UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE DOUTORADO EM TEORIA ECONÔMICA

ENSAIOS EM MICROECONOMIA DO DESENVOLVIMENTO: DEMOGRAFIA, EDUCAÇÃO E MERCADO DE TRABALHO

Vinícius de Araújo Mendes
Orientador: Prof. Dr. Naercio Aquino Menezes Filho

São Paulo
2017

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Hélio Nogueira da Cruz
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

VINÍCIUS DE ARAUJO MENDES

ENSAIOS EM MICROECONOMIA DO DESENVOLVIMENTO: DEMOGRAFIA, EDUCAÇÃO E MERCADO DE TRABALHO

Tese apresentada ao Programa de Pós-Graduação em Economia do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Naercio Aquino Menezes Filho

Versão Corrigida

(Versão original disponível na biblioteca da Faculdade de Economia, Administração e Contabilidade – FEA/USP)

São Paulo
2017

FICHA CATALOGRÁFICA
Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Mendes, Vinícius de Araújo
Ensaios em microeconomia do desenvolvimento: demografia, educação e mercado de trabalho / Vinícius de Araújo Mendes. – São Paulo, 2017.
150 p.

Tese(Doutorado) – Universidade de São Paulo, 2017.
Orientador: Naercio Aquino Menezes Filho.

1. Microeconomia 2. Demografia 3. Desenvolvimento econômico 4. Econometria I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 338.5

Para Rosa, Cido e Linda
(in memorian)

AGRADECIMENTOS

Todas as pessoas que passaram em minha vida e marcaram-me fazem parte deste trabalho, fazem parte de minha formação. A começar pelos meus pais, pessoas que sonharam comigo, ajudaram-me em todos os momentos e me formaram. Sou grato a cada dia de ensinamento. A minha irmã, companheira desde cedo, sempre presente em todas minhas etapas e a Aquiles, um irmão que a vida me deu. Minha esposa Naiana sempre que acompanhou com zelo, incentivo a cada momento. Esta tese tem a sua contribuição a cada parágrafo. A Orlando, Noélia, Nara, Tatiane e Cláudio, todos muito presentes. Lá no começo, Rosângela, Tião e o Murilo abriram a porta de sua casa e me receberam. O carinho e a dedicação, cuidando-me como um filho, fez com que as etapas do doutorado fossem conquistadas com muito mais alegria. Atitudes de gratidão como esta ajudam pessoas a conquistar o desejado. Agradeço à família de meu tio Roberto, sempre presente, sempre cuidadosa e à família de minha avó Antónia.

Agradeço ao meu orientador, professor Naercio Menezes Filho, que me ensinou o maior de todos os valores acadêmicos: o método é apenas um meio de se responder de maneira dedicada uma pergunta relevante. Cada contribuição e cada conselho guiaram minha formação e minha pesquisa. Agradeço a professora Renata Narita, que sempre gentil ajudou-me em minhas atividades acadêmicas. Aos professores Rodrigo Soares e André Portela por terem me acolhido na FGV-EESP enquanto aluno. Agradeço aos professores Fernanda Estevan e Ricardo Madeira pelas contribuições na qualificação da tese de doutorado. Agradeço aos funcionários da FEA-USP, em especial a Pinho, Alessandra e Ana Paula, sempre solícitos. Aproveito para agradecer ao CNPQ pelo financiamento de meu doutorado.

Diversos amigos eu construí em minha jornada na USP. Em especial agradeço a Gian Paulo por tantas ajudas e a Diana Gonzaga. Agradeço aos amigos da UFBa, em especial a Rodrigo Oliveira e Gervásio Ferreira pelos incentivos. Agradeço aos meus amigos de longa data, em especial Júlio Sanderson, Paulo Henrique e Lorena Fonseca, pela amizade e por ser solícita ajudando-me. Da UFMG, sou grato pelos ensinamentos e amizade de Lízia de Figueiredo, Cássio Turra, Eduardo Rios-Neto, Laura Schiavon e Rafael Klanfer.

RESUMO

Este estudo, dividido em três artigos, tem por objetivo analisar as contribuições da demografia, educação e mercado de trabalho no desenvolvimento brasileiro ao longo do século XX. Os três artigos estão encadeados em uma sequência temporal tal que o primeiro artigo concentra-se na transição demográfica brasileira para as coortes nascidas entre 1890 e 1960. O segundo artigo preocupa-se em entender como os efeitos da transição demográfica são absorvidos pelas novas coortes e seu efeito na taxa de matrícula. O terceiro artigo investiga como as coortes mais jovens, com mais capital humano, mudam a oferta relativa da economia e este processo impacta os salários relativos por grupos de qualificação. No primeiro artigo, os microdados dos Censos demográficos foram utilizados na investigação usando variáveis de educação, localidade e coorte de nascimento. Uma vez controlando por coorte e localidade, assumindo a hipótese que na localidade municipal as coortes são mais homogêneas quanto à exposição da oferta escolar, educação explicou aproximadamente 30% da queda da fecundidade. A simulação contrafactual evidenciou que caso não tivesse ocorrido o aumento da escolaridade das coortes, a transição demográfica seria mais lenta e gradual. No segundo artigo, dados da PNAD foram construídos para investigar a probabilidade de matrícula assumindo que a transição demográfica gera uma variação exógena no tamanho absoluto e no tamanho relativo das coortes. Os resultados evidenciaram que para as séries com maiores avanços na taxa de matrícula, diminuição do tamanho da coorte apresentou-se negativamente relacionada com a probabilidade de matrícula. Porém, a magnitude deste efeito não gerou mudanças significativas na simulação contrafactual e, para a primeira série do ensino fundamental, o avanço da escolaridade dos pais mostrou-se significativo. No terceiro artigo, dados da PNAD são utilizados para se investigar mudanças relativas no mercado de trabalho ocasionadas pela entrada de novas coortes com melhores indicadores educacionais. Os resultados apontam que mudanças na oferta relativa de grupos etários mais jovens conduziram a mudanças em seus salários relativos e as mudanças nos salários relativos de grupos etários mais velhos são explicadas por mudanças na oferta relativa agregada da economia. A simulação contrafactual demostrou que, caso não houvesse mudança na oferta relativa da economia, os salários de qualificados em relação a não qualificados apresentaria tendência crescente entre 1981 e 2013.

Palavras-chave: Microeconomia, Demografia, Desenvolvimento Econômico e Econometria.

ABSTRACT

This study, divided into three articles, aims to analyze the contributions of demography, education and the labor market in Brazilian development throughout the 20th century. The three articles are connected in a temporal sequence such that the first article focuses on the Brazilian demographic transition for the cohorts born between 1890 and 1960. The second article is concerned with understanding how the effects of the demographic transition are absorbed by the new cohorts and affect school attendance. The third article investigates how the younger cohorts, with more human capital, change the relative supply of the economy and this process impacts relative wages by qualification groups. In the first article, the microdata of demographic Census were used in the investigation using variables of education, municipality and birth cohort. Once the cohort and locality control, assuming the hypothesis that in the municipality the cohorts are more homogeneous regarding the exposure of the school supply, education explained approximately 30% of the decrease of the fertility rate. The counterfactual simulation showed that if there had not been an increase in cohort education, the demographic transition would be slower and gradual. In the second article, PNAD data were constructed to investigate the probability of enrollment assuming that the demographic transition generates an exogenous variation in the absolute size and relative size of the cohorts. The results showed that for the series with the greatest advances in enrollment, a decrease in cohort size was negatively related to the probability of enrollment. However, the magnitude of this effect did not generate significant changes in the counterfactual simulation, and for the first grade of elementary school, the progress of parents' schooling was significant. In the third article, PNAD data are used to investigate relative changes in the labor market caused by the entry of new cohorts with better educational indicators. The results indicate that changes in the relative supply of younger age groups have led to changes in relative wages and changes in the relative wages of older age groups are explained by changes in the aggregate relative supply of the economy. The counterfactual simulation showed that if there were no change in the relative supply of the economy, relative wages from skilled to unskilled workers would show a growing trend between 1981 and 2013.

Keywords: Microeconomics, Demography, Economic Development and Econometrics.

SUMÁRIO

INTRODUÇÃO.....15

TRANSIÇÃO DEMOGRÁFICA NO BRASIL: EVIDÊNCIAS POR COORTES, EDUCAÇÃO E LOCALIDADES.....17

TAMANHO DA COORTE, TAMANHO DA FAMÍLIA E MATRÍCULA ESCOLAR: EVIDÊNCIAS DO BRASIL.....65

A ESCOLARIDADE DAS COORTES E A EVOLUÇÃO SALARIAL BRASILEIRA ENTRE 1981 E 2013.....112

CONCLUSÃO.....171

REFERÊNCIAS.....172

INTRODUÇÃO

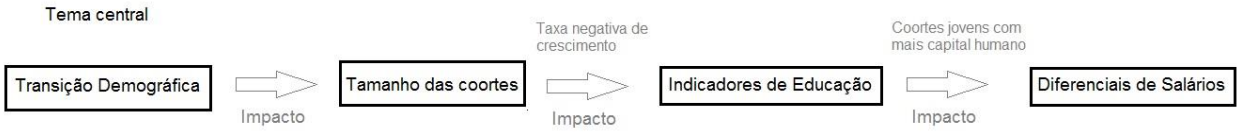
Esta tese tem como objetivo estudar três tópicos, de forma integrada, em microeconomia do desenvolvimento: demografia, educação e mercado de trabalho. A demografia investiga a transição demográfica em perspectiva histórica por coortes de nascimento, a educação investiga os efeitos da transição demográfica sobre a taxa de matrícula ao longo das décadas de 1980, 1990 e 2000 e o mercado de trabalho investiga os efeitos da escolarização e da oferta de trabalho mais qualificada sobre os diferenciais de salários principalmente na década de 2000, período de queda da desigualdade de renda.

O primeiro artigo, denominado *Transição demográfica no Brasil: evidências por coortes, educação e localidades*, reconstrói a história de coortes de mulheres desde 1890 a 1960, um período de profunda transformação no Brasil. Neste período o Brasil foi se urbanizando, estruturou suas redes de ensino primária e secundária, diminuiu a mortalidade infantil e as mulheres ampliaram sua participação no mercado de trabalho. Este artigo tem como propósito investigar como as transformações ocorridas ao longo deste período influenciaram na decisão de fecundidade a partir da educação e localidade. As evidências tanto teóricas quanto empíricas conduzem a uma diminuição no número de filhos quando as coortes de mulheres passam a ter mais acesso a educação, por exemplo. No entanto, a literatura empírica evidencia que, para se investigar o impacto causal da educação na decisão de filhos, é necessário estabelecer um experimento aleatório tal como Aaronson et al (2014) o fizeram. Este artigo tem por objetivo entender os movimentos históricos das coortes de mulheres e utiliza controles de efeito fixo de coorte, efeito fixo de localidade e interação entre efeito fixo de coorte e efeito fixo de localidade para se estimar um parâmetro menos endógeno possível, que explique os movimentos históricos do começo do século XX no Brasil.

Como causa da transição demográfica, tanto o tamanho da família quanto o tamanho das coortes jovens diminuem e estes fatos são utilizados para se capturar as consequências da transição demográfica na educação, proposta do segundo artigo. O artigo, denominado *Tamanho da coorte, tamanho da família e matrícula escolar: evidências do Brasil*, investiga como mudanças demográficas afetam a probabilidade de matrícula no ensino regular (Fundamental e Médio). Diversas contribuições teóricas e empíricas demostram que na margem famílias trocariam uma maior quantidade de filhos por uma menor quantidade, mas investindo mais na educação destes. Para um país com uma oferta de vagas escolares em escolas públicas, como o caso brasileiro, mudanças no tamanho da coorte provocam mudanças na demanda por educação e, uma vez que o Estado realiza o esforço de construir

escolas, contratar professores, uma queda na taxa de fecundidade provoca um aumento na quantidade de recursos *per capita*, criando a possibilidade de se aumentar a taxa de matrícula na margem por uma consequência da transição demográfica. Assim, o objetivo do segundo artigo é entender se a queda da taxa de fecundidade, estudada no primeiro artigo, contribuiu com o aumento da taxa de matrícula uma vez que os recursos *per capita* para oferta de vagas aumentam pela queda do tamanho das coortes (em valores absolutos ou relativos) a partir da década de 1990.

O terceiro artigo, denominado *A escolaridade das coortes e a evolução salarial brasileira entre 1981 e 2013*, apresenta as consequências do processo de transição demográfica e educação no mercado de trabalho. O aumento do percentual de coortes matriculadas nas décadas de 80 e 90 provocou uma mudança relativa da oferta da mão-de-obra na década de 2000, aumentando a proporção de trabalhadores qualificados em relação a trabalhadores pouco qualificados. A mudança da oferta relativa, em equilíbrio, conduz a uma mudança dos salários relativos. Assim, uma parte da mudança na desigualdade de renda é explicada por se ter mais indivíduos qualificados no mercado de trabalho, consequência da melhoria dos indicadores educacionais nas décadas de 80 e 90. Outro fator que impacta a desigualdade de renda é a diminuição do salário relativo dos trabalhadores muito qualificados em relação aos trabalhadores pouco qualificados, consequência das mudanças de oferta relativas por grupos de educação. O terceiro artigo terá como objetivo investigar como mudanças na desigualdade de renda são explicadas por mudanças relativas de trabalhadores qualificados e pouco qualificados. O fluxograma abaixo apresenta a conexão que há entre os três artigos por uma sucessão de eventos analisados por coortes de nascimento.



TRANSIÇÃO DEMOGRÁFICA NO BRASIL: EVIDÊNCIAS POR COORTES, EDUCAÇÃO E LOCALIDADES

1. INTRODUÇÃO

A relação entre demografia e economia é estabelecida desde Robert Malthus, economista clássico britânico, porém sua explicação sobre a dinâmica da população contempla apenas o período que posteriormente fora denominado época malthusiana, caracterizada por uma taxa de crescimento da renda *per capita* próximo de zero e crescimento populacional próximo de zero explicado pelas altas taxas de mortalidade. A primeira etapa da transição demográfica, a passagem da época malthusiana para o regime pós-malthusiano, está associada à Revolução Industrial (Galor, 2005). No regime pós-malthusiano, a renda per capita é crescente e cresce *pari passu* ao crescimento populacional provocado por uma queda da mortalidade. A terceira e última etapa da transição demográfica, denominada crescimento sustentável, ocorre quando a população decresce, mas a renda per capita mantém-se em trajetória crescente[1].

Diversos autores pesquisaram demografia econômica e os determinantes da transição demográfica. Gary Becker, em um artigo seminal de 1960, utilizou elementos da análise microeconômica para mostrar como indivíduos tomam decisão de tamanho da família quando preferências, qualidade do filho, expressa em educação, e renda interagem na escolha[2]. Na denominação microeconômica, Becker não acreditava que crianças eram um "bem inferior", mas que pais poderiam estar substituindo quantidade por qualidade de seus filhos[3]. Em um

[1] Oded Galor (2005) analisa os três períodos (época malthusiana, regime pós-malthusiano e crescimento sustentável) assim como as transições demográficas de um regime para o outro, com uma riqueza de dados históricos tais como renda *per capita*, população, taxas de natalidade e mortalidade e expectativa de vida ao nascer. Para muitas destas variáveis, a série de tempo contempla do ano zero ao ano 2000.

[2] Becker analisa a importância da tecnologia e dispersão de métodos contraceptivos na tomada de decisão do tamanho da família. Segundo o autor, este ponto não foi considerado na análise de Malthus. (Becker, 1960)

[3] A relação entre variação da renda e número de filhos foi investigada empiricamente por Black et al (2013). Os autores utilizam uma variação exógena na renda por um choque internacional no preço do carvão por conta da crise do petróleo. Os autores concluíram em seu experimento que famílias localizadas na região produtora de minério tiveram um aumento exógeno da renda e este aumentou causou, na margem, uma elevação do número de filhos.

artigo sobre transição demográfica, Becker, Murphy e Tamura (1990) apresentam um modelo formalizado onde sociedades com baixo capital humano escolheriam tamanho maiores de famílias enquanto sociedades com alto capital humano escolheriam famílias menores. Seguindo uma tradição iniciada por Becker, Soares (2005) apresenta um modelo teórico onde queda da mortalidade gera transição demográfica, com o aumento da expectativa de vida alterando decisões do indivíduo que escolhe reduzir a fecundidade e aumentar seu capital humano. Galor e Weil (2000) apresentam um modelo de crescimento endógeno com decisão de número de filhos. Segundo os autores, a transição para um regime sustentável é consequência da interação entre educação e tecnologia. A história contada pelo artigo narra que no período pós-malthusiano, a renda *per capita* era função do tamanho da família. A introdução da tecnologia poupadora de mão-de-obra possibilitou que indivíduos antes alocados no trabalho pudessem adquirir capital humano. Este aumento do capital humano provoca mudanças tecnológicas, por conta do componente endógeno da tecnologia na formulação dos autores, o que entraria em um ciclo de acúmulos de capita humano e tecnológico provocando a transição demográfica para um regime sustentável.

Galor (2005) apresenta diversas teorias de transição demográficas para além de suas formulações. Segundo o autor, os fatores que causariam transição demográfica seriam: 1) o aumento da demanda industrial por capital humano, 2) declínio da mortalidade e aumento da expectativa de vida[4], 3) suporte a reformas educacionais, 4) declínio do trabalho infantil, 5) evolução cultural e genética na valoração do capital humano, 6) o surgimento da formação do capital humano, 7) declínio da diferença de gênero. Aaronson et al (2014) apresentam como fontes para explicar a transição demográfica: 1) *skill-biased technical change*[5]: 2) declínio no custo de métodos contraceptivos, 3) crescimento do salário relativo das mulheres, 4) aumento da expectativa de vida, 5) declínio no valor do trabalho infantil. Os autores investigaram, em seu artigo, transição da fecundidade na margem extensiva, quando se opta por não ter filho, e na margem intensiva, quando quantidade e qualidade de filhos são substitutas. Usando um programa de construção de escolas para população negra rural, os autores demonstram que a fecundidade cai na margem intensiva e aumentou na margem extensiva segundo o

[4] Tal qual apresentado por Soares (2005)

[5] A literatura denomina "*skill-biased technical change*" a demanda por trabalho qualificado relativo ao trabalho pouco qualificado por conta de deslocamentos da função de produção. Goldin e Katz (2008) e Katz e Murphy (1992) são referências da literatura.

experimento. No entanto, ao comparar a fecundidade das mulheres que estavam em idade escolar, quando o programa de construção de escolas fora implementado, com as mulheres de coortes que não estavam mais em idade escolar, o modelo prediz que a fecundidade cai tanto na margem intensiva quanto na margem extensiva para mulheres beneficiadas pela construção de escolas, por conta do aumento do valor do tempo das mulheres beneficiadas[6].

Este trabalho investigará a transição demográfica no Brasil usando coortes com período reprodutivo completo para se analisar os elementos que conduziram a transição demográfica. Os dados são dos Censos demográficos de 1970 a 2010, separados por coortes de nascimento de mulheres de 50 a 80 anos, montando uma trajetória histórica de coortes nascidas de 1890 a 1960, período de profunda transformação no Brasil. Dado que a decisão de ter filho ocorre em um período anterior à entrevista ao Censo, foram utilizadas apenas variáveis com informações são estáveis no tempo, principalmente educação. A localidade foi compatibilizada ao longo dos censos por áreas mínimas comparáveis aos municípios do Censo de 1970, para se capturar efeitos fixos comuns a uma coorte nascida que teve acesso a educação desta localidade, por exemplo, além de se comparar fecundidade entre migrante e nativo dentro de uma mesma área[7]. Os resultados encontrados mostram que educação explica a diminuição do número de filhos e sua magnitude diminui ao se adicionar controles de localidade. Os efeitos fixos de coorte sugerem que a transição demográfica foi realizada por coortes nascidas entre o meio da década de 30 e o começo da década de 40 do século XX. Os efeitos fixos de localidade sugerem um aumento do número de filhos para municípios localizados nas regiões Norte e Nordeste.

[6] Becker (1965) apresenta um modelo de alocação do tempo com uma aplicação o tamanho de família. Willis (1973) aplica a teoria de alocação de tempo em um modelo de fecundidade com oferta feminina de trabalho. Em ambos os casos as formulações são consistentes com os resultados encontrados por Aaronson et al (2014).

[7] Incorporar migração ao modelo enfrenta as dificuldades de compatibilização da informação ao longo dos Censos. Nos Censos antigos, o indivíduo declarava se a migração foi feita a partir de uma zona rural ou uma zona urbana. Esta informação é mais imprecisa no Censo de 2010, que declara o código do município migrado. Outra dificuldade é caracterizar o tempo exato da mudança para o município atual. Assim, o "nativo" é o indivíduo que nasceu no município enquanto o "migrante" é o indivíduo que declarou não ter nascido no município que residia no ato da pesquisa censitária.

2. BASE DE DADOS

Os dados utilizados são os microdados dos Censos demográficos brasileiros de 1970, 1980, 1990, 2000 e 2010. Alguns critérios de compatibilização foram utilizados para se analisar as variáveis ao longo do tempo. Educação é analisada tanto por anos completos de estudo[8] quanto por faixas de escolaridade e a construção do algoritmo considerou Riani e Golgher (2004) e Rigotti (2004) como referências[9]. Para se capturar efeitos fixos de localidade, áreas mínimas comparáveis (AMC) foram utilizadas e usandas como referência para que uma localidade fosse identificada ao longo dos Censos. Assim, foram criadas 3660 AMC tendo como referência 1970, 4268 AMC tendo como referência 1991 e 5508 AMC tendo como referência 2010.

A fecundidade considera filhos nascidos vivos para uma mulher após seu período potencialmente reprodutivo, ou seja, mulheres com 50 anos ou mais[10]. Estas informações foram compatibilizadas para que a declaração do número de filhos tivesse valor máximo de 30, gerando informações padronizadas ao longo dos Censos. Tanto para as estatísticas descritivas quanto para as regressões foram utilizados dados de mulheres por coorte até 80 anos de idade, o que representa apenas uma parcela dos sobreviventes da coorte desde o nascimento até a entrevista ao Censo. Uma mulher foi considerada nativa caso ela sempre tenha morado no município que nasceu e, com base nesta informação, define-se migrante[11]. As Unidades da Federação foram compatibilizadas para um total de 25 estados analisados ao

[8] A variável "Anos completos de estudo" é gerada para os Censos de 1970, 1980, 1991 e 2000. As mudanças apresentadas pelo Censo de 2010, agregando as séries cursadas em intervalos maiores, impossibilita a criação de anos de estudo.

[9] O Censo de 2010 não permite construir a variável "anos de estudo" por apresentar a série que cursa, para os estudantes, ou a série cursada para quem não estuda mais, em intervalos e não efetivamente a última série concluída com aprovação. No entanto, é possível analisar educação por intervalos de escolaridade.

[10] Há uma tradição demográfica de se investigar transição pela taxa de fecundidade definida, segundo o IBGE, pelo número médio de filhos nascidos vivos, tidos por uma mulher de uma geração hipotética (15 e 49 anos de idade) ao final de seu período reprodutivo. Este artigo analisa a transição demográfica a partir do conceito de taxa de fecundidade por coorte, definida aqui pelo ano de nascimento. Em economia, Kai (2009) é uma referência para a análise da transição demográfica por coortes de mulheres em perspectiva histórica.

[11] O fluxo migratório rural-rural, rural-urbano, urbano-rural e urbano-urbano foi gerado apenas para os Censos de 1970, 1981 e 1991 e, portanto, realizado apenas algumas estatísticas descritivas para se entender os processos de mudanças durante o período de transição demográfica. A dificuldade de se estabelecer o período exato da migração nos Censos fez com que tal investigação não fosse adicionada durante a estratégia empírica.

longo de todos os censos demográficos, ou seja, o estado do Tocantins foi integrado ao estado de Goiás e o estado de Mato Grosso do Sul foi integrado ao estado de Mato Grosso. A declaração sobre religião não é homogênea ao longo dos Censos e foram compatibilizadas para quatro categorias: sem religião, católica, evangélica, espírita e outras religiões. Para os Censos de 1980, 1991, 2000 e 2010, a religião evangélica pentecostal foi agregada à religião evangélica tradicional, espírita kardecista e espírita afro-brasileira foram agregadas compondo a religião espírita e, por fim, religiões orientais e judaica/israelita foram agregadas a outras religiões.

As bases de dados foram agregadas adicionando as observações do Censo de 1980 abaixo das observações do Censo de 1970, as informações do Censo de 1991 abaixo das informações do Censo de 1980 e assim sucessivamente[12]. Para todas as bases apenas as observações de mulheres com 50 a 80 anos foram mantidas e foram geradas dummies de escolaridade, dummies de religião, dummies de coorte e dummies de localidade (unidades da federação e áreas minimamente comparáveis). Da base de dados final, uma subamostra de 5% do universo foi selecionada a partir de uma distribuição de binomial no nível do município e a probabilidade foi calculada, para cada censo específico, para que o número de observações selecionadas fosse exatamente 5% do número de observações do universo[13].

