixa de renda foi definida para garantir que o índice represente cerca de 50% das famílias chefiadas por assalariados que vivem em áreas urbanas contempladas pelo SNIPC (Sistema Nacional de Índices de Preços ao Consumidor).

privado. No Brasil, grande parte dos servidores públicos está submetida a estatutos e regimes próprios. Mesmo entre empregados regidos pela Consolidação das Leis do Trabalho (CLT) no setor público ou em sociedades de economia mista, há frequentemente maior estabilidade no emprego⁷, ou a substituições associadas a mudanças de governo, como ocorre em determinados cargos comissionados. Também se excluiu trabalhadores que, embora aparecessem pela primeira vez na RAIS no ano da coorte, registravam tempo de emprego superior a 12 meses, pois essa informação indicaria vínculo iniciado antes do ano definido como entrada no mercado formal.

Na próxima seção, serão apresentadas as estatísticas descritivas do painel construído a partir desses critérios. Essas estatísticas ajudam a caracterizar a composição da amostra e fornecem elementos importantes para o delineamento metodológico apresentado posteriormente.

4.3.2 Estatísticas Descritivas

Inicia-se esta seção apresentando, nas tabelas 46 e 47, estatísticas descritivas básicas para ambas as coortes nos anos de entrada. Nelas, apresentam-se a composição da coorte para o ano inicial, ou seja, como era a composição de indivíduos e empresas no momento de ingresso ao mercado de trabalho pela 1ª vez desses trabalhadores.

Tabela 46 – Estatísticas descritivas na entrada — Coorte de 1989

Variável	N	Missing*	%	Média	DP	Mín	Máx
Salário real em dezembro (R\$)	2,364,555	315,344	13.34	2,387.68	2,626.05	197.28	49,971.94
Remuneração de dezembro (em SM)	2,364,555	0	0.00	3.04	3.79	0	73.46
Remuneração média no ano (em SM)	2,364,555	0	0.00	2.93	2.88	0	119.73
ln salário real em dezembro	2,364,555	315,344	13.34	7.47	0.73	5.28	10.82
Escolaridade (nível)	2,364,555	20,326	0.86	4.44	1.79	1	9
Faixa etária (código)	2,364,555	0	0.00	3.00	0.00	3	3
Tamanho do estabelecimento	2,364,555	0	0.00	6.73	2.56	2	10
Tempo no emprego (anos)	2,364,555	0	0.00	2.03	1.92	0	12
Experiência-calendário (anos)	2,364,555	0	0.00	0.00	0.00	0	0
Experiência formal (anos com presença)	2,364,555	0	0.00	1.00	0.00	1	1
Intensidade de presença no formal	2,364,555	0	0.00	1.00	0.00	1	1
Dummy: troca de firma	2,364,555	0	0.00	0.00	0.00	0	0
Dummy: troca para firma de maior porte	2,364,555	0	0.00	0.00	0.00	0	0

Fonte: Elaboração Própria com dados da RAIS. **Notas:** Estatísticas calculadas apenas para o ano de entrada da coorte de 1989. N = número total de observações por variável. Missing* = número de observações sem informação na variável. *Salário real em dezembro igual a zero ou missing indica que o colaborador aparece na RAIS, mas não teve nenhum recebimento de valores em dezembro daquele ano.

Nas tabelas 46 e 47, é possível observar que o número de indivíduos que aparecem pela primeira vez na RAIS em 1989 é quase 3 vezes maior do que os indivíduos que

⁷ Como o objetivo desta análise também é compreender questões relacionadas à permanência dos profissionais no mercado de trabalho formal, optou-se pela exclusão dos trabalhadores da rede pública. A determinação dos salários nesse grupo não segue necessariamente a lógica de mercado e, em muitos casos, esses trabalhadores tendem a permanecer no emprego por longos períodos, o que poderia afetar a análise do grupo de interesse.

Tabela 47 – Estatísticas descritivas na entrada — Coorte de 1995

Variável	N	Missing*	% Missing	Média	DP	Mín	Máx
Salário real em dezembro (R\$)	873,607	137,396	15.73	1,221.03	1,385.24	117.28	48,529.41
Remuneração de dezembro (em SM)	873,607	0	0.00	2.54	3.33	0	120.00
Remuneração média no ano (em SM)	873,607	0	0.00	2.76	2.45	0	119.64
ln salário real em dezembro	873,607	137,396	15.73	6.87	0.62	4.76	10.79
Escolaridade (nível)	873,607	9,213	1.05	4.45	1.79	1	11
Faixa etária (código)	873,607	0	0.00	3.00	0.00	3	3
Tamanho do estabelecimento	873,607	0	0.00	5.67	2.52	1	10
Tempo no emprego (anos)	873,607	0	0.00	5.87	3.48	0	12
Experiência-calendário (anos)	873,607	0	0.00	0.00	0.00	0	0
Experiência formal (anos com presença)	873,607	0	0.00	1.00	0.00	1	1
Intensidade de presença no formal	873,607	0	0.00	1.00	0.00	1	1
Dummy: troca de firma	873,607	0	0.00	0.00	0.00	0	0
Dummy: troca para firma de maior porte	873,607	0	0.00	0.00	0.00	0	0

Fonte: Elaboração Própria com dados da RAIS. **Notas:** Estatísticas calculadas apenas para o ano de entrada da coorte de 1995. N = número total de observações por variável. Missing* = número de observações sem informação na variável. *Salário real em dezembro igual a zero (missing) indica que o colaborador aparece na RAIS, mas não teve nenhum recebimento de valores em dezembro daquele ano.

ingressaram na RAIS em 1995, considerando as demais variáveis observadas apresentadas na seção 4.3.1. Além da diferença no número de indivíduos, é possível verificar uma diferença significativa no salário real em dezembro, em média, aproximadamente 95% maior para ingressantes em 1989, o que já denota a diferença que se apresentará ao longo das duas coortes. Esse Δ será também objeto de acompanhamento.

Na tabela 48 apresentam-se as estatísticas sobre o perfil dos trabalhadores por nível de escolaridade. À medida que o grau de escolaridade avança, a remuneração média desses trabalhadores também cresce, com exceção dos profissionais que ingressaram no mercado de trabalho com Mestrado e Doutorado.

Tabela 48 – Estatísticas descritivas do salário real em dezembro por nível de escolaridade

Escolaridade (nível)	Coorte 1989			Coorte 1995		
	N	Média	DP	N	Média	DP
Analfabeto	57,086	1,721.62	2,110.00	26,390	913.77	1,135.18
1º ciclo fundamental incompleto	287,696	1,738.52	2,330.73	107,625	993.98	1,240.05
1º ciclo fundamental completo	446,288	1,785.36	1,889.64	138,556	1,061.10	1,197.35
2º ciclo fundamental incompleto	507,015	2,053.35	2,085.43	192,198	1,097.20	1,208.71
2º ciclo fundamental completo	388,456	2,245.33	2,206.38	158,396	1,172.85	1,228.15
2º grau incompleto	263,066	2,700.18	2,609.78	100,923	1,256.17	1,313.82
2º grau completo	303,363	3,589.34	3,309.32	109,959	1,571.73	1,535.69
Superior incompleto	64,518	4,554.84	4,101.47	20,649	2,333.17	2,326.03
Superior completo	26,741	6,080.63	5,452.13	9,441	3,279.79	3,157.12
Mestrado	0	0	0	199	2,017.99	2,347.93
Doutorado	0	0	0	58	1,755.32	896.76
NA — Não informado	20,326	1,690.42	1,919.87	9,213	891.45	896.26

Fonte: Elaboração Própria com dados da RAIS. **Notas:** Estatísticas referem-se ao salário real de dezembro (em R\$ correntes) das coortes de entrada de 1989 e 1995, calculadas por nível de escolaridade. N = Número de observações. Média = Média do salário real em dezembro. DP = Desvio-padrão. Códigos de escolaridade conforme dicionário RAIS/CAGED.

Uma vez apresentadas na tabela 48 a escolaridade com os respectivos rendimentos médios, apresenta-se na tabela 49 as estatísticas descritivas para todo o painel, que compreende as coortes de 1989 até 2017 e de 1995 até 2017.

Tabela 49 – Estatísticas descritivas no painel completo.

Variável	Coorte 1989					Coorte 1995				
	N	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx
Ano	50.534.443	2.001,84	8,41	1.989,00	2.017,00	15.957.263	2.005,32	6,61	1.995,00	2.017,00
Escolaridade (nível)	49.808.436	5,38	2,00	1,00	11,00	15.688.479	5,44	1,92	1,00	11,00
Faixa etária (código)	50.496.643	4,84	1,08	1,00	8,00	15.955.681	4,53	0,97	1,00	8,00
Remuneração média no ano (SM)	50.534.443	1.160,88	2.635,43	0,00	140.436,00	15.957.263	1.146,96	2.171,72	0,00	138.960,00
Remuneração de dezembro (SM)	50.534.443	1.071,26	2.759,91	0,00	251.815,50	15.957.263	985,63	2.257,76	0,00	234.461,55
Salário real em dezembro* (R\$)	39.237.918	3.290,74	4.289,15	0,02	612.030,22	11.693.282	2.503,51	3.333,73	1,26	569.851,95
ln salário real em dezembro*	39.237.917	7,69	0,84	0,78	13,32	11.693.282	7,48	0,75	0,23	13,25
Tamanho do estabelecimento	50.534.443	6,34	2,71	0,00	10,00	15.957.263	5,72	2,68	0,00	10,00
Tipo do estabelecimento	50.534.443	1,07	0,43	1,00	9,00	15.957.263	1,08	0,40	1,00	9,00
Tempo no emprego	50.534.443	57,58	77,30	-0,70	590,80	15.957.263	37,98	46,24	0,00	558,90
Experiência-calendário (anos)	50.534.443	12,84	8,41	0,00	28,00	15.957.263	10,32	6,61	0,00	22,00
Experiência formal (anos com presença)	50.534.443	12,56	7,74	1,00	29,00	15.957.263	10,40	6,17	1,00	23,00
Intensidade de presença no formal	50.534.443	0,93	0,13	0,07	1,00	15.957.263	0,93	0,12	0,09	1,00
Dummy: troca de firma	50.534.443	0,30	0,46	0,00	1,00	15.957.263	0,35	0,48	0,00	1,00
Dummy: troca para firma de maior porte	50.534.443	0,13	0,33	0,00	1,00	15.957.263	0,16	0,36	0,00	1,00

Fonte: Elaboração Própria com dados da RAIS. **Notas:** Estatísticas calculadas para o painel completo de cada coorte (1989–2017 e 1995–2017). N = número de observações não-missing por variável. SM = Salário Mínimo. *Salário real em dezembro igual a zero ou missing indica que o colaborador aparece na RAIS, mas não teve nenhum recebimento de valores em dezembro daquele ano.

Avançando agora por estatísticas descritivas mais avançadas, buscando analisar o salário dos indivíduos ao longo do tempo, como primeiro passo, busca-se compreender como os salários evoluíram entre as coortes de 1989 (que, por definição, chamamos de tratados) e 1995 (que, também por definição, chamaremos de controle). Optou-se por adotar como remuneração padrão para o trabalho a remuneração real em dezembro de cada ano, denotada por $y_{i,t}$ para o indivíduo i , considerando o ano-calendário t . Como a participação no mercado formal pode apresentar *entradas*, *saídas* e *retornos* ao longo do tempo, construíram-se três agregações complementares para caracterizar a evolução dos salários, dependendo das características desse trabalhador no mercado formal de trabalho.⁸ Na figura 15, inicia-se a especificação descritiva, considerando a evolução das médias salariais por ano-calendário, e todos os trabalhadores do painel receberão a denotação apresentada na equação 4.1.

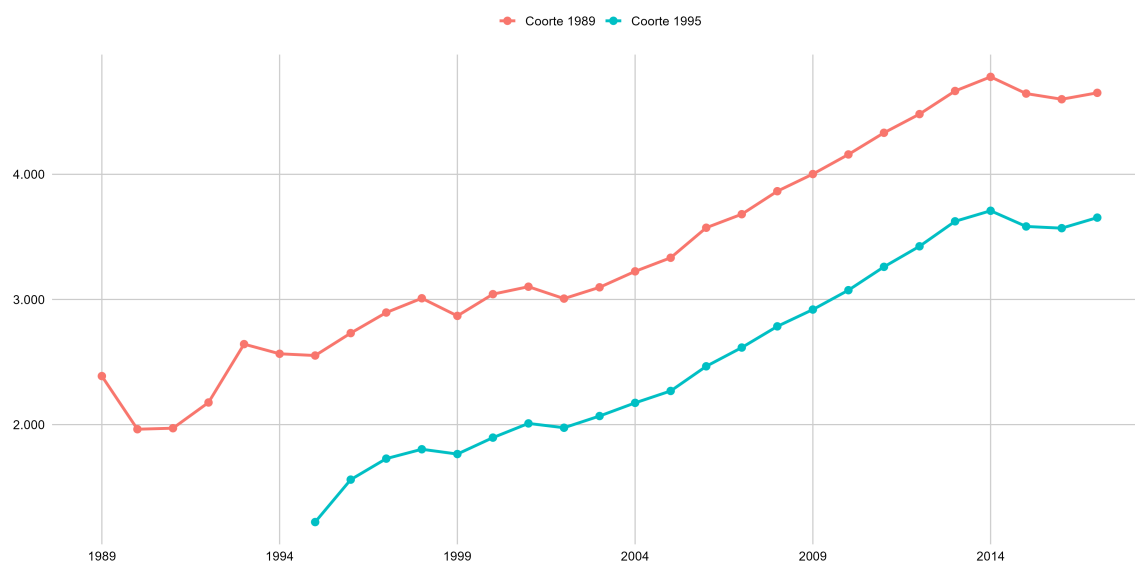
$$\bar{y}_{c,t} = \frac{1}{N_{c,t}} \sum_{i \in \mathcal{A}_{c,t}} y_{i,t}, \quad \mathcal{A}_{c,t} = \{i : i \in c, y_{i,t} \text{ observado e } y_{i,t} > 0\}. \quad (4.1)$$

onde $\mathcal{A}_{c,t}$ é o conjunto de indivíduos pertencentes à coorte c com salário real de dezembro observado e estritamente positivo no ano t , e $N_{c,t} = |\mathcal{A}_{c,t}|$. Dessa forma, a equação 4.1 calcula, para cada coorte $c \in \{1989, 1995\}$ e para cada ano t , a média da

⁸ Levou-se em conta que a experiência formal observada é um sinal importante emitido pelos trabalhadores e passível de checagem por potenciais empregadores, podendo influenciar diretamente a avaliação da experiência profissional. Esse atributo observável no Brasil ficou conhecido como carteira assinada.

remuneração real em dezembro entre os trabalhadores com remuneração econômica válida naquele ano. Em outras palavras, para cada combinação coorte-ano, considera-se apenas o subconjunto de indivíduos que aparecem na base com salário real de dezembro observado e maior que zero; somam-se seus rendimentos $y_{i,t}$ e divide-se esse total pelo número de observações válidas no respectivo grupo. Assim, trabalhadores sem remuneração observada em dezembro ou com remuneração igual a zero não entram nem na soma nem no denominador, sendo interpretadas essas estatísticas como uma média condicional à existência de remuneração válida em dezembro, e não como a média para todos os indivíduos da coorte registrados na RAIS naquele ano. A figura 15 apresenta a trajetória de $\bar{y}_{c,t}$ para as coortes de ingresso em 1989 e 1995 ao longo do painel observado.

Figura 15 – Média Salarial 1989 a 2017, para as duas coortes



Fonte: Elaboração própria com dados RAIS 1989 - 2017.

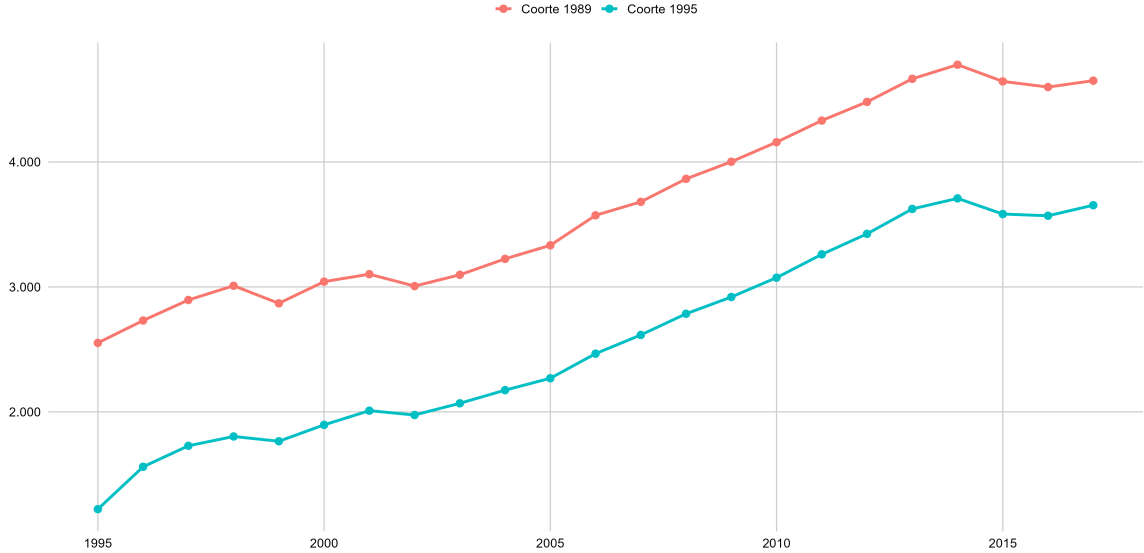
Ao adotar esse desenho metodológico inicial, busca-se resumir o nível salarial típico entre os trabalhadores com remuneração válida em dezembro, permitindo uma leitura descritiva das trajetórias anuais das duas coortes e de sua possível associação com o ambiente macroeconômico e com a dinâmica do mercado formal de trabalho.

Como limitação, essa primeira medida mistura variações no nível salarial com mudanças na composição da amostra ao longo do tempo. Em particular, alterações em $\bar{y}_{c,t}$ podem refletir tanto mudanças nos rendimentos individuais quanto mudanças no conjunto de trabalhadores com salário positivo em dezembro de cada ano. Assim, aumentos ou reduções em $\bar{y}_{c,t}$ não devem ser interpretados automaticamente como evolução salarial individual, pois também podem capturar seleção decorrente da entrada, permanência ou saída de trabalhadores do subconjunto com remuneração válida.

Na Figura 16 apresentam-se as mesmas estatísticas da Figura 15, mas restritas ao intervalo de anos em que ambas as coortes possuem observações, isto é, ao período de 1995

a 2017. Trata-se, portanto, de uma comparação em ano-calendário comum, construída para contrastar as trajetórias salariais das duas coortes no mesmo recorte temporal.

Figura 16 – Média Salarial 1995 a 2017, para ambas as coortes.



Fonte: Elaboração própria com dados RAIS 1989 - 2017, considerando apenas anos comuns às coortes, 1995 a 2017.

Essa sobreposição temporal, contudo, não implica comparabilidade em termos de estágio da trajetória laboral. Em 1995, a coorte de 1989 já havia ingressado no mercado formal seis anos antes, ao passo que a coorte de 1995 está em seu ano de entrada. Assim, a diferença observada entre as curvas no início do período comum é esperada e não deve ser interpretada apenas como diferença salarial entre coortes, pois também reflete distintos níveis de experiência acumulada, permanência no mercado formal e composição da amostra.

Após apresentar a evolução da média salarial real em dezembro nas Figuras 15 e 16, passa-se a uma medida complementar voltada à dispersão relativa dos salários no interior de cada coorte ao longo do tempo. Para isso, utilizou-se o *coeficiente de variação* (CV) dos salários, definido nas equações 4.2 e 4.3.

$$CV_{c,t} = \frac{\sigma_{c,t}}{\mu_{c,t}}, \quad c \in \{1989, 1995\}, \quad (4.2)$$

que pode ser expresso em termos percentuais, conforme a equação 4.3:

$$CV_{c,t}(\%) = 100 \times \frac{\sigma_{c,t}}{\mu_{c,t}}. \quad (4.3)$$

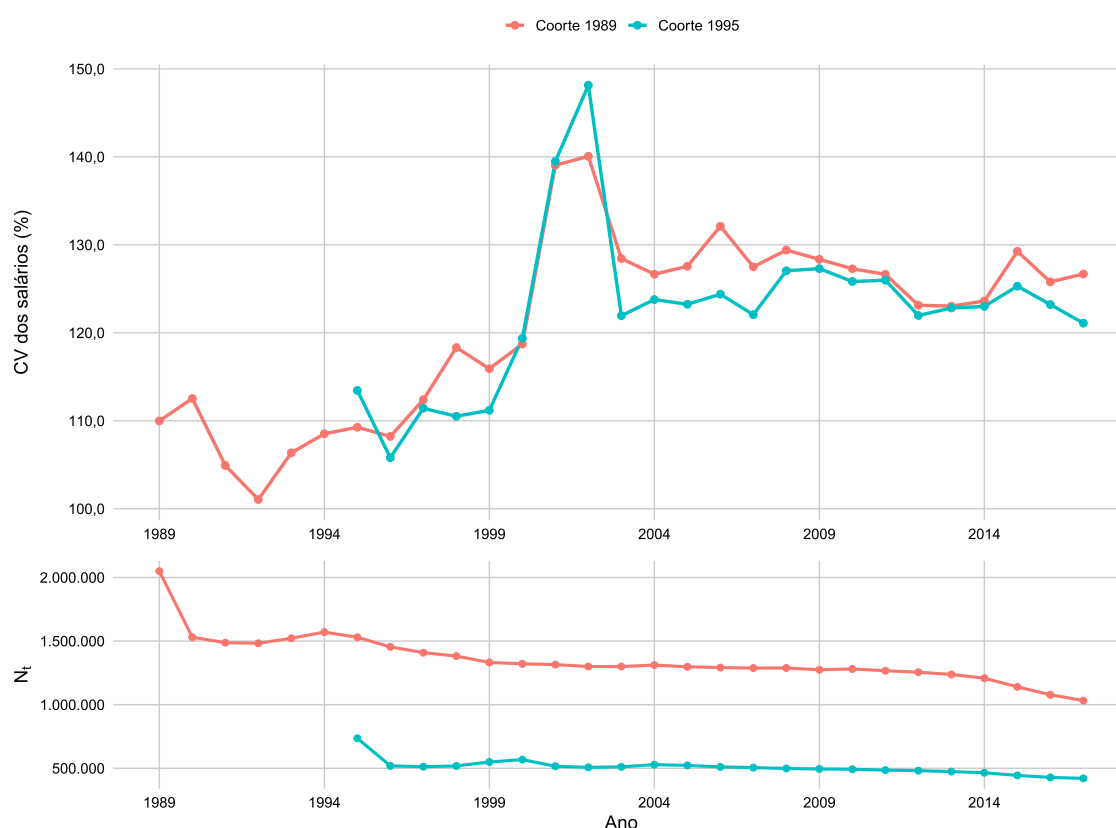
Nessas expressões, $\mu_{c,t}$ e $\sigma_{c,t}$ denotam, respectivamente, a média e o desvio-padrão dos salários reais de dezembro entre os indivíduos da coorte c com remuneração observada e estritamente positiva no ano t . O coeficiente de variação é uma medida adimensional

de dispersão relativa, o que permite comparar a heterogeneidade salarial entre grupos com diferentes níveis médios de remuneração. Essa propriedade é particularmente útil no presente contexto, dado que as coortes de 1989 e 1995 podem apresentar médias salariais distintas ao longo do tempo.

Assim, o $CV_{c,t}$ resume quão dispersos estão os salários em relação ao nível médio observado em cada célula coorte-ano, sem depender da unidade de medida. Em mercados de trabalho nos quais os níveis salariais variam substancialmente por atributos como ocupação, escolaridade e trajetória laboral, essa medida é útil para descrever mudanças na heterogeneidade salarial ao longo do tempo e entre coortes.

A Figura 17 apresenta, no painel superior, a evolução anual do coeficiente de variação dos salários reais de dezembro para as coortes de 1989 e 1995. Já o painel inferior mostra o número de observações válidas $N_{c,t}$ utilizadas no cálculo em cada ano. A leitura conjunta dos dois painéis é importante, pois mudanças no coeficiente de variação podem refletir tanto alterações efetivas na dispersão salarial quanto mudanças na composição do subconjunto de trabalhadores com remuneração positiva observada em dezembro.