3. TRANSIÇÃO DEMOGRÁFICA NO BRASIL

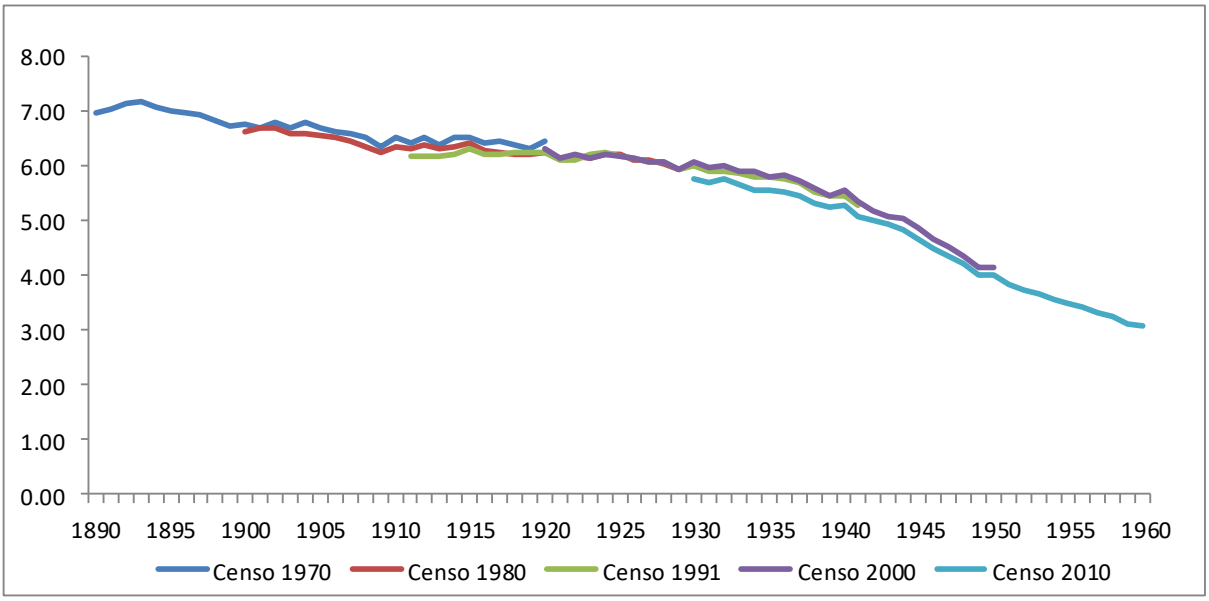
Usualmente a transição demográfica no Brasil é analisada a partir da taxa de fecundidade, definida como uma estimativa do número médio de filhos que uma mulher teria ao final de seu ciclo reprodutivo. Esta abordagem, comum em pesquisas demográficas, mostrou que o Brasil fez sua transição demográfica com tendência decrescente na fecundidade entre os censos de 1960 e 1970. Além da mudança na fecundidade, o censo de 1970 mostrou mudanças importantes em variáveis como urbanização, escolaridade, mercado de trabalho e

[12] As bases foram agregadas via "*append*" e uma variável denominada censo foi criada para identificar à qual pesquisa pertence cada observação. Além disto, foram criadas 5 dummies de censos que foram utilizadas como controle na estratégia empírica.

[13] A subamostra apresenta um ganho de eficiência para se estimar uma base de dados volumosa. Para que os resultados fossem representativos do universo, os pesos amostrais foram ponderados novamente e as comparações entre a amostra e a subamostra são apresentadas no Apêndice 1 deste capítulo. Foi utilizado como referência para se selecionar a subamostra o procedimento adotado pelo IPUMS, da *Minnesota Population Center at the University of Minnesota*.

inserção da mulher neste contexto. A taxa de fecundidade calculada pelo censo de 1970 considera mulheres entre 15 e 49 anos, correspondentes às coortes de 1921 e 1955, era 5,76[14]. Uma abordagem alternativa à taxa de fecundidade é calcular o número médio de filhos por coorte de nascimento para mulheres que já completaram seu ciclo reprodutivo. O Gráfico 1 apresenta a trajetória histórica das coortes de mulheres desde 1890 a 1960[15].

GRÁFICO 1: NÚMERO DE FILHOS POR COORTE DE NASCIMENTO



Fonte: Censos 1970, 1980, 1991, 2000 e 2010

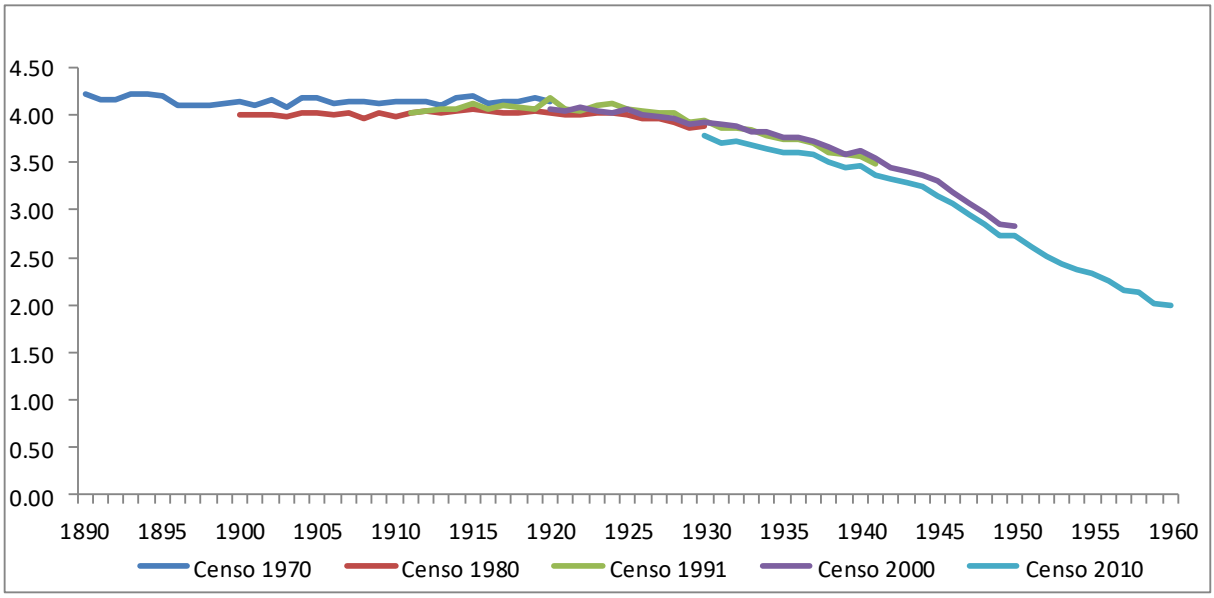
Ao se analisar por coorte de nascimento, o Brasil apresenta uma tendência decrescente desde as coortes nascidas em 1890. Entretanto, há uma mudança a partir das coortes nascidas na década de 1930 com forte declínio no número médio de filhos nascidos vivos. O gráfico sugere também que para uma coorte específica, a média do número de filhos é menor para pesquisas mais recentes quando os resultados são sobrepostos no gráfico e este resultado é

[14] Usando dados do IBGE, a taxa de fecundidade no Brasil em 1960 era 6,28, em 1980 era 4,35 e em 1991 era de 2,85 filhos nascidos vivos para uma coorte hipotética de mulheres entre 15 e 49 anos.

[15] Para cada censo demográfico define-se a coorte pela diferença entre o ano da pesquisa e a idade da pessoa. Informações mais precisas de ano de nascimento não são divulgadas pelo Censo (estão disponíveis na PNAD). Assim, para cada uma das pesquisas, é calculada a média do número de vivos nascidos vivos por coorte de nascimento. Esta técnica apresenta a vantagem de se calcular exatamente o número de filhos nascidos vivos após o ciclo reprodutivo, mas apresenta a desvantagem de, após longo período, o número de sobreviventes da coorte passa a ser cada vez menor. Na média, pessoas com maior probabilidade de sobrevivência têm menos filhos e isto precisa ser primeiro, documentado sobrepondo os resultados das coortes em diferentes pesquisas censitárias nas estatísticas descritivas, assim como precisa ser controlado na estratégia empírica.

mais evidente ao compararmos com o Censo de 2010. Por exemplo, para a coorte nascida em 1930, o Censo de 1980 apresenta media de 5,37 filhos nascidos vivos com 534.328 observações, o Censo de 1991 apresenta média de 5,36 filhos nascidos vivos com 412.423 observações, o Censo de 2000 apresenta média de 5,38 filhos nascidos vivos com 377.406 observações e, para o Censo de 2010, a média é 5,02 filhos nascidos vivos com 265.980 observações. Assim, mulheres com maior média de número de filhos têm menor probabilidade de sobrevivência entre períodos censitários. Este fator está correlacionado com outras variáveis, em especial educação e situação do setor censitário[16]. O gráfico 2 apresenta o desvio-padrão do número de filhos por coorte de nascimento e, assim como a média do número de filhos, o desvio-padrão apresenta mesma trajetória decrescente a partir das coortes nascidas na década de 1930 e isto, como será apresentado *a posteriori*, pode ser explicado pela queda no número de filhos de mulheres com baixa instrução educacional.

GRÁFICO 2: DESVIO-PADRÃO DO NÚMERO DE FILHOS POR COORTE DE NASCIMENTO

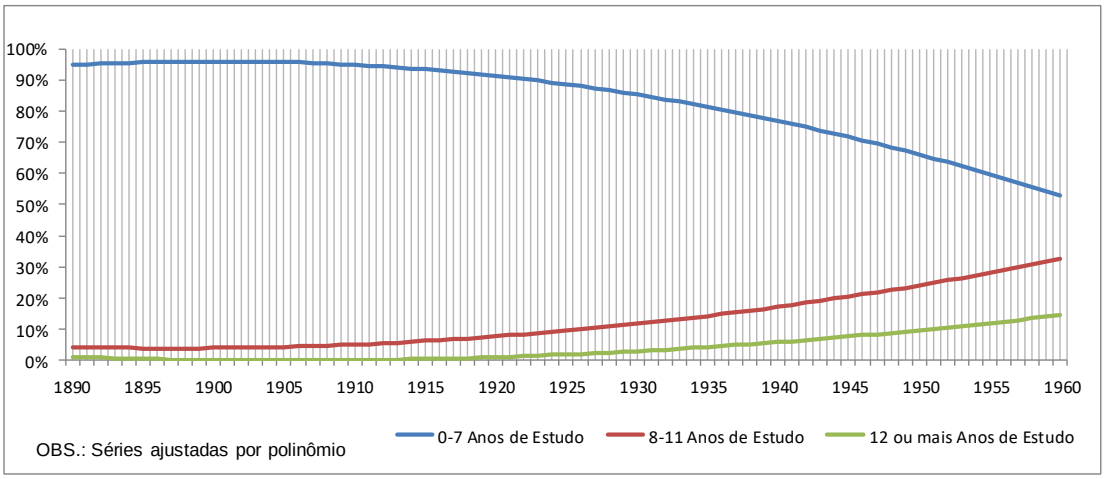


Fonte: Censos 1970, 1980, 1991, 2000 e 2010

[16] O IBGE define como situação de residência a declaração de moradia em zona rural ou zona urbana.

O Gráfico 3 apresenta o percentual de mulheres por anos de estudo para cada coorte de nascimento[17]. Na média, 95 por cento das mulheres nascidas no final do século XIX tinham até 7 anos de estudo[18]. A partir da década de 10 do século XX há uma tendência negativa para o percentual de mulheres com até 7 anos de estudo acompanhado por um aumento do percentual de mulheres com 8 a 11 anos de estudo. Para a coorte de 1960, 53% das mulheres apresentavam até 7 anos de estudo, o que representa uma profunda transformação nos indicadores educacionais ao longo deste período de transição demográfica brasileira. Para a última década do século XIX, 76% das mulheres do grupo com até 3 anos de estudo eram analfabetas. Este percentual cai para 70% para a primeira década do século XX e, para a coorte de 1945, que apresentava 51% das mulheres com até 3 anos de estudo, do total de mulheres com até 3 anos de estudo, 50% eram analfabetas. As coortes mais antigas no Brasil representavam uma parcela considerável da sociedade que nasceu em áreas rurais e, devido às restrições da época, tinham pouco acesso a escolas ou eram analfabetas.

GRÁFICO 3: PERCENTUAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – ANÁLISE POR COORTE DE NASCIMENTO



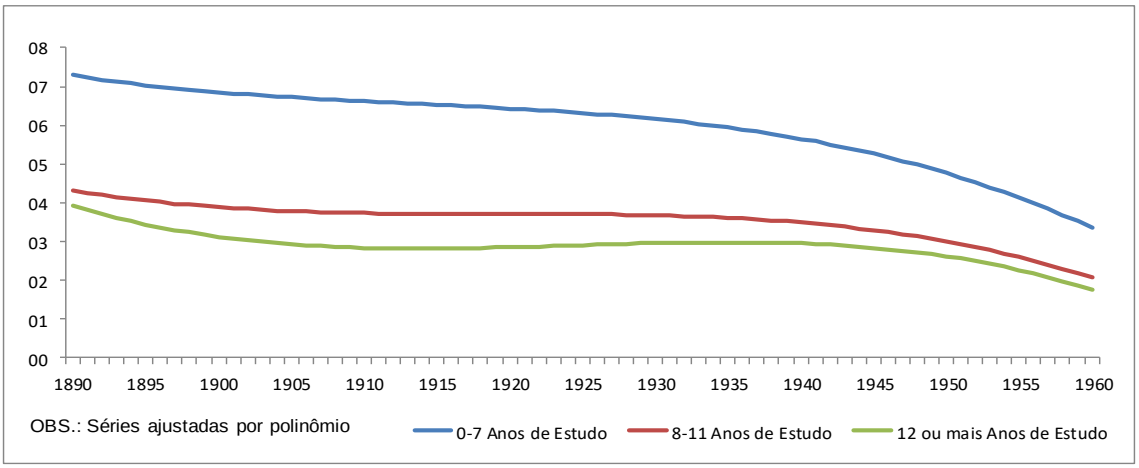
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

[17] As séries de número de filhos por educação, localidade e situação de residência apresentam muitos ruídos. Assim, todas as séries foram suavizadas por ajustes polinomiais e o grau do polinômio foi dado individualmente com o objetivo de maximizar o ajuste de cada série.

[18] Em geral, 60% das mulheres que nasceram no final do século XIX eram analfabetas. Para estatísticas mais desagregadas, 80% das mulheres nascidas no final do século XIX tinham até 3 anos de estudo. A desagregação dos grupos de escolaridade é apresentada no Apêndice 2.

Do total de mulheres nascidas em 1890, apenas 1% tinham 12 anos de estudo ou mais e, na média, 2,2% das mulheres tinham esta escolaridade no final do século XIX. No final da década de 1920, 5% das mulheres tinham de 12 ou mais anos de estudo e, ao longo da década de 30, os incrementos no percentual foram na margem. As coortes nascidas a partir da década de 40 apresentam uma tendência crescente no percentual de mulheres com melhores indicadores educacionais. Um fato histórico que pode estar associado à melhoria dos indicadores educacionais de mulheres para coortes nascidas a partir das décadas de 30 e 40 seria a estruturação do ensino secundário no Brasil através da Lei Orgânica do Ensino Secundário (1942), instituída no governo Getúlio Vargas e que define as finalidades do ensino secundário. Na lei de 1942, consta um artigo específico para o ensino secundário feminino, recomendando escolas femininas para o ensino de mulheres ou, no caso de escolas mistas, a separação de homens e mulheres em salas diferentes[19]. O número médio de filhos por coorte de nascimento varia em relação ao grupo de escolaridade, mas a tendência de queda da fecundidade é presenciada para todos os grupos de escolaridade para coortes nascidas a partir da década de 1930. Em especial, o declínio no número de filhos é maior, em magnitude, para as coortes menos qualificadas. Segundo o Gráfico 4, uma mulher nascida no final do século XIX e com até 7 anos de estudo tinha aproximadamente 7 filhos aproximadamente, e, para o mesmo período, uma mulher com 8 anos ou mais de estudo tinha, na média, 4 filhos.

GRÁFICO 4: Nº FILHOS POR GRUPOS DE ESCOLARIDADE – ANÁLISE POR COORTE DE NASCIMENTO

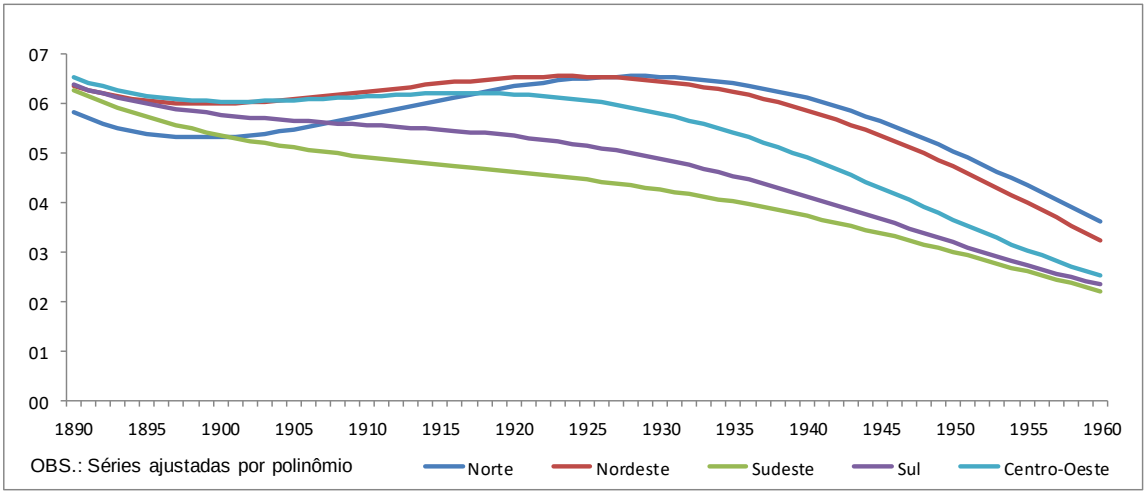


Fonte: Censos 1970, 1980, 1991, 2000 e 2010

[19] Diário Oficial da União de 1942. As estatísticas históricas do século XX documentadas pelo IBGE não apresentam dados mais desagregados para este período, não possibilitando uma investigação mais detalhada dos determinantes do aumento da escolaridade brasileira para coortes nascidas a partir da década de 1930.

Assim como mulheres pouco escolarizadas, as mulheres com melhores indicadores educacionais também apresentaram trajetória decrescente da taxa de fecundidade a partir da década de 1930[20]. Portanto a queda da taxa de fecundidade brasileira é impactada, na transição a partir das coortes nascidas na década de 1930, tanto pelo aumento de mulheres mais escolarizadas, com uma média menor do número de filhos na "*cross-section*", quanto pela queda do número de filhos para todos os grupos de estudo na "*tendência*", com maior magnitude de queda para os grupos menos escolarizados. Uma estratégia empírica convencional tenta estabelecer um experimento natural para identificar o impacto da educação na transição demográfica, como elaborado por Aaronson et al (2014) que usa um programa de construção de escolas para negros que habitam a zona rural, verificando o impacto da variação exógena na educação tanto na margem intensiva (*trade-off* quantidade-qualidade no sentido do Becker & Lewis[21]) quanto na margem extensiva (opção de não ter filho durante o período reprodutivo). No entanto, a queda da fecundidade para diversos grupos de escolaridade ao longo das coortes representa um "*puzzle*" na transição demográfica. Uma possível tentativa seria investigar variáveis relacionadas à transição, como urbanização e localidade.

GRÁFICO 5: Nº FILHOS POR REGIÕES GEOGRÁFICAS – ANÁLISE POR COORTE DE NASCIMENTO



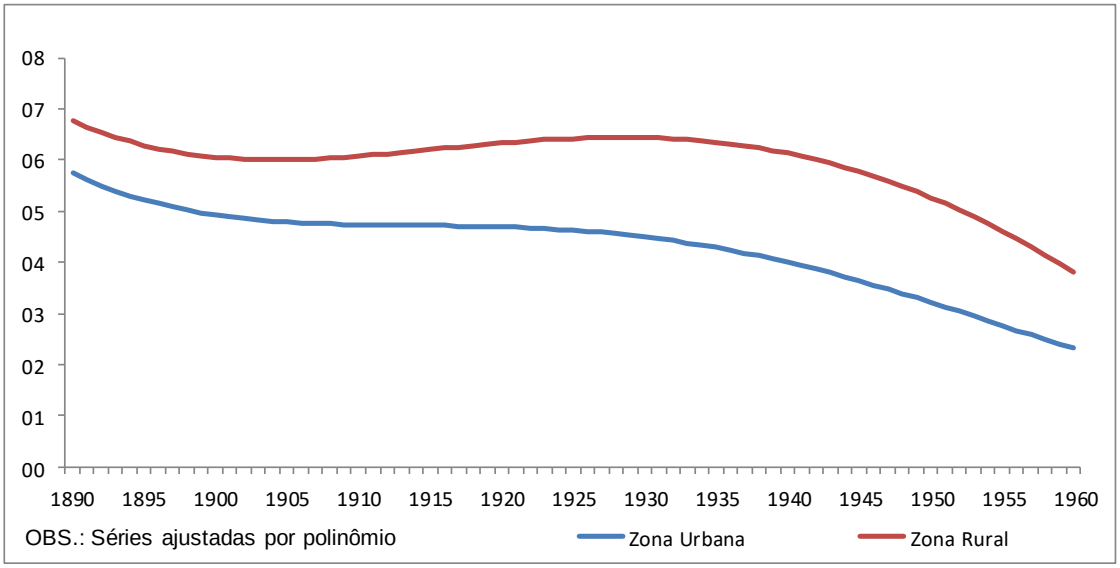
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

[20] Mulheres com 15 anos ou mais de estudo apresentam uma trajetória decrescente mais suave ao longo das coortes. Para uma coorte nascida em 1935, mulheres muito escolarizadas tinham 2,2 filhos, na média. A coorte nascida em 1965 tinha 1,5 filhos, uma variação negativa de 31%.

[21] Becker & Lewis (1974)

O Gráfico 5 apresenta a distribuição geográfica do número médio de filhos por coorte de nascimento. Mulheres residentes nas regiões sul e sudeste apresenta trajetória negativa em sua taxa de fecundidade desde o início da série, no final do século XIX. Mulheres residentes no centro-oeste apresentam trajetória negativa em sua taxa de fecundidade a partir da década de 1920. A região norte, que apresentava uma taxa de fecundidade abaixo da média brasileira para coortes nascidas até o final do século XIX, aumentou sua média de número de filhos para as coortes nascidas entre 1900 e 1930 e, após este período, reverteu o com a transição demográfica e queda da taxa de fecundidade. Para o período de transição com queda da fecundidade, a região nordeste passa a apresentar uma média menor do que a região norte. O Gráfico 6 separa a média do número de filhos por coortes nascidas apenas em situação urbana e coortes nascidas apenas em situação rural[22]. Nativos urbanos apresentam trajetória com queda da fecundidade ao longo da série, enquanto que nativos rurais realizam a transição demográfica a partir das coortes nascidas na década de 1930, como verificado para a média das coortes brasileiras.

GRÁFICO 6: Nº FILHOS POR RURAL-URBANA – ANÁLISE POR COORTE DE NASCIMENTO



Fonte: Censos 1970, 1980, 1991, 2000 e 2010
OBS. Apenas nativos

[22] O Brasil apresentou forte fluxo migratório do campo para a cidade principalmente para as coortes nascidas no começo do século XX, documentadas no Censo de 1970. Para se avaliar o efeito rural x urbano, foram selecionadas apenas mulheres nativas de ambas as situações.

As estatísticas descritivas recontam a história das coortes de mulheres brasileiras nascidas de 1890 a 1960, período de profunda transformação econômica e social. Assim, o Brasil vivenciou uma transição com queda da fecundidade principalmente para as coortes nascidas a partir da década de 1930. A queda do número de filhos não foi exclusiva de grupos mais educados ou moradores de áreas urbanas. As estatísticas descritivas mostraram que o número médio de filhos diminui para mulheres escolarizadas, porém diminui em maior magnitude para mulheres pouco escolarizadas. As regiões Norte e Nordeste, com maior média de filhos por mulheres realizou a transição demográfica posteriormente às regiões Sul e Sudeste e o número de filhos também diminui para nativas rurais.

3.1. Áreas Mínimas Comparáveis

A Tabela 1 apresenta as estatísticas descritivas das 3660 localidades denominadas área minimamente comparável (AMC), um conjunto de municípios comparados ao longo de todos os censos[23]. A média da população de AMC praticamente dobra entre os censos de 1970 e 2010. Este aumento populacional é acompanhado pelo aumento da participação das mulheres de 50 a 80 anos em relação ao total de mulheres de uma AMC que, na média, cresce de 10% para 21% entre os censos. As localidades tornaram-se mais urbanas assim como melhoraram seus indicadores médios de capital humano, com queda do percentual de analfabetos e aumento dos anos médios de estudo por AMC.

[23] A declaração do município de residência sofre mudanças ao longo das décadas por conta tanto da criação de novos municípios quanto pela incorporação de um município a outro. No Censo de 1970 existiam 3660 municípios catalogados pelo IBGE e seus respectivos códigos de identificação. No Censo de 1991, o IBGE catalogou 4268 municípios e, para o Censo de 2010, o IBGE catalogou 5508 municípios. Para que tais municípios fossem fixos e estáveis no tempo, a compatibilização implicaria em verificar a equivalência territorial dos 608 municípios catalogados a mais no Censo de 1991 e dos 1848 municípios catalogados a mais no Censo de 2010. Ao fixarmos os municípios de 1970 como referência, a construção das AMC com base na documentação da divisão territorial do Brasil implicou que, na média, cada AMC tivesse 1,5 municípios.

TABELA 1: ESTATÍSTICAS DESCRITIVAS DAS ÁREAS MINIMAMENTE COMPARÁVEIS

	Censo 1970	**Censo 1980**	**Censo 1991**	**Censo 2000**	**Censo 2010**
Média População	25775	32534	40125	46426	52133
Desvio-Padrão População	134075	182528	211961	235537	258983
% População mulheres 50-80	10%	12%	14%	17%	21%
% população urbana na AMC	33%	44%	55%	63%	68%
% população analfabeta na AMC	48%	49%	34%	22%	16%
Média Filhos por AMC	7.3	7.1	7.0	6.4	5.1
Desvio-Padrão Nº Filhos por AMC	1.1	1.1	1.3	1.4	1.2
ANOS DE ESTUDO					
Média Anos de Estudo por AMC	1.4	2.2	3.0	3.7	-
Média Desvio-Padrão por AMC	1.9	2.6	3.1	3.6	-
GRUPO ETÁRIO 50 - 80 ANOS					
Desvio-Padrão Anos de Estudo no Brasil	2.6	3.0	3.6	4.2	-
Média Desvio-Padrão por AMC	1.3	1.7	2.3	2.9	-
GRUPO ETÁRIO 50 - 60 ANOS					
Desvio-Padrão Anos de Estudo no Brasil	2.6	3.1	3.8	4.4	-
Média Desvio-Padrão por AMC	1.4	1.8	2.5	3.2	-
GRUPO ETÁRIO 60 - 70 ANOS					
Desvio-Padrão Anos de Estudo no Brasil	2.5	2.8	3.4	3.8	-
Média Desvio-Padrão por AMC	1.2	1.5	2.0	2.6	-
GRUPO ETÁRIO 70 - 80 ANOS					
Desvio-Padrão Anos de Estudo no Brasil	2.4	2.7	3.1	3.5	-
Média Desvio-Padrão por AMC	1.0	1.3	1.6	2.1	-
	AMC 1970		**AMC 1991**		**AMC 2010**
Total Municípios	3660		4268		5508
Média Municípios por AMC	1.5		1.3		1.0
Desvio-Padrão Municípios por AMC	2.4		1.2		0.1

Fonte: Censos 1970, 1980, 1991, 2000 e 2010

Diferentemente do apresentado no Gráfico 2, onde o desvio-padrão do número de filhos cai com a queda da media do número de filhos para o Brasil como um todo, o desvio-padrão do número de filhos dentro da AMC[24] apresentado na Tabela 1 varia pouco ao longo dos Censos. Para os grupos etários de 50 a 80 anos, 50 a 60 anos, 60 a 70 anos e 70 a 80 anos, o desvio-padrão dos anos de estudo das AMC são menores do que o desvio-padrão do Brasil como um todo. Este ponto será utilizado para justificar a necessidade de se controlar por efeitos fixos de localidade e a interação entre coorte e localidade, pois dentro do município, ou AMC, as coortes são mais homogêneas. No Apêndice 3 são apresentadas informações agregadas dos municípios correlacionando-os com o número médio de filhos nascidos vivos para mulheres com 50 a 80 anos e apresentados separadamente por censos. Para um conjunto de 3660 observações, quanto maior o percentual de mulheres de 50 a 80 anos analfabetas, maior a média do número de filhos nascidos vivos. Para urbanização, escolaridade e desvio-padrão da escolaridade na AMC, estas variáveis são negativamente relacionadas com a média do número de filhos[25]. Anos médios de estudo e o desvio-padrão dos anos médios de estudo não são apresentados para o Censo de 2010 por conta da impossibilidade de se construir tal variável. Para todos os gráficos de dispersão, o grau de ajuste melhora para os censos demográficos mais recentes.

4. ESTRATÉGIA EMPÍRICA

A estratégia empírica para decompor o número de filhos por grupo de escolaridade segue a literatura de estudo de múltiplos cortes transversais apresentado por Fienberg & Mason (1985)[26]. Estes artigos guiam a literatura que analisa seu objeto de pesquisa por um modelo de idade-período-coorte. A estrutura de dados fornecida pelos Censos demográficos não possibilita acompanhar indivíduos ao longo das pesquisas, como acontece quando há um

[24] Com o total de microdados e utilizando o fator de expansão da amostra, foram calculadas a média e o desvio-padrão das variáveis apresentadas na Tabela 1 para cada AMC. Com isto, estatísticas descritivas foram extraídas deste banco de dados de 3660 observações.

[25] Todas as variáveis analisadas nos gráficos de dispersão referem-se a mulheres de 50 a 80 anos de idade com valores calculados para cada um dos cinco censos demográficos.

[26] Fienberg & Mason apresentam estudos com diversas estruturas de dados. Greenberg, Wright e Sheps (1950), Farkas (1977) e Pullum (1980) são aplicações de modelos de idade-período-corte com múltiplos cortes transversais. Greenberg, Wright e Sheps (1950) analisam a incidência de sífilis com uma base de dados da Carolina do Norte, nos Estados Unidos, com dados disponíveis para os anos de 1941 a 1947 e idades de 15 a 29 anos.

painel. No entanto, ao se definir a coorte a partir da diferença entre o ano da pesquisa e a idade da pessoa entrevistada, podem-se obter dados de diversas coortes ao longo dos censos, ou seja, cada censo demográfico representa um corte transversal e as coortes são identificadas através da definição:[27]

$$coorte_i = censo_i - idade_i$$

Para cada censo foram selecionadas mulheres de 50 a 80 anos de idade. O IBGE considera período fértil para cálculo da taxa de fecundidade de 15 a 49 anos. Assim, potencialmente, uma mulher a partir de 50 anos já teria concluído seu período reprodutivo e o número de filhos declarado será o número de filhos definitivo. A tabela abaixo sintetiza as coortes geradas a partir dos censos e das idades dos indivíduos investigados. As coortes nascidas entre 1890 e 1899 aparecem apenas no censo de 1970 enquanto que as coortes nascidas entre 1951 e 1960 aparecem apenas no censo de 2010. As coortes nascidas entre 1900 e 1950 aparecem em diversos censos sendo que apenas uma parcela da coorte de 1900, por exemplo, sobrevive entre os censos de 1970 e 1981.

[27] A ausência da pergunta "ano de nascimento" impossibilita se determinar precisamente qual o ano exato que o indivíduo nasceu. Esta identificação seria factível para os dados da Pesquisa Nacional por Amostras Domiciliares (PNAD).

TABELA 2: COORTES POR MÚLTIPLOS CORTES TRANSVERSAIS NOS CENSOS 1970-1991

Idade	Período 1970	1980	1991	2000	2010
80	1890	1900	1911	1920	1930
79	1891	1901	1912	1921	1931
78	1892	1902	1913	1922	1932
77	1893	1903	1914	1923	1933
76	1894	1904	1915	1924	1934
75	1895	1905	1916	1925	1935
74	1896	1906	1917	1926	1936
73	1897	1907	1918	1927	1937
72	1898	1908	1919	1928	1938
71	1899	1909	1920	1929	1939
70	1900	1910	1921	1930	1940
69	1901	1911	1922	1931	1941
68	1902	1912	1923	1932	1942
67	1903	1913	1924	1933	1943
66	1904	1914	1925	1934	1944
65	1905	1915	1926	1935	1945
64	1906	1916	1927	1936	1946
63	1907	1917	1928	1937	1947
62	1908	1918	1929	1938	1948
61	1909	1919	1930	1939	1949
60	1910	1920	1931	1940	1950
59	1911	1921	1932	1941	1951
58	1912	1922	1933	1942	1952
57	1913	1923	1934	1943	1953
56	1914	1924	1935	1944	1954
55	1915	1925	1936	1945	1955
54	1916	1926	1937	1946	1956
53	1917	1927	1938	1947	1957
52	1918	1928	1939	1948	1958
51	1919	1929	1940	1949	1959
50	1920	1930	1941	1950	1960

Fonte: Elaboração própria com base em Fienberg & Mason (1985, p. 62)

O modelo empírico segue a equação:

$$Filhos_{i,c,s} = \beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + \sum_{s=1}^{3660} \beta_4^s . Local_i^s + \sum_{c=1890}^{1960} \sum_{s=1}^{3660} \beta_5^{c,s} . Coorte_i^c . Local_i^s + \sum_{j=1}^{5} \beta_6^j . Censo_i^j + e_i \quad (1)$$

Sendo que educação, $Educ_i^j$ é classificada em 0-3 anos de estudo ($j = 1$), 4-7 anos de estudo ($j = 2$), 8-10 anos de estudo ($j = 3$), 11-14 anos de estudo ($j = 4$), e 15 ou mais anos de estudo ($j = 5$). Cada uma destas categorias representa uma *dummy* que assume valores 0, caso o indivíduo não possua tal escolaridade, e 1 caso seja o grupo de escolaridade declarado

pelo indivíduo. O controle por efeito fixo de localidade deve-se às características das coortes serem mais próximas na localidade do que em relação ao conjunto da coorte, como apresentado por estatísticas descritivas para áreas minimamente comparáveis e para o Brasil. A interação entre coorte e localidade captura as mudanças dos efeitos fixos de localidades para diferentes coortes. Esta interação implica em um controle 256.200 dummies, algo computacionalmente inviável na estimação do modelo. Para tanto, os resultados de β_1 e β_2^j podem ser capturados estimando o modelo com desvio em relação à média da coorte na localidade (AMC). No Apêndice 1 são apresentados os resultados de modelos comparando a estimação da equação original e a mesma equação estimada como diferença tanto em relação à média da coorte quanto para a diferença em relação à média da localidade. Aplicando o operador de esperança em $Filhos_{i,c,s}$ condicional na coorte c e na localidade s, tem-se

$$
\begin{aligned}
&E\left[Filhos_{i,c,s}\,|c,s\right] \\
&\quad = E\left[\beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + \sum_{s=1}^{3660} \beta_4^s . Local_i^s \right. \\
&\quad \left. + \sum_{c=1890}^{1960} \sum_{s=1}^{3660} \beta_5^{c,s} . Coorte_i^c . Local_i^s + \sum_{j=1}^{5} \beta_6^j . Censo_i^j + \; e_i \mid c,s \right] \\
&\quad = \beta_0 + \beta_1 E[X_i|c,s] + \sum_{j=1}^{5} \beta_2^j . E\left[Educ_i^j|c,s\right] \\
&\quad + \sum_{c=1890}^{1960} \beta_3^c E[Coorte_i^c|c,s] \\
&\quad + \sum_{s=1}^{3660} \beta_4^s . E[Local_i^s|c,s] \\
&\quad + \sum_{c=1890}^{1960} \sum_{s=1}^{3660} \beta_5^{c,s} . E[Coorte_i^c . Local_i^s|c,s] + \sum_{j=1}^{5} \beta_6^j . E\left[Censo_i^j|c,s\right] \\
&\quad + \; E[e_i|c,s]
\end{aligned}
$$

Por linearidade da expectativa condicional[28], tem-se, $E[Coorte_i^c|c,s] = Coorte_i^c$, $E[Local_i^s|c,s] = Local_i^s$ e $E[Coorte_i^c.Local_i^s|c,s] = Coorte_i^c.Local_i^s$. A Lei Fraca dos Grandes Números[29] implica que $E\left[Filhos_{i,c,s}\,|c,s\right] \xrightarrow{p} \frac{1}{N_{c,s}}\sum_{i=1}^{N_{c,s}} Filhos_{i,c,s} = \overline{Filhos}_{c,s}$, onde $\xrightarrow{p}$ denota convergência em probabilidade, $N_{c,s}$ refere-se ao total de pessoas da coorte *c* na localidade *s* e $\overline{Filhos}_{c,s}$ é a média do número de filhos dos indivíduos nascidos na coorte *c* e residentes da localidade *s*, $E[X_i|c,s] \xrightarrow{P} \frac{1}{N_{c,s}}\sum_{i=1}^{N_{c,s}} X_i = \bar{X}_{c,s}$, onde $\bar{X}_{c,s}$ refere-se à média do vetor X em relação à coorte *c* e localidade *s*, $E\left[Educ_i^j|c,s\right] \xrightarrow{p} \frac{1}{N_{c,s}}\sum_{i=1}^{N_{c,s}} Educ_i^j = \overline{Educ}_{c,s}^j$ para todo $j = 1, \dots, 5$, $\overline{Educ}_{c,s}^j$ representa a média da variável $Educ^j$ em relação à coorte *c* e localidade *s*. A hipótese de exogeneidade implica que $E[e_i|c,s] = 0$. Assim, a equação (1) pode ser estimada como diferença em relação à média da coorte *c* na localidade *s*, $\left(Filhos_{i,c,s} - \overline{Filhos}_{c,s}\right)$

$$\begin{aligned}\left(Filhos_{i,c,s} - \overline{Filhos}_{c,s}\right) &= \beta_1\left(X_i - \bar{X}_{c,s}\right) + \sum_{j=1}^{5} \beta_2^j.\left(Educ_i^j - \overline{Educ}_{c,s}^j\right) \\ &+ \sum_{j=1}^{5} \beta_6^j.\left(Censo_i^j - \overline{Censo}_{c,s}^j\right) + e_i\end{aligned} \tag{2}$$

5. RESULTADOS

A Tabela 3 apresenta os resultados das regressões sem controlar por localidade de residência e sem interação entre localidade e coorte. As *dummies* de censos sugerem que uma vez fixada a coorte de nascimento, pertencer aos censos de 1980, 1991 e 2000 impacta negativamente em 0,15 o número médio de filhos se comparado com mulheres pertencentes ao censo de 1970. O Gráfico 1 demonstrou que a média dos anos de estudo de uma coorte são inferiores no censo de 2010 em relação à mesmo coorte em outros censos. No entando, uma vez controlado por educação, a diferença da coorte entre 2010 e 1970 diminui. Na coluna (2) são adicionados, às

[28] Como define Hayashi (2000, p.9), linearidade da expectativa condicional implica que $E[f(x)y|x] = f(x)E[y|x]$

[29] Wooldridge (2010, p. 42) é uma referência para Lei Fraca dos Grandes Números e convergência em probabilidade.

dummies de coorte e censos, os controles de educação. Uma vez controlado por educação, os efeitos fixos de coorte diminuem principalmente para as coortes nascidas a partir da década de 1920, quando as estatísticas descritivas evidenciam melhoras nos indicadores de capital humano. Uma vez controlado por coorte e censo, ter de 4 a 7 anos de estudo diminui o número de filho em 1,7, na média, se comparado com mulheres com 0 a 3 anos de escolaridade. As coortes nascidas a partir da década de 1920 apresentaram um aumento percentual de mulheres tanto com 8 a 10 anos de estudo quanto de 11 a 14 anos de estudo.

Em relação ao grupo de escolaridade de 8 a 10 anos de estudo, aproximadamente 4,6% das mulheres nascidas em 1920 tinham de 8 a 10 anos de estudo e 4% das mulheres da mesma coorte tinham de 11 a 14 anos de estudo. Estes mesmos grupos de escolaridade representavam 6,4% e 6% das mulheres pertencentes às coortes de 1930, respectivamente, 8,8% e 10% das coortes nascidas em 1940, 11,7% e 15% das coortes nascidas em 1950 e, por fim, 15,1% e 20% das coortes nascidas em 1960. A coluna 3 da tabela 3 acrescenta à regressão religião e *dummy* de estrangeiro, que é mais representativo nas coortes nascidas na última década do século XIX, período em que a história econômica caracteriza a população como aberta dado o fluxo migratório internacional. Na média, 12% das mulheres nascidas entre 1890 e 1900 eram estrangeira, 8% das mulheres nascidas entre 1900 e 1910 eram estrangeiras e, este percentual, cai para 4% das mulheres nascidas entre 1910 e 1920, segundo o Censo de 1970. Os resultados mostram que ser estrangeiro impacta negativamente em 1,6 na média do número de filhos se comparado com brasileiras[30], resultado estatisticamente significativo a 1%. Ser católico ou evangélico aumenta a média do número de filhos se comparado com quem declara não ter religião, com magnitude maior para evangélicos do que para católicos. Outras religiões, definidas como espírita, budista e religiões africanas, na média tem 0,24 filhos a menos do que quem declara não ter religião. O aumento do R^2 de 0,008 para 0,174 entre os modelos (1) e (2) demostra o quão importante educação é na explicação do número de filhos por mulheres. No entanto, a introdução de outros controles não impacta tanto o R^2 na comparação dos modelos (2) e (3).

[30] São consideradas brasileiras apenas declarantes brasileiras natas. Brasileiras naturalizadas são consideradas estrangeiras tanto nas estatísticas descritivas quanto nas regressões.

TABELA 3: REGRESSÕES COM EFEITOS FIXOS DE COORTE

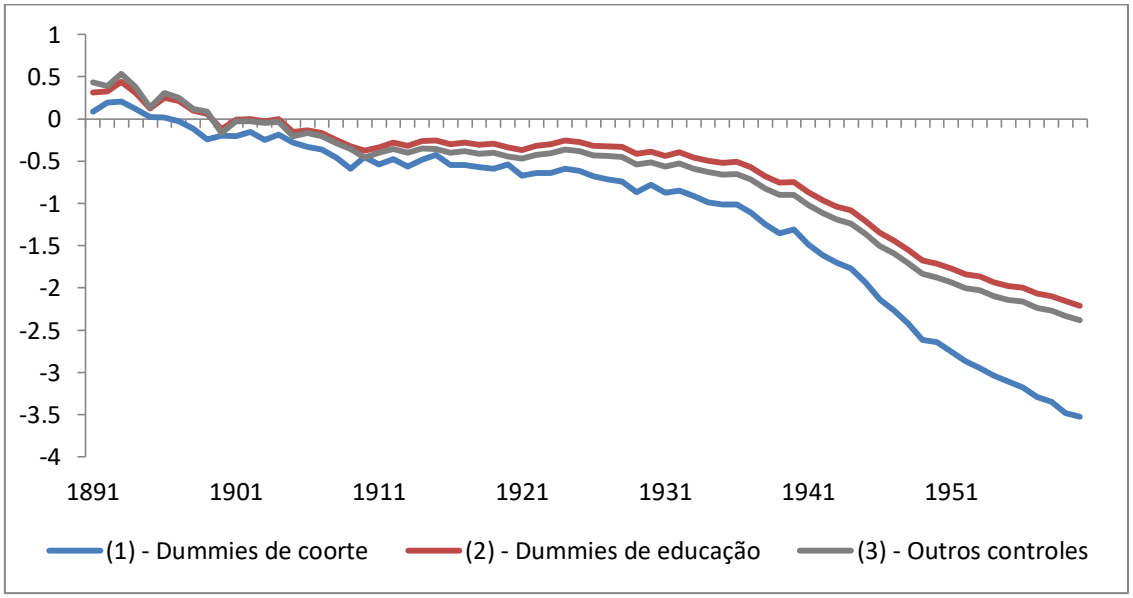
VARIÁVEIS	(1) Dummies de Coorte	(2) Dummies Educacão	(3) Outros controles
Censo 1970	*Omitido*	*Omitido*	*Omitido*
Censo 1980	-0.148***	-0.123***	-0.129***
	(0.00260)	(0.00247)	(0.00246)
Censo 1991	-0.178***	-0.115***	-0.125***
	(0.00297)	(0.00283)	(0.00282)
Censo 2000	-0.155***	-0.0862***	-0.106***
	(0.00319)	(0.00304)	(0.00303)
Censo 2010	-0.371***	-0.210***	-0.216***
	(0.00339)	(0.00323)	(0.00323)
Educação 0-3		*Omitido*	*Omitido*
Educação 4-7		-1.738***	-1.695***
		(0.00123)	(0.00123)
Educação 8-10		-2.448***	-2.369***
		(0.00195)	(0.00195)
Educação 11-14		-2.772***	-2.682***
		(0.00186)	(0.00187)
Educação 15+		-3.064***	-2.955***
		(0.00247)	(0.00248)
Sem Religião			*Omitido*
Católico			0.218***
			(0.00339)
Evangélico			0.425***
			(0.00356)
Outras religiões			-0.239***
			(0.00384)
Estrangeiro			-1.593***
			(0.00364)
Constante	6.972***	7.177***	7.108***
	(0.0148)	(0.0140)	(0.0144)
Nº Observações	47,141,183	47,141,183	47,141,183
R^2	0.088	0.174	0.179
Dummy de Censo	**SIM**	**SIM**	**SIM**
Dummy de Região	NÃO	NÃO	NÃO
Dummy de UF	NÃO	NÃO	NÃO
Dummy de AMC	NÃO	NÃO	NÃO
Coorte x AMC	NÃO	NÃO	NÃO

Erro-Padrão entre parênteses

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

O gráfico 7 apresenta os efeitos fixos de coorte da regressões (1), apenas controlando por coortes de nascimento, (2), adicionando educação e (3), adicionando outros controles. Nas três regressões, os efeitos fixos de coorte sugerem uma tendência à queda do número médio de filhos para coortes nascidas a partir da década de 1930. No entanto, uma vez controlando por educação, a magnitude do efeito fixo da coorte é menor para estas coortes. Assim, não controlando por educação, nascer na coorte de 1958 implica em ter, na média, 3,5 filhos a menos do que uma mulher que nasceu em 1890. Controlando por educação, o impacto de nascer na coorte de 1958 é de 2,1 filhos a menos do que mulheres que nasceram em 1890.

GRÁFICO 7: EFEITOS FIXOS DE COORTE – REGRESSÕES (1) – (3)[31]



OBS.: O resultado é relativo à coorte de 1890

[31] Para o modelo (1), *dummies* de coorte, a coorte de 1891 apresentou *p-value* de 0,03, a coorte de 1895 *p-value* de 0,202, a coorte de 1896 *p-value* de 0,522 e a coorte de 1897 *p-value* de 0,219. Para este mesmo modelo, as outras coortes apresentaram *p-value* de zero.
Para o modelo (2), *dummies* de educação, a coorte de 1899 apresentou *p-value* de 0,03, a coorte de 1901 apresentou *p-value* de 0,639, a coorte de 1902 apresentou *p-value* de 0,851, a coorte de 1903 *p-value* de 0,074 e a coorte de 1904 *p-value* de 0,83. Para o mesmo modelo, as outras coortes apresentaram *p-value* zero.
Para o modelo (3), outros controles, a coorte de 1901 apresentou *p-value* de 0,123, a coorte de 1902 *p-value* de 0,066, a coorte de 1903 *p-value* de 0,008 e a coorte de 1904 *p-value* de 0,044. As outras coortes apresentaram *p-value* zero.

A tabela 4 apresenta os resultados das regressões com grupos de educação e adicionando tipos diferentes de controles de localidade. A última coluna, denominada *coorte x amc*, representa a regressão da equação (2) acima definida no modelo econométrico. Ao controlar por efeito fixo de localidade, definidos como regiões geográficas, unidades da federação ou áreas mínimas comparáveis, o impacto da educação diminui em magnitude para todos os grupos de escolaridade. Comparado com a região norte, uma mulher com a mesma escolaridade e mesma coorte tem 0.159 filhos a mais no Nordeste, 1,54 filhos a menos no Sudeste, 1,33 filhos a menos no Sul e 0,71 filhos a menos no Centro-Oeste. Ao se comparar os modelos, verifica-se que a introdução do controle por localidade aumenta o R^2 para os modelos (5) e (6) se comparados com o modelo (4). O modelo (7), se estimado com a subamostra de 5% do universo sem desvios em relação à média, como apresentado no apêndice, apresenta um R^2 de 0,25. Ao interagir *dummies* de coorte com *dummies* de AMC, os resultados dos parâmetros de educação são marginalmente próximos do modelo sem interação (7)[32].

[32] Os resultados apresentados na Tabela 4 para as estimações (7) e (8) não controlam por *dummy* de nativo e por *dummy* de urbano para se preservar a comparação com os resultados apresentados no apêndice. Ao adicionar dummies de nativo e urbano ao modelo (7), os parâmetros Educação 4-7, Educação 8-10, Educação 11-14 e Educação 15+ são, respectivamente, -1.09, -1.613, -1.971 e -2.159, variando na margem. O R2 ajustado do modelo aumenta de 0,1419 para 0,1434 adicionando tais controles. Os controles sinalizam que uma mulher da mesma coorte, no mesmo município e com a mesma escolaridade tem, na média, menos 0,4 filhos se nativa e menos 0,41 filhos se morar na parte urbana de tal município. Ao adicionar dummies de nativo e urbano ao modelo (8) os resultados encontrados foram similares na margem ao modelo (7). Todas as *dummies* de educação, nativo e urbano apresentaram p-valor zero nas duas estimações.