Figura 17 – Dispersão relativa dos salários reais de dezembro e tamanho amostral anual, por coorte



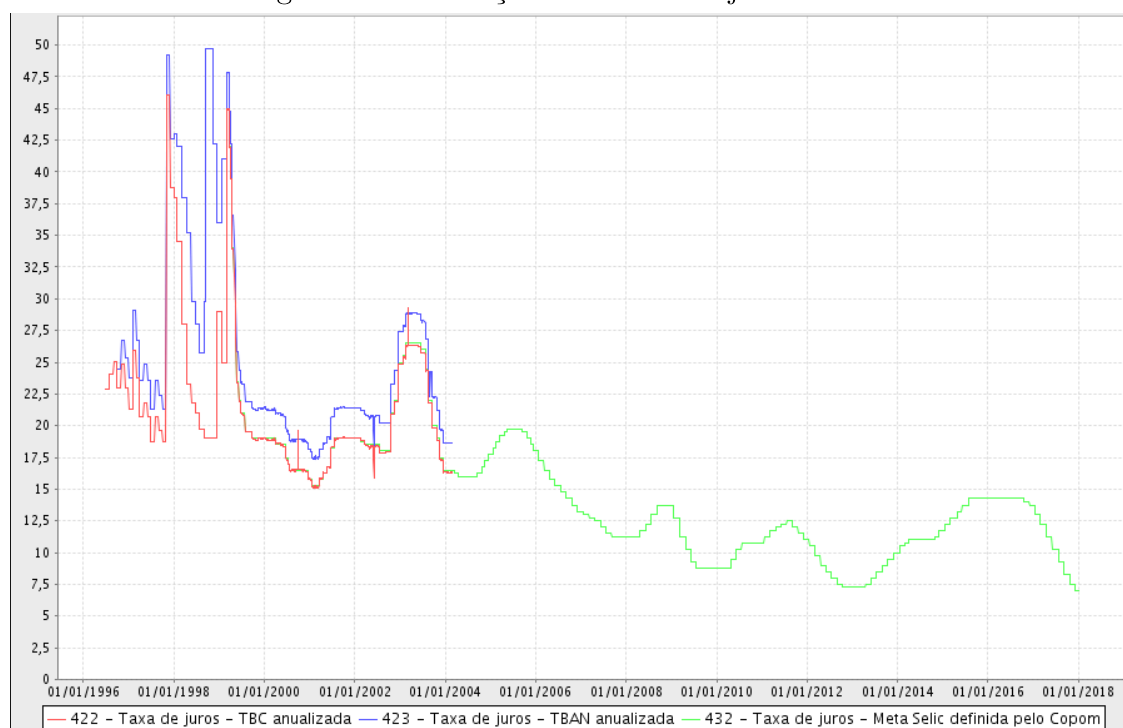
Fonte: Elaboração própria com dados da RAIS de 1989 - 2017.

Como se pode observar na figura 17, em 1999, observou-se um movimento acentuado que coincidiu com o primeiro choque macroeconômico no país após a implementação

do plano real em 1994. O Brasil enfrentou um ataque especulativo que resultou em uma grande desvalorização do real, exigindo um ajuste nos níveis de taxas de juros e acesso a linhas de financiamento externo. Essas situações exigiram que, por parte do governo, fossem instituídos mecanismos que salvaguardassem o plano real, mecanismos que em conjunto receberam o nome de Tripé Macroeconômico.⁹

Após 1999, o país viveu uma série de desafios, inclusive ao longo dos anos em um calendário pré eleitoral acirrado, quando, em 2002, novamente o país se viu em um cenário macroeconômico desafiador. A figura 18 apresenta uma evolução dos juros no Brasil para o período, com dados extraídos do sistema gerador de séries temporais do Banco Central do Brasil (BACEN, 2025). É possível observar variações significativas da taxa Selic ao longo do período.

Figura 18 – Evolução das taxas de juros no Brasil



Fonte: Sistema gerador de séries temporais, (BACEN, 2025).

Ao se analisar os movimentos entre a taxa de juros apresentada na figura 18 e o coeficiente de variação apresentado na figura 17, é possível verificar que o efeito foi positivo em ambas as coortes, mas foi mais acentuado na coorte dos ingressantes em 1995. Para verificar o comportamento N_t , acrescentou-se a informação no próprio gráfico,

⁹ O tripé macroeconômico foi um conjunto de políticas implementadas no Brasil a partir de 1999, durante o governo de Fernando Henrique Cardoso, com o objetivo de garantir a estabilidade econômica. Ele se baseava em três pilares: o regime de metas de inflação, para controlar os preços por meio de uma política monetária ativa; o superávit primário, que visava manter o equilíbrio das contas públicas; e o câmbio flutuante, que permitia ao mercado determinar o valor da moeda.

onde foi possível verificar que a tendência em N_t não parece ter se alterado de forma demasiadamente diferente entre as coortes ao longo do tempo.

Para deixar de olhar apenas para o ano-calendário e aproximar a análise da trajetória laboral, organizou-se também os salários em função da experiência. Em um primeiro passo, foi considerada a *experiência-calendário* (denotada por t_{cal}), definida como o número de anos decorridos desde o ingresso do trabalhador na RAIS, independentemente de ele permanecer continuamente empregado no mercado formal após a entrada. Formalmente,

$$t_{\text{cal},it} = t - \text{ano_ingresso}_i. \quad (4.4)$$

Assim, $t_{\text{cal},it} = 0$ no ano de ingresso, $t_{\text{cal},it} = 1$ no ano seguinte, e assim sucessivamente. Essa medida não exige permanência contínua no mercado formal; ainda que o indivíduo passe alguns anos sem vínculo formal, o relógio de experiência-calendário continua avançando. A série de médias salariais por experiência-calendário é construída, para cada coorte $c \in \{1989, 1995\}$ e cada valor de $k = t_{\text{cal}}$, a partir da agregação

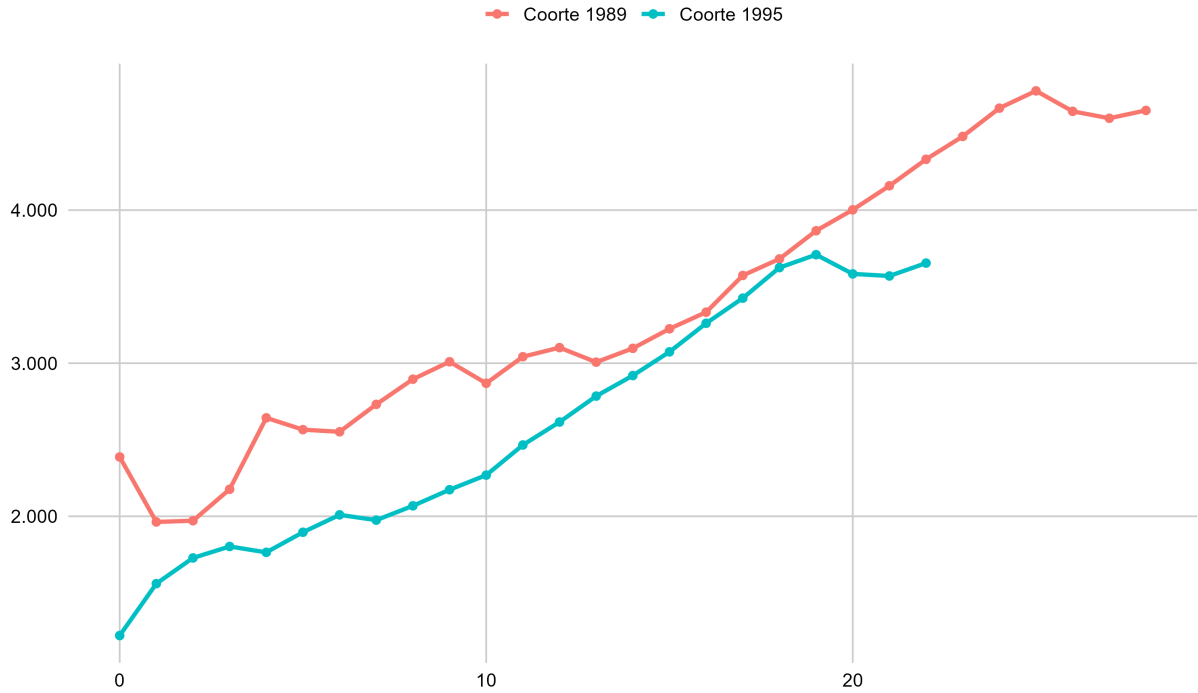
$$\bar{y}_k^{(c)} = \frac{1}{N_k^{(c)}} \sum_{(i,t) \in \mathcal{A}_k^{(c)}} y_{i,t}, \quad (4.5)$$

em que $y_{i,t}$ representa o salário real de dezembro, $\mathcal{A}_k^{(c)}$ é o conjunto de observações pessoa-ano da coorte c com $t_{\text{cal},it} = k$ e salário observado estritamente positivo, e $N_k^{(c)} = |\mathcal{A}_k^{(c)}|$ é o número de observações válidas utilizadas no cálculo.

O gráfico correspondente (Figura 19) mostra, portanto, o perfil médio dos salários reais de dezembro ao longo da trajetória medida em anos desde a entrada no mercado formal. Diferentemente das figuras organizadas por ano-calendário, essa representação alinha as coortes pelo tempo transcorrido desde o ingresso na RAIS, permitindo comparar de forma mais direta o perfil salarial ao longo da carreira formal.

Ainda assim, a estatística permanece condicional à observação de salário positivo em dezembro em cada ponto da trajetória, de modo que diferenças entre coortes podem refletir tanto diferenças nos rendimentos quanto mudanças de composição entre os indivíduos com remuneração válida em cada nível de experiência-calendário. Como se pode observar, a coorte de 1989 é acompanhada por 28 anos enquanto a coorte de 1995 por 22 anos, ou seja, são 6 anos de diferença.

Figura 19 – Variação média dos salários conforme anos de experiência, considerando os anos de experiência desde o nascimento do trabalhador para o mercado formal.



Fonte: Elaboração própria com dados da RAIS de 1989 - 2017.

Como próximo passo, se avança nas estatísticas descritivas sobre a evolução salarial, considerando a experiência formal observada, aqui denotada por t_{for} , interpretada como o número acumulado de anos em que o indivíduo aparece com presença formal na RAIS. Diferentemente das estatísticas organizadas por ano-calendário, essa medida reindexa o tempo pela trajetória observada do trabalhador no mercado formal, acomodando saídas e retornos. Define-se a experiência formal observada como

$$t_{\text{for},it} = \sum_{s \leq t} \mathbb{1}(i \in \mathcal{R}_s), \quad (4.6)$$

em que $\mathcal{R}_s$ representa o conjunto de indivíduos com presença formal observada na RAIS no ano s . Assim, o primeiro ano em que o indivíduo i aparece recebe $t_{\text{for},it} = 1$, o segundo recebe $t_{\text{for},it} = 2$, e assim sucessivamente. Anos em que o indivíduo não aparece na RAIS não incrementam essa contagem.

O eixo horizontal foi reindexado pela experiência formal observada, de modo que o ponto $k = 1$ corresponde ao primeiro ano de presença formal do indivíduo, $k = 2$ ao segundo ano de presença, e assim por diante. Para cada coorte $c \in \{1989, 1995\}$ e cada valor de k , foram calculadas as médias salariais entre as observações com salário real de dezembro observado e estritamente positivo:

$$\bar{y}_k^{(c)} = \frac{1}{N_k^{(c)}} \sum_{(i,t) \in \mathcal{A}_k^{(c)}} y_{i,t}, \quad (4.7)$$

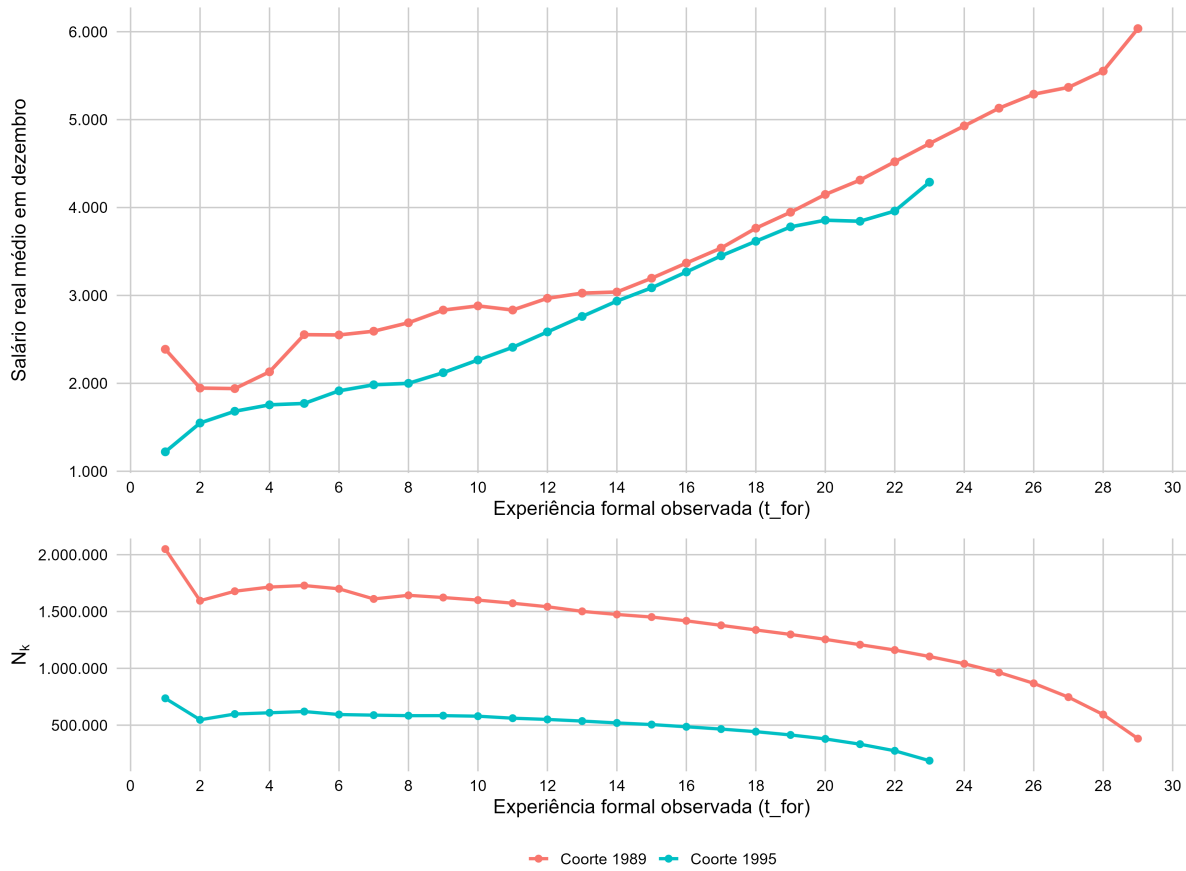
em que:

$$\mathcal{A}_k^{(c)} = \{(i, t) : i \in c, t_{\text{for},it} = k, y_{i,t} \text{ observado e } y_{i,t} > 0\}, \quad N_k^{(c)} = |\mathcal{A}_k^{(c)}|.$$

Dessa forma, $\bar{y}_k^{(c)}$ representa o salário real médio em dezembro entre as observações da coorte c que se encontram exatamente no k -ésimo ano de presença formal observada, independentemente do ano-calendário. Com o eixo indexado por t_{for} , a comparação entre as coortes de 1989 e 1995 passa a ser feita em termos de estágio de experiência observada no mercado formal, e não mais em termos de ano civil.

A Figura 20 apresenta, no painel superior, a evolução de $\bar{y}_k^{(c)}$ ao longo da experiência formal observada; e, no painel inferior, o número de observações válidas $N_k^{(c)}$ utilizado no cálculo da média em cada valor de k .

Figura 20 – Salário real médio de dezembro e número de observações válidas por experiência formal observada, por coorte



Fonte: Elaboração própria com dados da RAIS de 1989 - 2017.

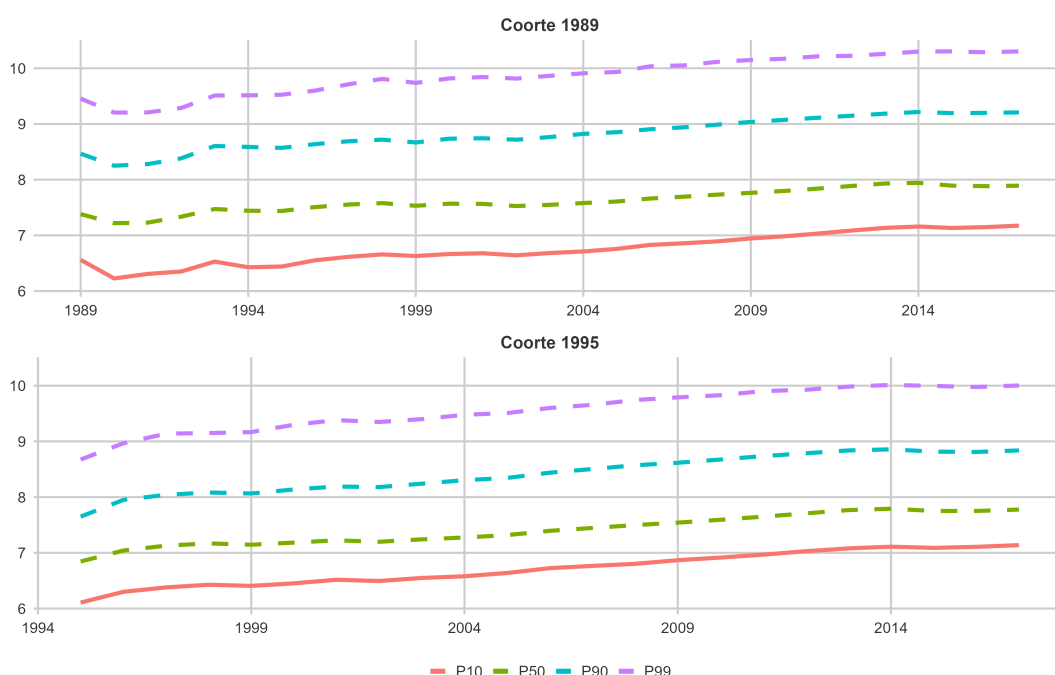
A parte inferior da Figura 20 mostra que o suporte amostral diminui à medida que a experiência observada avança, de modo que a interpretação dos valores mais altos

de t_{for} deve levar em conta a redução do número de observações. Em relação aos rendimentos médios, observa-se uma aproximação dos rendimentos médios em torno do 15^o ano¹⁰ de experiência formal observada. A partir desse ponto, a coorte de 1989 passa a exibir crescimento mais acentuado dos rendimentos médios, ao passo que a coorte de 1995 apresenta uma trajetória mais moderada. Para avaliar as desigualdades intra-coortes, são apresentadas na próxima seção estatísticas descritivas que correspondem a distribuição dos rendimentos ao longo do ciclo de vida dos trabalhadores.

4.3.3 Distribuição salarial e desigualdade ao longo do ciclo de vida

A distribuição dos rendimentos dentro de cada coorte foi analisada por meio de estatísticas-resumo da distribuição do log do salário real de dezembro. Esse exercício permite observar não apenas a tendência central, mas também a heterogeneidade ao longo do tempo, em linha com a literatura que enfatiza a dinâmica distributiva dos rendimentos. Foram utilizadas, para cada coorte e ano-calendário, medidas da média e dos percentis p_{10} , p_{50} , p_{90} e p_{99} do logaritmo do salário real. A Figura 21 apresenta a evolução temporal dessas estatísticas para as coortes de 1989 e 1995.

Figura 21 – Percentis do log do salário real de dezembro por ano-calendário e coorte



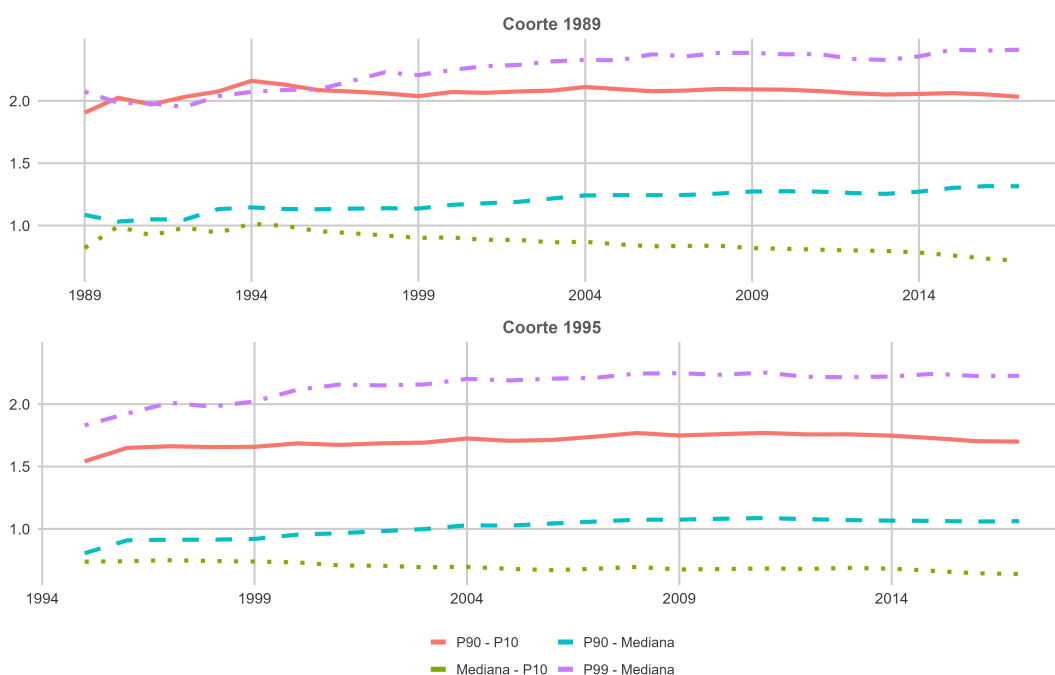
Fonte: Elaboração própria com dados da RAIS de 1989 a 2017. **Notas:** A figura não representa a distribuição completa dos salários, mas sim, um resumo de sua posição central e de sua dispersão ao longo do tempo. Cálculo baseado no log do salário real de dezembro, valores não *missing*. $p_{50} = \text{mediana}$.

¹⁰ Importante destacar que essa, inclusive, é uma marca importante para o trabalhador quando se pensa em contribuição previdenciária. Somente a partir de 180 meses de contribuição para a previdência social que o trabalhador pode ingressar, por exemplo, com pedido de aposentadoria por idade.

Essa representação, inspirada em Guvenen *et al.* (2021), permite avaliar como diferentes partes da distribuição evoluem dentro de cada coorte, verificando se o topo da distribuição salarial cresce mais rapidamente do que o centro e a base, bem como comparando os níveis e a inclinação das trajetórias distributivas entre as coortes. Como a figura está indexada por ano-calendário, sua leitura deve ser feita como evolução temporal no painel, e não como perfil de ciclo de vida ou de experiência.

A figura 22 apresenta os *spreads* entre os percentis, trazendo à tona os extremos da desigualdade. As medidas de desigualdade utilizadas são as diferenças entre percentis do logaritmo do salário real. Por construção, um spread como $p_{90} - p_{10}$ mede a diferença em $\ln$ salário entre o percentil 90 e o percentil 10 da distribuição. Diferenças em log podem ser interpretadas como razões aproximadas de salário; ou seja, se Δ é a diferença em $\ln$, a razão entre os salários é e^{Δ} .

Figura 22 – Spreads como medida de desigualdade – desigualdade geral, e desigualdade nos topos – em $\log$ do salário real em dezembro



Fonte: Elaboração própria com dados da RAIS de 1989 - 2017, inspirado em Guvenen *et al.* (2021)

No início do painel da coorte de 1989, a distância entre o percentil 90 e o percentil 10 do $\log$ do salário real em dezembro, $p_{90} - p_{10}$, situa-se em torno de 1,9 a 2,0. Em termos multiplicativos, isso implica que o salário no percentil 90 é aproximadamente entre $e^{1,9} \approx 6,7$ e $e^{2,0} \approx 7,5$ vezes o salário no percentil 10. Já a diferença entre o percentil 90 e a mediana, $p_{90} - p_{50}$, começa próxima de 1,1, o que corresponde a um salário no topo dos 10% cerca de $e^{1,1} \approx 3$ vezes superior ao salário mediano.

Ao longo do tempo, observa-se que a desigualdade na cauda superior se amplia. Enquanto $p_{50} - p_{10}$ recua de algo entre 0,8 e 1,0 para valores próximos de 0,7, o *spread*

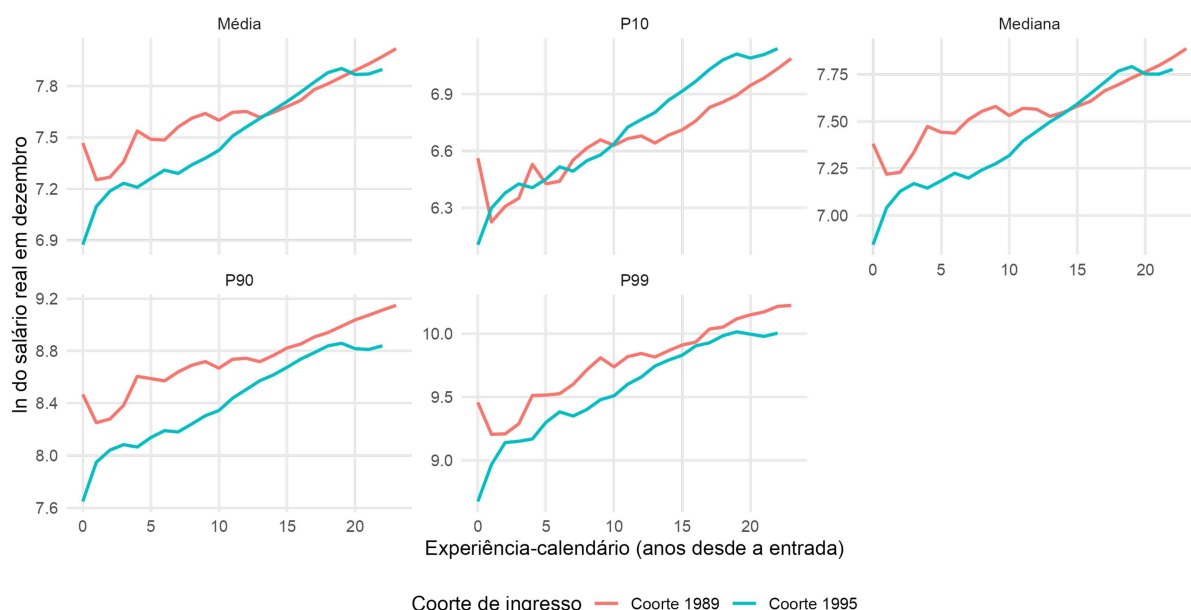
$p_{90} - p_{50}$ aumenta de cerca de 1,1 para algo próximo de 1,3. Isso indica que, na coorte de 1989, o aumento da desigualdade decorre sobretudo do afastamento do topo em relação ao centro da distribuição. Quando se considera a distância entre o percentil 99 e o percentil 10, obtida por $(p_{99} - p_{50}) + (p_{50} - p_{10})$, o diferencial inicial já corresponde a aproximadamente $e^{2,89} \approx 18$ vezes, alcançando ao longo do painel valores acima de 20 vezes.

Para a coorte de 1995, os *spreads* começam em níveis claramente mais baixos. Em 1995, $p_{90} - p_{10}$ é da ordem de 1,54, o que equivale a um salário no percentil 90 que é cerca de $e^{1,54} \approx 4,7$ vezes maior que no percentil 10. Já $p_{90} - p_{50}$ inicia em torno de 0,81, isto é, o percentil 90 corresponde a pouco mais de $e^{0,81} \approx 2,2$ vezes a mediana, enquanto $p_{99} - p_{50}$ fica em torno de 1,83, equivalendo a aproximadamente $e^{1,83} \approx 6,2$ vezes a mediana.

Esses diferenciais aumentam com a evolução do tempo, sobretudo na cauda superior. O *spread* $p_{90} - p_{50}$ sobe gradualmente para valores próximos de 1,06–1,08, enquanto $p_{50} - p_{10}$ diminui de cerca de 0,74 para algo entre 0,64 e 0,68. Assim como na coorte de 1989, a desigualdade cresce principalmente pelo distanciamento do topo em relação ao centro da distribuição. Ainda assim, os níveis da coorte de 1995 permanecem sistematicamente abaixo dos observados para a coorte de 1989 ao longo de quase todo o período, de modo que há aumento da desigualdade, mas não convergência plena entre as duas coortes.

A figura 23 apresenta a trajetória da média e dos percentis do log do salário real em dezembro em função dos anos desde a entrada no mercado formal t_{cal} , separadamente para as coortes de 1989 e 1995.

Figura 23 – Trajetória de percentis de salários (em logaritmo) por coorte por t_{cal}



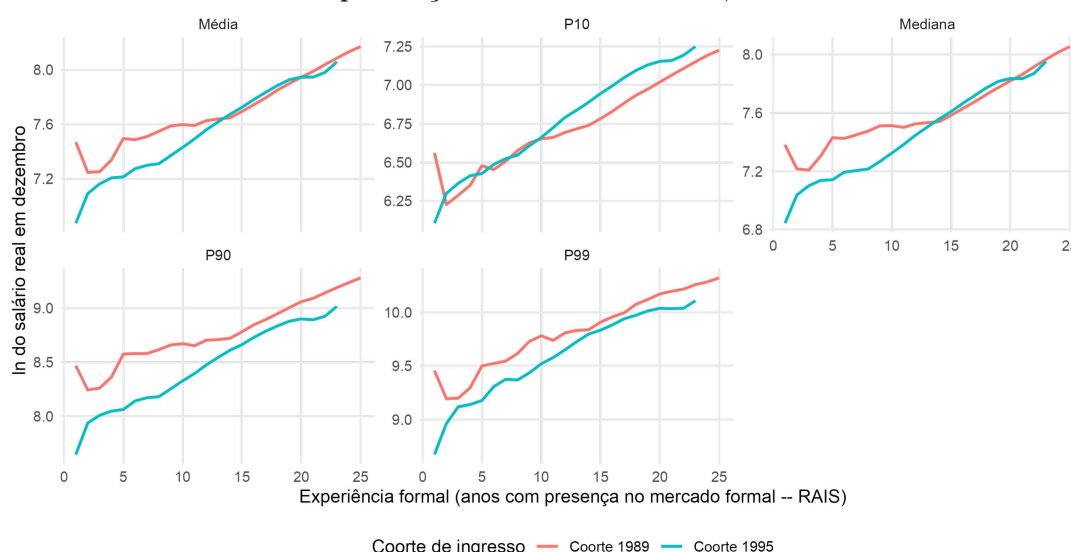
Fonte: Elaboração própria com dados da RAIS de 1989 - 2017. **Notas:** Cada curva reflete a posição do percentil na distribuição do logaritmo do salário real em dezembro naquele ano e coorte.

Essa evolução em função da experiência-calendário t_{cal} (anos desde a entrada no mercado formal) permite verificar que ambas as coortes apresentam perfis fortemente crescentes. Tanto a média quanto os percentis 10, 50, 90 e 99 aumentam de forma quase que monotônica com t_{cal} , refletindo ganhos salariais significativos à medida que os trabalhadores acumulam anos de carreira.

A comparação entre as coortes para um mesmo t_{cal} mostra que uma parte relevante das diferenças observadas por ano-calendário se deve a diferenças de estágio no ciclo de vida. Na coorte de 1989, por exemplo, a mediana do log do salário real em dezembro passa de aproximadamente 7,38 no ano de entrada para cerca de 7,89 após vinte anos de experiência, enquanto o percentil 99 cresce de 9,46 para pouco mais de 10,15. A coorte de 1995 exibe um padrão semelhante, partindo de níveis mais baixos (mediana em torno de 6,84 e p_{99} em torno de 8,67 no ingresso) e convergindo gradualmente para valores próximos aos da coorte anterior. Após cerca de vinte anos de carreira, as medianas das duas coortes são praticamente indistinguíveis. Esses resultados sugerem que as diferenças de nível salarial vistas na análise por ano-calendário refletem, em larga medida, o fato de que a coorte de 1989 encontra-se sistematicamente em estágios mais avançados do ciclo de vida, enquanto diferenças mais estruturais entre coortes parecem permanecer, sobretudo, entre os trabalhadores de maior remuneração.

Apresenta-se o mesmo exercício da figura 23 na figura 24, mas em função da experiência formal t_{for} , definida como o número de anos em que o trabalhador aparece empregado formalmente na RAIS.

Figura 24 – Trajetória de percentis de salários (em logaritmo) por coorte, considerando anos de presença no mercado formal, na RAIS.



Fonte: Elaboração própria com dados da RAIS de 1989 - 2017. **Notas:** Cálculo baseado em log do salário real de dezembro (apenas observações com log do salário real em dezembro finito). Cada curva reflete a posição do percentil na distribuição do logaritmo do salário real em dezembro naquele ano e coorte.

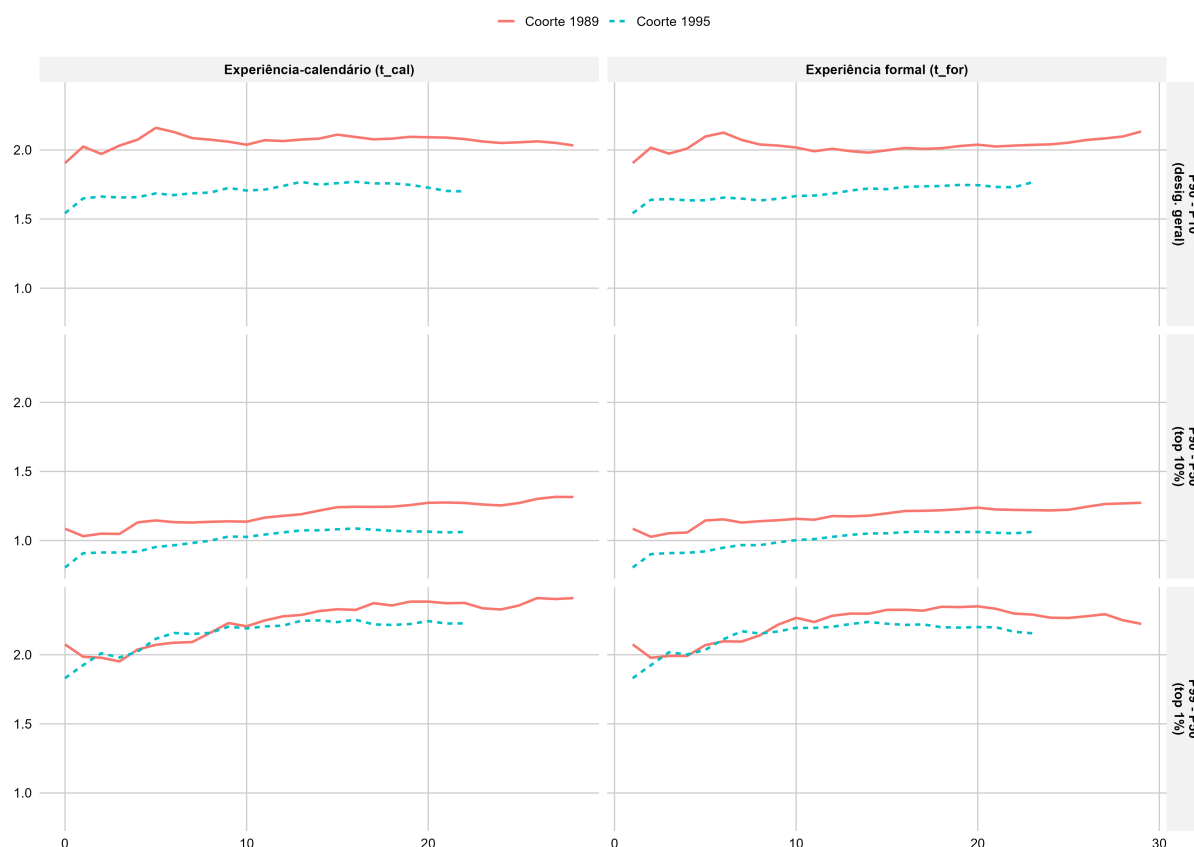
Avaliando agora os cenários apresentados na figura 24, à medida que os indivíduos acumulam anos de presença no setor formal, tanto a média quanto os diferentes quantis da distribuição salarial se deslocam de forma quase monotônica para cima. No primeiro ano de experiência formal ($t_{\text{for}} = 1$), a coorte de 1989 apresenta salários substancialmente mais altos do que a coorte de 1995 em toda a distribuição; a mediana do log salário é cerca de 0,54 pontos log acima na coorte de 1989, o que corresponde a um salário mediano aproximadamente 70% maior.

À medida que a experiência formal aumenta, as diferenças entre coortes na base e no centro da distribuição se reduzem de forma acentuada. Para valores intermediários de t_{for} (por volta de 10 a 15 anos), as medianas e os percentis inferiores das duas coortes tornam-se muito semelhantes e, a partir de cerca de vinte anos de presença no mercado formal, as médias e medianas são praticamente indistinguíveis, demonstrando que essas trajetórias salariais médias convergem, mesmo partindo de pontos distintos sob a ótica do salário real.

No topo da distribuição, contudo, permanece uma diferença relevante. Mesmo para níveis elevados de t_{for} , os percentis 90 e, sobretudo, 99 da coorte de 1989 situam-se acima dos da coorte de 1995. Por exemplo, com cerca de vinte anos de experiência formal, o percentil 99 da coorte de 1989 ainda é cerca de 0,15 pontos log superior ao da coorte de 1995, o que implica um salário no 1% superior aproximadamente 15–16% maior. Assim, alinhar as coortes pelo número de anos de presença no mercado formal elimina quase por completo as diferenças no centro da distribuição, mas a vantagem da coorte de 1989 persiste entre os trabalhadores de maior remuneração, sugerindo que as diferenças entre coortes são particularmente marcantes na cauda superior da distribuição salarial.

Já para analisar os *spreads* salariais entre os percentis, a Figura 25 resume a evolução da desigualdade salarial ao longo da carreira, comparando a experiência-calendário (t_{cal}) e a experiência formal (t_{for}). Em todas as especificações, a diferença ($p_{90} - p_{10}$) é elevada, sendo que para a coorte de 1989 ela se mantém em torno de dois pontos log, o que corresponde a salários no percentil 90 cerca de sete vezes maiores do que no percentil 10, enquanto na coorte de 1995 varia de aproximadamente 1,5 para algo em torno de 1,7 ao longo da carreira (cerca de cinco a cinco vezes e meia o salário do percentil 10). Já o *spread* $p_{90} - p_{50}$ mostra um padrão semelhante, com o top 10% ganhando algo entre duas e três vezes o salário mediano, e indica que a coorte de 1995 parte de níveis de desigualdade menores, mas converge gradualmente para valores próximos aos da coorte de 1989 à medida que a experiência aumenta.

Figura 25 – Spreads como medida de desigualdade – desigualdade geral, e desigualdade nos topos – em log do salário real em dezembro para t_{cal} e t_{for} .



Fonte: Elaboração própria com dados da RAIS (1989–2017). **Notas:** Cálculos baseados no log do salário real de dezembro, considerando apenas observações com valores finitos. Cada curva representa a posição do respectivo percentil na distribuição do log salário real em dezembro, calculada por anos de presença no mercado formal (t_{for}) ou por anos desde o ingresso no mercado formal (t_{cal}).

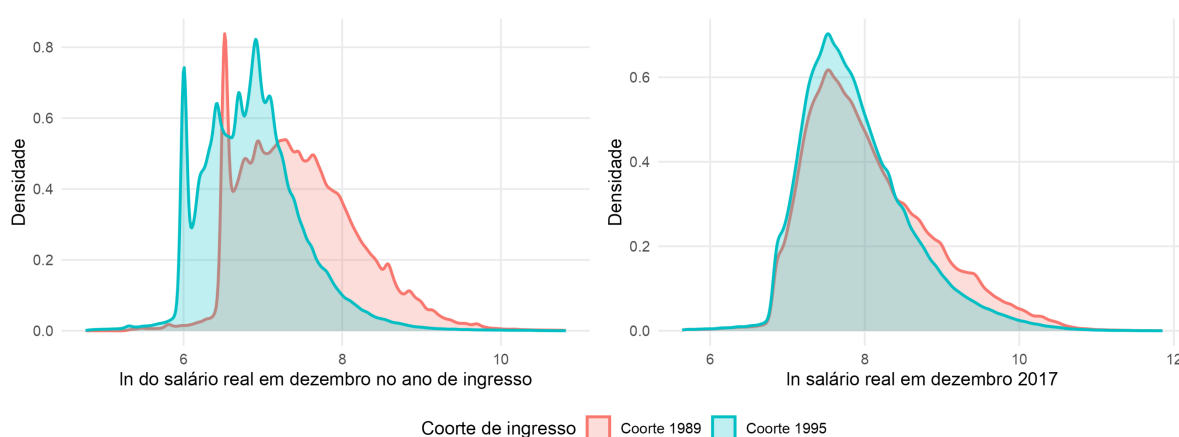
As medidas focadas no topo da distribuição revelam uma dinâmica mais acentuada. O *spread* entre o percentil 99 e a mediana ($p_{99} - p_{50}$) cresce ao longo da carreira para ambas as coortes, passando de cerca de dois pontos log para valores acima de 2,3 em vários momentos, o que implica que o 1% mais rico saiu de algo como sete a oito vezes o salário mediano para aproximadamente dez vezes esse valor. A comparação entre t_{cal} e t_{for} mostra que reindexar o tempo por anos de presença formal reduz as diferenças entre coortes nas medidas $p_{90} - p_{10}$ e $p_{90} - p_{50}$, sugerindo que parte da desigualdade aparente decorre de trajetórias de participação distintas no mercado formal observado.

Dessa forma, percebe-se que a coorte de 1989 permanece sistematicamente mais desigual no topo, onde os *spreads* $p_{99} - p_{50}$ são superiores aos da coorte de 1995, indicando que as diferenças entre coortes se concentram, sobretudo, na cauda superior da distribuição salarial. Os resultados mostram que a desigualdade intra-coorte aumenta ao longo do tempo, impulsionada principalmente pelo topo ($p_{99} - p_{50}$). Esse padrão é consistente com a evidência internacional documentada por Guvenen *et al.* (2021), segundo a qual

choques persistentes e heterogeneidade individual geram forte dispersão ao longo da vida laboral.

A Figura 26 apresenta densidades do log-salário no ano de entrada e em 2017. No ingresso, a coorte de 1995 tem distribuição mais concentrada e ligeiramente deslocada à direita. Em 2017, há uma maior convergência entre as coortes, embora a cauda direita permaneça mais pesada, indicando um subconjunto persistente de trabalhadores de alta renda.

Figura 26 – Densidade dos salários em dezembro, anos de entrada (1989 e 1995) *vs.* último ano observado (2017).



Fonte: Elaboração própria com dados da RAIS de 1989 - 2017. Cálculo baseado em log do salário real de dezembro; apenas observações com log do salário real em dezembro finito.

As estatísticas descritivas revelam padrões salariais bastante estáveis entre as duas coortes analisadas, mas com diferenças sistemáticas de nível e de dispersão. Para todas as métricas, sejam médias, quantis e spreads, observa-se uma trajetória crescente ao longo do ciclo de vida, tanto quando a experiência é medida em anos desde a entrada no mercado formal (t_{cal}) quanto quando a medida utilizada para avaliar a experiência é a presença efetiva no emprego formal (t_{for}).

A coorte de 1995 apresenta salários iniciais mais baixos, porém taxas de crescimento semelhantes às da coorte de 1989; adicionalmente, seus indicadores de desigualdade (especialmente $p_{90} - p_{50}$ e $p_{99} - p_{50}$) são consistentemente menores, sugerindo menor dispersão tanto no centro quanto na cauda superior da distribuição. A persistência dessas diferenças ao longo da vida laboral, bem como o comportamento mais volátil dos quantis superiores, indica que o processo gerador de salários combina fatores determinísticos, sugerindo que estão ligados a experiência e ao retorno à permanência no mercado formal, com forte heterogeneidade individual e componentes estocásticos persistentes e transitórios. Esses padrões motivam o arcabouço metodológico que guiará a análise econométrica, buscando compreender as dinâmicas salariais observadas nas estatísticas descritivas, me-

metodologia parcialmente inspirada em Guvenen *et al.* (2021).

4.4 Metodologia

O objetivo central desta etapa é descrever como a distribuição do crescimento salarial no mercado formal de trabalho varia ao longo do tempo de carreira e ao longo da distribuição de rendimentos recentes Guvenen *et al.* (2021), e como esses padrões diferem entre coortes. Nesta seção, organizam-se as definições, as escolhas operacionais e as estatísticas descritivas específicas que sustentam nossa análise.

4.4.1 Construção da amostra elegível para o mercado formal brasileiro

Inspirados em Guvenen *et al.* (2021), esta seção documenta a amostra construída para o contexto brasileiro a partir dos microdados da RAIS, buscando documentar a heterogeneidade na dinâmica de rendimentos. Conclui-se que não apenas a variância do crescimento salarial muda ao longo do ciclo de vida, como também a *assimetria* (se quedas são mais frequentes/intensas do que altas) e questões relacionadas a *ocorrência de eventos extremos*, que mudam sistematicamente com a idade/experiência e com a posição do trabalhador na distribuição de rendimentos recentes. Diferente de Guvenen *et al.* (2021), estimamos a idade não como a idade do trabalhador, mas sim como uma idade desde o nascimento deste no mercado de trabalho formal, que denotamos como t_{cal} .

O propósito deste processo é, ao invés de resumirmos o processo salarial por poucos parâmetros, descrever diretamente a superfície de momentos do crescimento salarial no mercado formal de trabalho em células definidas por tempo de carreira e posição em rendimentos recentes, estimando assim o comportamento de seus rendimentos e a permanência no mercado formal no futuro.

Enquanto na seção 4.3 se descreveu de forma mais geral as estatísticas descritivas para a amostra dos trabalhadores ingressantes no mercado de trabalho nos anos de 1989 e 1995, para facilitar o entendimento da construção das variáveis, serão apresentados, em subtópicos nesta seção, um passo a passo da construção das variáveis utilizadas no estudo.

Unidade de análise e tempo de carreira.

A unidade de análise é o trabalhador i em um ano-calendário a em que ele aparece na RAIS (isto é, em que há vínculo formal registrado). Para tornar comparáveis trajetórias que começam em anos diferentes, o tempo foi reindexado por tempo de carreira τ , definido como o número de anos transcorridos desde o primeiro ingresso observado no formal (variável t_{cal}). Um aspecto importante desse índice é que ele continua avançando mesmo que o trabalhador fique alguns anos fora do mercado formal; assim, τ mede o tempo desde a entrada e não o tempo estritamente presente como empregado no mercado

formal. Essa escolha é deliberada, pois o interesse é comparar trabalhadores em um mesmo estágio de carreira desde o ingresso no formal.

Para garantir comparabilidade “igual com igual”, restringiu-se o suporte ao horizonte comum observado nas duas coortes, que contém 23 pontos de carreira e vai de $\tau = 0$ (ingresso) até $\tau = 22$ (último ano que será observado). Dessa forma, evita-se que diferenças entre coortes sejam confundidas com diferenças mecânicas de idade/experiência decorrentes de janelas de observação desiguais. Além disso, conforme demonstrado na seção 4.3, o perfil do trabalhador ingressante em relação à idade¹¹, apresentando um perfil semelhante entre as duas coortes.

Salário formal, ausência de rendimentos em dezembro e transformação logarítmica.

O salário formal observado é o real de dezembro, $Y_{i,a(\tau)}$ ¹². No contexto da RAIS, é comum existirem observações em que o trabalhador aparece no registro, mas sem remuneração positiva em dezembro. Nesses casos, adota-se como convenção

$$Y_{i,a(\tau)} = 0$$

quando a remuneração de dezembro é ausente ou não positiva. Assim, optou-se por manter essas observações no painel como casos de renda formal nula em dezembro, evitando que a dinâmica do mercado formal seja descrita apenas a partir dos trabalhadores com remuneração positiva naquele mês.

Além disso, ao construir a medida de crescimento futuro dos rendimentos, atribuiu-se valor zero ao rendimento formal de dezembro quando o trabalhador não é observado na RAIS no período futuro *lead* $\tau + k$, isto é, definiu-se $Y_{i,a(\tau+k)} = 0$. Essa decisão é consistente com o escopo do capítulo, que se restringe ao mercado de trabalho formal.

Para permitir a inclusão de observações com rendimento formal igual a zero sem comprometer a ordenação da variável, assume-se a medida de rendimentos como a transformação $\log(1 + Y)$ ¹³:

$$y_{i,a(\tau)} = \log(1 + Y_{i,a(\tau)}).$$

Essa transformação¹⁴ Ela reduz a influência de valores extremos e permite interpretar

¹¹ Restringiu-se à amostra aos trabalhadores que ingressaram no mercado de trabalho em 1989 e 1995, na faixa etária entre 18 e 24 anos e com rendimentos de no máximo R\$50.000,00.

¹² Variável *sm_dez_real* na RAIS.

¹³ A transformação $\log(1 + Y)$ permite manter observações com rendimento formal igual a zero, preservando a ordenação dos rendimentos e reduzindo a assimetria da distribuição. Essa escolha, contudo, implica que variações em y podem refletir tanto mudanças nos rendimentos formais positivos quanto transições de entrada ou saída do mercado formal.

¹⁴ Embora a transformação $\log(1 + Y)$ seja frequentemente utilizada para acomodar observações com salário igual a zero, é importante ressaltar que o efeito médio do tratamento estimado por essa transformação não possui uma interpretação direta como efeito percentual (Chen; Roth, 2024).

variações em y como mudanças proporcionais aproximadas quando Y assume valores suficientemente elevados. Como a RAIS não observa vínculos e rendimentos no setor informal, valores iguais a zero devem ser lidos como ausência de rendimento formal registrado em dezembro, e não como ausência de qualquer fonte de renda do trabalhador.

Piso anual e admissibilidade.