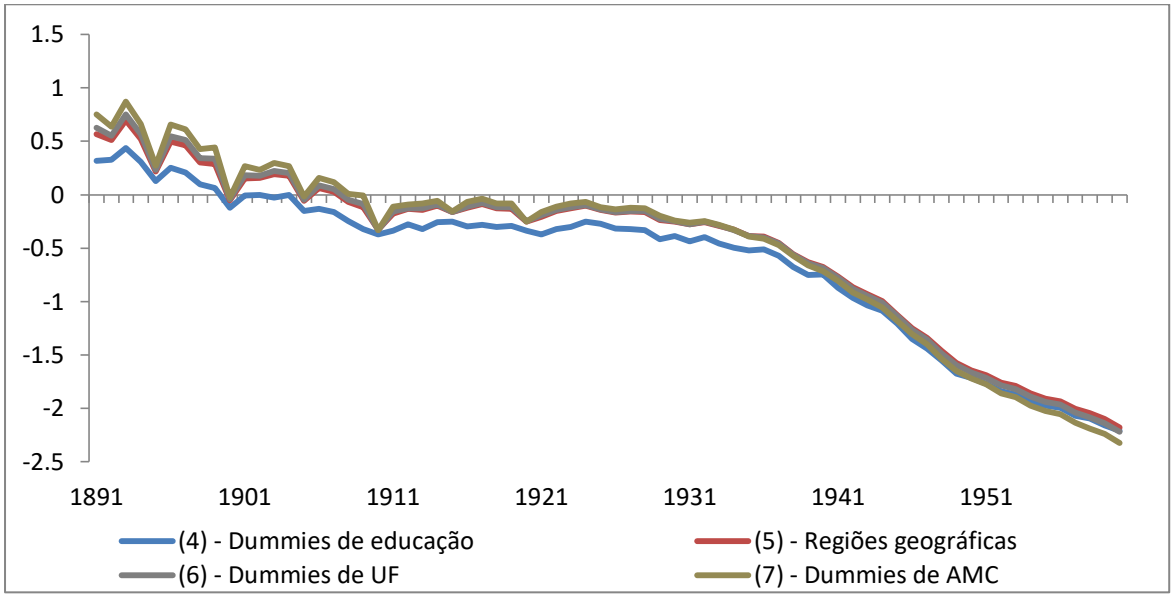
TABELA 4: REGRESSÕES COM EFEITOS FIXOS DE COORTE E LOCALIDADE

VARIÁVEIS	(4) Dummies Educação	(5) Regioes Geográficas	(6) Dummies UF	(7) Dummies AMC	(8) Coorte x AMC
Censo 1970	*Omitido*	*Omitido*	*Omitido*	*Omitido*	*Omitido*
Censo 1980	-0.123***	-0.147***	-0.140***	-0.125***	-0.118***
	(0.00247)	(0.00241)	(0.00239)	(0.00236)	(0.00234)
Censo 1991	-0.115***	-0.151***	-0.148***	-0.130***	-0.134***
	(0.00283)	(0.00276)	(0.00274)	(0.0027)	(0.0027)
Censo 2000	-0.0863***	-0.126***	-0.122***	-0.103***	-0.111***
	(0.00304)	(0.00296)	(0.00294)	(0.0029)	(0.0029)
Censo 2010	-0.210***	-0.257***	-0.265***	-0.257***	-0.260***
	(0.00323)	(0.00315)	(0.00313)	(0.00309)	(0.00305)
Educação 0-3	*Omitido*	*Omitido*	*Omitido*	*Omitido*	*Omitido*
Educação 4-7	-1.738***	-1.422***	-1.358***	-1.134***	-1.081***
	(0.00123)	(0.00122)	(0.00121)	(0.00122)	(0.00120)
Educação 8-10	-2.448***	-2.123***	-1.999***	-1.665***	-1.636***
	(0.00195)	(0.00191)	(0.00191)	(0.00191)	(0.00188)
Educação 11-14	-2.772***	-2.507***	-2.409***	-2.032***	-2.027***
	(0.00186)	(0.00183)	(0.00182)	(0.00184)	(0.00181)
Educação 15+	-3.064***	-2.738***	-2.636***	-2.222***	-2.228***
	(0.00247)	(0.00242)	(0.00241)	(0.00242)	(0.00238)
Norte		*Omitido*			
Nordeste		0.159***			
		(0.00248)			
Sudeste		-1.541***			
		(0.00241)			
Sul		-1.330***			
		(0.00258)			
Centro-Oeste		-0.711***			
		(0.00302)			
Constante	7.177***	7.916***	8.064***	*Omitido*	*Omitido*
	(0.0140)	(0.0139)	(0.0153)		
Nº Observações	47,140,075	47,140,075	47,140,075	47,140,075	47,140,075
R^2	0.174	0.214	0.227	0.140	0.047
Dummy de Censo	**SIM**	**SIM**	**SIM**	**SIM**	**SIM**
Dummy de Região	NÃO	**SIM**	NÃO	NÃO	NÃO
Dummy de UF	NÃO	NÃO	**SIM**	NÃO	NÃO
Dummy de AMC	NÃO	NÃO	NÃO	**SIM**	**SIM**
Coorte x AMC	NÃO	NÃO	NÃO	NÃO	**SIM**

Erro-Padrão entre parênteses

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

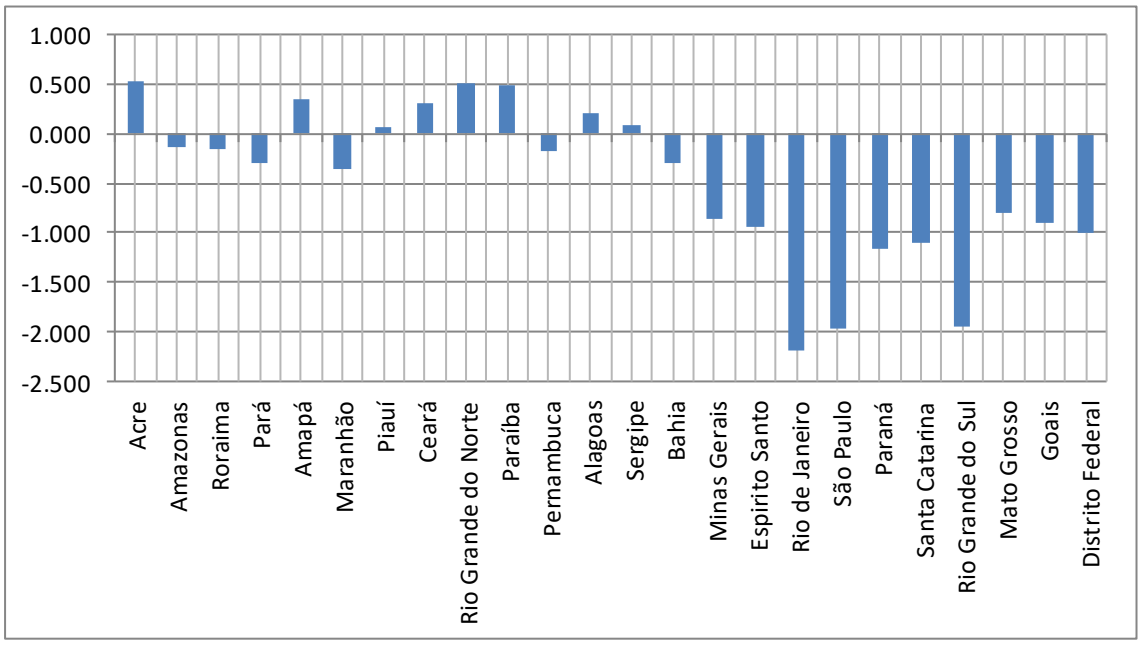
GRÁFICO 8: EFEITOS FIXOS DE COORTE – REGRESSÕES (4) – (7)



OBS.: O resultado é relativo à coorte de 1890

O gráfico 8 apresenta dos resultados dos efeitos fixos de coorte para as regressões (4) a (7). Ressalta-se aqui que ao se estimar a equação como desvio em relação à média da coorte na localidade, especificação da equação (2) acima, perdem-se os efeitos fixos de coortes. Os resultados mostram que uma vez controlando por educação, a escolha do controle adicional de localidade apresenta resultados similares. Assim, a magnitude do parâmetro do efeito da coorte é próxima se controlando a localidade por regiões geográficas, se controlando por unidades da federação ou se controlando por áreas minimamente comparáveis. Em comparação com o efeito da coorte sem controle de localidade, estes se mostram diferentes para as coortes nascidas antes da década de 1940. O gráfico 9 apresenta os resultados das *dummies* de unidade da federação para o modelo (6), na tabela 4. Os resultados são relativos ao estado de Rondônia e controlando por educação e coorte, ser residente de um estado do Norte ou do Nordeste impacta no máximo em 0,5 filhos a mais na média (casos do Acre, Rio Grande do Norte e Paraíba) ou em 0,3 filhos a menos na média (casos de Pará, Maranhão, e Bahia). Já as regiões Sudeste e Sul apresentam um impacto negativo considerável no número médio de filhos. Na média, residir no Rio de Janeiro, em São Paulo ou no Rio Grande do Sul, para uma mesma coorte com mesma escolaridade, tem 2 filhos a menos comparado com Rondônia.

GRÁFICO 9: EFEITOS FIXOS DE UNIDADE DA FEDERAÇÃO – REGRESSÃO (6)

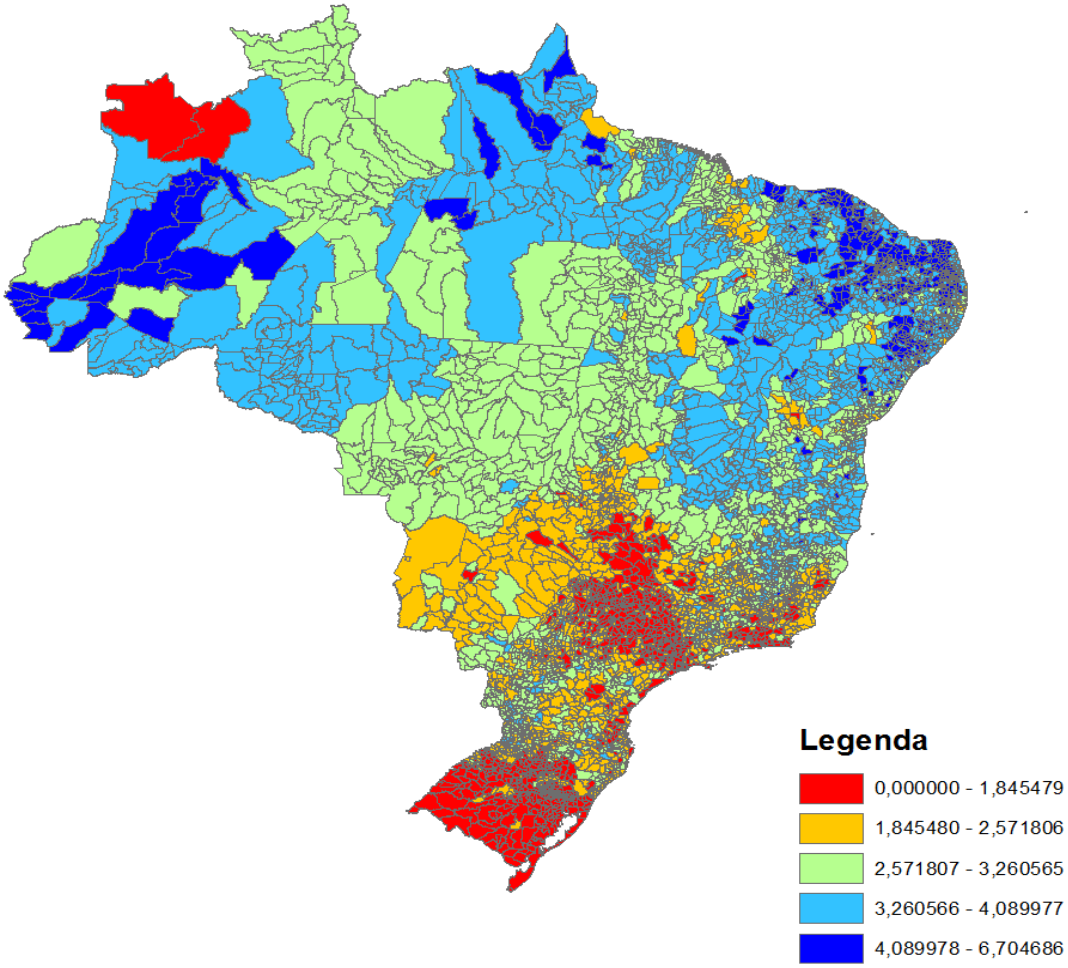


OBS.: O resultado é relativo a Rondônia

Os resultados dos efeitos fixos de localidades, aqui expressas por áreas minimamente comparáveis (AMC), são apresentados no mapa abaixo[33]. A referência para análise é a AMC de Cândido Mendes, escolhida por ser o menor valor relativo na estimação. Assim, fixado Cândido Mendes como referência, todos os valores relativos de *dummies* de AMC terão valores maiores ou iguais a zero. Uma vez controlado por educação, coorte de nascimento efeito fixo de censo, residir na região Norte do Brasil aumenta, na média, em 3,33 o número de filhos e residir na região Nordeste aumenta, na média, em 3,74 o número de filhos. O mapa dos efeitos fixos, apresentado no Gráfico 10, mostra como estes efeitos são distribuídos em todos os municípios.

[33] Resultados dos parâmetros das 3559 áreas minimamente comparáveis na regressão (7) da Tabela 4. Cândido Mendes foi escolhida por apresentar o menor valor relativo e, com isto, os efeitos fixos variam entre 0 e 6,7.

GRÁFICO 10 – RESULTADOS DOS EFEITOS FIXOS DE LOCALIDADE DA REGRESSÃO (MODELO 5)



OBS.: Resultado relativo a Cândido Mendes (controlado por coorte, educação e censo).

TABELA 5: ESTATÍSTICAS DESCRITIVAS DOS EFEITOS FIXOS DE AMC

Parâmetros	Nº Obs	Média	Desvio-Padrão	Mínimo	Máximo
Brasil	5618	2.8586	0.9511	0.0000	6.7047
Região Norte	449	3.3349	0.4295	1.4765	4.8629
Região Nordeste	1795	3.7420	0.7026	1.6522	6.7047
Região Sudeste	1668	2.3191	0.7584	0.0000	4.2734
Região Sul	1188	2.1946	0.6529	0.5963	3.7442
Região Centro-Oeste	518	2.6442	0.4656	1.0425	3.8477

OBS.; Resultado relativo a Cândido Mendes (Controlado por coorte, educação e censo)

* Cândido Mendes apresenta o menor valor relativo da dummy de AMC

A variável “anos médios de estudo” é construída apenas para os censos de 1970, 1980, 1991 e 2000 e, por mudanças na agregação da série cursada, não é construída para o censo de 2010. Assim, ao se utilizar “anos médios de estudo” em vez de grupos de escolaridade, a amostra exclui as observações do censo de 2010 e, consequentemente, as coortes nascidas entre 1951 e 1960. A tabela 6 refaz os exercícios econométricos da tabela 4 controlando por anos médios de estudo e, de imediato, verifica-se que a *dummy* de censo de 2010 não é utilizada assim como o número de observações diminui de 47.140.075 para 30.446.902, uma redução de 35%. A magnitude de um ano de estudo a mais em relação ao número médio de filhos diminui ao se controlar por localidades mais desagregadas. No entanto, ao se interagir coorte com localidade, coluna referente ao modelo (8), o parâmetro de anos de estudo é marginalmente próximo do parâmetro apresentado quando não se faz a interação, coluna referente ao modelo (7). Assim, controlando por coorte, *dummies* de pesquisas censitárias, localidade expressa em áreas minimamente comparáveis, que representam a menor unidade de desagregação comparável no tempo, e interagindo as coortes nas localidades, um ano a mais de estudo implica, segundo o modelo especificado, em 0,24 filhos a menos. Na média, uma mulher nascida na coorte de 1890 tinha 0,77 anos de estudo enquanto que uma mulher nascida na coorte de 1950 tinha, na média, 5,29 anos de estudo. Esta variação de 4,52 filhos na média entre as coorte, segundo o efeito marginal do modelo, resultaria em uma queda de 1,08 filhos, o que representaria 33% da variação do número de filhos entre as coortes de 1890 e 1950.

TABELA 6: EDUCAÇÃO EM ANOS MÉDIOS DE ESTUDO

VARIÁVEIS	(4) Dummies Educação	(5) Regioes Geográficas	(6) Dummies UF	(7) Dummies AMC	(8) Coorte x AMC
Censo 1970	*Omitido*	*Omitido*	*Omitido*	*Omitido*	*Omitido*
Censo 1980	-0.111***	-0.136***	-0.129***	-0.117***	-0.110***
	(0.00270)	(0.00264)	(0.00261)	(0.00258)	(0.00256)
Censo 1991	-0.0894***	-0.130***	-0.129***	-0.121***	-0.120***
	(0.00309)	(0.00303)	(0.00300)	(0.00296)	(0.00294)
Censo 2000	0.00741**	-0.0440***	-0.0431***	-0.0432***	-0.0463***
	(0.00335)	(0.00328)	(0.00324)	(0.00321)	(0.00317)
Censo 2010	-	-	-	-	-
Anos Estudo	-0.339***	-0.296***	-0.282***	-0.239***	-0.235***
	(0.000193)	(0.000192)	(0.000193)	(0.000202)	(0.000199)
Norte		*Omitido*			
Nordeste		0.314***			
		(0.00350)			
Sudeste		-1.434***			
		(0.00341)			
Sul		-1.118***			
		(0.00364)			
Centro-Oeste		-0.491***			
		(0.00431)			
Constante	7.218***	7.821***	8.280***	*Omitido*	*Omitido*
	(0.0153)	(0.0153)	(0.0182)		
Nº Observações	30,446,902	30,446,902	30,446,902	30,446,902	30,446,902
R^2	0.119	0.156	0.174	0.077	0.044
Dummy de Censo	**SIM**	**SIM**	**SIM**	**SIM**	**SIM**
Dummy de Região	NAO	**SIM**	NAO	NAO	NAO
Dummy de UF	NAO	NAO	**SIM**	NAO	NAO
Dummy de AMC	NAO	NAO	NAO	**SIM**	**SIM**
Coorte x AMC	NAO	NAO	NAO	NAO	**SIM**

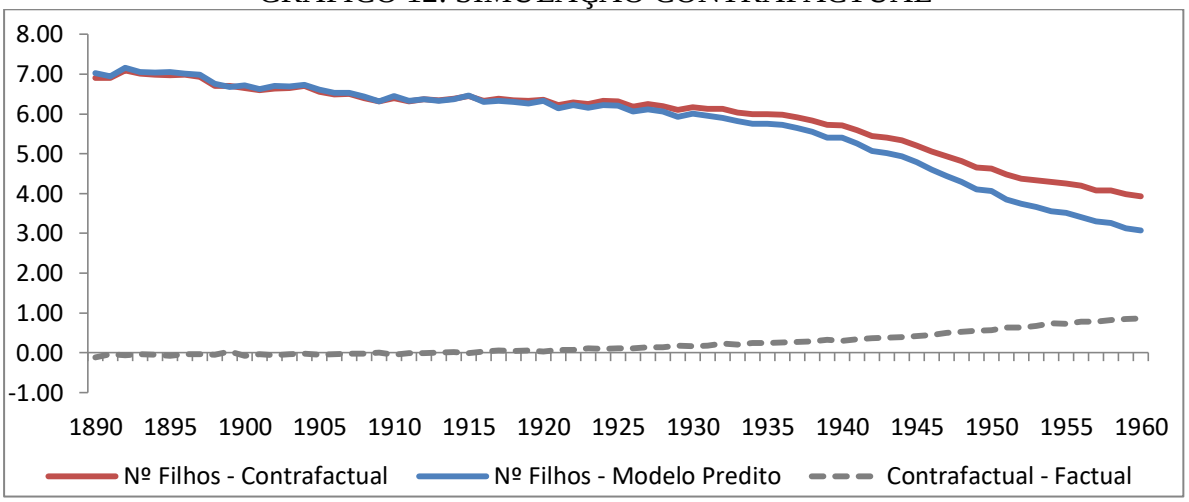
Erro-Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

Obs.: Anos de estudo não disponível no censo de 2010

6. SIMULAÇÃO CONTRAFACTUAL

Pare se realizar a simulação do efeito da escolaridade na transição demográfica brasileira, um exercício de Monte Carlo foi realizado para cada município (ou AMC) utilizando a distribuição por escolaridade da coorte de 1890 e mantendo esta proporção fixa ao longo das coortes futuras. No Apêndice 2 são apresentadas as duas situações, a factual que é observada nos dados dos censo e a contrafactual, em que todas as coortes entre 1890 e 1960 mantém 80% de suas mulheres com até 3 anos de estudo[34]. O modelo factual é gerado pela estimação (7) da Tabela 4 com os valores observados das *dummies* de escolaridade e o modelo contrafactual é gerado a partir da mesma estimação, porém as *dummies* de escolaridade contrafactual foram geradas a partir de distribuições binomiais que reproduzem a escolaridade da coorte de 1890 no município. O Gráfico 12 apresenta o resultado do modelo factual, que reproduz a taxa de fecundidade por coorte e o modelo contrafactual, supondo que a distribuição por escolaridade das mulheres brasileiras não mudasse. A coorte nascida em 1960, na média, tinha 3,07 filhos e, para esta mesma coorte, a situação contrafactual ensejaria em 3,99 filhos na média. Assim, o diferencial de educação explica 30% da diminuição do número de filhos, na média.

GRÁFICO 12: SIMULAÇÃO CONTRAFACTUAL



OBS.: Contrafactual mantém fixa a distribuição da escolaridade da coorte de 1890 para as coortes posteriores.
OBS2: Linha tracejada: resultado contrafactual menos o resultado factual (observado)

[34] Para cada um dos 3660 municípios, calculou-se a média das distribuições de mulheres com até 3 anos de estudo, de 4 a 7 anos de estudo, de 8 a 10 anos de estudo, de 11 a 14 anos de estudo e com 15 ou mais anos de estudo, todas estas médias calculadas para a coorte de 1890 no município. Estas médias foram utilizadas como probabilidade da distribuição binomial gerada para cada município. Assim, a situação contrafactual mantém a distribuição da escolaridade para média da coorte de 1890 em cada município. O Gráfico A2.8 apresenta compara a distribuição factual e a distribuição contrafactual.

7. CONCLUSÃO

Este artigo investigou a transição demográfica no Brasil reconstruindo as coortes a partir do final de século XIX. Os dados mostram que o Brasil fez a transição da taxa de fecundidade com as coortes nascidas entre as décadas de trinta e quarenta do século XX. Neste período houve uma profunda transformação nos indicadores educacionais, a probabilidade de nascimento de um filho morto, além dos fluxos migratórios que tornaram o Brasil mais urbano na passagem dos Censos de 1970 para o de 1980. Os resultados empíricos mostram que ao se controlar por localidade, o impacto da educação diminui em magnitude, motivados pelas estatísticas descritivas que mostraram que a heterogeneidade, medida pelo desvio-padrão, apresenta-se menor na área minimamente comparável do que no Brasil como um todo. Ao se analisar anos médios de estudo, o modelo restringe-se às coortes nascidas entre 1890 e 1950 e, como resultado, educação explica 33% da queda da fecundidade entre estas coortes. Se o modelo sem controle de localidade fosse utilizado para se verificar o impacto da educação no número médio de filhos, o resultado seria de 47%. Portanto, este diferencial de 14 pontos percentuais representa parte das características locais que educação está capturando. As simulações contrafactuais evidenciam que mesmo se o Brasil não tivesse alterado a composição educacional das coortes, mantendo as características das coortes de 1890 com 80% das mulheres com até 3 anos de estudo, o Brasil realizaria a transição demográfica a partir das coortes nascidas em 1930, porém a transição seria feita de maneira mais gradual.

APÊNDICE 1: ESPECIFICAÇÕES DOS MODELOS ECONOMÉTRICOS

1.1.EFEITO FIXO DE COORTE

Seja o modelo com efeito fixo de coorte de nascimento dado por:

$$Filhos_{i,c} = \beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + e_i \qquad (1')$$

Aplicando a esperança condicional na coorte, tem-se:

$$E\left[Filhos_{i,c} \,|c\right] = E\left[\beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + e_i \mid c\right]$$

$$= \beta_0 + \beta_1 E[X_i|c] + \sum_{j=1}^{5} \beta_2^j . E\left[Educ_i^j|c\right] + \sum_{c=1890}^{1960} \beta_3^c E[Coorte_i^c|c] + E[e_i|c]$$

Por linearidade da expectativa condicional tem-se $E[Coorte_i^c|c] = Coorte_i^c$. A Lei Fraca dos Grandes Números[35] implica que $E\left[Filhos_{i,c}\,|c\right] \xrightarrow{p} \frac{1}{N_c}\sum_{i=1}^{N_c} Filhos_{i,c} = \overline{Filhos}_c$, onde N_c refere-se ao total de pessoas da coorte *c* e $\overline{Filhos}_c$ é a média do número de filhos dos indivíduos nascidos na coorte *c*, $E[X_i|c] \xrightarrow{P} \frac{1}{N_c}\sum_{i=1}^{N_c} X_i = \bar{X}_c$, onde $\bar{X}_c$ refere-se à média do vetor X em relação à coorte *c*, $E\left[Educ_i^j|c\right] \xrightarrow{p} \frac{1}{N_c}\sum_{i=1}^{N_c} Educ_i^j = \overline{Educ}_c^j$ para todo $j = 1, \ldots, 5$, $\overline{Educ}_c^j$ representa a média da variável $Educ^j$em relação à coorte *c*. A hipótese de exogeneidade implica que $E[e_i|c] = 0$. Assim, a equação (1') pode ser estimada como diferença em relação à média da coorte *c*, $\left(Filhos_{i,c} - \overline{Filhos}_c\right)$

$$\left(Filhos_{i,c} - \overline{Filhos}_c\right) = \beta_1 (X_i - \bar{X}_c) + \sum_{j=1}^{5} \beta_2^j . \left(Educ_i^j - \overline{Educ}_c^j\right) + e_i \qquad (2')$$

[35] Hayashi (2000, p.9) é uma referência para linearidade da expectativa condicional. Wooldridge (2010, p. 42) é referência para Lei Fraca dos Grandes Números.