Imponho um piso anual $Y_{\min,a}$ para definir observações com vínculo formal minimamente relevante e evitar que variações residuais sejam dominadas por remunerações muito pequenas ou por ruído de mensuração. Para aproximar esse piso, utilizou-se meio salário mínimo real anual. A *proxy* do salário mínimo real no ano a é dada por

$$\widehat{MW}_a = \text{mediana}_i \left(\frac{\text{sm_dez_real}_{i,a}}{\text{vlremundezembros}_{i,a}} \right), \quad Y_{\min,a} = 0,5 \times \widehat{MW}_a.$$

Definiu-se a variável indicadora de admissibilidade como

$$\text{admissivel}_{i,a(\tau)} = 1\{Y_{i,a(\tau)} \geq Y_{\min,a}\}.$$

A admissibilidade não é um resultado em si, mas um critério de construção amostral para a etapa de rendimentos recentes. Ainda assim, ela já descreve o quanto do painel se encontra acima de um limiar mínimo de relevância formal no fim do ano.

Rendimentos Recentes (RE) e amostra elegível.

Inspirado em Guvenen *et al.* (2021), construíram-se as medidas de rendimentos recentes como a média dos cinco anos anteriores de carreira, usando *lags* em τ e impondo o piso anual corrente em cada um desses termos:

$$\widehat{RE}_{i,\tau} = \frac{1}{5} \sum_{j=1}^5 \max\{Y_{i,a(\tau-j)}, Y_{\min,a(\tau)}\}.$$

A observação (i, τ) pertence ao universo RE-elegível quando é admissível em τ , admissível em $\tau - 1$ e admissível em pelo menos dois dos quatro períodos entre $\tau - 2$ e $\tau - 5$. Formalmente, a amostra RE-elegível exige presença salarial mínima no presente e um histórico recente suficiente para construir uma medida informativa de rendimentos recentes.

A Tabela 50 reporta as frações `admissível` e `RE_elegível` por coorte.

Tabela 50 – Frações `admissível` e `RE_elegível` por coorte

Coorte	% admissível	% RE-elegível
1989	0.772	0.465
1995	0.732	0.418

Fonte: Elaboração Própria com dados da RAIS 1989 - 2017.

Como se pode observar, a coorte de 1989 tem uma maior fração de observações acima do piso e, conseqüentemente, uma maior fração RE-elegível. Isso antecipa que a base efetiva para construir RE e estimar momentos condicionais é relativamente mais ampla na coorte de 1989, algo que foi levado em conta ao interpretar diferenças entre coortes.

Residualização por células (ano $\times$ carreira).

Para remover a variação comum associada ao ano-calendário e ao estágio de carreira, subtraiu-se as médias por célula (a, τ) :

$$y_{i,a(\tau)}^{res} = y_{i,a(\tau)} - \hat{y}_{a,\tau}, \quad \hat{y}_{a,\tau} = \mathbb{E}[y_{i,a(\tau)} \mid a, \tau].$$

Essa etapa é central para a interpretação: ao se analisar $\Delta^k y^{res}$, se está examinando variações relativas ao padrão típico do ano e do estágio de carreira, o que reduz o risco de que mudanças macroeconômicas ou tendências gerais do ciclo de vida dominem a leitura dos momentos. Também se centralizam $\widehat{RE}_{i,\tau}$ pela sua média nas células (a, τ) , usando apenas observações RE-elegíveis.

Assim, obtém-se uma medida de posição relativa dos rendimentos recentes dentro do mesmo ano e do mesmo estágio de carreira, modelo inspirado no empregado na literatura de dinâmica de rendimentos¹⁵(Guvenen *et al.*, 2021).

Faixas percentílicas de RE , denotadas por RE_{bin} .

Em cada coorte c e tempo de carreira τ , ranqueou-se trabalhadores por $RE_{i,\tau}$ e dividiu-se esses trabalhadores em $B = 100$ faixas (*bins*) percentílicas. Assim, cada RE_bin representa uma fatia de 1% da distribuição de RE dentro de (c, τ) . O $RE_bin = 1$ agrupa trabalhadores com rendimentos recentes relativamente mais baixos naquele estágio de carreira, enquanto o $RE_bin = 100$ agrupa os relativamente mais altos. Essa definição é importante para a comparação, pois não se trata do “1% da variação absoluta de renda”, mas sim de 1% da distribuição dos rendimentos recentes.

Crescimento em k anos, ausência no lead e exit_rate.

Para cada horizonte $k \in \{1, 5\}$, definiu-se

$$\Delta^k y_{i,\tau}^{res} = y_{i,a(\tau+k)}^{res} - y_{i,a(\tau)}^{res}.$$

Quando o trabalhador não aparece na RAIS no *lead* $\tau + k$, se imputou $Y_{i,a(\tau+k)} = 0$ naquele dezembro futuro; logo, $y_{i,a(\tau+k)} = 0$ e

$$y_{i,a(\tau+k)}^{res} = -\hat{y}_{a(\tau+k),\tau+k}.$$

¹⁵ Diferentemente de Guvenen *et al.* (2021), como já explicado anteriormente, utilizou-se (a, τ) como medida de idade desde o nascimento para o mercado de trabalho formal, e não como idade em anos do indivíduo.

Definiu-se o indicador de saída do formal no horizonte k como

$$exit_{i,\tau}^{(k)} = 1\{i \text{ não observado no formal em } \tau + k\},$$

e sua média em cada célula como $exit_rate$. A Tabela 51 resume esse objeto no universo RE-elegível.

Tabela 51 – Resumo de $exit_rate$ (ausência no formal no lead) no universo RE-elegível

k	Coorte	média de $exit_rate$	mediana de $exit_rate$
1,000	1989	0,161	0,161
1,000	1995	0,190	0,190
5,000	1989	0,357	0,365
5,000	1995	0,400	0,404

Fonte: Elaboração Própria com dados da RAIS 1989 - 2017.

No horizonte de 1 ano, a ausência no mercado formal no futuro já é relevante, situando-se em torno de 16% na coorte de 1989 e 19% na coorte de 1995. Em 5 anos, essa fração acumulada cresce substancialmente, alcançando aproximadamente 36% e 40%, respectivamente. Essa estatística antecipa que a dinâmica de médio prazo no mercado formal envolve uma margem extensiva expressiva, o que justifica reportar $exit_rate$ junto aos momentos de crescimento salarial.

Momentos por célula e interpretação do número de “grupos” (células).

Os momentos do crescimento são calculados em células (c, τ, RE_bin, k) : tamanho (n), dispersão (sd e $P90 - P10$), assimetria (Kelley), curtose (Crow-Siddiqui) e $exit_rate$. Como o suporte comum de carreira vai de $\tau = 0$ até $\tau = 22$, o número máximo de células por coorte em cada horizonte é dado por

$$(22 - k + 1) \times 100.$$

Assim, para $k = 1$, o máximo teórico é

$$(22 - 1 + 1) \times 100 = 2200$$

células por coorte, enquanto para $k = 5$ o máximo teórico é

$$(22 - 5 + 1) \times 100 = 1800$$

células por coorte.

A cobertura resultante é reportada na Tabela 56. Esse método, inspirado em Guvenen *et al.* (2021) produz uma superfície extensa de estatísticas, e as tabelas e figuras da seção de Resultados funcionam justamente como formas de visualizar e resumir essa superfície.

4.4.2 Momentos da distribuição e leitura econômica das superfícies

A construção das superfícies de momentos permite caracterizar não apenas o nível médio do crescimento dos rendimentos, mas também a forma da distribuição de $\Delta^k y^{res}$ em cada célula da superfície (Guvenen *et al.*, 2021). Define-se uma célula como

$$g \equiv (\tau, RE_bin, k),$$

em que τ representa o estágio de carreira, RE_bin indica a posição do trabalhador na distribuição de rendimentos recentes, e k é o horizonte de crescimento. Para cada coorte $c \in \{1989, 1995\}$, seja $\mathcal{I}_{g,c}$ o conjunto de trabalhadores pertencentes à célula g , com tamanho $n_{g,c}$. Para cada trabalhador $i \in \mathcal{I}_{g,c}$, o crescimento residual¹⁶ no horizonte k é definido como

$$\Delta^k y_{i,g,c}^{res} = y_{i,t+k}^{res} - y_{i,t}^{res}.$$

Quando o trabalhador não é observado no registro formal no horizonte futuro, o evento é incorporado à margem extensiva por meio de uma variável indicadora de saída do formal.¹⁷ Para cada célula g e coorte c , calcularam-se momentos da distribuição de $\Delta^k y_{i,g,c}^{res}$. A média do crescimento residual, quando utilizada, é dada por

$$\overline{\Delta^k y_{g,c}^{res}} = \frac{1}{n_{g,c}} \sum_{i \in \mathcal{I}_{g,c}} \Delta^k y_{i,g,c}^{res}.$$

A margem extensiva é medida pela taxa de saída do formal,

$$exit_rate_{g,c}^k = \frac{1}{n_{g,c}} \sum_{i \in \mathcal{I}_{g,c}} \mathbf{1}\{i \text{ não é observado no formal em } t+k\}.$$

Essa medida captura a probabilidade de não permanência no setor formal observado.

Já a dispersão do crescimento residual é medida pelo desvio-padrão,

$$sd_{g,c}^k = \sqrt{\frac{1}{n_{g,c} - 1} \sum_{i \in \mathcal{I}_{g,c}} \left(\Delta^k y_{i,g,c}^{res} - \overline{\Delta^k y_{g,c}^{res}} \right)^2}.$$

Valores mais altos de $sd_{g,c}^k$ indicam maior heterogeneidade dos resultados futuros dentro da célula. Contudo, o desvio-padrão não informa se essa volatilidade está associada a perdas ou ganhos; ele apenas mede a amplitude geral da dispersão.

Para verificar se essa dispersão também aparece no corpo da distribuição, utiliza-se a medida robusta de dispersão, o intervalo interdecílico $P90 - P10$, amplamente empregado na literatura de dinâmica de rendimentos (Guvenen *et al.*, 2021).

$$P90-P10_{g,c}^k = Q_{0.90,g,c}^k - Q_{0.10,g,c}^k,$$

¹⁶ Guvenen *et al.* (2021)

¹⁷ A saída do formal deve ser interpretada como ausência do trabalhador no registro formal observado no horizonte k , e não necessariamente como ausência de renda total. O trabalhador pode ter migrado para a informalidade, estar desempregado, trabalhando por conta própria ou estar em outra situação não observada na RAIS.

em que $Q_{p,g,c}^k$ denota o percentil p da distribuição de $\Delta^k y^{res}$ na célula g e coorte c . Como essa medida é baseada em quantis, ela é menos sensível a poucos valores extremos isolados. Quando sd e $P90 - P10$ aumentam conjuntamente, a interpretação é de uma abertura mais generalizada da distribuição.

Já a forma como se mensura a assimetria se dá pela assimetria de Kelley (1947), conforme sua aplicação recente à distribuição de variações dos rendimentos (Guvenen *et al.*, 2021).

$$Skew_{g,c}^{Kelley} = \frac{(Q_{0.90,g,c}^k - Q_{0.50,g,c}^k) - (Q_{0.50,g,c}^k - Q_{0.10,g,c}^k)}{Q_{0.90,g,c}^k - Q_{0.10,g,c}^k}.$$

Onde valores negativos indicam que a distribuição está mais alongada para a esquerda, isto é, que quedas grandes têm maior peso relativo do que altas elevadas. Valores positivos indicam maior peso relativo da cauda direita, associada a ganhos elevados. Valores próximos de zero sugerem uma distribuição mais equilibrada entre perdas e ganhos extremos.

A curtose pode ser mensurada por alternativas robustas à medida convencional baseada no quarto momento, como a razão de Crow–Siddiqui. Kim e White (2004) destacam que essa razão, fundamentada em quantis, apresenta maior robustez frente à sensibilidade a outliers, característica marcante das medidas tradicionais de curtose. Dessa forma, o uso de medidas baseadas em quantis, como a proposta por Crow e Siddiqui (1967), contribui para uma avaliação mais confiável da distribuição dos retornos de mercado. Assim, a curtose é medida pela razão de Crow–Siddiqui (Crow; Siddiqui, 1967)

$$Kurt_{g,c}^{CS} = \frac{Q_{0.975,g,c}^k - Q_{0.025,g,c}^k}{Q_{0.75,g,c}^k - Q_{0.25,g,c}^k}.$$

Essa medida compara a amplitude das caudas externas com a amplitude interquartílica. Valores mais altos indicam caudas mais espessas, ou seja, maior incidência de mudanças muito grandes em relação ao comportamento típico da célula. A curtose, isoladamente, não informa se os extremos são perdas ou ganhos; por isso, sua interpretação deve ser combinada com a assimetria.

Para comparar as coortes, definiu-se, para cada momento m , a diferença entre 1989 e 1995 como

$$d_g^m = \hat{m}_{g,1989} - \hat{m}_{g,1995}.$$

Assim, valores negativos de d_g^m indicam que o momento m é maior na coorte de 1995, enquanto valores positivos indicam que o momento é maior na coorte de 1989.

A construção das superfícies de momentos permite caracterizar não apenas o nível médio do crescimento dos rendimentos, mas também a forma da distribuição de $\Delta^k y^{res}$ em cada célula (τ, RE_bin, k) . Para orientar a interpretação econômica, a Tabela 52 resume como cada momento deve ser lido e o que se busca identificar empiricamente. A ideia central é que sd , $P90 - P10$, a assimetria e a curtose respondem a perguntas diferentes: o desvio-padrão mede a amplitude geral da dispersão; $P90 - P10$ verifica se essa dispersão

aparece no corpo da distribuição; a assimetria indica se o risco pesa mais para quedas ou para ganhos; e a curtose indica se há maior incidência de eventos extremos.

Tabela 52 – Momentos calculados nas células da superfície e interpretação econômica

Momento	Interpretação econômica
$exit_rate$	Probabilidade de saída do registro formal no horizonte k . Valores maiores indicam maior fragilidade na margem extensiva, isto é, maior risco de o trabalhador não reaparecer no setor formal observado.
$sd(\Delta^k y^{res})$	Mede a dispersão geral do crescimento residual dos rendimentos. Valores maiores indicam trajetórias futuras mais heterogêneas: trabalhadores semelhantes dentro da mesma célula podem experimentar resultados muito diferentes. Essa medida não informa, isoladamente, se o risco está associado a quedas ou ganhos.
$P90 - P10$	Mede a abertura robusta da distribuição do crescimento residual. Valores maiores indicam maior distância entre resultados relativamente bons e ruins. Como é baseada em quantis, essa medida é menos sensível a poucos valores extremos isolados do que o desvio-padrão.
Assimetria de Kelley	Indica a direção dominante do risco. Valores negativos sugerem maior peso relativo de quedas severas, isto é, cauda à esquerda. Valores positivos sugerem maior peso relativo de ganhos elevados, isto é, cauda à direita. Valores próximos de zero indicam distribuição mais equilibrada entre perdas e ganhos extremos.
Curtose de Crow-Siddiqui	Indica a espessura das caudas da distribuição. Valores maiores significam maior incidência de eventos extremos, isto é, mudanças muito grandes em relação ao comportamento típico da célula. A curtose não informa sozinha se esses extremos são perdas ou ganhos; por isso, deve ser lida junto com a assimetria.

Fonte: Elaboração Própria com base em Kelley (1947), Crow e Siddiqui (1967) e Guvenen *et al.* (2021).

Antes de apresentar os resultados, é útil explicitar como as diferentes estatísticas de superfície devem ser lidas em conjunto. A Tabela 53 organiza essa interpretação ao combinar medidas de dispersão, assimetria, curtose e margem extensiva. Essa leitura conjunta é necessária porque um mesmo aumento de dispersão pode ter significados econômicos distintos dependendo da forma da distribuição, podendo refletir maior heterogeneidade no corpo da distribuição, poucos eventos extremos nas caudas, maior peso relativo de quedas severas ou maior incidência de ganhos excepcionais. A tabela funciona como um guia de leitura dos resultados apresentados nas seções seguintes, permitindo interpretar os momentos não apenas como estatísticas descritivas, mas como indicadores de risco, instabilidade e mobilidade ao longo do ciclo de vida.

Tabela 53 – Guia de interpretação conjunta dos momentos

Dimensão	Padrão observado	Interpretação econômica
Dispersão geral	<i>sd</i> alto	Maior volatilidade do crescimento residual. Trabalhadores em células semelhantes enfrentam trajetórias futuras mais heterogêneas.
Dispersão robusta	<i>sd</i> alto e $P90 - P10$ alto	A dispersão aparece também no corpo da distribuição, não apenas em poucos valores extremos. Indica abertura ampla entre trajetórias boas e ruins.
Extremos nas caudas	<i>sd</i> alto e $P90 - P10$ moderado	A volatilidade pode estar sendo puxada por observações extremas. Nesse caso, a curtose é central para avaliar a importância das caudas.
Direção do risco	Assimetria negativa	A distribuição é inclinada para a esquerda. Quedas grandes têm maior peso relativo do que ganhos grandes, indicando risco de perdas severas.
Direção favorável	Assimetria positiva	A distribuição é inclinada para a direita. Ganhos elevados têm maior peso relativo do que quedas severas.
Intensidade dos extremos	Curtose alta	Há maior incidência de eventos extremos. A curtose indica a presença de caudas espessas, mas a direção dos extremos deve ser lida junto com a assimetria.
Vulnerabilidade elevada	<i>exit_rate</i> alto, <i>sd</i> alto e assimetria negativa	A célula combina maior saída do formal, maior volatilidade e risco de quedas. É o padrão mais associado à fragilidade no mercado formal observado.
Instabilidade intensiva	<i>exit_rate</i> baixo, <i>sd</i> alto e curtose alta	A permanência formal é relativamente maior, mas os rendimentos dos que permanecem são mais instáveis e sujeitos a eventos extremos.

Fonte: Elaboração Própria com base em Kelley (1947), Crow e Siddiqui (1967) e Guvenen *et al.* (2021).

Notas: As classificações “alto”, “baixo” ou “moderado” são relativas às demais células da mesma superfície, estágio de carreira, faixa de *RE* ou coorte. O desvio-padrão mede a dispersão total; $P90 - P10$ mede a dispersão no corpo da distribuição; a assimetria indica a direção dominante do risco; e a curtose indica a presença de eventos extremos. A leitura econômica dos extremos exige combinar curtose e assimetria.

Em síntese, a leitura das superfícies segue uma lógica cumulativa. Primeiro, *exit_rate* identifica a fragilidade na margem extensiva, isto é, o risco de não permanência no registro formal. Em seguida, *sd* e $P90 - P10$ indicam se o crescimento residual é mais ou menos disperso. Por fim, a assimetria e a curtose qualificam essa dispersão: a assimetria informa se o risco pesa mais para quedas ou ganhos, enquanto a curtose informa se a distribuição apresenta eventos extremos. Essa decomposição permite distinguir, por exemplo, uma célula com volatilidade ampla e simétrica de uma célula com risco concentrado em perdas severas, ou ainda uma célula cuja instabilidade aparece principalmente como saída do formal.

4.4.3 Estatísticas descritivas do painel construído considerando o mercado formal.

Antes de passar aos momentos condicionais, reportam-se as estatísticas descritivas para a base do painel com os trabalhadores no mercado formal por coorte. A Tabela 54 apresenta o tamanho do painel (observações e trabalhadores únicos), o número médio de observações por trabalhador e a fração de $Y = 0$ (ausência de remuneração formal em dezembro). Esses números ajudam a contextualizar a escala do painel e deixam claro que a presença de zeros (dezembro sem remuneração formal) não é marginal, sendo parte importante do objeto empírico.

Tabela 54 – Estatísticas descritivas do painel formal por coorte

Coorte	N obs	N trabalhadores	obs/trabalhador	% $Y = 0$	% admissível
1989	41 774 906	2 364 555	17.667	0.227	0.772
1995	15 957 263	873 607	18.266	0.267	0.732

Fonte: Elaboração Própria com dados da RAIS 1989 - 2017.

A coorte de 1995 apresenta uma maior fração de $Y = 0$ (ausência de remuneração formal em dezembro), consistente com a maior rotatividade/saídas do formal ou maior incidência de vínculos sem pagamento em dezembro. Ao mesmo tempo, o número médio de observações por trabalhador é semelhante entre coortes, sugerindo acompanhamento longitudinal comparável dentro do horizonte comum de carreira.

A Tabela 55 complementa a descrição ao reportar quantis de salários positivos em dezembro (em reais) e quantis de $y = \log(1 + Y)$. Essa tabela busca tornar palpáveis magnitudes típicas de salários no setor formal por coorte e mostrar que a transformação $\log(1 + Y)$ preserva a ordenação e comprime a escala, preparando o terreno para comparações de crescimento residual.

Tabela 55 – Distribuição do salário formal positivo (dezembro) e de $y = \log(1 + Y)$ por coorte

Coorte	N com $Y > 0$	$P10(Y)^*$	$P50(Y)^*$	$P90(Y)^*$	$P10(y)^*$	$P50(y)^*$	$P90(y)^*$
1989	32 285 443	756.973	1 870.680	6 135.318	6.631	7.535	8.722
1995	11 693 282	741.966	1 619.098	4 751.995	6.611	7.390	8.467

Fonte: Elaboração própria com dados da RAIS (1989–2017). **Notas:** Y é o salário real de dezembro no mercado formal (em R\$). $y = \log(1 + Y)$. * Os quantis $P10(\cdot)$, $P50(\cdot)$ e $P90(\cdot)$ são calculados apenas entre observações com salário positivo em dezembro ($Y > 0$); observações com $Y = 0$ (ausência de rendimentos formais em dezembro/ausência no registro naquele dezembro) não entram nesses quantis.

A partir dessas definições, a seção de Resultados explora a superfície de momentos do crescimento salarial residual $\Delta^k y^{res}$ por (τ, RE_bin) , comparando ambas as coortes. Fixou-se um horizonte k e compararam-se estágios de carreira τ (início, meio e fim), observando como a dinâmica muda ao caminhar do RE_bin baixo para o alto. Assim,

evoluiu-se de estatísticas descritivas de nível (Tabelas 54 e 55) para estatísticas descritivas da *dinâmica* (Tabela 51) e, finalmente, para os momentos condicionais inspirados em Guvenen *et al.* (2021) que serão apresentados por fatias da distribuição de *RE*. Na próxima seção, serão discutidos os resultados encontrados para ambas as coortes e como se comportaram ao longo do tempo.

4.5 Resultados

Esta seção apresenta os resultados empíricos discutidos na seção 4.4.1 da metodologia, onde foi adaptada parcialmente a estratégia de Guvenen *et al.* (2021) ao mercado formal brasileiro. O foco agora é comparar as coortes de ingresso em 1989 e 1995 ao longo do ciclo de carreira, tomando como referência o tempo desde a entrada no formal. Como o suporte comum da nossa base vai de $\tau = 0$ a $\tau = 22$, essa é a janela máxima observável para a análise. No caso específico dos resultados com horizonte de cinco anos ($k = 5$), o suporte efetivo termina em $\tau = 17$, pois é preciso observar o lead $\tau + 5$; por isso, o último estágio agregado reportado adiante é 16–17.

Os resultados foram organizados em três blocos. Primeiro, apresenta-se a composição da amostra efetiva. Em seguida, examina-se a margem extensiva, isto é, a probabilidade de o trabalhador não aparecer no registro formal no futuro, capturada por nossa *proxy exit_rate*. Por fim, demonstra-se como os momentos do crescimento dos rendimentos formais variam ao longo da distribuição de rendimentos recentes e ao longo da carreira, destacando as diferenças fundamentais entre as duas coortes.

4.5.1 Composição do painel formal e base efetiva do procedimento

A Tabela 54, apresentada na seção 4.4.1, descreve a base longitudinal no mercado formal para cada coorte. A coorte de 1989 contém 41,8 milhões de observações e 2,36 milhões de trabalhadores únicos, ao passo que a coorte de 1995 contém 15,96 milhões de observações e 873 mil trabalhadores únicos. Em ambos os casos, a quantidade média de observações por trabalhador é elevada, em torno de 18, reforçando que o registro administrativo permite acompanhar trajetórias longas no mercado formal.

Um ponto relevante para a interpretação do restante dos resultados é a parcela de observações com $Y = 0$ em dezembro, dada nossa convenção empírica. Essa fração é de aproximadamente 22,7% na coorte de 1989 e 26,7% na de 1995 (Tabela 54). A Tabela 55 complementa essa leitura ao mostrar a distribuição dos salários positivos. Em termos de níveis, a coorte de 1989 apresenta salários positivos mais elevados ao longo da distribuição, com mediana e percentis superiores aos da coorte de 1995, padrão que também aparece quando se observa $y = \log(1 + Y)$. Esses resultados, por si só, já sugerem que as coortes não diferem apenas em *nível* de rendimentos formais, mas potencialmente também em *dinâmica*, o que motiva a análise dos momentos condicionais do crescimento.

4.5.2 Admissibilidade e RE-elegibilidade: tamanho da amostra utilizada

A estratégia inspirada em Guvenen *et al.* (2021) exige um histórico mínimo para construir *Recent Earnings* (RE) e definir células por posição na distribuição de rendimentos recentes. Por isso, nem todas as observações do painel entram diretamente no cálculo dos momentos.

A Tabela 50 reporta duas estatísticas-chave. A primeira é a fração de observações *admissíveis*, isto é, aquelas com salário de dezembro acima do piso anual. A segunda é a fração *RE-elegível*, que impõe, além da admissibilidade no presente, um histórico mínimo nos períodos anteriores para a construção de RE . A coorte de 1989 apresenta maior participação na base efetiva: cerca de 77,2% das observações são admissíveis, e 46,5% são RE-elegíveis. Na coorte de 1995, esses valores caem para 73,2% e 41,8%, respectivamente.

Esse resultado é importante porque antecipa que a superfície de momentos condicionais é estimada com base em uma amostra efetiva relativamente maior na coorte de 1989. Ao mesmo tempo, a coorte de 1995 apresenta uma maior incidência de observações com remuneração nula em dezembro e menor cobertura no universo RE-elegível. Não se observa isso como um problema; ao contrário, trata-se da comparação entre as duas coortes, já que o objetivo do estudo é justamente entender como elas diferem em um contexto de entrada no formal marcado por cenários macroeconômicos distintos.

4.5.3 Cobertura de células e leitura do número de “grupos”

A estratégia inspirada em Guvenen *et al.* (2021) reorganiza o painel original, com dezenas de milhões de registros por trabalhador e por ano, em células comparáveis definidas por (c, τ, RE_bin, k) : coorte c , tempo de carreira τ , posição na distribuição de *Recent Earnings* (RE_bin) e horizonte de crescimento k . Cada célula descreve um grupo específico de trabalhadores, isto é, indivíduos no mesmo estágio de carreira e na mesma posição relativa de renda recente; dentro dela, calculam-se momentos da distribuição do crescimento dos rendimentos formais, além da margem extensiva medida por *exit_rate*.

A Tabela 56 reporta a cobertura de células após impor um tamanho mínimo por célula ($n \geq 200$), buscando garantir precisão estatística para os momentos. Para $k = 1$, existem 2.200 células potenciais por coorte e foram mantidas 1.900 (86,4% do máximo), totalizando 18,45 milhões de observações válidas na coorte de 1989 e 6,35 milhões na de 1995. Para $k = 5$, há 1.800 células potenciais por coorte e foram mantidas 1.500 (83,3%), com 14,56 milhões e 5,01 milhões de observações válidas, respectivamente.

Tabela 56 – Cobertura de células (τ, RE_bin) por coorte e horizonte k

Horizonte (k)	Coorte	Células mantidas	N (observações)	Células máximas	Proporção mantida
1	1989	1,900	18,450,736	2,200	0.864
1	1995	1,900	6,350,517	2,200	0.864
5	1989	1,500	14,564,898	1,800	0.833
5	1995	1,500	5,013,948	1,800	0.833

Fonte: Elaboração própria com dados da RAIS (1989–2017). **Notas:** “Células máximas” corresponde ao número potencial de combinações (τ, RE_bin) disponíveis no suporte efetivo para cada horizonte k (com $RE_bin \in \{1, \dots, 100\}$). “Células mantidas” conta apenas células com tamanho $n \geq 200$. N é a soma do tamanho n das células mantidas, isto é, o total de observações trabalhador-carreira efetivamente usado no cálculo dos momentos.

Dois pontos ajudam a interpretar esses números. Primeiro, uma célula não é uma observação; cada célula reúne um grande número de indivíduos em (τ, RE_bin) , e por isso o N total permanece elevado, mesmo quando o número de células é relativamente pequeno. Segundo, a queda de cobertura ao alterar o eixo observacional de $k = 1$ para $k = 5$ é esperada, já que horizontes mais longos reduzem o conjunto de τ utilizável e aumentam a chance de ausência do indivíduo no futuro.

Em suma, a Tabela 56 indica que a análise está ancorada em uma malha densa e bem coberta de células ao longo de τ e RE , o que fornece sustentação empírica para examinar, na sequência, como os momentos do crescimento variam sistematicamente ao longo da carreira e ao longo da distribuição de renda recente.

4.5.4 Margem extensiva: ausência no formal no futuro (*exit_rate*)

No desenho empírico, a medida de ausência do trabalhador no mercado de trabalho formal é capturada por *exit_rate*, definido como a fração de indivíduos sem observação no lead $\tau + k$. A Tabela 51 resume *exit_rate* no universo RE-elegível. Mas o ponto mais interessante aparece quando essa probabilidade é desagregada por posição na distribuição de renda recente e por estágio de carreira. A Tabela 57, que reporta o caso $k = 5$, mostra um padrão bastante claro: em ambas as coortes, trabalhadores no *Bottom 10%* de RE exibem maior probabilidade de desaparecer do mercado formal do que trabalhadores no *Top 10%*. Em outras palavras, a permanência no mercado formal é fortemente desigual, e a posição relativa em rendimentos recentes antecipa o risco de saída do registro nos anos seguintes.

Há, contudo, uma diferença importante entre as coortes. Na coorte de 1995, a fragilidade da base da distribuição é mais persistente ao longo da carreira. No estágio inicial ($\tau = 0-7$), o *exit* é muito alto nas duas coortes, cerca de 48,2% em 1989 e 49,5% em 1995 no *Bottom 10%*. No entanto, à medida que a carreira avança, a coorte de 1989 reduz mais fortemente essa probabilidade, enquanto a de 1995 permanece mais vulnerável. No estágio final observável ($\tau = 16-17$), por exemplo, a diferença chega a quase 10 pontos

percentuais no *Top 10%*: 20,7% na coorte de 1989 contra 30,5% na de 1995. Isso sugere que, na coorte de 1995, a fragilidade na margem extensiva persiste por mais tempo, mesmo para trabalhadores relativamente melhor posicionados na distribuição de renda recente.

Esse resultado já prepara o terreno para os próximos blocos. Se a probabilidade de “sumir do formal” varia fortemente com *RE* e com a coorte, então a própria distribuição de $\Delta^k y$ tende a refletir essa heterogeneidade, seja por meio de maior dispersão, seja por meio de maior peso das quedas severas ou ainda por maior incidência de eventos extremos.

Além da heterogeneidade na margem extensiva, a Tabela 57¹⁸ também antecipa padrões importantes na margem intensiva do crescimento dos rendimentos formais em cinco anos. Primeiro, a coluna *sd* mostra que a instabilidade típica do crescimento tende a ser maior nas fatias superiores de *RE*. No início da carreira ($\tau = 0-7$), por exemplo, o desvio-padrão aumenta do *Bottom 10%* para o *Top 10%* em ambas as coortes, passando na coorte de 1989 de 3,454 para 4,276 e na coorte de 1995 de 3,458 para 4,138. Isso sugere que trabalhadores posicionados no topo da distribuição de renda recente enfrentam trajetórias futuras mais heterogêneas, ainda que tenham menor risco de sair do formal.

Segundo, a coluna *skew* é negativa em praticamente todas as células, indicando uma assimetria para baixo, na qual quedas grandes pesam mais do que altas grandes. Aqui também há uma diferença interessante entre as coortes. Na coorte de 1995, especialmente na base da distribuição, a assimetria é menos negativa do que na coorte de 1989. Isso sugere que, em 1995, uma parte importante da fragilidade dos trabalhadores de baixa renda recente aparece menos como grandes quedas condicionais e mais diretamente como desaparecimento do registro formal. Já em 1989, mesmo fora da margem extensiva, as distribuições seguem mais inclinadas para o lado das perdas.

Terceiro, a coluna *kurt* mostra que a principal diferença entre as coortes está concentrada nas caudas da distribuição. Na maior parte das células, a curtose permanece próxima de 1,2 em ambas as coortes. Mas, no topo da distribuição de *RE*, especialmente em estágios intermediários e finais, a coorte de 1989 se destaca fortemente. No estágio 8–15, o *Top 10%* apresenta curtose de 4,503 em 1989 contra 1,191 em 1995. No estágio 16–17, a diferença se torna ainda mais expressiva: 12,873 em 1989 contra 1,149 em 1995. Esse resultado indica que a coorte de 1989, exposta a um ambiente macroeconômico mais instável no início da trajetória, concentra muito mais episódios extremos entre trabalhadores relativamente bem posicionados na distribuição de rendimentos recentes.

¹⁸ A inferência formal sobre a sistematicidade dessas diferenças é apresentada no Apêndice A. Enquanto a Tabela 57 resume os padrões em fatias agregadas de *RE* para $k = 5$, o apêndice testa as diferenças no nível mais desagregado das células da superfície, $g = (\tau, RE_bin, k)$, para $k = 1$ e $k = 5$. Assim, a tabela principal deve ser lida como da magnitude, da direção e da localização das diferenças, enquanto o apêndice avalia se essas diferenças são estatisticamente sistemáticas ao longo da superfície.

Tabela 57 – Momentos do crescimento em 5 anos e margem extensiva por estágio de carreira e fatia de *Recent Earnings* (RE): coortes 1989 e 1995 lado a lado ($k = 5$)

Estágio (τ)	Fatia de RE	Coorte 1989						Coorte 1995						Diferença (1989 – 1995)			
		Células	n	$exit$ (%)	sd	$skew$	$kurt$	Células	n	$exit$ (%)	sd	$skew$	$kurt$	$\Delta exit$ (p.p.)	Δsd	$\Delta skew$	$\Delta kurt$
0–7	Bottom 10%	50	486,883	48.16	3.454	-0.452	1.272	50	156,099	49.51	3.458	-0.119	1.229	-1.35	-0.004	-0.333	0.043
0–7	Middle (P46–55)	50	486,889	41.85	3.851	-0.784	1.209	50	156,093	43.66	3.727	-0.799	1.200	-1.80	0.124	0.015	0.009
0–7	Top 10%	50	486,890	35.64	4.276	-0.830	1.201	50	156,095	35.21	4.138	-0.843	1.195	0.43	0.137	0.014	0.006
8–15	Bottom 10%	80	774,896	44.08	3.484	-0.817	1.217	80	276,766	49.36	3.619	-0.165	1.194	-5.28	-0.135	-0.652	0.022
8–15	Middle (P46–55)	80	774,898	31.85	3.657	-0.865	1.193	80	276,767	38.90	3.814	-0.843	1.186	-7.06	-0.157	-0.022	0.007
8–15	Top 10%	80	774,903	25.98	4.052	-0.885	4.503	80	276,772	29.56	4.166	-0.880	1.191	-3.59	-0.114	-0.005	3.312
16–17	Bottom 10%	20	194,701	42.43	3.569	-0.841	1.190	20	68,531	49.01	3.669	-0.070	1.169	-6.58	-0.101	-0.772	0.021
16–17	Middle (P46–55)	20	194,703	28.17	3.625	-0.870	1.190	20	68,533	36.82	3.830	-0.873	1.166	-8.65	-0.205	0.003	0.025
16–17	Top 10%	20	194,703	20.65	3.870	-0.904	12.873	20	68,533	30.45	4.285	-0.900	1.149	-9.80	-0.414	-0.004	11.724

Fonte: Elaboração própria com dados da RAIS (1989–2017). **Notas:** $exit$ (%) = $100 \times exit_rate$. sd é o desvio-padrão de $\Delta^5 y^{res}$. $skew$ é a assimetria de Kelley, e $kurt$ é a curtose de Crow–Siddiqui. As colunas Δ reportam 1989 – 1995; portanto, valores negativos indicam valores maiores na coorte de 1995. n corresponde ao número de observações utilizadas no cálculo dos momentos, e “Células” corresponde ao número de combinações de (τ, RE_bin) agregadas em cada fatia. Como t_cal vai de 0 a 22, o suporte efetivo do baseline em $k = 5$ termina em 17; por isso, o último estágio agregado é 16–17.

4.5.5 Evidências sobre heterogeneidade no crescimento formal

Antes de explorar as superfícies completas por *RE_bin*, é útil sintetizar, em uma única estatística, o quanto os momentos de crescimento variam dentro de cada estágio de carreira. A Tabela 58 faz exatamente isso; para cada coorte e estágio, reporta a amplitude (máx.–mín.) observada entre as células (τ, RE_bin) pertencentes a cada estágio.

Essa amplitude não captura apenas a desigualdade de *nível* salarial, mas a desigualdade na *dinâmica* do crescimento, ou seja, o quão instável é a trajetória, quão assimétrica é a distribuição, quão relevantes são os extremos e como isso se relaciona à margem extensiva. A leitura geral da tabela é bastante coerente com o restante dos resultados, onde a heterogeneidade em dispersão é relevante nas duas coortes, sugerindo que a instabilidade do crescimento em cinco anos varia ao longo do tempo (τ, RE). A heterogeneidade em assimetria existe, mas aparece com magnitude menor. Já a heterogeneidade em caudas se destaca mais claramente, sobretudo na coorte de 1989, onde os eventos extremos se tornam mais concentrados em segmentos específicos da distribuição de renda recente.

Também vale notar que, como no caso $k = 5$, o suporte efetivo termina em $\tau = 17$; assim, o estágio final relevante nesta parte da análise é 16–17. Essa correção é importante para manter a consistência entre o texto, as figuras e a própria construção da amostra utilizada no cálculo dos momentos.

Tabela 58 – Heterogeneidade ao longo de *RE_bin*. Amplitude (máx.–mín.) dos momentos por estágio de carreira ($k = 5$)

Coorte	Estágio (τ)	Células	<i>sd_range</i>	<i>skew_range</i>	<i>kurt_range</i>	<i>exit_range</i> (p.p.)
1989	0–7	150	1.317	1.666	0.163	20.4
1989	8–15	240	1.163	0.184	11.375	28.6
1989	16–17	60	0.897	0.144	15.081	29.8
1995	0–7	150	1.272	1.762	0.165	24.4
1995	8–15	240	1.259	1.824	0.083	29.3
1995	16–17	60	1.215	1.812	0.053	27.6

Fonte: Elaboração própria com dados da RAIS (1989–2017). **Notas:** como o maior valor observável de *t_cal* é 22, o suporte efetivo para $k=5$ termina em 17; por isso o último estágio nesta tabela é 16–17.

No próximo bloco, serão apresentadas as figuras por *RE_bin* e serão comparadas as coortes de 1989 e 1995, enfatizando onde elas divergem mais: na margem extensiva, na dispersão do crescimento ou no comportamento das caudas.

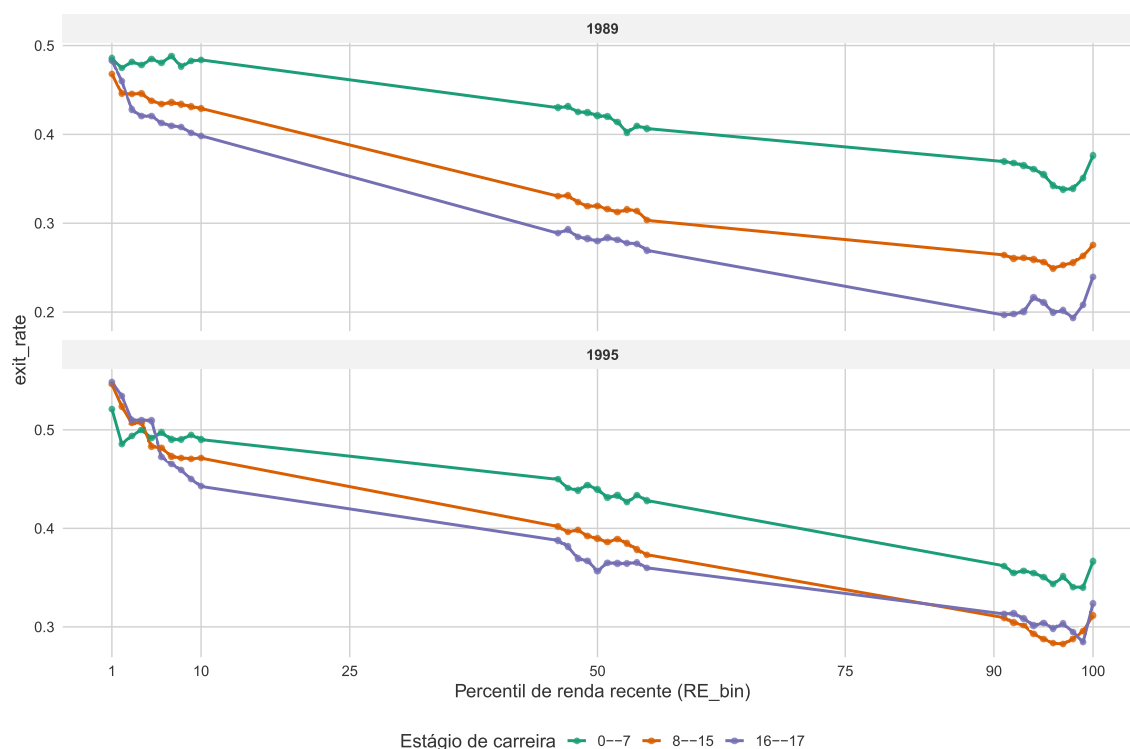
4.5.6 Superfícies de momentos por *RE_bin* e evidência visual da heterogeneidade

A Tabela 58 é um resumo compacto que quantifica o tamanho da heterogeneidade dentro de cada estágio, mas não mostra onde ela se concentra ao longo dos percentis de renda recente. Por isso, a seguir, as superfícies de momentos são analisadas por *RE_bin*,

inspiradas em Guvenen *et al.* (2021). Em todas as figuras, mantém-se $k = 5$ e são plotados os momentos de $\Delta^5 y^{res}$ ao longo de RE_bin , separando estágios de carreira e comparando as duas coortes.

A Figura 27 mostra o gradiente de $exit_rate$ ¹⁹ ao longo de RE_bin . O padrão é consistente com as médias por fatias discutidas anteriormente a partir da Tabela 57: trabalhadores em posições mais baixas de renda recente exibem maior probabilidade de não aparecer na RAIS no horizonte de cinco anos.

Figura 27 – Margem extensiva no horizonte de 5 anos, $exit_rate$ por percentil de renda recente (RE_bin) e estágio de carreira ($k = 5$)



Fonte: Elaboração própria com dados da RAIS (1989–2017).

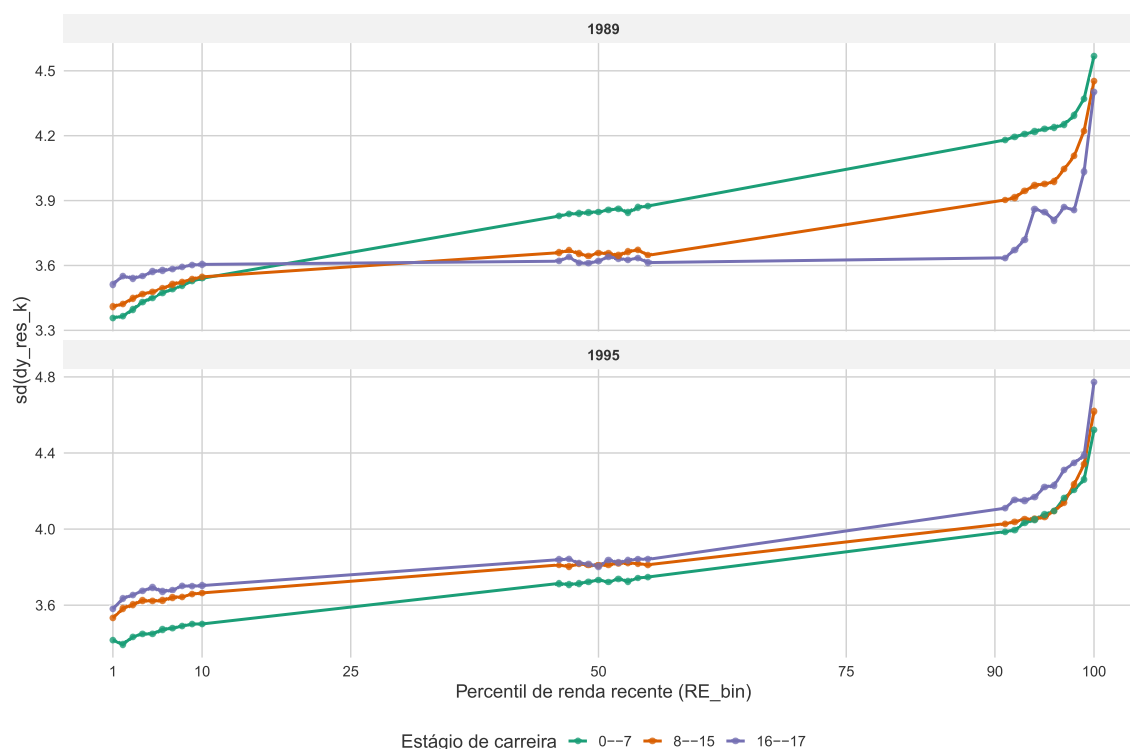
Nota: Linhas representam estágios de carreira e painéis comparam coortes.

A leitura da Figura 27 é bastante direta. Em ambas as coortes, as curvas são decrescentes, deixando claro que quanto maior a posição em RE , menor a probabilidade de saída do formal no horizonte de cinco anos. Ao mesmo tempo, a coorte de 1995 revela maior persistência de fragilidade, sobretudo nos estágios intermediário e final. A base da distribuição continua muito exposta ao risco de não reaparecer no formal, e mesmo no topo essa probabilidade permanece relativamente elevada. Na coorte de 1989, por outro lado, esse gradiente também existe, mas a queda é mais acentuada à medida que se avança na carreira, principalmente nas fatias média e superior.

¹⁹ Para fins de interpretação, $exit_rate$ deve ser entendido como saída do registro formal, e não como ausência de renda total, uma vez que não se acompanha o trabalhador no mercado informal.

Após visualizar quem mais some do formal em cinco anos, passa-se à dispersão do crescimento dos rendimentos. A Figura 28 plota o desvio-padrão de $\Delta^5 y^{res}$ ao longo de RE_bin . A leitura econômica é que trabalhadores com a mesma “idade de carreira” podem enfrentar níveis muito distintos de risco de crescimento, dependendo da posição em RE . Esse resultado dialoga diretamente com a evidência resumida em sd_range na Tabela 58.

Figura 28 – Dispersão do crescimento residual. $sd(\Delta^5 y^{res})$ por RE_bin e estágio de carreira ($k = 5$)



Fonte: Elaboração própria com dados da RAIS (1989–2017).

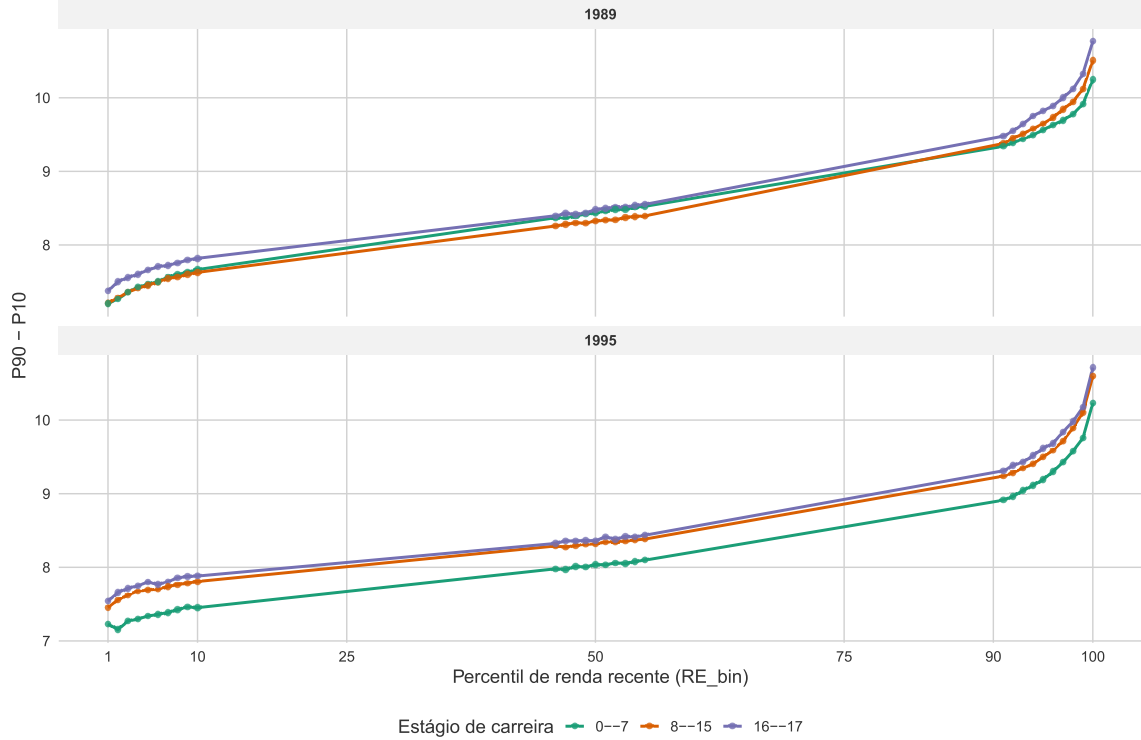
Nota: Linhas representam estágios de carreira e painéis comparam coortes.