A tabela abaixo sintetiza o resultado da regressão usando a equação (1'), denominado modelo 1, e a equação (2'), denominado modelo 2. Ambas as regressões são estimadas com a amostra completa e com uma subamostra aleatória equivalente a 5% do universo[36]. Em ambos os casos, foram utilizados o fator de expansão da amostra e o fator de expansão da subamostra, calculado a partir da probabilidade do sorteio por uma distribuição binomial por município. Os resultados na comparação entre os modelos 1 e 2 são idênticos para os parâmetros estimados, o modelo 2 não tem o intercepto (constante) por construção matemática. A soma dos quadrados dos resíduos (SSR: *sum of squares residual*) é praticamente idêntica entre os modelos, mas o R^2 difere com 0.174 para o modelo 1 em ambas as amostras e 0.095 para o modelo 2.

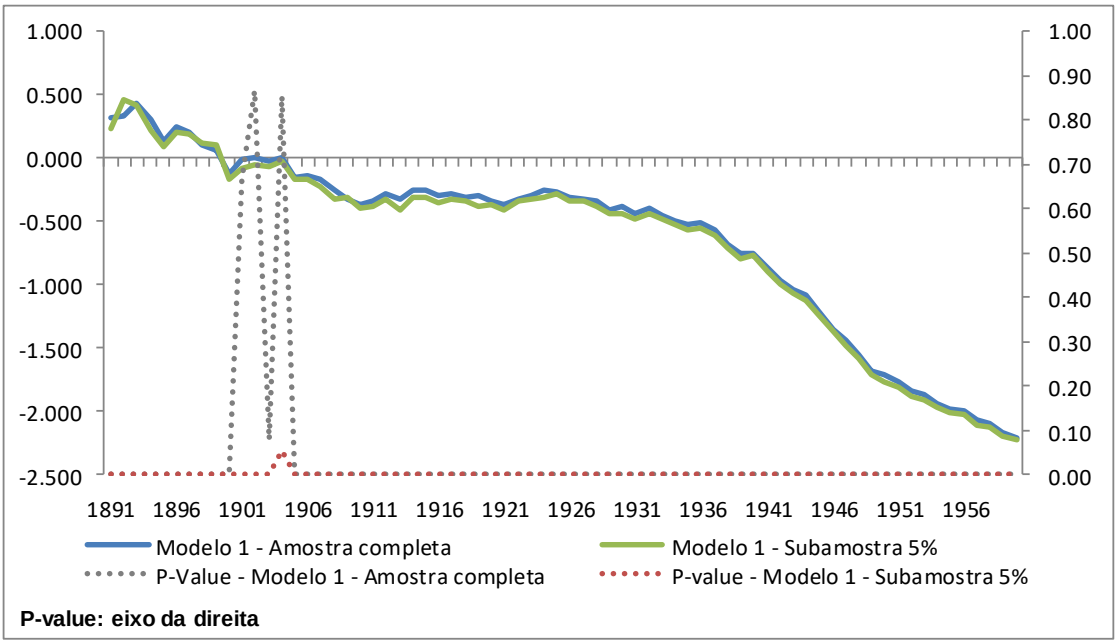
[36] O sorteio aleatório da amostra foi gerado por uma distribuição binomial para cada município com probabilidade calibrada para que o resultado final equivalha a 5% do universo. Para os censos de 1970 e 1980, a probabilidade calibrada foi 0.20, para o censo de 1991, a probabilidade calibrada foi 0.43, para o censo de 2000 a probabilidade calibrada foi 0.42 e, para o censo de 2010, a probabilidade foi 0.46. Para cada censo, o novo peso amostral foi definido pela divisão do peso amostral original pela probabilidade calibrada. Os resultados finais, com o novo fator de expansão amostral, demonstraram-se representativos do universo.

	AMOSTRA COMPLETA			SUBAMOSTRA 5% UNIVERSO	
	MODELO 1	MODELO 2		MODELO 1	MODELO 2
0-3 anos de estudo	Omitido	Omitido		Omitido	Omitido
4-7 anos de estudo	-1.736***	-1.736***		-1.735***	-1.735***
	(0.0012)	(0.0012)		(0.0012)	(0.0012)
8-10 anos de estudo	-2.444***	-2.444***		-2.437***	-2.437***
	(0.0019)	(0.0019)		(0.0019)	(0.0019)
11-14 anos de estudo	-2.769***	-2.769***		-2.766***	-2.766***
	(0.0019)	(0.0019)		(0.0019)	(0.0019)
15+ anos de estudo	-3.062***	-3.062***		-3.060***	-3.060***
	(0.0025)	(0.0025)		(0.0025)	(0.0025)
Censo 1970	*Omitido*	*Omitido*		*Omitido*	*Omitido*
Censo 1980	-0.124***	-0.124***		-0.122***	-0.122***
	(0.0025)	(0.0025)		(0.0025)	(0.0025)
Censo 1991	-0.117***	-0.117***		-0.123***	-0.123***
	(0.0028)	(0.0028)		(0.0028)	(0.0028)
Censo 2000	-0.085***	-0.085***		-0.094***	-0.094***
	(0.0030)	(0.0030)		(0.0030)	(0.0030)
Censo 2010	-0.209***	-0.209***		-0.213***	-0.213***
	(0.0032)	(0.0032)		(0.0032)	(0.0032)
Constante	7.177***	*Omitido*		7.211***	*Omitido*
	(0.0140)	-		(0.0141)	
Nº Obs	47429103	47429103		47409851	47409851
R^2	0.174	0.095		0.174	0.095
SSR	521109528	521109536		520040206	520040206

Erro-Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

A comparação entre a amostra completa e a subamostra equivalente a 5% do universo apresenta valores aproximados para os parâmetros estimados. Os resultados dos efeitos fixos de coortes, apresentados no gráfico abaixo são idênticos para os parâmetros estimados, mas diferem a significância estatística para as coortes nascidas entre 1898 e 1905. Para estas coortes, o *p-value* da regressão da amostra completa mostrou-se não significativo enquanto o *p-value* da regressão da subamostra de 5% do universo não rejeitava a hipótese nula para as mesmas coortes.

GRÁFICO 7.1 – RESULTADO DOS EFEITOS FIXOS DE COORTES PARA A AMOSTRA E PARA A SUBAMOSTRA



OBS.: O resultado é relativo à coorte de 1890.

1.2. EFEITO FIXO DE LOCALIDADE (AMC)

Definindo-se localidade como área minimamente comparável, uma delimitação de 3659 localidades comparáveis entre os censos, o modelo com efeito fixo de localidade é dado por[37]:

$$Filhos_{i,s} = \beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{s=1}^{3660} \beta_4^s . Local_i^s + e_i \qquad (3')$$

Aplicando a esperança condicional na localidade, tem-se:

[37] Nota-se que não há controle para a coorte de nascimento do indivíduo.

$$E\left[Filhos_{i,s}\,|s\right] = E\left[\beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{s=1}^{3660} \beta_4^s . Local_i^s \; + e_i \mid s\right]$$

$$= \beta_0 + \beta_1 E[X_i|s] + \sum_{j=1}^{5} \beta_2^j . E\left[Educ_i^j|s\right] + \sum_{s=1}^{3660} \beta_4^s . E[Local_i^s|s] + E[e_i|s]$$

Por linearidade da expectativa condicional tem-se $E[Local_i^s|s] = Local_i^s$. A Lei Fraca dos Grandes Números implica que $E\left[Filhos_{i,s}\,|s\right] \xrightarrow{p} \frac{1}{N_s}\sum_{i=1}^{N_s} Filhos_{i,s} = \overline{Filhos_s}$, onde N_s refere-se ao total de pessoas na localidade s e $\overline{Filhos_s}$ é a média do número de filhos dos indivíduos residentes na localidade s, $E[X_i|s] \xrightarrow{P} \frac{1}{N_s}\sum_{i=1}^{N_s} X_i = \bar{X}_s$, onde $\bar{X}_s$ refere-se à média do vetor X em relação ao residentes da localidade s, $E\left[Educ_i^j|s\right] \xrightarrow{p} \frac{1}{N_s}\sum_{i=1}^{N_s} Educ_i^j = \overline{Educ}_s^j$ para todo $j = 1, \ldots, 5$, $\overline{Educ}_s^j$ representa a média da variável $Educ^j$ em relação à localidade s. A hipótese de exogeneidade implica que $E[e_i|s] = 0$. Assim, a equação (3') pode ser estimada como diferença em relação à média da localidade s, $\left(Filhos_{i,s} - \overline{Filhos_s}\right)$

$$\left(Filhos_{i,s} - \overline{Filhos_s}\right) = \beta_1(X_i - \bar{X}_s) + \sum_{j=1}^{5} \beta_2^j . \left(Educ_i^j - \overline{Educ}_s^j\right) + e_i \qquad (4')$$

A tabela abaixo apresenta os resultados das regressões comparando o modelo 3 (equação 3') com o modelo 4 (equação 4') apenas para a subamostra de 5% do universo. A comparação entre a amostra e a subamostra é feita apenas com o modelo 4[38]. Os modelos 3 e 4 apresentam resultados idênticos tanto no parâmetro estimado quanto no desvio-padrão, resultado similar na soma dos quadrados dos resíduos (SSR: *sum of squares residual*) mas o R2 apresentou inferior no modelo 4 se comparado com o modelo 3.

[38] O modelo 3 possui 3668 variáveis em uma base de dados com 2.759.722 observações que, com fator de expansão da subamostra, apresenta 47.408.656 observações. A estimação do modelo 3 na amostra completa é computacionalmente custosa, pois a amostra completa possui 8.001.696 observações. O modelo 4 possui 8 variáveis e, devido a adequação do número de variáveis, este modelo foi estimado tanto com a base completa quanto com a subamostra de 5% do universo. Nota-se que, no entanto, o desvio em relação à média foi construído respeitando os fatores de expansão da amostra, no caso da amostra completa, e da subamostra, quando se usa 5% dos dados do universo.

	AMOSTRA COMPLETA			SUBAMOSTRA 5% UNIVERSO	
	MODELO 3	MODELO 4		MODELO 3	MODELO 4
0-3 anos de estudo	-	Omitido		Omitido	Omitido
4-7 anos de estudo	-	-1.244***		-1.244***	-1.244***
		(0.0012)		(0.0012)	(0.0012)
8-10 anos de estudo	-	-1.889***		-1.884***	-1.884***
		(0.0019)		(0.0019)	(0.0019)
11-14 anos de estudo	-	-2.321***		-2.317***	-2.317***
		(0.0018)		(0.0018)	(0.0018)
15+ anos de estudo	-	-2.572***		-2.568***	-2.568***
		(0.0024)		(0.0024)	(0.0024)
Censo 1970	-	*Omitido*		*Omitido*	*Omitido*
Censo 1980	-	-0.211***		-0.2055***	-0.2055***
		(0.0020)		(0.0020)	(0.0020)
Censo 1991	-	-0.419***		-0.4175***	-0.4175***
		(0.0019)		(0.0019)	(0.0019)
Censo 2000	-	-0.844***		-0.8429***	-0.8429***
		(0.0019)		(0.0019)	(0.0019)
Censo 2010	-	-1.669***		-1.662***	-1.662***
		(0.0018)		(0.0018)	(0.0018)
Constante	-	*Omitido*		6.343***	*Omitido*
		-		(0.2147)	
Nº Obs	-	47427995		47408656	47408656
R^2	-	0.120		0.232	0.120
SSR	-	484700000		483105597	483105804

Erro-Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

1.3.EFEITO FIXO DE COORTE E DE LOCALIDADE

Seja o modelo com efeito fixo de coorte de nascimento e efeito fixo de localidade dado por:

$$Filhos_{i,c,s} = \beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + \sum_{s=1}^{3660} \beta_4^s . Local_i^s + e_i \quad (5')$$

Aplicando a esperança condicional na localidade[39], tem-se:

$$E\left[Filhos_{i,c,s}\,|s\right] = E\left[\beta_0 + \beta_1 X_i + \sum_{j=1}^{5} \beta_2^j . Educ_i^j + \sum_{c=1890}^{1960} \beta_3^c . Coorte_i^c + e_i \mid s\right]$$

$$= \beta_0 + \beta_1 E[X_i|s] + \sum_{j=1}^{5} \beta_2^j . E\left[Educ_i^j|s\right] + \sum_{c=1890}^{1960} \beta_3^c E[Coorte_i^c|s]$$

$$+ \sum_{s=1}^{3660} \beta_4^s . E[Local_i^s|s] + E[e_i|s]$$

Por linearidade da expectativa condicional tem-se $E[Local_i^s|s] = Local_i^s$. Assim como no modelo de efeito fixo de localidade, $E\left[Filhos_{i,c,s}\,|s\right] \xrightarrow{p} \frac{1}{N_s}\sum_{i=1}^{N_s} Filhos_{i,c,s} = \overline{Filhos}_s$ representa a média do número de filhos dos indivíduos residentes na localidade s, $E[X_i|s] \xrightarrow{P} \frac{1}{N_s}\sum_{i=1}^{N_s} X_i = \bar{X}_s$, refere-se à média do vetor X em relação ao residentes da localidade s, $E\left[Educ_i^j|s\right] \xrightarrow{p} \frac{1}{N_s}\sum_{i=1}^{N_s} Educ_i^j = \overline{Educ}_s^j$ para todo $j = 1, \ldots, 5$, representa a média de $\overline{Educ}_s^j$ na localidade s e, para as coortes *c* nascidas entre 1890 e 1960, $E[Coorte_i^c|s] \xrightarrow{p} \frac{1}{N_s}\sum_{i=1}^{N_s} Coorte_i^c = \overline{Coorte}_i^c$ representa o percentual da coorte *c* residente na localidade *s*. Assumindo exogeneidade, $E[e_i|s] = 0$, a equação (5') pode ser estimada como diferença em relação à média da localidade s, $\left(Filhos_{i,c,s} - \overline{Filhos}_s\right)$

$$\left(Filhos_{i,c,s} - \overline{Filhos}_s\right)$$

$$= \beta_1(X_i - \bar{X}_s) + \sum_{j=1}^{5} \beta_2^j . \left(Educ_i^j - \overline{Educ}_s^j\right) \qquad (6')$$

$$+ \sum_{c=1890}^{1960} \beta_3^c . (Coorte_i^c - . \overline{Coorte}_s^c) + e_i$$

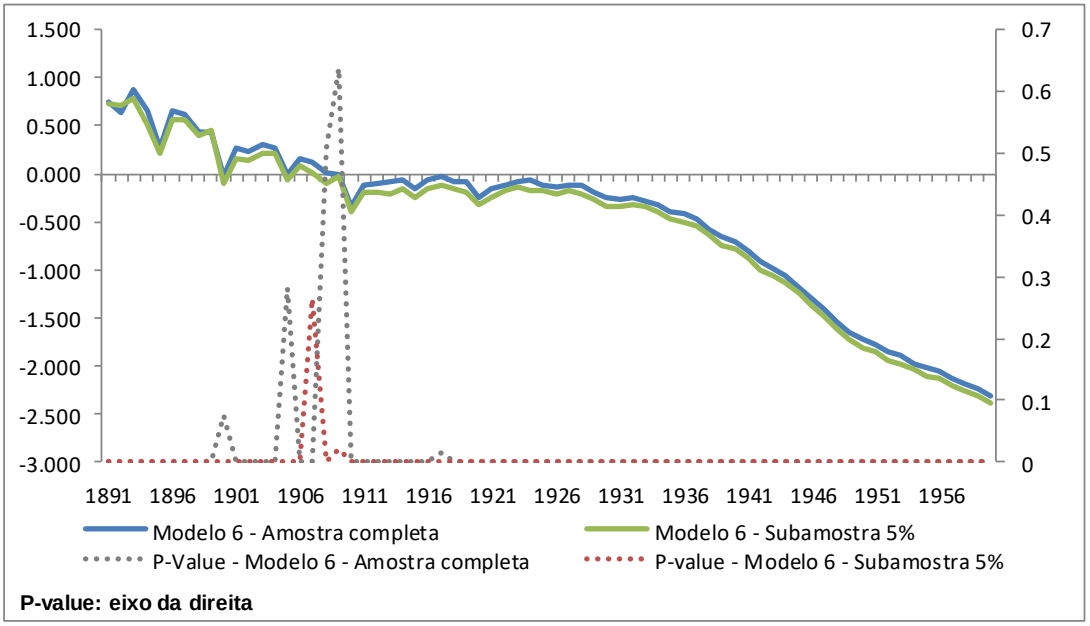
[39] O operador esperança condicional à coorte e localidade, $E\left[Filhos_{i,c,s}\,|c,s\right]$, resolveria o custo computacional de se controlar por 3659 *dummies* de localidades mas eliminaria como consequência o efeito fixo das coortes (70 *dummies* de coortes de nascimento). A equação (5') tem 3.738 variáveis explicativas após a exclusão de 4 variáveis por problemas de multicolinearidade perfeita e computacionalmente é estimada apenas na subamostra de 5% do universo. A equação (6') tem 77 variáveis explicativas após a exclusão de 3 variáveis por problemas de multicolinearidade perfeita, não apresenta intercepto (constante) por construção matemática e, devido à diminuição do custo computacional da estimação, este modelo é estimado tanto para a amostra completa quanto para a subamostra de 5% do universo.

	AMOSTRA COMPLETA			SUBAMOSTRA 5% UNIVERSO	
	MODELO 5	MODELO 6		MODELO 5	MODELO 6
0-3 anos de estudo	-	Omitido		Omitido	Omitido
4-7 anos de estudo	-	-1.133***		-1.132***	-1.132***
		(0.0012)		(0.0012)	(0.0012)
8-10 anos de estudo	-	-1.663***		-1.657***	-1.657***
		(0.0019)		(0.0019)	(0.0019)
11-14 anos de estudo	-	-2.031***		-2.027***	-2.027***
		(0.0018)		(0.0018)	(0.0018)
15+ anos de estudo	-	-2.222***		-2.218***	-2.218***
		(0.0024)		(0.0024)	(0.0024)
Censo 1970	-	*Omitido*		*Omitido*	*Omitido*
Censo 1980	-	-0.125***		-0.119***	-0.119***
		(0.0024)		(0.0024)	(0.0024)
Censo 1991	-	-0.132***		-0.1321***	-0.1321***
		(0.0027)		(0.0027)	(0.0027)
Censo 2000	-	-0.101***		-0.1053***	-0.1053***
		(0.0029)		(0.0029)	(0.0029)
Censo 2010	-	-0.254***		-0.2531***	-0.2531***
		(0.0031)		(0.0031)	(0.0031)
Constante	-	*Omitido*		6.489***	*Omitido*
		-		(0.2127)	
Nº Obs	-	47427995		47408656	47408656
R^2	-	0.140		0.250	0.140
SSR	-	473741316		472126898	472127096

Erro-Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

Os resultados das estimações dos modelos 5 e 6 para a subamostra são idênticos, a diferença entre a soma dos quadrados dos resíduos (SSR) do modelo 6 em relação ao modelo 5 é 198, ou $4.19e^{-5}$% do valor apresentado pelo modelo 5. O R2 do modelo 6 é inferior ao modelo 5, representa 56% do R^2 do modelo 5, o que impossibilita comparações entre modelos caso um seja estimado expondo todas as variáveis e o outro modelo seja estimado com desvios em relação à média da localidade. Os resultados dos efeitos fixos de coortes, apresentados no gráfico abaixo são praticamente idênticos para os parâmetros estimados, mas diferem a significância estatística para as coortes nascidas entre 1898 e 1911. No entanto, mesmo com diferença em termos da significância dada pelo *p-value* das regressões, os parâmetros para os quais o estimador é estatisticamente não significante são marginalmente próximos de zero.

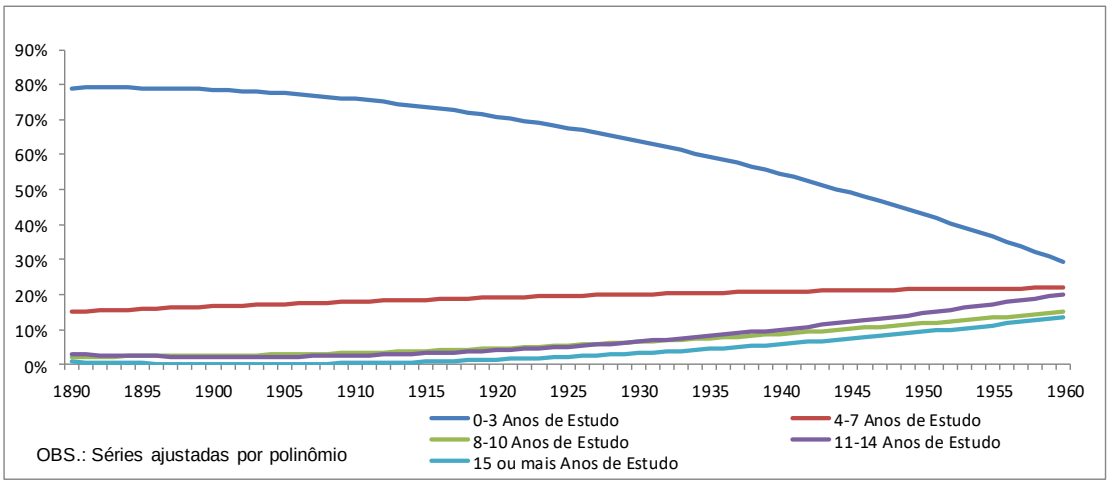
GRÁFICO 7.2 – RESULTADO DOS EFEITOS FIXOS DE COORTES PARA A AMOSTRA E PARA A SUBAMOSTRA



OBS.: O resultado é relativo à coorte de 1890.