Em ambas as coortes, a dispersão é maior nas fatias superiores de RE , o que indica que trabalhadores melhor posicionados em rendimentos recentes enfrentam trajetórias futuras mais abertas e heterogêneas. Mas a comparação entre coortes também revela algo importante: na coorte de 1989, a dispersão é relativamente mais alta no início da carreira, sobretudo no meio e no topo da distribuição; já na coorte de 1995, ela se mantém elevada e, em vários pontos, supera a de 1989 nos estágios intermediário e final. A coorte de 1995 parece mais marcada por uma volatilidade persistente, ao passo que a de 1989 concentra parte importante da heterogeneidade em segmentos mais específicos da superfície.

Para verificar se esse padrão é robusto à influência de poucos valores extremos, a Figura 29 apresenta $P90 - P10$, isto é, a diferença entre o 90º e o 10º percentil de $\Delta^5 y^{res}$. Quando sd e $P90 - P10$ se movem de forma semelhante ao longo de RE_bin , reforça-se a interpretação de que a heterogeneidade observada não decorre apenas de eventos

muito raros nas caudas, mas sim de uma abertura mais generalizada da distribuição do crescimento dos rendimentos.

Figura 29 – Dispersão robusta do crescimento. $P90 - P10$ de $\Delta^5 y^{res}$ por RE_bin ($k = 5$)



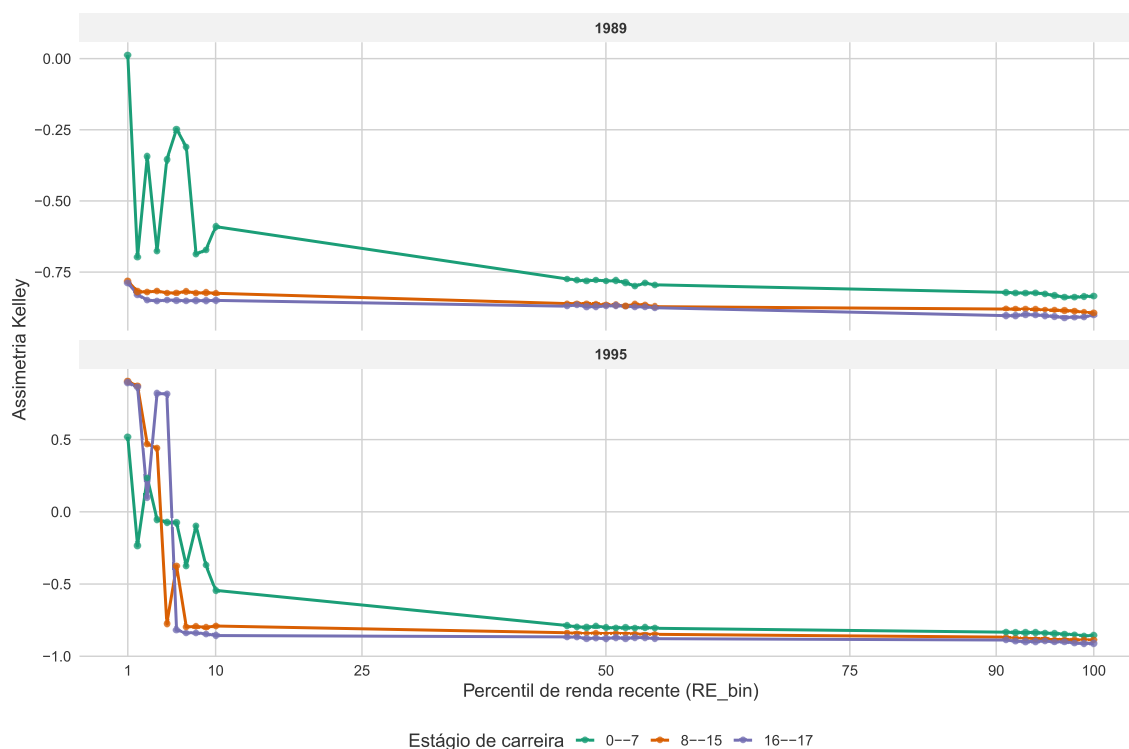
Fonte: Elaboração própria com dados da RAIS (1989–2017).

Nota: Medida baseada em quantis, menos sensível a extremos isolados.

Quando comparadas as medidas de sd e $P90 - P10$, a distribuição do crescimento se abre à medida que se avança para posições mais altas de renda recente. Isso sugere que a maior instabilidade não se resume a poucos casos anômalos, mas reflete uma heterogeneidade mais ampla entre bons e maus resultados futuros. Ao mesmo tempo, em alguns trechos da superfície, a distância entre sd e $P90 - P10$ sugere que a dispersão pode estar sendo puxada por extremos mais severos, o que nos leva naturalmente à leitura das caudas.

A Figura 30 mostra a assimetria de Kelley da distribuição de $\Delta^5 y^{res}$ ao longo de RE_bin . Em termos simples, valores mais negativos indicam que a distribuição do crescimento é “puxada para baixo”: quedas grandes têm mais peso relativo do que altas grandes. No desenho proposto neste artigo, isso é particularmente informativo porque $\Delta^5 y^{res}$ incorpora, além das variações salariais dentro do formal, episódios em que o trabalhador deixa de aparecer na RAIS no lead, tratados como $Y = 0$ no dezembro futuro.

Figura 30 – Assimetria do crescimento residual. Assimetria (Kelley) de $\Delta^5 y^{res}$ por RE_bin e estágio de carreira ($k = 5$)



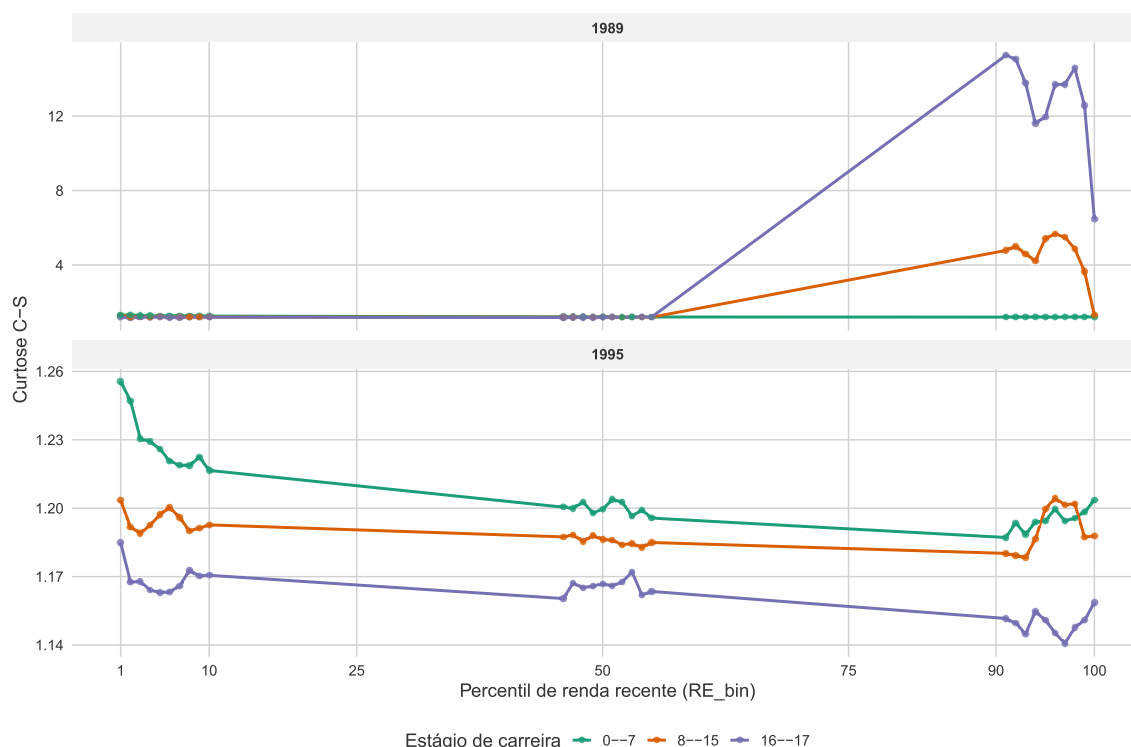
Fonte: Elaboração própria com dados da RAIS (1989–2017).

Nota: Valores mais negativos sugerem maior peso relativo da cauda de quedas.

Essa assimetria é negativa em praticamente toda a superfície. Isso significa que, mesmo quando a trajetória típica é relativamente estável, as perdas severas pesam mais do que os ganhos severos. No entanto, a forma como isso aparece nas duas coortes não é a mesma. Na coorte de 1995, especialmente na base da distribuição, a assimetria é menos negativa, o que é compatível com a ideia de que uma parte relevante do risco se manifesta diretamente como saída do formal. Já na coorte de 1989, sobretudo nos estágios intermediário e final, a distribuição permanece mais inclinada para o lado das quedas, mesmo entre trabalhadores que continuam sendo observados.

Antes de olhar diretamente para a curtose por percentis, vale notar que a principal diferença entre as coortes não está apenas na dispersão, nem apenas na assimetria, mas na intensidade dos eventos extremos em segmentos muito específicos da distribuição. É justamente isso que a Figura 31 permite visualizar.

Figura 31 – Eventos extremos no crescimento residual. Curtose (Crow–Siddiqui) de $\Delta^5 y^{res}$ por RE_bin ($k = 5$)



Fonte: Elaboração própria com dados da RAIS (1989–2017).

Nota: Valores maiores indicam caudas mais espessas, isto é, maior prevalência de extremos na distribuição do crescimento.

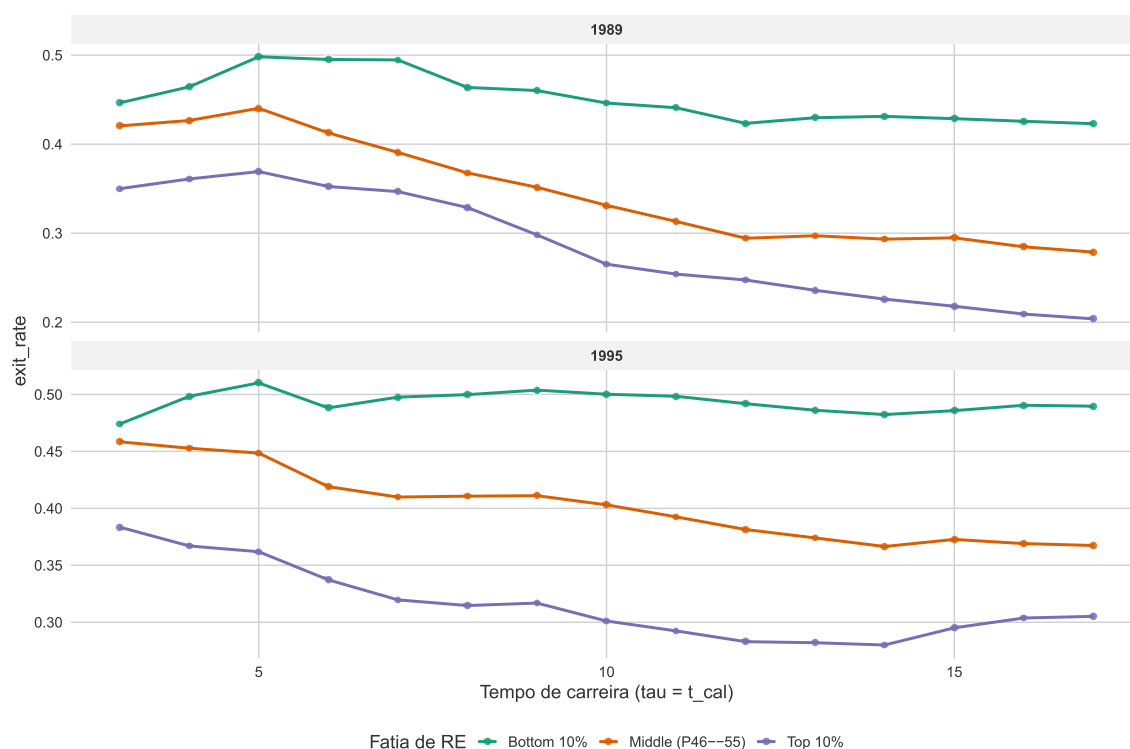
A Figura 31 apresenta a curtose de Crow–Siddiqui, uma medida robusta da espessura das caudas em $\Delta^5 y^{res}$. Em linguagem direta, uma curtose alta significa que, em algum ponto da distribuição, uma fração pequena de trabalhadores experimenta mudanças muito grandes em cinco anos, muito além do comportamento típico do restante da amostra.

É aqui que a diferença entre as coortes se torna mais nítida. Na coorte de 1995, a curtose permanece relativamente baixa e estável ao longo de quase toda a superfície. Já na coorte de 1989, ela sobe de forma abrupta em pontos específicos, sobretudo no topo da distribuição de RE e em estágios mais avançados da carreira. Esse é um resultado importante porque sugere que a coorte exposta ao período macroeconômico mais instável não apenas enfrentou níveis distintos de risco, mas também uma distribuição muito mais sujeita a eventos extremos em certos segmentos da carreira. Assim, a comparação entre as coortes não é um simples deslocamento paralelo, destacando-se a coorte de 1995 pela fragilidade persistente na margem extensiva, enquanto 1989 se destaca mais pela presença localizada de extremos.

A evidência por RE_bin mostra onde esses padrões se concentram dentro de cada estágio. Para completar a leitura do ciclo de vida, é útil olhar a dinâmica ao longo

de τ , fixando fatias interpretáveis de RE . As Figuras 32 e 33 fazem exatamente isso, demonstrando como a margem extensiva e a instabilidade do crescimento evoluem ao longo da carreira para três posições da distribuição de renda recente, usando médias ponderadas pelo tamanho das células.

Figura 32 – Saída do formal ao longo da carreira. *exit_rate* por fatias de RE ao longo de τ ($k = 5$)

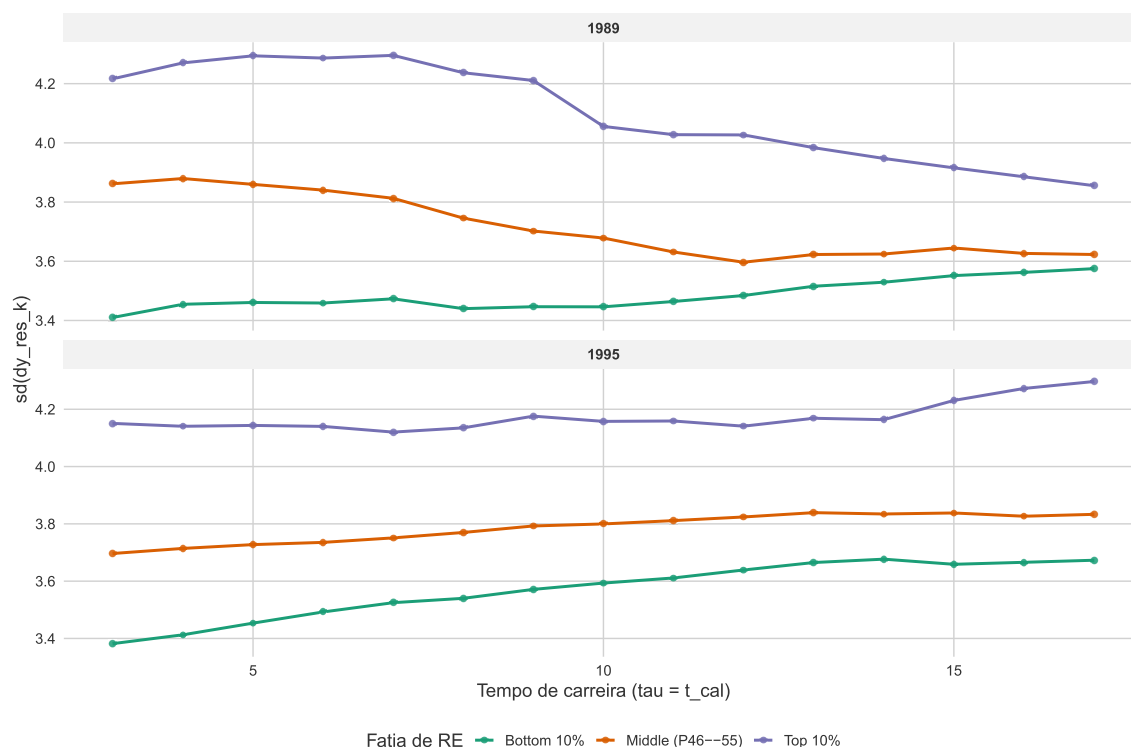


Fonte: Elaboração própria com dados da RAIS (1989–2017).

Nota: Curvas são médias ponderadas por n dentro de cada fatia de RE .

A Figura 32 mostra que a probabilidade de “sumir” do registro formal é estruturalmente maior para trabalhadores com menor renda recente. O *Bottom 10%* aparece sistematicamente acima das demais fatias nas duas coortes. Mas, novamente, a diferença entre 1989 e 1995 importa. Na coorte de 1989, essa probabilidade tende a cair ao longo da carreira, principalmente no meio e no topo. Na de 1995, a fragilidade é mais persistente, sobretudo na base da distribuição, o que reforça a leitura de que essa coorte permanece mais vulnerável na margem extensiva. Já a Figura 33 complementa essa leitura na margem intensiva.

Figura 33 – Instabilidade do crescimento ao longo da carreira. $sd(\Delta^5 y^{res})$ por fatias de RE ao longo de τ ($k = 5$)



Fonte: Elaboração própria com dados da RAIS (1989–2017).

Nota: Curvas são médias ponderadas por n dentro de cada fatia de RE .

Pode-se observar que a dispersão do crescimento também varia ao longo da carreira e conforme a posição em RE . Na coorte de 1989, a volatilidade aparece mais fortemente associada ao topo da distribuição, sobretudo no início e no meio da carreira, mas de forma menos uniforme ao longo do tempo. Na coorte de 1995, a instabilidade tende a permanecer mais elevada à medida que a carreira avança, inclusive em grupos que, a princípio, estariam melhor posicionados na distribuição de renda recente. Em conjunto, as duas figuras mostram que a heterogeneidade não se restringe ao nível salarial, envolvendo também a probabilidade de saída, a dispersão do crescimento e a incidência de extremos.

Em síntese, as Tabelas 57 e 58 e as Figuras 27–33 mostram que a dinâmica do crescimento no mercado formal é fortemente heterogênea ao longo da carreira e ao longo dos rendimentos recentes dos trabalhadores. Essa heterogeneidade aparece simultaneamente na margem extensiva, na dispersão do crescimento e, sobretudo, nas caudas da distribuição. A principal diferença entre as coortes está no fato de que a coorte de 1995 apresenta maior persistência de fragilidade na margem extensiva, especialmente entre trabalhadores com baixa renda recente, enquanto a coorte de 1989 concentra maior incidência de eventos extremos, sobretudo no topo da distribuição e nos estágios mais avançados observáveis da carreira. Para facilitar a interpretação econômica dos resultados e das estimativas, é apresentada na seção apêndice B a tabela 60, onde se apresenta um dicionário de leitura

das diferenças entre coortes.

A conclusão retoma esses padrões para discutir como esses resultados fornecem importantes achados sobre a comparação entre coortes expostas a ambientes macroeconômicos distintos e suas implicações para risco e mobilidade no mercado formal brasileiro.

4.6 Conclusão

Este artigo investigou como a dinâmica dos rendimentos no mercado de trabalho formal brasileiro evoluiu ao longo do ciclo de vida e variou de forma sistemática conforme a posição do trabalhador na distribuição de rendimentos recentes. O principal resultado do artigo é que a dinâmica dos rendimentos no setor formal é fortemente heterogênea, e essa heterogeneidade não aparece apenas no nível salarial. Ela se manifesta também na margem extensiva, isto é, na probabilidade de o trabalhador deixar de aparecer no registro formal no futuro, na dispersão do crescimento dos rendimentos, na assimetria da distribuição e na ocorrência de eventos extremos. Trabalhadores em posições distintas da distribuição de renda recente não enfrentam apenas salários diferentes, mas riscos distintos, trajetórias distintas e formatos diferentes de evolução dos rendimentos ao longo da carreira.

Os resultados mostraram, em primeiro lugar, que a saída do registro formal não é um fenômeno residual. Trabalhadores situados na base dos rendimentos recentes tendem, em geral, a apresentar maior probabilidade de não aparecer no formal no futuro do que aqueles posicionados no topo. Isso sugere que a vulnerabilidade no mercado formal não se expressa apenas em salários mais baixos no presente, mas também em maior instabilidade ocupacional ao longo do tempo.

Em segundo lugar, os resultados indicaram que a heterogeneidade na margem intensiva também é relevante. A dispersão do crescimento em cinco anos, medida tanto pelo desvio-padrão quanto por $P90 - P10$, varia ao longo dos rendimentos recentes do trabalhador e do estágio de carreira em que ele se encontra. De forma geral, observou-se que trabalhadores em posições superiores da distribuição de renda recente tendem a apresentar maior instabilidade do crescimento salarial, sugerindo que trajetórias com rendimentos mais altos no setor formal não são necessariamente mais estáveis, sendo acompanhadas de maior volatilidade e maior exposição a movimentos mais amplos ao longo do horizonte analisado.

Em terceiro lugar, a assimetria negativa encontrada em praticamente toda a superfície de momentos indica que a distribuição do crescimento futuro é sistematicamente inclinada para baixo. Isso significa que perdas severas têm um peso relativamente maior do que ganhos severos. Esse resultado reforça que a dinâmica de rendimentos no formal não pode ser bem resumida apenas por médias e variâncias. Mesmo em segmentos com maior renda recente, o risco não desaparece; ele muda de forma. Em vários casos, o topo

da distribuição de rendimentos recentes combina menor probabilidade de saída do formal com maior volatilidade e distribuições mais assimétricas para baixo.

Por fim, a análise das caudas mostrou que os eventos extremos também variam ao longo da carreira e da distribuição de renda recente. A curtose de Crow–Siddiqui revelou que, em determinados segmentos, sobretudo na coorte de 1989, as caudas do crescimento em cinco anos são substancialmente mais espessas. Esse resultado sugere que a coorte que ingressou no mercado formal em um ambiente macroeconômico mais instável pode ter carregado, em partes específicas da trajetória, uma dinâmica mais sujeita a extremos.

Ainda assim, a comparação entre coortes não aponta para um deslocamento uniforme de 1989 em relação a 1995. As diferenças aparecem de forma localizada, dependendo do estágio de carreira e da posição em rendimentos recentes ocupadas pelo trabalhador. Em alguns trechos da distribuição e da carreira, a coorte de 1989 apresenta maior risco; em outros, a coorte de 1995 se aproxima ou até a supera em determinadas dimensões. Isso recomenda cautela analítica, pois a comparação entre coortes precisa ser feita olhando para a superfície completa de momentos e não apenas para médias agregadas.

Os resultados sugerem que o ambiente macroeconômico vivido no início da trajetória laboral pode deixar marcas persistentes sobre a dinâmica de rendimentos formais. Essas marcas, porém, não parecem operar de maneira simples nem por um único canal. Elas podem aparecer na maior probabilidade de saída do emprego formal, em maior dispersão do crescimento, em maior assimetria para baixo ou em maior incidência de extremos, dependendo do ponto da carreira e da posição relativa do trabalhador na distribuição de rendimentos recentes. Assim, o artigo reforça a ideia de que tratar o risco idiossincrático como homogêneo ao longo do ciclo de vida tende a esconder dimensões importantes da desigualdade dinâmica.

Ao adaptar ao caso brasileiro a lógica empírica de Guvenen *et al.* (2021) aplicada a duas coortes de ingressantes no mercado de trabalho em momentos tão distintos, 1989 e 1995, este trabalho contribuiu documentando e evidenciando como o risco e a desigualdade ocorrem de forma dinâmica e persistente em um contexto de elevada segmentação entre formalidade e informalidade, em um cenário de hiperinflação e estabilidade no ingresso no mercado formal de trabalho. Mais do que diferenças de nível entre coortes, os resultados sugerem diferenças na própria forma como o risco se organiza ao longo da trajetória dos trabalhadores.

Algumas limitações neste estudo precisam ser mencionadas. A primeira delas é que a RAIS observa apenas o mercado formal de trabalho, ou seja, que $Y = 0$ e *exit_rate* devem ser lidos como ausência de rendimentos ou de registro formal em dezembro, e não como ausência de renda total do trabalhador. A segunda limitação é metodológica, pois a imputação de $Y = 0$ no *lead* e o uso da transformação $y = \log(1 + Y)$ são escolhas operacionais que preservam observações e permitem acompanhar a margem extensiva, mas

também afetam a cauda inferior da distribuição do crescimento. Em particular, parte da forte assimetria negativa e dos eventos extremos pode refletir justamente transições para fora do mercado formal. A terceira limitação diz respeito à construção de rendimentos recentes *RE*, pois a inclusão de zeros nos *lags*, o uso de um piso anual aproximado a partir do salário mínimo e a própria definição de elegibilidade podem influenciar os resultados, especialmente para trabalhadores com trajetórias intermitentes no mercado formal. Por fim, trata-se de uma análise descritiva e não causal. Assim, a comparação entre as duas coortes não permite atribuir as diferenças observadas a um mecanismo único, como inflação, estabilização monetária, mudanças institucionais ou transformações setoriais.

Embora seja importante destacar tais limitações, ao mesmo tempo, vale a pena apresentar caminhos promissores para pesquisas futuras. Um primeiro passo seria aproximar a análise do mercado de trabalho total, articulando os registros formais com bases que permitam observar, ainda que parcialmente, a transição para o setor informal. O CAD único poderia auxiliar nesse contexto. Um segundo avanço possível seria explorar a heterogeneidade dos resultados por sexo, escolaridade, setor de atividade, região, ocupação e tipo de firma, verificando se os padrões documentados aqui se concentram em grupos específicos. Também seria útil testar a robustez dos resultados a definições alternativas de *RE*, do piso de admissibilidade e da própria imputação de rendimentos nulos no *lead*.

Referências do capítulo

BANCO CENTRAL DO BRASIL. **30 anos do Plano Real**. [S. l.: s. n.], 2024.
<https://www.bcb.gov.br/controleinflacao/30anosreal>. Acesso em: 11 ago. 2025.

BANCO CENTRAL DO BRASIL. **Plano Real**. [S. l.: s. n.], 2024.
<https://www.bcb.gov.br/controleinflacao/planoreal>. Acesso em: 11 ago. 2025.

BANCO CENTRAL DO BRASIL. **Planos econômicos fracassaram em derrotar a superinflação até a chegada do Real**. [S. l.: s. n.], 2019. Publicado em 28/06/2019 e atualizado em 28/08/2019. Acessado em: 27 jul. 2025. Disponível em:
<https://www.bcb.gov.br/detalhenoticia/355/noticia>.

BANCO CENTRAL DO BRASIL. **SGS – Sistema Gerenciador de Séries Temporais: Taxas de Juros (TBC, TBAN, Meta Selic)**. Período de 01/07/1996 a 31/12/2017. Versão 2.1. Acesso em: 8 nov. 2025. 2025. Disponível em:
<https://www3.bcb.gov.br/sgspub/consultarvalores/consultarValoresSeries.do?method=consultarGrafico>.

BENJAMINI, Yoav; HOCHBERG, Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 57, n. 1, p. 289–300, 1995. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

CARD, David. **Earnings, Schooling, and Ability Revisited**. [S. l.], 1994. Acessado em 21 de outubro de 2025. DOI: 10.3386/w4832. Disponível em: <https://doi.org/10.3386/w4832>.

CARD, David. The Causal Effect of Education on Earnings. *In*: ASHENFELTER, Orley; CARD, David (ed.). **Handbook of Labor Economics**. [S. l.]: Elsevier, 1999. 3A. p. 1801–1863. DOI: 10.1016/S1573-4463(99)03011-4.

CHEN, Jiafeng; ROTH, Jonathan. Logs with Zeros? Some Problems and Solutions. **The Quarterly Journal of Economics**, v. 139, n. 2, p. 891–936, 2024. DOI: 10.1093/qje/qjad054. Disponível em: <https://doi.org/10.1093/qje/qjad054>.

CROW, Edwin L.; SIDDIQUI, M. M. Robust Estimation of Location. **Journal of the American Statistical Association**, v. 62, n. 318, p. 353–389, 1967. DOI: 10.1080/01621459.1967.10482914.

DVORKIN, Maximiliano A.; MARKS, Cassandra. Nominal Wage Adjustments during High Inflation and Tight Labor Markets. **Federal Reserve Bank of St. Louis Review**, Federal Reserve Bank of St. Louis, v. 106, n. 10, p. 1–19, 2024. Fourth Quarter. DOI: 10.20955/r.2024.10. Disponível em: <https://doi.org/10.20955/r.2024.10>.

ENGBOM, Niklas; MOSER, Christian. Earnings Inequality and the Minimum Wage: Evidence from Brazil. **American Economic Review**, v. 112, n. 12, p. 3803–3847, 2022. DOI: 10.1257/aer.20181506. Disponível em: <https://doi.org/10.1257/aer.20181506>.

FREGUGLIA, Ricardo; PROCÓPIO, Thais. Efeitos da mudança de emprego e da migração interestadual sobre os salários no Brasil formal: evidências a partir de dados em painel. **Pesquisa e Planejamento Econômico**, v. 43, n. 2, p. 255–278, 2013. Ipea.

GAMBETTI, Luca; MESSINA, Julián. Evolução da Ciclicidade Salarial na América Latina. **The World Bank Economic Review**, v. 32, n. 3, p. 709–726, 2018.

GUVENEN, Fatih *et al.* What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics? **Econometrica**, v. 89, n. 5, p. 2303–2339, 2021. Acessado em 21 de outubro de 2025. DOI: 10.3982/ECTA14603. Disponível em: <https://www.jstor.org/stable/44250423>.

HECKMAN, James J.; LOCHNER, Lance J.; TODD, Petra E. **Fifty Years of Mincer Earnings Regressions**. [S. l.], 2003. Acessado em 21 de outubro de 2025. DOI: 10.3386/w9732. Disponível em: <https://doi.org/10.3386/w9732>.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Tabela 1737: IPCA – Série histórica com número-índice, variação mensal e acumulados (3 meses, 6 meses, ano, 12 meses)**. [S. l.: s. n.], 2025. SIDRA – Sistema IBGE de Recuperação

Automática. Disponível em: <https://sidra.ibge.gov.br/tabela/1737> (acessado em 10 ago. 2025).

KELLEY, Truman Lee. **Fundamentals of Statistics**. Cambridge, MA: Harvard University Press, 1947.

KIM, Tae-Hwan; WHITE, Halbert. On More Robust Estimation of Skewness and Kurtosis. **Finance Research Letters**, v. 1, n. 1, p. 56–73, 2004. DOI: 10.1016/S1544-6123(03)00003-5.

LEMIEUX, Thomas. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? **The American Economic Review**, American Economic Association, v. 96, n. 3, p. 461–498, 2006. Accessed: 8 November 2025. Disponível em: <http://www.jstor.org/stable/30034058>.

MELLO, João M. P. de; WAISMAN, Caio; ZILBERMAN, Eduardo. Os efeitos da exposição à hiperinflação na escolha ocupacional. **Journal of Economic Behavior & Organization**, v. 106, p. 109–123, 2014.

APÊNDICE

Apêndice A - Testes de Robustez: significância estatística das diferenças entre coortes nas superfícies de momentos

Este apêndice apresenta um exercício de robustez para avaliar se as diferenças entre as coortes de 1989 e 1995, documentadas nas tabelas e figuras de superfícies de momentos, são estatisticamente sistemáticas. A Tabela 57, por exemplo, mostra diferenças relevantes entre coortes no horizonte de cinco anos para fatias agregadas de rendimentos recentes e estágios de carreira. O exercício deste apêndice complementa essa evidência ao testar se as diferenças persistem no nível mais desagregado das células da superfície, definidas por (τ, RE_bin, k) . A motivação para esse procedimento decorre do fato de que a base utilizada é uma base administrativa de grande escala. Nesse contexto, testes convencionais aplicados diretamente às observações individuais trabalhador-ano tendem a rejeitar a hipótese nula mesmo diante de diferenças economicamente pequenas. Por isso, a inferência é conduzida no nível das células da superfície, que são o objeto empírico central da análise.

Definiu-se uma célula da superfície como

$$g \equiv (\tau, RE_bin, k),$$

em que τ denota o estágio de carreira, RE_bin representa a posição do trabalhador na distribuição de rendimentos recentes e k é o horizonte de crescimento. Para cada célula g e para cada momento m , calculou-se a diferença entre as coortes como

$$d_g^m = \hat{m}_{g,1989} - \hat{m}_{g,1995}.$$

Os momentos considerados são a taxa de saída do registro formal, $exit_rate$, o desvio-padrão do crescimento residual, $sd(\Delta^k y^{res})$, a dispersão robusta $P90 - P10$, a assimetria de Kelley e a curtose de Crow-Siddiqui.

O teste de significância avalia se a diferença média entre as coortes ao longo das células da superfície é estatisticamente distinta de zero. A hipótese nula é

$$H_0 : E[d_g^m] = 0,$$

contra a alternativa bicaudal

$$H_1 : E[d_g^m] \neq 0.$$

Assim, o exercício não testa diferenças médias nas observações individuais, mas sim se as diferenças entre coortes são sistemáticas ao longo da superfície estimada.

Também foram reportadas duas medidas agregadas. A primeira é a diferença média simples entre células:

$$\bar{d}^m = \frac{1}{G} \sum_{g=1}^G d_g^m,$$

em que G é o número de células válidas da superfície. Essa medida atribui o mesmo peso a cada célula e responde se a diferença entre coortes aparece de forma sistemática ao longo da superfície.

A segunda medida é a diferença média ponderada pelo tamanho das células:

$$\bar{d}_w^m = \frac{\sum_{g=1}^G w_g d_g^m}{\sum_{g=1}^G w_g}, \quad w_g = n_{g,1989} + n_{g,1995},$$

em que $n_{g,c}$ é o número de observações da coorte c na célula g . A versão ponderada dá maior peso às regiões da superfície estimadas com maior suporte amostral.²⁰

A inferência é obtida por *bootstrap* com reamostragem das células da superfície. Em cada réplica $b = 1, \dots, B$, foram reamostradas com reposição as células g e recalculadas a média simples e a média ponderada:

$$\bar{d}^{m,(b)} = \frac{1}{G} \sum_{g \in \mathcal{G}^{(b)}} d_g^m,$$

e

$$\bar{d}_w^{m,(b)} = \frac{\sum_{g \in \mathcal{G}^{(b)}} w_g d_g^m}{\sum_{g \in \mathcal{G}^{(b)}} w_g}.$$

Os intervalos de confiança de 95% são construídos a partir dos percentis 2,5% e 97,5% da distribuição *bootstrap*. Também se reportam p -valores bicaudais baseados na distribuição *bootstrap* ponderada.

Como são avaliados simultaneamente diferentes momentos da distribuição e horizontes de crescimento, os (p)-valores obtidos pelo *bootstrap* ponderado foram ajustados pelo procedimento de Benjamini e Hochberg (1995). O ajuste foi realizado conjuntamente sobre a família de testes formada pelas combinações entre momentos e horizontes analisados. Esse procedimento controla a taxa de falsas descobertas (*false discovery rate*, FDR), isto é, a proporção esperada de rejeições incorretas entre as hipóteses rejeitadas, reduzindo o risco de atribuir significância estatística a padrões decorrentes apenas da multiplicidade de comparações. Nas tabelas, os valores resultantes são reportados como (q)-valores BH, e as estrelas de significância são definidas a partir desses valores ajustados.

A interpretação dos sinais segue diretamente da definição $d_g^m = \hat{m}_{g,1989} - \hat{m}_{g,1995}$. Para *exit_rate*, valores negativos indicam maior saída do registro formal na coorte de 1995. Para $sd(\Delta^k y^{res})$ e $P90 - P10$, valores negativos indicam maior dispersão do crescimento residual na coorte de 1995. Para a assimetria, valores negativos indicam que a assimetria é menor na coorte de 1989. Quando as assimetrias são negativas, isso significa que a distribuição da coorte de 1989 está mais inclinada para a cauda esquerda; isto é, apresenta maior peso relativo de quedas severas. Para a curtose, valores positivos indicam

²⁰ A comparação entre a média simples e a média ponderada é útil porque permite verificar se os resultados são conduzidos por células pequenas ou se permanecem quando as regiões mais densas da superfície recebem maior peso.

caudas mais espessas na coorte de 1989, sugerindo maior incidência de eventos extremos nessa coorte.

A Tabela 59 apresenta os resultados do exercício. A coluna “Dif. média” reporta $\bar{d}^m$, enquanto a coluna “Dif. média pond.” reporta $\bar{d}_w^m$. O intervalo de confiança é construído por bootstrap com reamostragem das células, e a coluna q -valor BH apresenta o p -valor ajustado para múltiplos testes. A significância estatística deve ser interpretada como evidência de que as diferenças entre coortes são sistemáticas ao longo das células da superfície. Ela não substitui a interpretação substantiva baseada na magnitude, no sinal e na localização das diferenças ao longo de τ e de RE_bin , nem deve ser interpretada isoladamente como evidência causal.

Tabela 59 – Significância estatística das diferenças entre coortes nas superfícies de momentos

Horizonte	Momento	Células	Dif. média	Dif. média pond.	IC 95% bootstrap	p-valor	q-valor BH
$k = 1$	Exit rate	1.900	-0.0289	-0.0288***	[-0.0297; -0.0278]	0.002	0.002
$k = 1$	Desvio-padrão	1.900	-0.1580	-0.1572***	[-0.1664; -0.1477]	0.002	0.002
$k = 1$	P90-P10	1.900	-1.0679	-1.0698***	[-1.1894; -0.9544]	0.002	0.002
$k = 1$	Assimetria de Kelley	1.900	0.0741	0.0740***	[0.0674; 0.0811]	0.002	0.002
$k = 1$	Curtose Crow-Siddiqui	1.900	4.1691	4.2295***	[3.8845; 4.5832]	0.002	0.002
$k = 5$	Exit rate	1.500	-0.0429	-0.0424***	[-0.0443; -0.0405]	0.002	0.002
$k = 5$	Desvio-padrão	1.500	-0.0609	-0.0618***	[-0.0694; -0.0538]	0.002	0.002
$k = 5$	P90-P10	1.500	0.1262	0.1222***	[0.1119; 0.1339]	0.002	0.002
$k = 5$	Assimetria de Kelley	1.500	-0.0626	-0.0631***	[-0.0816; -0.0464]	0.002	0.002
$k = 5$	Curtose Crow-Siddiqui	1.500	1.0194	1.0224***	[0.8882; 1.1585]	0.002	0.002

Fonte: Elaboração Própria com dados da RAIS 1989-2017.

Nota: As estrelas são baseadas no q -valor ajustado: *** $q < 0,01$, ** $q < 0,05$, * $q < 0,10$. Como foram usadas 1.000 réplicas *bootstrap* com correção finita no cálculo do p -valor, valores iguais a 0,002 correspondem ao menor p -valor reportável neste exercício.

Os resultados da Tabela 59²¹ indicam que as diferenças entre as coortes são estatisticamente sistemáticas ao longo das células da superfície. Em todos os momentos e nos dois horizontes considerados, os intervalos de confiança *bootstrap* não incluem zero, e os q -valores ajustados pelo procedimento de Benjamini-Hochberg permanecem abaixo de 1%. Isso sugere que os padrões documentados nas figuras e na Tabela 57 não decorrem de flutuações isoladas em poucas células, mas refletem diferenças persistentes entre as superfícies de momentos das coortes de 1989 e 1995.

No horizonte $k = 5$, que é o foco da Tabela 57, a coorte de 1995 apresenta maior saída do formal e maior desvio-padrão médio ponderado, reforçando a interpretação de maior fragilidade na margem extensiva e maior volatilidade dos rendimentos. Por outro lado, a curtose permanece maior na coorte de 1989, indicando maior presença de eventos

²¹ Cabe notar que a Tabela 57 resume a evidência para $k = 5$ em fatias agregadas de RE , enquanto a Tabela 59 utiliza todas as células disponíveis da superfície para $k = 1$ e $k = 5$. Assim, o exercício de significância não replica apenas a tabela agregada, mas testa a sistematicidade das diferenças em uma malha mais desagregada de percentis de rendimentos recentes e estágios de carreira.

extremos nessa coorte. Assim, o apêndice confirma estatisticamente a leitura das tabelas principais: 1995 se destaca por maior saída do formal e maior instabilidade média, enquanto 1989 concentra caudas mais espessas e eventos extremos localizados.

Apêndice B - Dicionário de interpretação dos sinais e momentos

Para facilitar a interpretação econômica dos sinais estimados, a Tabela 60 apresenta um dicionário de leitura das diferenças entre coortes. Como todas as diferenças são definidas como 1989 – 1995, um sinal negativo indica que o respectivo momento é maior na coorte de 1995, enquanto um sinal positivo indica que ele é maior na coorte de 1989. Essa convenção é especialmente importante para os momentos de ordem superior, pois o sinal da diferença em assimetria e curtose não tem a mesma interpretação direta de uma diferença em médias ou taxas.

A leitura conjunta mostra que, no curto prazo ($k = 1$), a coorte de 1995 apresenta maior saída do formal, maior dispersão e maior abertura robusta da distribuição do crescimento. Isso sugere uma maior instabilidade generalizada no curto prazo. Ao mesmo tempo, a curtose positiva em favor de 1989 indica que os eventos mais extremos são mais concentrados nessa coorte. Assim, o contraste de curto prazo combina maior fragilidade ampla em 1995 com caudas mais extremas em 1989.

No horizonte de cinco anos ($k = 5$), a maior saída do formal e o maior desvio-padrão em 1995 permanecem, reforçando a interpretação de fragilidade persistente dessa coorte. No entanto, $P90 - P10$, assimetria e curtose passam a apontar maior peso de caudas e maior abertura robusta em 1989. Isso sugere que, no horizonte mais longo, a coorte de 1989 concentra mais eventos extremos e maior risco de quedas severas em segmentos específicos da superfície, enquanto a coorte de 1995 se destaca principalmente pela maior saída do formal e maior volatilidade média.

Tabela 60 – Dicionário de leitura dos sinais nas diferenças entre coortes

Horizonte	Momento	Sinal estimado	Interpretação econômica
$k = 1$	<i>exit_rate</i>	Negativo	A saída do registro formal é maior na coorte de 1995. Como a diferença é 1989–1995, $d < 0$ indica maior fragilidade na margem extensiva para 1995.
$k = 1$	Desvio-padrão	Negativo	A dispersão do crescimento residual é maior em 1995. Isso indica maior volatilidade dos rendimentos no curto prazo.
$k = 1$	$P90 - P10$	Negativo	A dispersão robusta é maior em 1995. A distância entre bons e maus resultados de crescimento é maior nessa coorte.
$k = 1$	Assimetria de Kelley	Positivo	A assimetria é relativamente maior em 1989. Se a assimetria é negativa em ambas as coortes, esse sinal sugere que 1995 tem distribuição mais inclinada para quedas no curto prazo.
$k = 1$	Curtose Crow–Siddiqui	Positivo	As caudas são mais espessas em 1989. Isso indica maior presença de eventos extremos nessa coorte no horizonte de um ano.
$k = 5$	<i>exit_rate</i>	Negativo	A saída do registro formal é maior em 1995 também no horizonte de cinco anos. A fragilidade na permanência formal é persistente.
$k = 5$	Desvio-padrão	Negativo	A dispersão do crescimento residual é maior em 1995. A coorte de 1995 apresenta trajetórias futuras mais voláteis.
$k = 5$	$P90 - P10$	Positivo	A dispersão robusta é maior em 1989. No horizonte de cinco anos, a distância entre o 90 ^o e o 10 ^o percentil do crescimento é maior nessa coorte.
$k = 5$	Assimetria de Kelley	Negativo	A assimetria é mais baixa em 1989. Quando os valores são negativos, isso indica maior inclinação para a cauda esquerda, isto é, maior peso relativo de quedas severas em 1989.
$k = 5$	Curtose Crow–Siddiqui	Positivo	As caudas são mais espessas em 1989. Isso sugere maior incidência de eventos extremos no crescimento dos rendimentos no horizonte de cinco anos.

Notas: A diferença é definida como 1989 – 1995. Portanto, sinais negativos indicam que o momento é maior na coorte de 1995, enquanto sinais positivos indicam que o momento é maior na coorte de 1989. Para *exit_rate*, valores maiores significam maior saída do registro formal. Para *sd* e $P90 - P10$, valores maiores indicam maior dispersão do crescimento. Para a assimetria, a interpretação depende do nível da medida: valores mais negativos indicam maior peso relativo da cauda de quedas. Para a curtose, valores maiores indicam caudas mais espessas, isto é, maior incidência de eventos extremos.

5 Conclusão

Esta tese teve como objetivo investigar, ao longo de três ensaios, a presença de efeitos heterogêneos na educação e no mercado de trabalho brasileiro, explorando como fatores familiares, institucionais e macroeconômicos moldam trajetórias individuais ao longo do ciclo de vida. Partindo da literatura de capital humano e mobilidade intergeracional, apresentada no Capítulo II, o estudo tratou a educação e o mercado de trabalho como dimensões complementares e interdependentes, fundamentais para a compreensão da persistência das desigualdades e das oportunidades econômicas no país.

A partir desse enquadramento, os capítulos da tese examinaram, em sequência, como o contexto familiar influencia o desempenho escolar, como fricções de acesso, em particular o deslocamento residência–escola, afetam o aprendizado em grandes centros urbanos e, por fim, como a dinâmica de rendimentos no mercado de trabalho formal evolui ao longo do ciclo de vida sob diferentes contextos macroeconômicos. Em conjunto, os resultados mostram que a heterogeneidade relevante não se manifesta apenas em níveis médios de desempenho ou rendimentos, mas também na forma como riscos, assimetrias e eventos extremos se distribuem entre indivíduos, gerações e estágios da vida.

O Capítulo II (*Tal pai, tal filho? Um ensaio sobre a ocupação dos pais e o desempenho acadêmico de seus filhos*) examinou como o contexto familiar se traduz em diferenças sistemáticas de desempenho escolar já no início da trajetória educacional. A principal contribuição do capítulo é mostrar que a ocupação dos pais, em particular a inserção em ocupações gerenciais e profissionais, aqui tratadas como colarinho branco, está associada a uma vantagem persistente no desempenho em Matemática e Língua Portuguesa, mesmo após controlar pela escolaridade dos responsáveis, tipo de escola e medidas de engajamento familiar, além de testar efeitos fixos de escola. A evidência aponta, portanto, para um padrão de heterogeneidade que não é apenas *entre* escolas, mas também *dentro* de contextos escolares, consistente com a ideia de que a transmissão intergeracional opera por mecanismos diversos, seja via recursos materiais, seja via práticas e repertórios familiares que afetam o processo de aprendizagem.

Do ponto de vista empírico, o capítulo documenta que estudantes com ambos os pais em ocupações de colarinho branco apresentam desempenho superior na avaliação diagnóstica do início da alfabetização, e esse diferencial se mantém quando se exploram as estruturas em painel do GERES ao longo das cinco ondas. Na análise transversal, a vantagem associada à condição de ambos os pais em colarinho branco é positiva nas duas disciplinas e, no painel, o mesmo padrão se preserva; a desvantagem de não possuir nenhum dos pais nessa condição é igualmente robusta. Além disso, a análise evidencia um gradiente hierárquico na escolaridade dos responsáveis, estando níveis mais elevados de escolaridade materna e paterna associados a ganhos progressivos no desempenho, com

magnitudes particularmente relevantes para a escolaridade materna, sugerindo que dimensões de presença e acompanhamento podem mediar parte do efeito observado. Em complemento, o engajamento escolar mostra um contraste, com idas à escola associadas a reuniões e atividades comunitárias se correlacionando com melhores resultados, enquanto idas motivadas por conversas com docentes aparecem com sinal negativo, interpretação compatível com visitas reativas a dificuldades de aprendizagem.

Um aspecto adicional enfatizado no Capítulo II é que as desigualdades se expressam não apenas em médias, mas também na probabilidade de ocupar as extremidades da distribuição de desempenho. As estimações de modelos não lineares indicam que a ausência de pais em ocupações de colarinho branco está associada a uma maior chance de figurar nos quantis inferiores e a uma menor chance de estar nos quantis superiores em Matemática e Língua Portuguesa, reforçando a leitura de que o contexto familiar influencia tanto o nível quanto o posicionamento relativo do estudante. Embora o capítulo esteja ancorado em um conjunto de municípios e em um período específico do sistema educacional brasileiro, os resultados são consistentes em diferentes especificações e convergem para uma implicação central, onde diferenças de origem familiar se transformam, cedo, em diferenças de desempenho mensurável, com potencial para se acumular ao longo do ciclo escolar.

O Capítulo III (*Distância importa? Impactos da Proximidade Escolar no Desempenho Acadêmico dos Alunos*) avançou para um mecanismo distinto, mas complementar, estudando fricções de acesso à escola em grandes centros urbanos. A contribuição central do capítulo é mostrar que a distância entre residência e escola não é um mero detalhe logístico, mas uma dimensão relevante de desigualdade educacional, capaz de gerar perdas de aprendizagem. O capítulo evidencia que estimativas por mínimos quadrados ordinários sugerem um efeito pequeno e positivo da distância, padrão compatível com seleção (famílias com maior capital socioeconômico escolhendo escolas melhores e mais distantes). Quando se isolou a variação exógena da distância induzida pela política municipal de subsídio ao transporte escolar, o sinal se inverte, passando a distância a ter efeito negativo e estatisticamente significativo sobre o desempenho em Matemática e Língua Portuguesa.