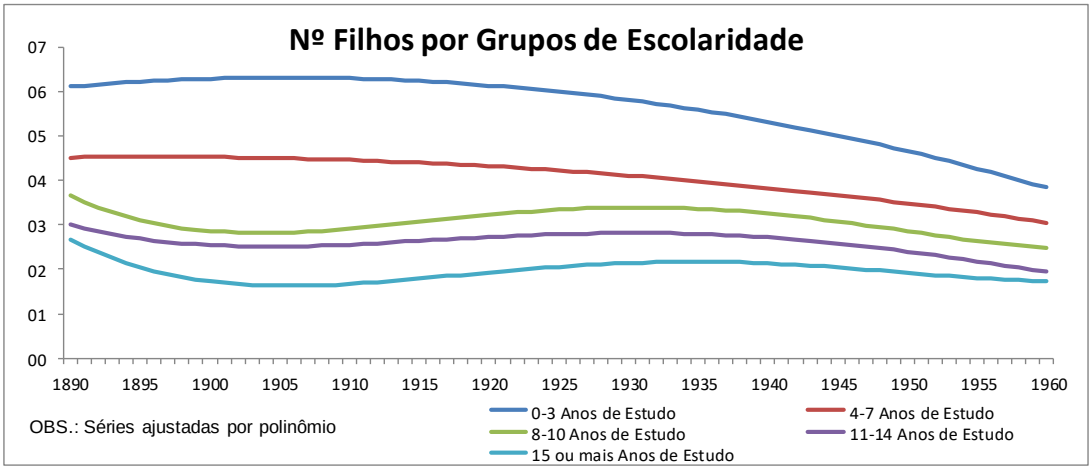
APÊNDICE 2: OUTRAS ESTATÍSTICAS DESCRITIVAS

GRÁFICO A2.1: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – ANÁLISE POR COORTE DE NASCIMENTO



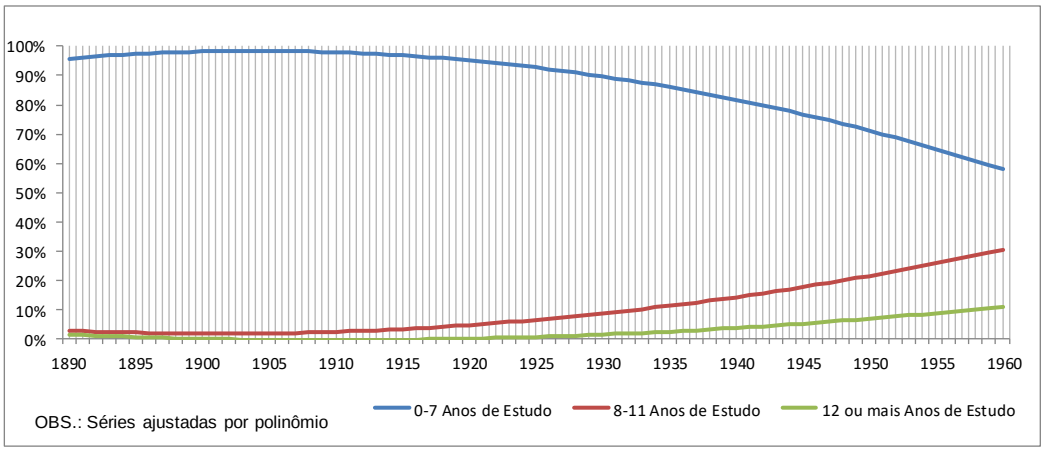
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.2: Nº FILHOS POR GRUPOS DE ESCOLARIDADE – ANÁLISE POR COORTE DE NASCIMENTO



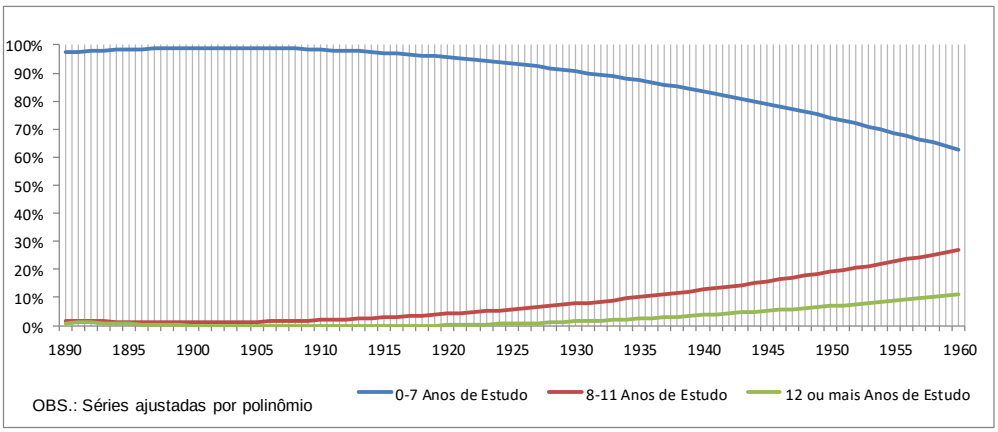
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.3: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – REGIÃO NORTE



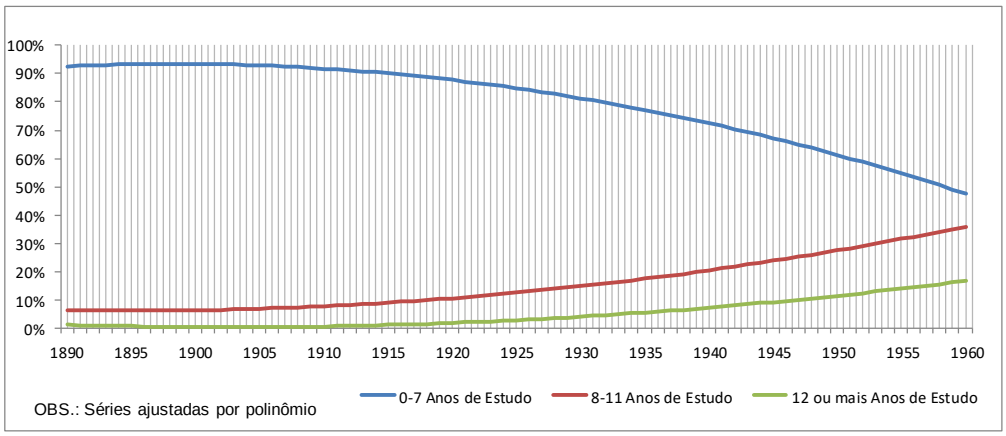
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.4: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – REGIÃO NORDESTE



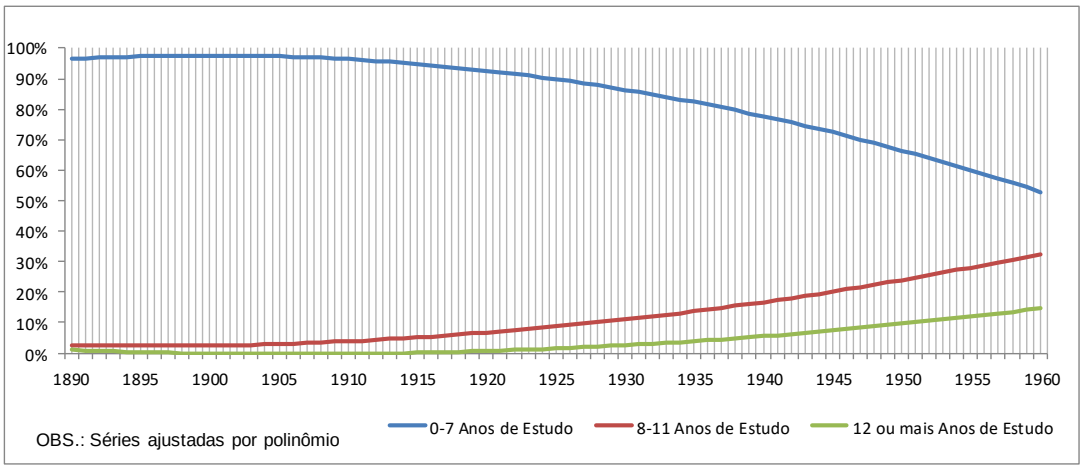
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.5: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – REGIÃO SUDESTE



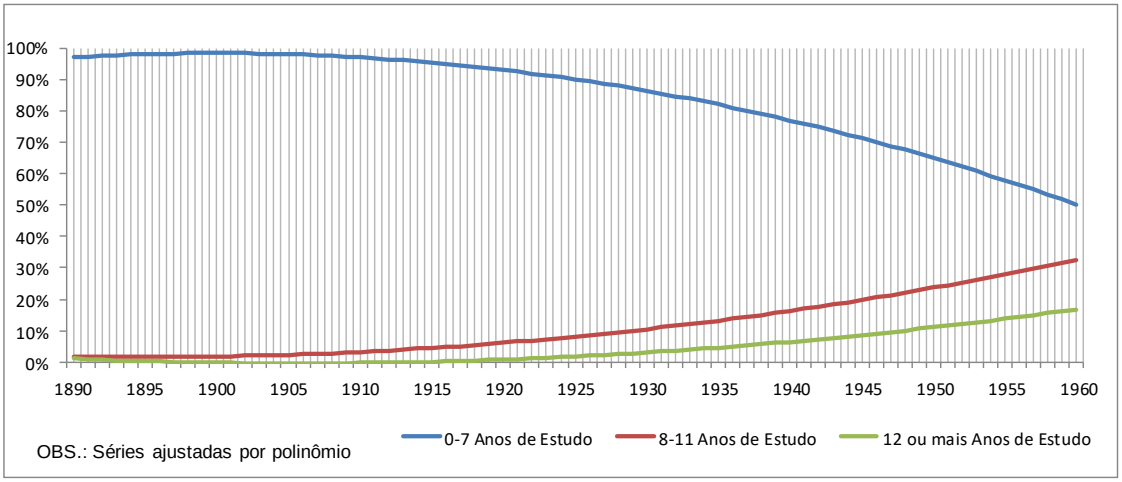
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.6: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – REGIÃO SUL



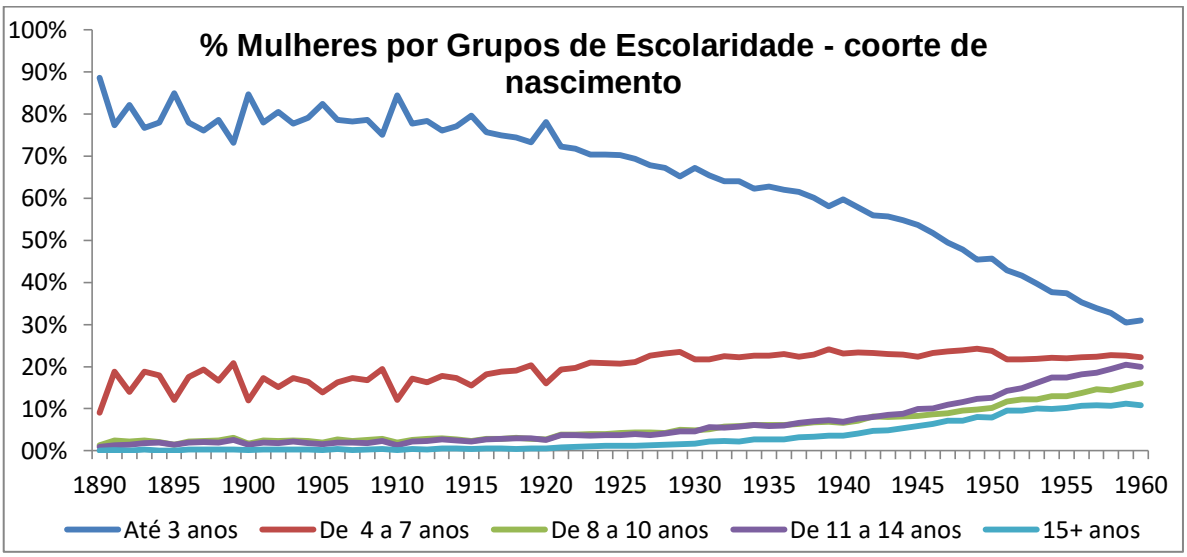
Fonte: Censos 1970, 1980, 1991, 2000 e 2010

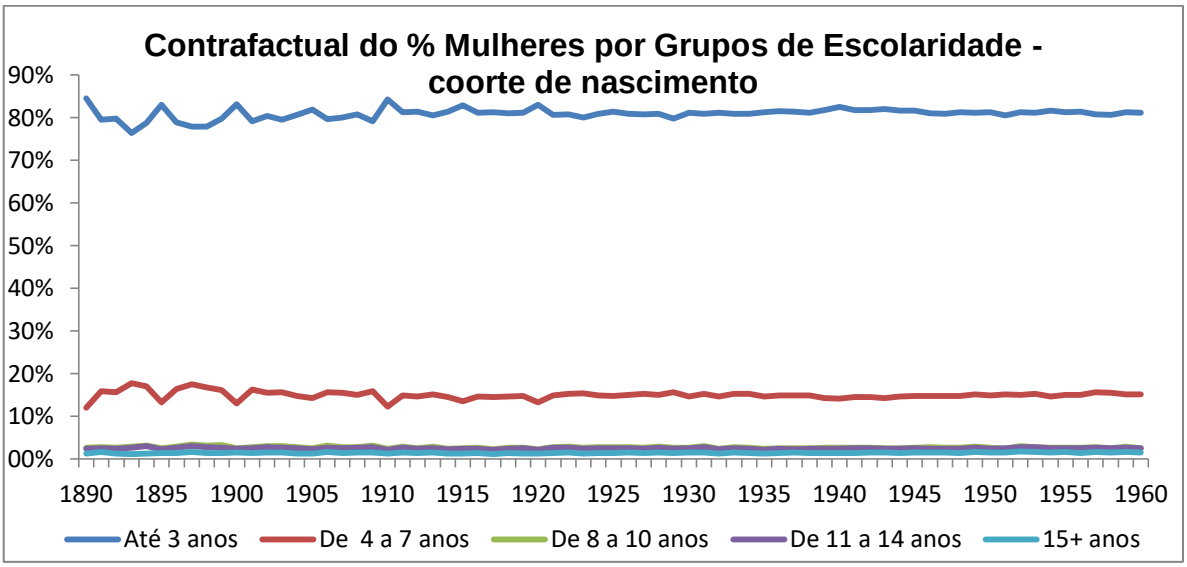
GRÁFICO A2.7: PERCEUTAL DE MULHERES POR GRUPOS DE ESCOLARIDADE – REGIÃO CENTRO-OESTE



Fonte: Censos 1970, 1980, 1991, 2000 e 2010

GRÁFICO A2.8: DISTRIBUIÇÃO DAS COORTES POR ESCOLARIDADE – FACTUAL E CONTRAFACTUAL

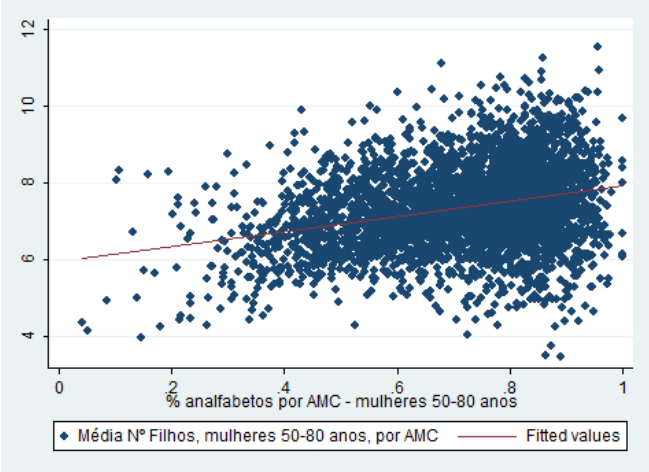




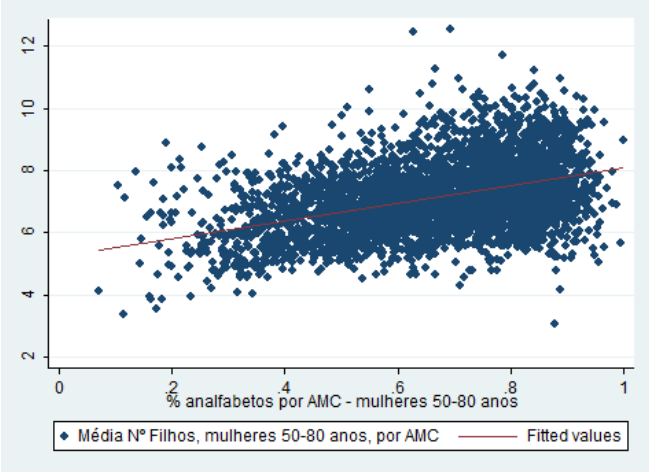
Fonte: Censos 1970, 1980, 1991, 2000 e 2010.