A interpretação econômica do Capítulo III é que a política de passe estudantil amplia o raio de escolha e altera o *trade-off* enfrentado pelas famílias, permitindo deslocamentos maiores; porém, uma vez isolada essa variação, distâncias mais longas se traduzem em perdas de aprendizado. Em termos de magnitude, as estimativas indicam quedas relevantes nos escores quando a distância aumenta, e os resultados em escala padronizada reforçam esse diagnóstico, com um aumento de um desvio-padrão na distância estando associado a reduções substanciais no desempenho em ambas as disciplinas. Os testes de diagnóstico corroboram a necessidade do estimador por variáveis instrumentais, rejeitando a exogeneidade da distância e indicando um instrumento forte. Assim, o capítulo sustenta a leitura de que a distância atua como uma barreira efetiva à aprendizagem,

particularmente em contextos escolares e familiares já desfavorecidos, e que decisões de escolha escolar podem incorporar custos de deslocamento que não são plenamente internalizados pelas famílias.

Por fim, o Capítulo III destaca que os resultados devem ser lidos na lógica de um efeito local, pois a identificação captura o impacto da distância para o subconjunto de estudantes cuja acessibilidade escolar responde à introdução do subsídio municipal ao transporte. Ainda assim, em grandes centros urbanos, fricções espaciais e de mobilidade têm um impacto educacional relevante. Em conjunto, os resultados dos Capítulos II e III reforçam o argumento mais amplo da tese, de que desigualdades de origem e desigualdades de acesso se combinam e se retroalimentam, moldando diferenças de desempenho já na base do ciclo de vida, com potencial para influenciar oportunidades futuras no mercado de trabalho.

O Capítulo IV, intitulado *Exposição à hiperinflação e ciclo de vida laboral: evidências de longo prazo sobre risco, rendimentos e permanência no mercado formal*, deslocou o foco da análise para a fase de inserção no mercado de trabalho e mostrou que a dinâmica dos rendimentos no mercado formal brasileiro é fortemente heterogênea ao longo do ciclo de vida, variando sistematicamente conforme a posição do trabalhador na distribuição de rendimentos recentes. Mesmo quando se comparam indivíduos em estágios semelhantes de carreira e em posições próximas nessa distribuição, observam-se importantes diferenças na evolução futura dos rendimentos. A desigualdade relevante, portanto, não se restringe ao nível salarial, manifestando-se também no risco associado às trajetórias e no próprio formato da distribuição do crescimento.

Na margem extensiva, os resultados indicam que a probabilidade de o trabalhador deixar de aparecer no registro formal cresce de forma expressiva à medida que o horizonte temporal considerado se amplia, sendo sistematicamente maior entre os trabalhadores localizados nas faixas inferiores da distribuição de rendimentos recentes, sobretudo nos estágios iniciais e intermediários da carreira. Em um mercado de trabalho segmentado entre formalidade e informalidade, esse resultado sugere que uma parcela importante do risco ao longo do ciclo de vida se materializa como intermitência e saída do registro formal, e não apenas como variação salarial entre aqueles que permanecem empregados formalmente.

A comparação entre as coortes de 1989 e 1995 mostrou, ainda, que as diferenças entre gerações não podem ser resumidas por simples deslocamentos médios. As divergências aparecem de forma localizada, variando simultaneamente com o estágio de carreira e com a posição na distribuição de rendimentos recentes. Em alguns momentos do ciclo de vida, a coorte exposta a maior instabilidade macroeconômica no início da trajetória (1989) apresentou maior risco de saída do mercado formal, enquanto em outros, esse padrão se enfraqueceu ou mesmo se inverteu. Esse resultado reforça a importância de analisar a

heterogeneidade ao longo de toda a superfície carreira–renda recente, em vez de depender exclusivamente de estatísticas agregadas.

Na margem intensiva, os achados mostram que o risco do crescimento futuro não se resume à volatilidade típica. A distribuição do crescimento apresentou assimetria sistematicamente negativa, com maior peso relativo de quedas expressivas do que de altas de mesma magnitude. Esse padrão não parece ser apenas um reflexo mecânico das saídas do mercado formal. Enquanto os trabalhadores posicionados nas faixas inferiores da distribuição enfrentam uma maior probabilidade de não serem observados no registro formal no futuro, aqueles situados nas faixas superiores tendem a apresentar trajetórias com maior assimetria para baixo; isto é, quando ocorrem choques de maior magnitude, estes aparecem relativamente mais como perdas do que como ganhos. O risco, portanto, assume formas distintas ao longo da distribuição, estando na base principalmente ligado a uma maior chance de saída do mercado formal e, no topo, a uma maior exposição a quedas severas condicionais à permanência.

Por fim, a dimensão mais marcante da heterogeneidade observada aparece nas caudas da distribuição do crescimento dos rendimentos formais. A incidência de eventos extremos varia intensamente ao longo do ciclo de vida e da distribuição de rendimentos recentes, com diferenças relevantes entre as coortes analisadas. Em particular, a coorte de 1989 apresentou, em segmentos específicos da carreira, especialmente nos estágios intermediários e finais, e nas posições intermediárias e superiores da distribuição, uma prevalência mais elevada de episódios extremos do que a coorte de 1995. Em contraste, a coorte mais recente exibiu, em geral, caudas mais contidas nesses mesmos estágios, ainda que, no início da carreira, existam pontos de convergência e algumas inversões pontuais. Esses padrões sugerem que a exposição a ambientes macroeconômicos adversos no início da trajetória pode deixar marcas persistentes, não necessariamente por meio de um aumento uniforme da volatilidade, mas pela alteração localizada da probabilidade de ocorrência de eventos extremos.

Em termos mais amplos, a tese contribui para a compreensão de como fatores familiares, institucionais e macroeconômicos interagem na reprodução das desigualdades no Brasil. As evidências sugerem a relevância de políticas integradas que combinem educação de qualidade, redução de fricções de acesso e condições que favoreçam trajetórias mais estáveis no mercado de trabalho, como parte de uma estratégia de desenvolvimento que concilie crescimento econômico e justiça social.

Referências gerais da tese

BALIYAN, Som Pal; KHAMA, Dira. How Distance to School and Study Hours after School Influence Students' Performance in Mathematics and English: A Comparative Analysis. **Journal of Education and e-Learning Research**, Asian Online Journal Publishing Group, v. 7, n. 2, p. 209–217, 2020. Received: 29 April 2020; Revised: 5 June 2020; Accepted: 7 July 2020; Published: 20 July 2020. DOI: 10.20448/journal.509.2020.72.209.217.

BAMMOU, Hamid; EL BOUHALI, Mohammed; EL ALAOUI, Aicha. Impact of the Distance Travelled to School on Students' Performance, Tardiness, and Absenteeism in Public Schools in Morocco. **European Scientific Journal, ESJ**, European Scientific Institute, v. 20, n. 34, p. 117, 2024. Submitted: 16 Sep 2024; Accepted: 17 Dec 2024; Published: 31 Dec 2024. DOI: 10.19044/esj.2024.v20n34p117. Disponível em: <https://doi.org/10.19044/esj.2024.v20n34p117>.

BANCO CENTRAL DO BRASIL. **30 anos do Plano Real**. [S. l.: s. n.], 2024. <https://www.bcb.gov.br/controleinflacao/30anosreal>. Acesso em: 11 ago. 2025.

BANCO CENTRAL DO BRASIL. **Plano Real**. [S. l.: s. n.], 2024. <https://www.bcb.gov.br/controleinflacao/planoreal>. Acesso em: 11 ago. 2025.

BANCO CENTRAL DO BRASIL. **Planos econômicos fracassaram em derrotar a superinflação até a chegada do Real**. [S. l.: s. n.], 2019. Publicado em 28/06/2019 e atualizado em 28/08/2019. Acessado em: 27 jul. 2025. Disponível em: <https://www.bcb.gov.br/detalhenoticia/355/noticia>.

BANCO CENTRAL DO BRASIL. **SGS – Sistema Gerenciador de Séries Temporais: Taxas de Juros (TBC, TBAN, Meta Selic)**. Período de 01/07/1996 a 31/12/2017. Versão 2.1. Acesso em: 8 nov. 2025. 2025. Disponível em: <https://www3.bcb.gov.br/sgspub/consultarvalores/consultarValoresSeries.do?method=consultarGrafico>.

BARBOSA, Felipe Alves Saraiva. **Impacto da distância casa-escola e do transporte escolar no desempenho escolar dos alunos da rede pública de Maracanaú**. 2024. f. 31. Dissertação de Mestrado em Economia do Setor Público – Faculdade de Economia, Administração, Atuária e Contabilidade (FEAAC), Universidade Federal do Ceará (UFC), Fortaleza, CE.

BECKER, Gary S. **Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education**. New York: Columbia University Press, 1964.

BENJAMINI, Yoav; HOCHBERG, Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 57, n. 1, p. 289–300, 1995. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

BLACK, Sandra E.; DEVEREUX, Paul G.; SALVANES, Kjell G. **The more the merrier? The effect of family composition on children's education.** [S. l.], 2004. (NBER Working Paper Series, 10720). DOI: 10.3386/w10720. Disponível em: <http://www.nber.org/papers/w10720>.

BLACK, Sandra E.; DEVEREUX, Paul J. **Recent Developments in Intergenerational Mobility.** Massachusetts, 2010. DOI: 10.3386/w15889. Disponível em: <https://doi.org/10.3386/w15889>.

BLACK, Sandra E.; DEVEREUX, Paul J.; SALVANES, Kjell G. Like father, like son? A note on the intergenerational transmission of IQ scores. **Economics Letters**, v. 105, n. 1, p. 138–140, 2009. ISSN 0165-1765. DOI: 10.1016/j.econlet.2009.06.022. Disponível em: <https://doi.org/10.1016/j.econlet.2009.06.022>.

BLACK, Sandra E.; DEVEREUX, Paul J.; SALVANES, Kjell G. Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. **The American Economic Review**, American Economic Association, v. 95, n. 1, p. 437–449, 2005. ISSN 00028282. Disponível em: <http://www.jstor.org/stable/4132690>. Acesso em: 3 ago. 2025.

BROOKE, N.; BONAMINO, A. **Geres 2005: Razões e Resultado de uma pesquisa longitudinal sobre eficácia escolar.** Rio de Janeiro: WalPrint, 2011. Disponível em: https://www.laedpuc.net.br/_files/ugd/fc35a5_a0f50bf2665448af8c06be6b3c089460.pdf.

CARD, David. **Earnings, Schooling, and Ability Revisited.** [S. l.], 1994. Acessado em 21 de outubro de 2025. DOI: 10.3386/w4832. Disponível em: <https://doi.org/10.3386/w4832>.

CARD, David. The Causal Effect of Education on Earnings. *In*: ASHENFELTER, Orley; CARD, David (ed.). **Handbook of Labor Economics.** [S. l.]: Elsevier, 1999. 3A. p. 1801–1863. DOI: 10.1016/S1573-4463(99)03011-4.

CARD, David. **Using geographic variation in college proximity to estimate the return to schooling.** [S. l.], 1993. DOI: 10.3386/w4483. Disponível em: <https://doi.org/10.3386/w4483>.

CHEN, Jiafeng; ROTH, Jonathan. Logs with Zeros? Some Problems and Solutions. **The Quarterly Journal of Economics**, v. 139, n. 2, p. 891–936, 2024. DOI: 10.1093/qje/qjad054. Disponível em: <https://doi.org/10.1093/qje/qjad054>.

CROW, Edwin L.; SIDDIQUI, M. M. Robust Estimation of Location. **Journal of the American Statistical Association**, v. 62, n. 318, p. 353–389, 1967. DOI: 10.1080/01621459.1967.10482914.

CUNHA, Flavio; HECKMAN, James. The Technology of Skill Formation. **American Economic Review**, v. 97, n. 2, p. 31–47, 2007. DOI: 10.1257/aer.97.2.31.

DEJONG, Peter Y.; MILTON, Brawer J.; STANLEY, Robin S. Patterns of Female Intergenerational Occupational Mobility: A Comparison with Male Patterns of Intergenerational Occupational Mobility. **American Sociological Review**, v. 36, n. 6, p. 1033–1042, 1971. DOI: 10.2307/2093763. Disponível em: <https://doi.org/10.2307/2093763>.

DING, Peng; FENG, Shuaizhang. How School Travel Affects Children's Psychological Well-Being and Academic Achievement in China. **International Journal of Environmental Research and Public Health**, MDPI, v. 19, n. 21, p. 13881, 2022. DOI: 10.3390/ijerph192113881. Disponível em: <https://doi.org/10.3390/ijerph192113881>.

DUFLO, Esther. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. **American Economic Review**, v. 91, n. 4, p. 795–813, 2001. DOI: 10.1257/aer.91.4.795. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.91.4.795>.

DUNCAN, Otis D. Occupation Trends and Patterns of Net Mobility in the United States. **Demography**, v. 3, n. 1, p. 1–18, 1966. DOI: 10.2307/2060060. Disponível em: <https://doi.org/10.2307/2060060>.

DUSTAN, Andrew; NGO, Diana K. L. Commuting to Educational Opportunity? School Choice Effects of Mass Transit Expansion in Mexico City. **Economics of Education Review**, v. 63, p. 116–133, 2018. ISSN 0272-7757. DOI: 10.1016/j.econedurev.2018.01.006. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0272775717303369>.

DVORKIN, Maximiliano A.; MARKS, Cassandra. Nominal Wage Adjustments during High Inflation and Tight Labor Markets. **Federal Reserve Bank of St. Louis Review**, Federal Reserve Bank of St. Louis, v. 106, n. 10, p. 1–19, 2024. Fourth Quarter. DOI: 10.20955/r.2024.10. Disponível em: <https://doi.org/10.20955/r.2024.10>.

ENGBOM, Niklas; MOSER, Christian. Earnings Inequality and the Minimum Wage: Evidence from Brazil. **American Economic Review**, v. 112, n. 12, p. 3803–3847, 2022. DOI: 10.1257/aer.20181506. Disponível em: <https://doi.org/10.1257/aer.20181506>.

ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, Brasil. Instituto Nacional de. **Índice de Desenvolvimento da Educação Básica (Ideb)**. [S. l.: s. n.], 2022. Acesso em: 10 ago. 2025. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/ideb>.

FRANCO, C.; BROOKE, N.; ALVES, F. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 16, n. 61, p. 625–637, 2008.

FREGUGLIA, Ricardo; PROCÓPIO, Thais. Efeitos da mudança de emprego e da migração interestadual sobre os salários no Brasil formal: evidências a partir de dados em painel. **Pesquisa e Planejamento Econômico**, v. 43, n. 2, p. 255–278, 2013. Ipea.

GAMBETTI, Luca; MESSINA, Julián. Evolução da Ciclicidade Salarial na América Latina. **The World Bank Economic Review**, v. 32, n. 3, p. 709–726, 2018.

GERES. **Banco de Dados do GERES, 2005: GERES – Estudo Longitudinal da Geração Escolar 2005**. [S. l.: s. n.], 2005. Banco de dados, notas técnicas e informações complementares. Acesso em: Wed, 08 Feb 2023 12:53:03 GMT. Disponível em: <https://www.laedpuc.net.br/projetogeres>. Acesso em: 8 fev. 2023.

GIANNELLI, Gianna Claudia; RAPALLINI, Chiara. Parental occupation and children's school outcomes in math. **Research in Economics**, v. 73, n. 4, p. 293–303, 2019. ISSN 1090-9443. DOI: 10.1016/j.rie.2019.08.003. Disponível em: <https://doi.org/10.1016/j.rie.2019.08.003>.

GOULART, Livia Mara Lima; MORAIS, Elaine Andrade de; VIEIRA JÚNIOR, Niltom. Tempo de permanência no transporte escolar sobre o desempenho estudantil. **Revista Interterritórios**, Instituto Federal de Roraima, v. 5, n. 9, 2019. DOI: 10.33052/inter.v5i9.243594. Disponível em: <https://doi.org/10.33052/inter.v5i9.243594>.

GU, Xiang *et al.* Like father, like son? Parental input, access to higher education, and social mobility in China. **China Economic Review**, Elsevier, v. 72, p. 101761, 2022.

GUNDERSON, Morley; OREOPOLOUS, Philip. Chapter 3 - Returns to education in developed countries. In: BRADLEY, Steve; GREEN, Colin (ed.). **The Economics of Education (Second Edition)**. Second Edition. [S. l.]: Academic Press, 2020. p. 39–51. ISBN 978-0-12-815391-8. DOI: <https://doi.org/10.1016/B978-0-12-815391-8.00003-3>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780128153918000033>.

GUVENEN, Fatih *et al.* What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics? **Econometrica**, v. 89, n. 5, p. 2303–2339, 2021. Acessado em 21 de outubro de 2025. DOI: 10.3982/ECTA14603. Disponível em: <https://www.jstor.org/stable/44250423>.

HECKMAN, James J. Skill Formation and the Economics of Investing in Disadvantaged Children. **Science**, v. 312, p. 1900–1902, 2006.

HECKMAN, James J.; LOCHNER, Lance J.; TODD, Petra E. **Fifty Years of Mincer Earnings Regressions**. [S. l.], 2003. Acessado em 21 de outubro de 2025. DOI: 10.3386/w9732. Disponível em: <https://doi.org/10.3386/w9732>.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Tabela 1737: IPCA – Série histórica com número-índice, variação mensal e acumulados (3 meses, 6 meses, ano, 12 meses)**. [S. l.: s. n.], 2025. SIDRA – Sistema IBGE de Recuperação Automática. Disponível em: <https://sidra.ibge.gov.br/tabela/1737> (acessado em 10 ago. 2025).

IMBENS, Guido W.; ANGRIST, Joshua D. Identification and Estimation of Local Average Treatment Effects. **Econometrica**, Wiley e Econometric Society, v. 62, n. 2, p. 467–475, 1994. ISSN 0012-9682. Disponível em: <https://www.jstor.org/stable/2951620>. Acesso em: 10 fev. 2026.

INEP. **Histórico do Saeb (Sistema de Avaliação da Educação Básica)**. [S. l.: s. n.], 2023. Publicado em 31 ago. 2020. Atualizado em 22 set. 2023. Acesso em: 3 ago. 2025. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/historico>.

INEP. **Programa Internacional de Avaliação de Estudantes (Pisa)**. [S. l.: s. n.], 2024. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa>.

KELLEY, Truman Lee. **Fundamentals of Statistics**. Cambridge, MA: Harvard University Press, 1947.

KIM, Tae-Hwan; WHITE, Halbert. On More Robust Estimation of Skewness and Kurtosis. **Finance Research Letters**, v. 1, n. 1, p. 56–73, 2004. DOI: 10.1016/S1544-6123(03)00003-5.

LABAND, David N.; LENTZ, Bernard F. Like Father, like Son: Toward an Economic Theory of Occupational Following. **Southern Economic Journal**, v. 50, n. 2, p. 474–493, 1983. DOI: 10.2307/1058220. Disponível em: <https://doi.org/10.2307/1058220>.

LEE, Jihyun; BORGONOV, Francesca. Relationships between family socioeconomic status and mathematics achievement in OECD and non-OECD countries. **Comparative Education Review**, The University of Chicago Press Chicago, IL, v. 66, n. 2, p. 199–227, 2022.

LEMIEUX, Thomas. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? **The American Economic Review**, American Economic Association, v. 96, n. 3, p. 461–498, 2006. Accessed: 8 November 2025. Disponível em: <http://www.jstor.org/stable/30034058>.

LENTZ, Bernard F.; LABAND, David N. Why So Many Children of Doctors Become Doctors: Nepotism vs. Human Capital Transfers. **The Journal of Human Resources**, v. 24, n. 3, p. 396–413, 1989. DOI: 10.2307/145820. Disponível em: <https://doi.org/10.2307/145820>.

MARTINS, Ana Paula Antunes. **Análise dos impactos das condições do transporte escolar rural no rendimento escolar dos alunos**. 2010. xi, 118 f. Dissertação de Mestrado – Universidade de Brasília, Brasília. Orientador: José Matsuo Shimoishi. Disponível em: <https://repositorio.unb.br/handle/10482/8791>.

MCFADDEN, Daniel. Conditional Logit Analysis of Qualitative Choice Behavior. *In*: ZAREMBKA, Paul (ed.). **Frontiers in Econometrics**. New York: Academic Press, 1974. p. 105–142.

MELLO, João M. P. de; WAISMAN, Caio; ZILBERMAN, Eduardo. Os efeitos da exposição à hiperinflação na escolha ocupacional. **Journal of Economic Behavior & Organization**, v. 106, p. 109–123, 2014.

MINCER, Jacob A. **Schooling, Experience, and Earnings**. New York: Columbia University Press for the National Bureau of Economic Research, 1974. ISBN 0-87014-265-8.

MINCER, Jacob A. **Schooling, Experience, and Earnings**. New York: Columbia University Press for NBER, 1974. Disponível em: <https://www.nber.org/books-and-chapters/schooling-experience-and-earnings>.

MORENO-MONROY, Ana I.; LOVELACE, Robin; RAMOS, Frederico R. Public transport and school location impacts on educational inequalities: Insights from São Paulo. **Journal of Transport Geography**, v. 67, p. 110–118, 2018. ISSN 0966-6923. DOI: <https://doi.org/10.1016/j.jtrangeo.2017.08.012>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0966692316303453>.

OECD. Do Parents' Occupations Have an Impact on Student Performance? **PISA in Focus**, OECD Publishing, Paris, n. 36, 2014. DOI: 10.1787/5jz8mr7kp026-en. Disponível em: <https://doi.org/10.1787/5jz8mr7kp026-en>.

PATRINOS, Harry Anthony; PSACHAROPOULOS, George. Returns to education in developing countries. *In*: THE economics of education. [S. l.]: Elsevier, 2020. p. 53–64.

PEPE, Marco Antonio Monteiro Villela. **Passe livre estudantil e frequência escolar: uma avaliação dos efeitos do Passe Livre Estudantil na frequência escolar dos educandos da rede pública municipal de São Paulo**. 2017. Dissertação de Mestrado – Escola de Administração de Empresas de São Paulo (EAESP), São Paulo. Orientador: Ciro Biderman. Membros da banca: Ricardo Paes de Barros, André Portela. Disponível em: <https://hdl.handle.net/10438/18192>.

PETEROS, Emerson D. *et al.* Effects of School Proximity on Students' Performance in Mathematics. **Open Journal of Social Sciences**, v. 10, p. 365–376, 2022. ISSN 2327-5960. DOI: 10.4236/jss.2022.107027. Disponível em: <https://www.scirp.org/journal/jss>.

PLASMAN, Jay *et al.* Parents' Occupations and Students' Success in STEM Fields: A Systematic Review and Narrative Synthesis. **Adolescent Research Review**, v. 6, n. 1, p. 33–44, 2021. ISSN 2363-8354. DOI: 10.1007/s40894-020-00136-z. Disponível em: <https://doi.org/10.1007/s40894-020-00136-z>.

RAMÍREZ-HASSAN, Andrés *et al.* What kind of schools parents choose when they have more options? Effects of school transport subsidies. **Socio-Economic Planning Sciences**, v. 87, p. 101509, 2023. ISSN 0038-0121. DOI: <https://doi.org/10.1016/j.seps.2023.101509>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012123000022>.

SANCHES, Ligia González. **O Ideb em pesquisas acadêmicas brasileiras entre 2007 e 2015**. 2017. f. 203. Dissertação (Mestrado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo. Acessado em: 13 jul. 2024. Disponível em: https://www.teses.usp.br/teses/disponiveis/48/48134/tde-06122017-115438/publico/LIGIA_GONCALEZ_SANCHEZ_rev.pdf.

SILVA, A. T. F. da. Aplicação de Regressão Logística na Associação entre Ocupação e Saúde Cardiovascular - Pesquisa Nacional de Saúde 2019. Projeto (Bacharelado em Estatística) – Departamento de Estatística, Universidade de Brasília. Brasília, 2019.

SOLON, Gary. Intergenerational Mobility in the Labor Market. *In*: HANDBOOK of Labor Economics. [S. l.]: Elsevier Science B.V., 1999. v. 3.A. p. 1761–1800. DOI: 10.1016/S1573-4463(99)03010-22. Disponível em: [https://doi.org/10.1016/S1573-4463\(99\)03010-22](https://doi.org/10.1016/S1573-4463(99)03010-22).

WANG, Xiaofang Sarah *et al.* Factors predicting mathematics achievement in PISA: a systematic review. **Large-scale Assessments in Education**, v. 11, n. 1, p. 24, 2023. ISSN 2196-0739. DOI: 10.1186/s40536-023-00174-8. Disponível em: <https://doi.org/10.1186/s40536-023-00174-8>.

ZERAGABER, Tsinat Yemane; TEAME, Ghirmai Tesfamariam; TSIGHE, Zemenfes. Assessing the effect of home-to-school distance on student dropout rate in Adi-Keyih sub-zone, Eritrea. **International Journal of Educational Research Open**, v. 7, p. 100340, 2024. ISSN 2666-3740. DOI: <https://doi.org/10.1016/j.ijedro.2024.100340>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666374024000220>.