APÊNDICE 3.1. MÉDIA DO NÚMERO DE FILHOS POR AMC E PERCENTUAL ANALFABETOS POR AMC

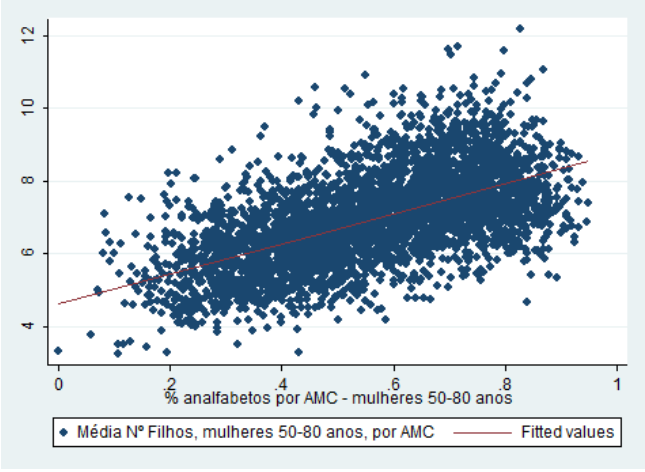
Censo 1970



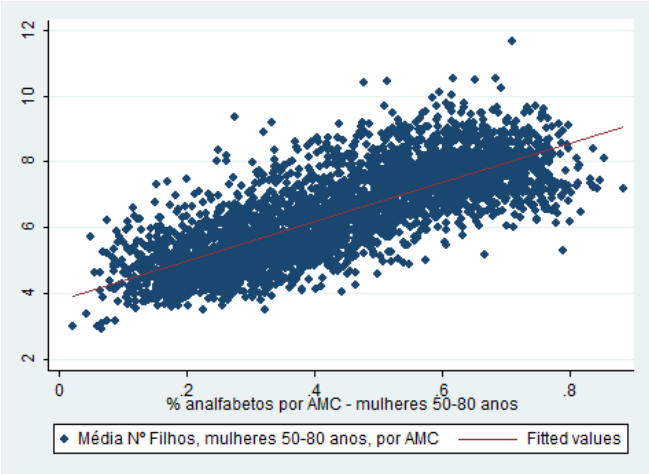
Censo 1980



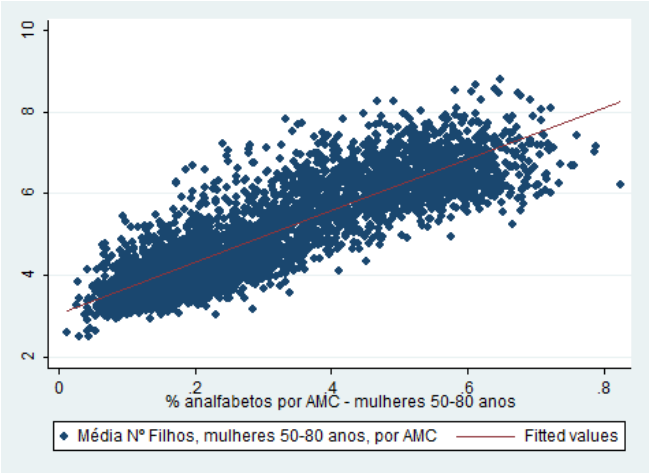
Censo 1991



Censo 2000

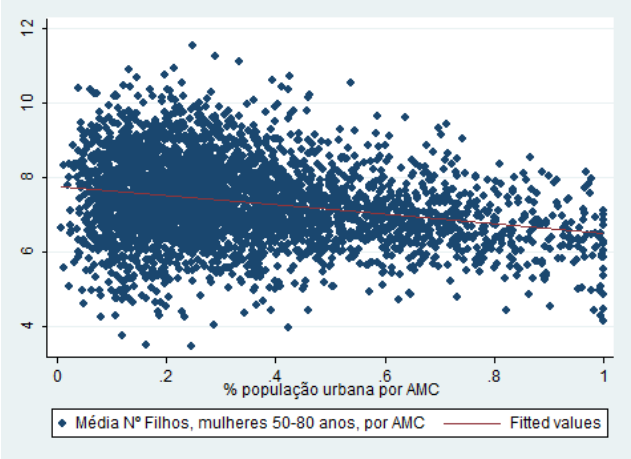


Censo 2010

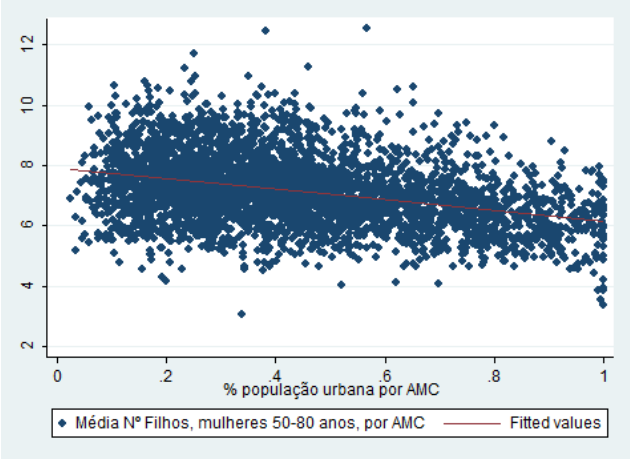


APÊNDICE 3.2. MÉDIA DO NÚMERO DE FILHOS POR AMC E PERCENTUAL URBANIZAÇÃO POR AMC

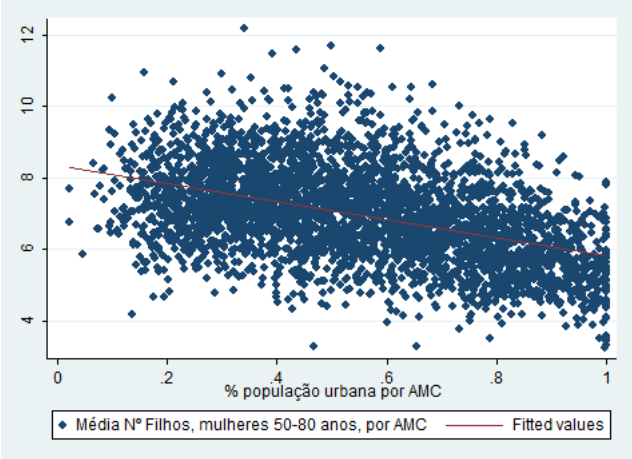
Censo 1970



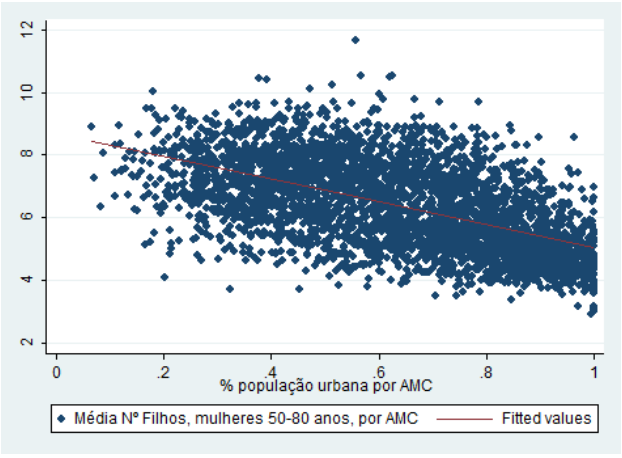
Censo 1980



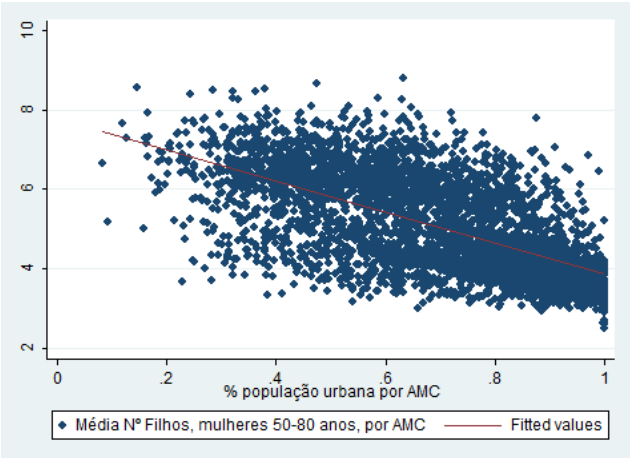
Censo 1991



Censo 2000

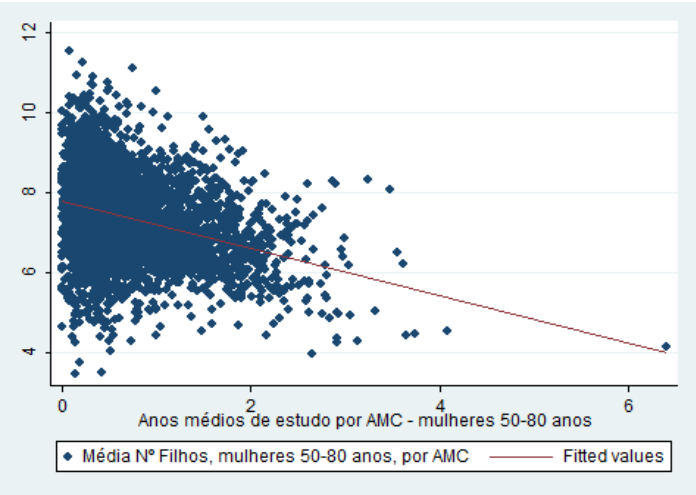


Censo 2010

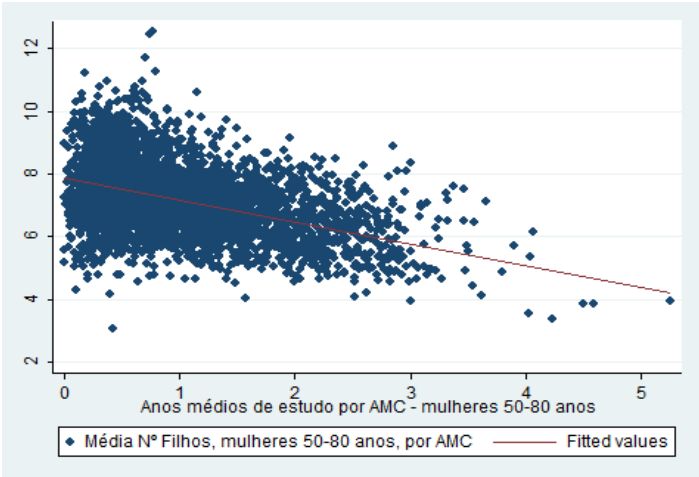


APÊNDICE 3.3. MÉDIA DO NÚMERO DE FILHOS POR AMC E MÉDIA DA ESCOLARIDADE POR AMC

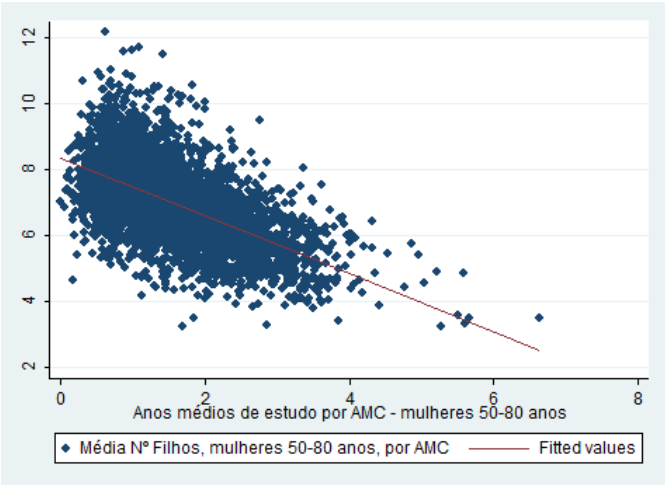
Censo 1970



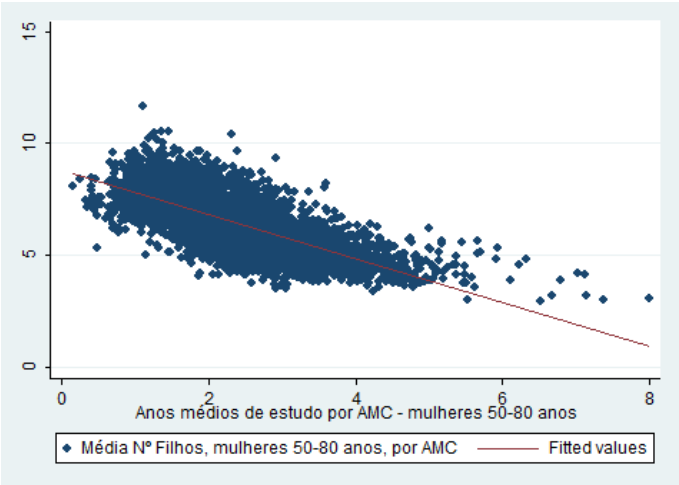
Censo 1980



Censo 1991

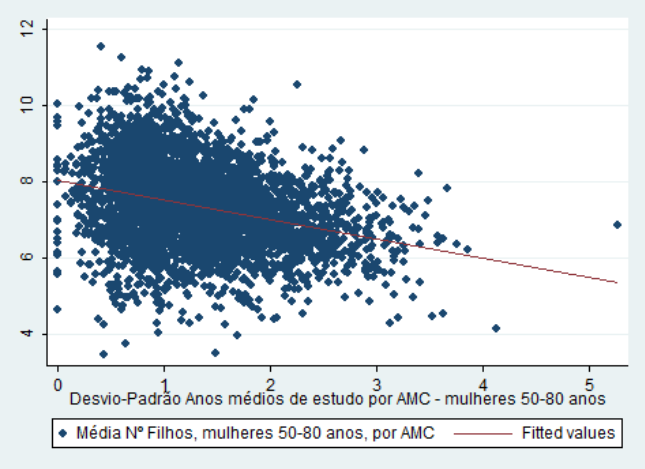


Censo 2000

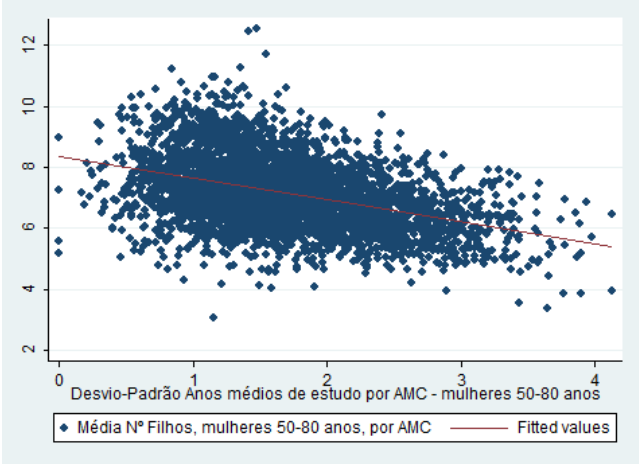


APÊNDICE 3.4. MÉDIA DO NÚMERO DE FILHOS POR AMC E DESVIO-PADRÃO DA ESCOLARIDADE POR AMC

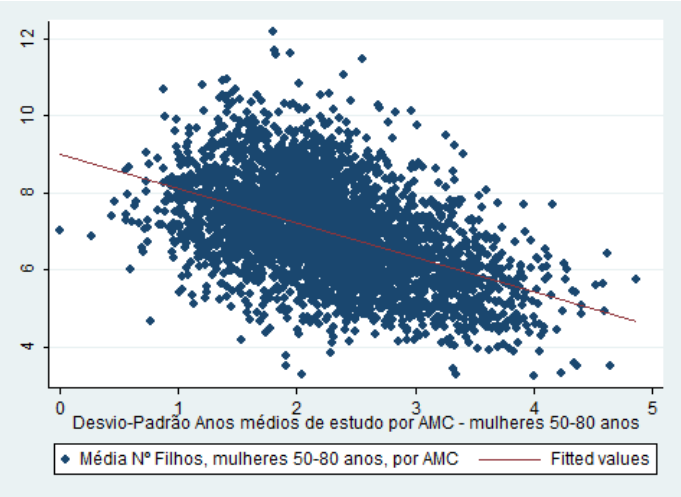
Censo 1970



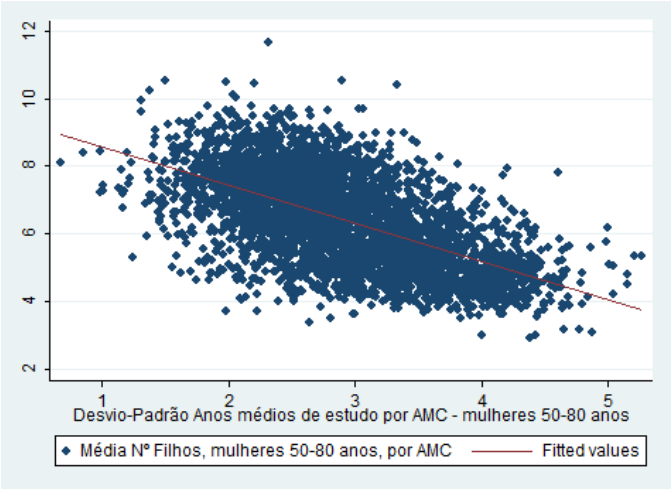
Censo 1980



Censo 1991



Censo 2000



TAMANHO DA COORTE, TAMANHO DA FAMÍLIA E MATRÍCULA ESCOLAR: EVIDÊNCIAS DO BRASIL.

1. INTRODUÇÃO

Desde a formulação do *trade-off* quantidade-qualidade no artigo seminal de Becker & Lewis (1974), a literatura empírica investiga se a escolha de um número menor de filhos implica em um aumento na qualidade destes[40]. Do ponto de vista teórico, Becker e Lewis demonstraram que o *trade-off* quantidade-qualidade é uma consequência da restrição orçamentária domiciliar onde aumentar qualidade é custoso quando se tem muitos filhos assim como aumentar o número de filhos é custoso quando se escolhe ter mais qualidade. Uma parte da literatura empírica utilizou diversos experimentos para se investigar se, na margem, famílias que decidiram ter menos filhos investiram mais na qualidade destes, como o caso Rosenzweig & Wolpin (1980). O *working paper* de Angrist, Lavy & Schlosser (2005)[41] utiliza uma variação exógena do tamanho da família para identificar uma relação causal entre quantidade de filhos e qualidade. Black et al (2005) utilizaram dados administrativos da Noruega para testar uma relação causal entre quantidade de filhos e qualidade, medida por escolaridade. Os autores encontram uma relação negativa entre quantidade e qualidade de filhos, porém, ao estabelecer um experimento usando gêmeos como variável instrumental, o efeito passa a ser não significativo. Para o Brasil, Ponczek & Souza (2012) utilizam a mesma metodologia de Angrist, Lavy & Schlosse para os dados do Censo de 1991 e mostram que há uma relação causal entre quantidade de filhos e qualidade.

Uma tradição mais demográfica investiga o efeito do tamanho da família e o tamanho da coorte sobre matrícula. Lam & Marteleto (2006) usam a taxa de crescimento populacional de crianças de 7 a 14 anos, que se mantém negativa ao longo da década de 1990, para verificar este efeito sobre a probabilidade de matrícula de crianças de 7 a 14 anos usando um modelo de escolha discreta. O tamanho da família é avaliado pelo número de irmãos e, para se

[40] Qualidade dos filhos está associada a indicadores educacionais e saúde, por exemplo.

[41] O *working-paper* é citado e utilizado como referência metodológica em Ponczek e Souza (2012). A versão final do artigo de Angrist, Lavy e Schlosser é publicada em 2010.

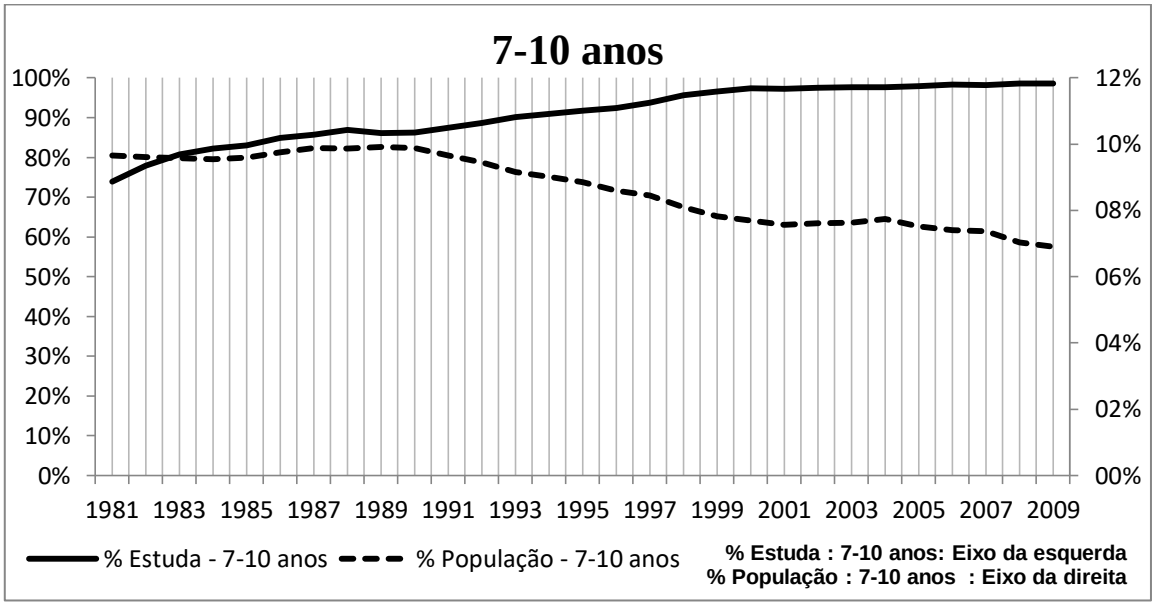
capturar efeitos heterogêneos, separa-se entre grupos de idade de 0 a 6 anos e de 7 a 14 anos. Lam & Marteleto encontram uma relação negativa entre tamanho da família e probabilidade de matricula, mas em relação ao tamanho da coorte, o resultado apresentou-se positivo para a regressão que considera apenas indivíduos de 8 a 9 anos. Para a regressão que considera apenas indivíduos com 10 a 13 anos, o resultado foi negativo, de acordo com a hipótese enunciada[42].

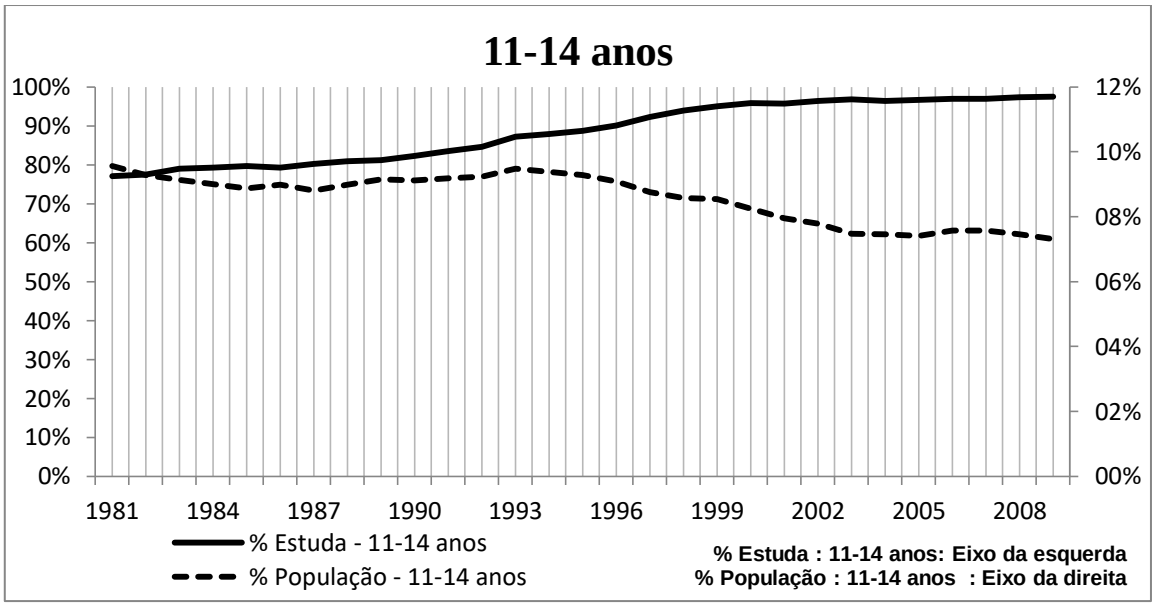
Autores de economia utilizam modelos de escolha discreta para se investigar probabilidade de matrícula, mas com outros propósitos como Cameron e Heckman (2001), que utilizam um modelo de matrícula e investigam educação para minorias (disparidades raciais e étnicas na matrícula) e como fatores associados ao *backgraound* familiar, ambiente familiar e restrição de crédito interagem. No entanto, diferente da literatura seguida por Cameron e Heckman, este presente artigo investiga a probabilidade de matrícula a partir dos dados da PNAD para as décadas de 1980, 1990 e 2000, período com profundas mudanças no tamanho da coorte em decorrência da transição demográfica iniciada no Brasil nas décadas anteriores[43]. O Brasil tem uma estrutura de oferta pública de vagas escolares e a natureza desta oferta possibilita uma análise a partir de mudanças na demanda por conta dos efeitos da transição demográfica, principalmente em um período da história onde o acesso ao ensino básico não era universalizado. O movimento da evolução da taxa de matrícula para indivíduos de 7 a 10 anos e indivíduos de 11 a 14 anos é apresentado no Gráfico 2.1 assim como as mudanças na coorte relativa como decorrência da queda da taxa de fecundidade.

[42] A análise de tamanho de coorte tem sua aplicação em mercado de trabalho, como Welch (1979) e Connelly (1986). No paper de 1995, Conelly & Gottschal estimam a probabilidade de matrícula, condicional na educação do pai, mas utilizando o conceito de composição de coorte, e não o tamanho da coorte. Assim, para os autores, coorte relativa é o tamanho da coorte em relação às coortes que nasceram 5 anos antes e 5 anos depois.

[43] No capítulo1, denominado *Transição Demográfica No Brasil: Evidências Por Coortes, Educação e Localidades*, é feita uma analise detalhada sobre o processo de transição demográfica no Brasil a partir de dados de coortes de mulheres que nasceram entre 1890 e 1960.

GRÁFICO 2.1: EVOLUÇÃO DA TAXA DE MATRÍCULA E COORTE RELATIVAPOR GRUPO ETÁRIO





Fonte: Microdados da PNAD de 1981 a 2009
OBS.: Interpolação linear para os anos de 1991, 1994 e 2000.

Os resultados encontrados mostram que tamanho da família é negativamente correlacionado com probabilidade de matrícula assim como o tamanho da coorte[44]. No entanto, mesmo sendo negativamente relacionado com a probabilidade de matrícula do indivíduo e estatisticamente significativo, o tamanho da coorte afeta na margem e sua magnitude, segundo as simulações, não explicam o aumento da taxa de matrícula. A educação dos pais mostrou efeito significativo nas simulações para a 1ª série do Ensino Fundamental, a série que apresentava a menor taxa de matrícula na década de 1980. Na próxima seção será apresentada a base de dados e, em seguida, a metodologia. Os resultados são apresentados por grupos etários e por séries do ensino regular. No apêndice são encontradas outras estatísticas descritivas e regressões que embasam a pesquisa.

2. BASE DE DADOS E METODOLOGIA

Foram utilizados os microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1981 a 2009. Informações sobre educação, mercado de trabalho, fecundidade e características do indivíduo no domicílio foram extraídas para se estruturar a base de dados[45]. Para se associar a cada indivíduo as características de seus pais, apenas filhos de chefes de família foram utilizados. Para domicílios com mais de uma família, o número de identificação da família foi associado ao número de identificação do domicílio para que cada família tivesse sido avaliada individualmente. Foi considerado "pai" o indivíduo que foi declarado chefe de família do sexo masculino ou o cônjuge do chefe de família, quando o sexo do cônjuge fosse masculino e o sexo do chefe de família fosse feminino. Considerou-se "mãe" o indivíduo

[44] Lam e Marteleto não encontraram esta evidência para a coorte de 8 a 9 anos. Utilizando taxas de crescimento do tamanho da coorte desagregadas por Unidade da Federação e situação de moradia, os resultados se mantiveram robustos em todas as regressões. No apêndice são apresentadas diversas regressões com variações no tamanho da coorte.

[45] Inicialmente foram utilizados dados da PNAD de 1976 a 2009. No entanto, após agregarmos os dados para a década de 1970, diversas distorções surgiram tanto pelo excesso de dados faltantes ("*missing data*") em anos específicos quanto pela perda da abrangência geográfica. Na década de 1970, a PNAD abrangia as atuais regiões Nordeste, Sudeste e Sul e as áreas urbanas das Regiões Norte e Centro-Oeste. A partir da PNAD de 1981, somente as áreas rurais de Rondônia, Acre, Amazonas, Roraima, Pará e Amapá não foram contempladas. Para uma análise histórica e metodológica da PNAD desde a sua formulação, Dedeca (1998) é uma referência bibliográfica.

chefe de família do sexo feminino ou cônjuge do chefe de família de sexo feminino. A partir da década de 90, duas informações novas informações foram utilizadas: a mulher declarar ter tido filho vivo até a data de referência da pesquisa e o número de identificação da mãe no domicílio para os filhos. Em nossa amostra, não existe a possibilidade de se diferenciar filho biológico de filho adotivo[46] e, para os anos onde o número de identificação da mãe não é disponibilizado, considerou-se como mãe apenas a declaração de chefe de família ou cônjuge do chefe de família[47].

Como as informações obtidas da PNAD não são uniformes ao longo dos anos, os dados precisaram ser compatibilizados. "Anos de estudo" é disponibilizado para a década de 80 com valor máximo de 12 anos e não desagregando para o ensino médio. Para as décadas de 90 e 2000, o valor máximo é 15 anos. Assim, definiu-se um algoritmo para a variável 'anos de estudo' com base na última série de um respectivo grau declarado pelo indivíduo na amostra e tal variável apresenta um intervalo discreto entre 0 e 17 anos de estudo. Este algoritmo foi aplicado a todos os anos da base de dados gerando informações padronizadas ao longo das décadas para escolaridade. O ensino regular brasileiro apresentou diversos nomes para a estrutura de séries do que hoje denominamos Ensino Fundamental e Ensino Médio. Na década de 1980, as denominações Elementar, Médio 1º Ciclo, Médio 2º Ciclo, 1º Grau e 2º Grau compunham o ciclo de ensino básico. Na década de 1990, a PNAD trata o ensino regular como 1º Grau e 2 º Grau e, a partir de 2003, Ensino Fundamental e Ensino Médio passam a ser a denominação para o grau do ensino regular a ser cursado. A base de dados foi toda compatibilizada para o seu equivalente a Ensino Fundamental de 8 séries e Ensino Médio de 3 séries e estas informações são harmônicas e estáveis para se comparar os dados ao longo das

[46] Após gerar *dummies* para "pai", "mãe" e "filho" seguindo a declaração de condição no domicílio e, quando necessária, na família dentro do domicílio, foram geradas como exercício de robustez duas variáveis: total de "pais" e total de "mães" por família, que representa o somatório das variáveis "*dummy* de pai" e "*dummy* de mãe" respectivamente. Para todos os anos, as variáveis total de "pais" e total de "mães" teve como valor mínimo zero e valor máximo 1, o que possibilitou associar a cada indivíduo definido como filho as características de seus pais. Foram selecionadas as famílias que apresentaram somatório de tais *dummies* igual a 1 na família, ou seja, apenas os indivíduos na amostra de famílias com pai e mãe identificados.

[47] Para os anos onde o número de identificação da mãe foi disponibilizado, realizou-se o "*matching*" entre mãe e filhos usando duas técnicas. A primeira com a utilização de tal número, que é um casamento preciso das informações. A segunda técnica, com a utilização apenas das indicações '*filho do chefe do domicílio*' e '*cônjuge do chefe do domicílio*' (técnica usada para os dados da década de 1980, quando não havia informações adicionais). As duas formas de "*matching*" mostraram-se equivalentes e as estatísticas geradas após o pareamento são idênticas. Infelizmente este exercício não pode ser feito para o "*matching*" entre filhos e pais, do sexo masculino, quando estes não são chefes do domicílio.

décadas[48]. Com base nestas declarações, duas outras variáveis foram criadas: série matriculada do filho e instrução dos pais[49].

Foi classificado como indivíduo que estuda aquele que respondeu que frequentava o ensino regular fundamental (ou primeiro grau regular para as séries mais antigas). Creches e Pré-Escolas, frequentes para indivíduos de 7 anos, foram excluídas da amostra, assim como supletivos e programas de ensino não regular, frequentes para indivíduos entre 12 e 17 anos. Em geral, mudanças na PNAD ao longo dos anos provocam mudanças na tendência dos dados quando os agregamos e analisamos ao longo do tempo, menos para os dados associados ao ensino regular. Supletivos e alfabetização de adultos, para a amostra construída, representa menos de 1% das declarações totais para 12 anos ou mais e representa 0% para indivíduos mais jovens, com exceção da década de 1980. Em especial o ano de 1982 tem 19% dos indivíduos de 7 anos, por exemplo, matriculados em supletivos ou alfabetização de adultos. Para este mesmo grupo, o percentual é 0,02% em 1983. Se considerássemos tal informação na investigação da taxa de matrícula, teríamos uma descontinuidade nos dados agregados todas as vezes que tais problemas aparecessem. Creche e pré-escola foram excluídas por dois motivos. O primeiro refere-se à natureza dos dados, onde creche era investigada apenas para crianças com até 6 anos nas PNAD até 1986 e, após, sendo investigada para outros grupos etários (nas PNAD da década de 1990, mais de 1% dos indivíduos de 11 anos declararam frequentar creche ou pré-escola e tal valor diminuiu a partir da segunda metade de tal década). O segundo motivo refere-se à decomposição da matrícula por séries que será realizada na investigação empírica.

[48] A partir de 2007 observa-se nos dados a presença de indivíduos pertencentes ao Ensino Fundamental com 9 anos de duração contínua. Este foi compatibilizado para o padrão de 8 séries tal que o primeiro ano foi considerado pré-escola, o segundo ano foi considerado 1ª série e assim sucessivamente, com o nono ano equivalente a 8ª série. Em relação aos dados da PNAD, 39,1% dos estudantes de 7 anos do Ensino Fundamental cursavam a nova modalidade (ensino com nove anos de duração) no ano de 2007 e este percentual sobe para 70% em 2009. A mudança deveu-se a um projeto de lei de 2006 e tinha a previsão de todas as escolas implementarem o sistema com 9 anos até 2010.

[49] Para se construir o algoritmo, duas referências foram utilizadas: Riani & Golgher (2004); Rigotti (2004)

Para se associar a cada indivíduo o capital humano de seu pai e de sua mãe, uma amostra para regressão foi desenhada excluindo-se famílias com dados faltantes para escolaridade tanto do pai quanto da mãe. Além disto, o número de irmãos é criado através das informações familiares e esta informação será decomposta entre número de irmãos mais novos, de 0 a 6 anos, e número de irmãos com idade próxima, de 7 a 14 anos. A amostra de regressão se constitui em indivíduos de 7 a 17 anos com os pais identificados e com suas informações sobre escolaridade, rendimento e escolaridade. O Gráfico 2.2 compara o percentual de indivíduos de 7 a 10 anos e 11 a 14 anos que estudam entre 1981 e 2009 segundo dados da PNAD para matrículas no Ensino Regular tanto para a população quanto para a amostra utilizada nas regressões[50]. Na média, 80% da base de dados foram utilizadas e esta se constitui uma amostra representativa principalmente para os indivíduos de 7 a 10 anos, pois fornece informações idênticas às informações populacionais.[51] Para os indivíduos de 11 a 17 anos, o percentual de matriculados no ensino regular é levemente superestimado na amostra de regressão em comparação ao dado populacional e pode ser explicado por indivíduos não pareados pelo "*matching*" familiar terem menor probabilidade de matrícula. Os gráficos ainda demonstram que mesmo pareando um percentual cada vez menor do total de indivíduos (linha tracejada), a amostra ainda manteve representativa para a tendência de taxa de matrícula do ensino regular no Brasil. Para a população de 7 a 10 anos, 85% desta população teve seu pareamento com as informações de seus pais em 1981 e este percentual diminui para 71% em 2009.

Para a análise do impacto do tamanho da coorte sobre a taxa de matrícula, utilizaremos duas medidas para coortes: a taxa de crescimento do tamanho da coorte[52] e o tamanho relativo da coorte no total da população. A taxa de crescimento do tamanho da coorte c no ano t e localidade s é definida por:

[50] Para todas as estatísticas descritivas assim como para todas as regressões, o fator de expansão amostral foi utilizado como peso para que a amostra representasse a população como um todo.

[51] Quando, a priori, os dados da PNAD da década de 1970 foram utilizados, o ano de 1979 gerou uma descontinuidade nos dados agregados muito pelo excesso de dados faltantes ("*missing data*"). A não utilização dos dados da década de 1970 não compromete os resultados, pois este artigo preocupa-se com os movimentos demográficos de tamanho de coorte que surgiram no final da década de 1980 e se propagaram para as décadas seguintes.

[52] A taxa de crescimento do tamanho da coorte é utilizada para atualizar o trabalho de Lam & Materleto, que é uma referência para este trabalho. Os resultados são apresentados no Apêndice 2.

$$Taxa\ de\ crescimento_{c,t,s} = Ln\left(\frac{Pop_{c,t,s}}{Pop_{c,t-1,s}}\right) \tag{2.1}$$

A população da coorte c, no ano t e localidade s é calculada, a priori, pelos microdados da PNAD com a utilização do fator de expansão amostral. Para que o tamanho da população não sofresse influência das mudanças metodológicas ocorridas ao longo da série, as unidades federativas foram compatibilizadas para não se gerar distorções no tamanho da coorte em uma localidade. Assim, o estado de Tocantins foi integrado ao estado de Goiás a partir das PNAD de 1992 e a área rural da região norte foi excluída da amostra a partir das PNAD de 2004[53]. A localidade s é definida por 26 Unidades da Federação o que constitui a menor desagregação para se analisar os microdados da PNAD ao longo da série. A coorte c é definida pelo ano de nascimento. Para se evitar variações bruscas da contagem da população na PNAD, foi utilizada a projeção populacional do IBGE com dados agregados para cada Unidade da Federação entre 1981 e 2009[54]. Assim, para cada Unidade da Federação, foi imputada a estimação da população total em cada ano entre 1981 e 2009 usando os dados da projeção populacional do IBGE (2008). O tamanho relativo da coorte foi calculado para cada UF e para cada ano a partir dos microdados da PNAD. A nova estimativa do tamanho da coorte para se calcular a equação (2.1) foi gerada com a combinação da projeção populacional com o tamanho relativo da coorte, ambos avaliados por UF e por ano[55].

---

[53] O estado de Tocantins foi criado em 05 de outubro de 1988 e foi delimitado no desenho amostral da PNAD a partir da pesquisa de 1992. Em relação à área rural da Região Norte (Rondônia, Acre, Amazonas, Roraima, Pará e Amapá), esta somente foi investigada a partir da PNAD de 2004. Assim, definindo localidade por Unidade da Federação (UF), a menor unidade geográfica desagregada na análise dos microdados, o tamanho da coorte em um determinado ano será o número de habitantes da UF pertencentes a tal coorte. Mudanças geográficas como a separação do Tocantins e Goiás geraria uma queda no total da coorte de Goiás após a separação e a taxa de variação populacional refletiria tanto mudanças no tamanho da coorte como também refletiria o processo de separação territorial. O mesmo pode ser analisado para a incorporação da área rural na Região Norte. Assim, para que o tamanho da coorte seja analisado sem viés, fez-se necessário a compatibilização das Unidades da Federação ao longo das pesquisas domiciliares.

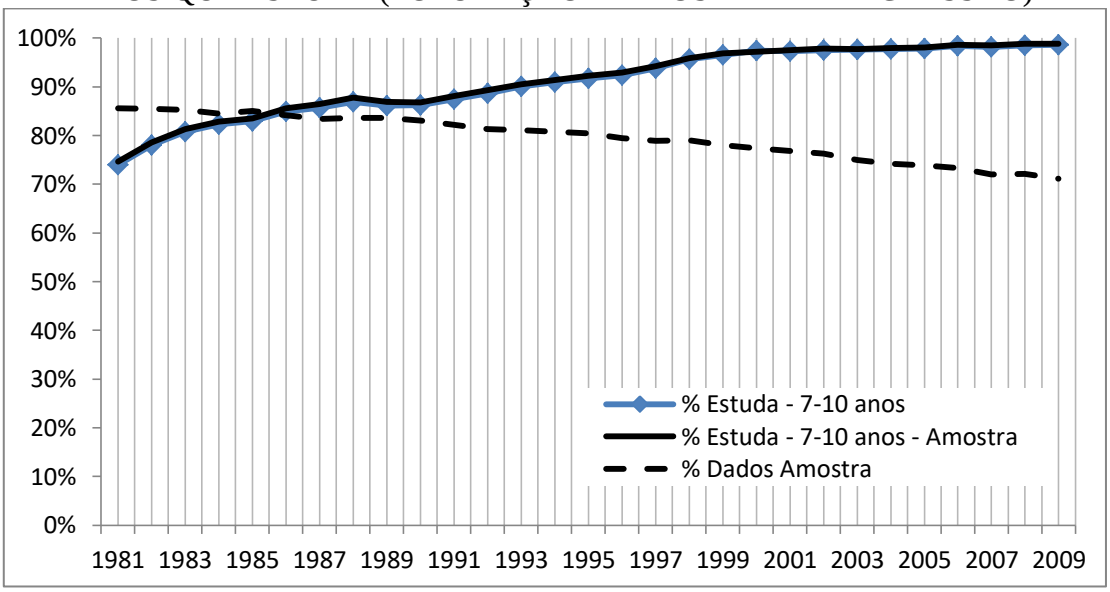
[54] IBGE (2008) foi uma revisão da projeção populacional de 1980 a 2030 com os dados populacionais tanto para o Brasil quanto por Unidade da Federação.

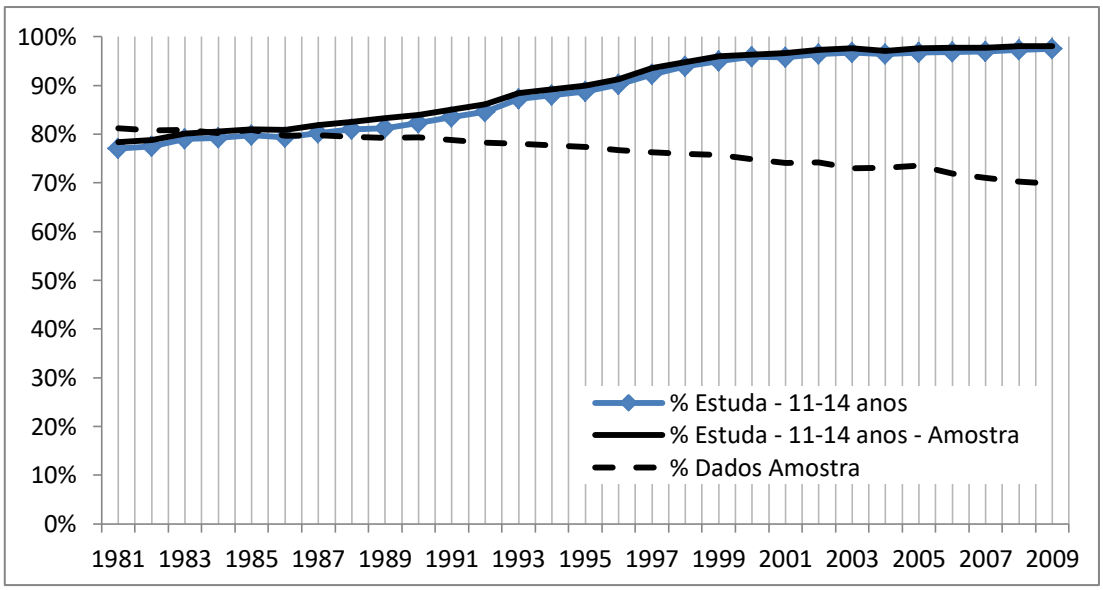
[55] Para cada ano t, localidade s e coorte c, a coorte relativa é calculada por

$$Coorte\ Relativa_{c,t,s} = \frac{Pop_{c,t,s}}{Pop_{t,s}}$$

Onde $Pop_{c,t,s}$ representa o total de habitantes da coorte c na localidade s e ano t e $Pop_{t,s}$ representa o total de habitantes na localidade s e ano t e são calculadas com os microdados da PNAD usando-se o fator de expansão

GRÁFICO 2.3: PERCENTUAL DAS POPULAÇÕES DE 7 A 10 ANOS E 11 A 14 ANOS QUE ESTUDA (POPULAÇÃO E AMOSTRA DA REGRESSÃO)





Fonte: Microdados da PNAD de 1981 a 2009
OBS.: Interpolação linear para os anos de 1991, 1994 e 2000.

O Gráfico 2.4 apresenta o tamanho absoluto da coorte e sua taxa de variação, calculado a partir da equação (2.1), ambos calculados tanto pelos microdados da PNAD com expansão da amostra quanto com a imputação da projeção populacional a partir dos dados do IBGE

---

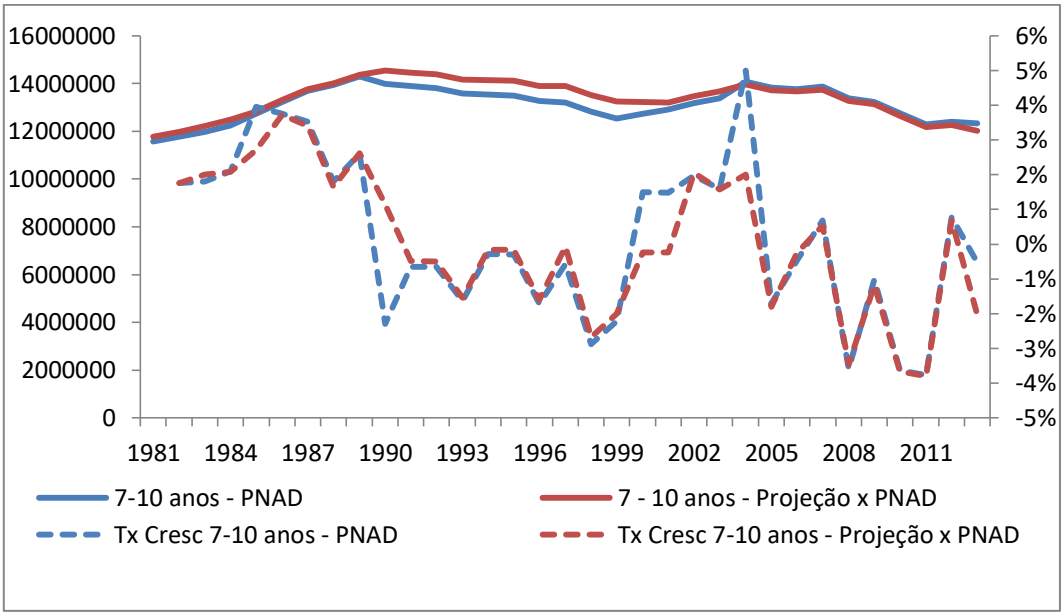
da amostra. Assim, dada a projeção populacional imputada à base de dados por localidade e ano, $Projeção\ Populaciona_{t,s}$, a estimativa da população por coorte será

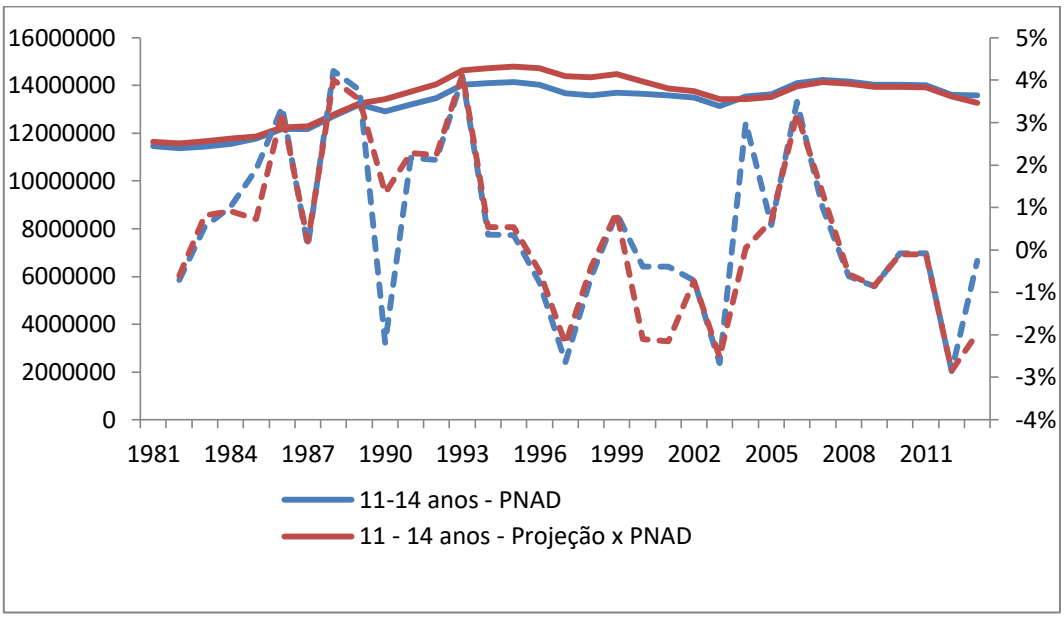
$$População_{c,t,s} = Projeção\ Populaciona_{t,s}\ x\ Coorte\ Relativa_{c,t,s}$$

(2008). A taxa de variação do tamanho absoluto da coorte apresenta mesmo comportamento em ambas estimativas, mas é mais suave com os dados gerados a partir da projeção populacional, principalmente entre os anos 1898 e 1990 e os anos 2003 e 2004. No geral, a taxa de crescimento de 7 a 10 anos apresentou valores negativos ao longo da década de 1990 e a taxa de crescimento de 11 a 14 anos apresentou valores negativos entre 1995 e 2005, na média[56]. Quando a estratégia empírica utilizar a taxa de variação do tamanho da coorte, será utilizada a variável construída a partir da projeção populacional e quando a estratégia empírica utilizar a coorte relativa, esta será definida por $Coorte\ Relativa_{c,t,s} = Pop_{c,t,s}/Pop_{t,s}$.

GRÁFICO 2.4: POPULAÇÃO E TAXA DE CRESCIMENTO PARA 7-10 ANOS E 11- 14 ANOS – PNAD E PROJEÇÃO POPULACIONAL

[56] Em seu artigo original, LAM, D.; MARTELETO (2006) utilizam a taxa de crescimento da coorte de 7 a 14 anos para cada Unidade da Federação. A desagregação tanto para a taxa de crescimento de grupos de 7 a 10 anos e grupos de 11 a 14 anos, assim como para cada idade individualmente se justifica pelo efeito de composição atenuar a magnitude do crescimento quando a série apresenta-se negativa ou quando a taxa de crescimento de 7 a 14 anos apresentar-se positiva. Assim, para o ano de 1995, por exemplo, a taxa de crescimento da coorte de 7 anos foi de -1,5%, a taxa de crescimento da coorte 7-10 anos foi -0,3% e a taxa de crescimento da coorte 7 – 14 anos foi nula.





Fonte: Microdados da PNAD 1981 – 2009
OBS. População absoluta é analisada no eixo da esquerda e a taxa de crescimento é analisada no eixo da direita.

3. ESTRATÉGIA EMPÍRICA E RESULTADOS

A estratégia empírica consiste em estabelecer como fatores demográficos e familiares influenciam a probabilidade de matricula de um indivíduo entre os anos de 1981 e 2009. Para isto, partiremos da metodoligia de Lam & Marteleto como referência, porém com modificações[57]. Consideramos $N_t = \sum_i i_t$ o total de indivíduos nascidos na geração t e $N_{t-1} = \sum_i i_{t-1}$ o total de indivíduos nascidos na geração $t-1$. Assim, podemos definir a taxa de fecundidade como sendo ${N_t}/{N_{t-1}}$, ou seja, se a taxa de fecundidade for, por exemplo,

1, isto impla que N_{t-1} tiveram, na média, 2 filhos por casal. Apenas os indivíduos na geração $t-1$ trabalham e recebem uma renda líquida $(1-\tau)w_i$, onde τ representa a parcela da renda tributada pelo governo que é devolvida em oferta pública de vagas. A receita governamental para oferta de vagas públicas é $\tau N_{t-1}E(w)$, onde $E(w)$ representa o salário médio recebido pelos indivíduos pertencentes à geração $t-1$. O custo de se ofertar N vagas públicas para indivíduos pertencentes a geração t, N_t^e, é dado por $C(N_t^e)$. Logo, o probema do governo é escolher N_t^e tal que

$$\widetilde{N}_t^e = argmax\ C(N_t^e) \leq \ \tau N_{t-1}E(w) \tag{2.2}$$

Por hipótese, assumiremos que $C(N_t^e) = N_t^e.c$, onde c é um parâmetro positivo que representa um custo unitário[58]. Além disto, $\widetilde{N}_t^e$ é escolhido tal que $C(\widetilde{N}_t^e) = \tau N_{t-1}E(w)$, ou seja, a receita tributária é devolvida por completo em oferta pública de vagas escolares e, portanto, $N_t^e.c = \tau N_{t-1}E(w)$. Dividindo ambos os lados por N_t, temos

$$\frac{N_t^e}{N_t} = \left(\frac{N_t}{N_{t-1}}\right)^{-1} \frac{\tau E(w)}{c} \tag{2.3}$$

[57] A metodologia de Lam & Marteleto é aplicada à base de dados utilizada neste artigo e os resultados são apresentados no Apêndice 2.

[58] Se assumirmos que $C(N_t^e)$ é homogênea de grau α, teríamos $C(\lambda N_t^e) = \lambda^\alpha C(N_t^e)$. Fazendo $\lambda = \frac{1}{(N_t^e)^{\frac{1}{\alpha}}}$, teríamos $C(N_t^e) = N_t^e C\left((N_t^e)^{\frac{\alpha-1}{\alpha}}\right)$. Se a função custo for homogênea de grau 1, $\alpha = 1$, então $C(N_t^e) = N_t^e C(1) = N_t^e c$

A probabilidade de estudar, pelo lado da oferta pública de vagas escolares, $\frac{N_t^e}{N_t}$, depende de um componente demográfico, a taxa de fecundidade N_t/N_{t-1}, a alíquota do tributo, da renda média e do custo unitário da educação[59]. Assim, uma queda da taxa de fecundidade provocaria um aumento na probabilidade de matrícula dada por

$$\frac{\partial\left(\frac{N_t^e}{N_t}\right)}{\partial\left(\frac{N_t}{N_{t-1}}\right)} = -1\left(\frac{N_t}{N_{t-1}}\right)^{-2}\frac{\tau E(w)}{c} < 0 \qquad (2.3)$$

A probabilidade de matrícula ainda pode aumentar com o aumento da renda média, um aumento da alíquota ou mesmo uma queda do custo unitário médio da oferta de vaga. Empiricamente, estes efeitos de renda média e custo uniário podem ser capturados por efeitos fixos de localidade ou efeito fixo de tempo. A equação empírica a ser estimada segue um modelo de escolha discreta probit, controlado por características familiares como educação do pai, educação da mãe e seus respectivos termos quadráticos para se capturar uma concavidade da educação na taxa de matrícula, componentes demográficos familiares como número de irmãos. Mudanças na taxa de fecundidade gera mudanças tanto no tamanho absoluto da coorte quanto na distribuição etária da população como verificamos nos gráficos 2.1 e 2.4. Os resultados apresentados referem-se à utilização da coorte relativa como variável resultante das mudanças demográficas ocorridas no Brasil entre as décadas de 1980 e 2000 e, no Apêndice B são apresentados os resultados da atualização do artigo de Lam & Marteleto com a taxa de

[59] O resultado do modelo não alteraria caso não se assumisse homogeneidade de grau 1 para a função de produção. Neste caso, teríamos $C(N_t^e) = N_t^e C\left((N_t^e)^{\frac{\alpha-1}{\alpha}}\right) = N_t^e c(N_t^e)$ e

$$\frac{N_t^e}{N_t} = \left(\frac{N_t}{N_{t-1}}\right)^{-1}\frac{\tau E(w)}{c(N_t^e)}$$

Note que se $\tau N_{t-1}E(w)$ é invariante em N_t^e, basta que $C(N_t^e)$ seja função injetora para que exista um único $\widetilde{N}_t^e$ tal que $C(N_t^e) = \tau N_{t-1}E(w)$. Além disto, $C(N_t^e)$ tem de ser assumido contínuo e, dado que N_t^e está contido em um conjunto compacto (ínfimo é zero e o supremo é N_t), o teorema de Weierstrass garante a existência de um máximo em (2.2)

crescimento do tamanho da coorte (equação 2.1) como componente da mudança demográfica[60]. O modelo segue a especificação:

$$P\left(E_{i,c,s,t} = 1\right) = \Phi\left(\beta_0 + \beta_1 Coorte_Relativa_{c,t,s} + \beta_2 X_{i,s,t} + \beta_3 Urbano + \beta_4 Idade + \beta_5 Ano + \beta_6 S\right) \quad (2.4)$$

Onde:

- $P\left(E_{i,c,s,t} = 1\right)$ é a probabilidade de o indivíduo de i da coorte c na localidade s e ano t estudar
- $Coorte_Relativa_{c,t,s}$ é o percentual da coorte c, no ano t e na localidade s no total da população
- $X_{i,s,t}$ é o vetor de características domiciliares definidos como Educação da Mãe, Educação do Pai e seus respectivos valores ao quadrado, Número de irmãos de 0 a 6 anos e Número de irmãos de 7 a 14 anos
- $Urbano$ é uma *dummy* com valor 1 se o domicílio que o indivíduo i habita é localizado na zona urbana
- $Idade$ é o vetor que representam *dummies* de idade
- Ano é um vetor com dummies de ano 1981 - 2009
- S é um vetor de dummies de localidade: uma *dummy* para cada uma das 26 nidades da federação

[60] A regressão da coorte relativa contra variáveis domiciliares, variáveis relativas aos indivíduos, controle de ano e controle de localidade, conforme a seguinte especificação

$$Coorte_{Relativa_{c,t,s}} = \beta_0 + \beta_1 X_{i,s,t} + \beta_2 Urbano + \beta_3 Idade + \beta_4 Ano + \beta_6 S + \epsilon_{c,t,s}$$

Sugere que Coorte Relativa não está correlacionada com as variáveis independentes do modelo. Mudanças no tamanho da coorte ou mudanças na coorte relativa são exógenas aos outros fatores que explicam taxa de matrícula.

TABELA 1: COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA

VARIÁVEIS	Regressão 1 8-9 anos	Regressão 2 10-13 anos	Regressão 3 14-17 anos
Coorte Relativa	-6.119***	2.616***	-2.665***
	(0.0999)	(0.0475)	(0.0454)
Mulher	0.107***	0.0873***	0.197***
	(0.000463)	(0.000351)	(0.000354)
Urbana	0.226***	0.292***	0.361***
	(0.000521)	(0.000392)	(0.000404)
Nº Irmãos 0-6 anos	-0.117***	-0.116***	-0.0830***
	(0.000233)	(0.000182)	(0.000209)
Nº Irmãos 7-17 anos	-0.0216***	-0.00644***	-0.0187***
	(0.000156)	(0.000125)	(0.000126)
Educação Pai	0.103***	0.0910***	0.0904***
	(0.000209)	(0.000164)	(0.000161)
Educação Mãe	0.167***	0.142***	0.112***
	(0.000209)	(0.000163)	(0.000161)
Educação Pai quadrado	-0.00434***	-0.00318***	-0.00185***
	(1.92e-05)	(1.59e-05)	(1.46e-05)
Educação Mãe quadrado	-0.00727***	-0.00617***	-0.00331***
	(2.06e-05)	(1.69e-05)	(1.56e-05)
9 anos	0.195***		
	(0.000466)		
11 anos		-0.0540***	
		(0.000504)	
12 anos		-0.198***	
		(0.000498)	
13 anos		-0.428***	
		(0.000494)	
15 anos			-0.265***
			(0.000477)
16 anos			-0.530***
			(0.000489)
17 anos			-0.793***
			(0.000516)
Constante	0.550***	0.304***	0.135***
	(0.00674)	(0.00583)	(0.00550)
Observações	72,616,393	111,995,442	65,812,053
Controle Ano	SIM	SIM	SIM
Controle UF	SIM	SIM	SIM
Amostra 1981 - 2009	SIM	SIM	SIM

Erro padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

Os resultados são apresentados na Tabela 1 para três regressões diferentes, uma amostra com indivíduos de 8 a 9 anos, uma amostra com indivíduos de 10 a 13 anos e uma amostra com indivíduos de 14 a 17 anos. Para a amostra com indivíduos de 8 a 9 anos, a coorte relativa representa o percentual de indivíduos de 8 a 9 anos no total da população por localidade e ano. Para a amostra com indivíduos de 10 a 13 anos, a coorte relativa é o percental deste grupo etário no total da população por localidade e ano e, por fim, para a amostra com indivíduos de 14 a 17 anos, a coorte relativa representa o percentual de indivíduos de 14 a 17 anos no total da população[61]. Os resultados mostram que o tamanho relativo da coorte é negativamente correlacionado com a probabilidade de matrícula para indivíduos com 8 a 9 anos e para indivíduos com 14 a 17 anos, como sugere o modelo, porém o tamanho relativo da coorte apresentou uma relação positiva com a probabilidade de matrícula para indivíduos com 10 a 13 anos e estes parâmetros são estatisticamente significativos a 1%. O componente demográfico familiar calculado a partir do número de irmãos apresentou relação negativa, ou seja, quanto maior o número de irmãos, menor a probabilidade de matrícula para o indivíduo em análise. A magnitude do número de irmãos mais novos, representado por irmãos de 0 a 6 anos, é maior do que a magnitude do número de irmãos de 7 a 17 anos. Educação dos pais está positivamente relacionada com probabilidade de matrícula do filho e o sinal negativo da educação ao quadrado sinaliza uma concâvidade da função, ou seja, um ano a mais de educação dos pais aumenta a probabilidade de matrícula mas com incrementos menores conforme aumenta-se a educação dos pais.

A Tabela 2 apresenta os efeitos marginais das variáveis analisadas nas regressões anteriores. A coorte relativa apresentou média de 0,0437 na amostra de regressão de 8 a 9 anos, 0,0436 na amostra de regressão de 10 a 13 anos e 0,0.432 na amostra de regressão de 14 a 17 anos. Para a regressão de 8 a 9 anos, a coorte relativa tinha valor 0,05 em 1981 e valor 0,035 em 2009, uma variação negativa de 0,0149 implicando em um efeito marginal de 1,28 pontos percentuais na probabilidade de matrícula. Para a regressão de 14 a 17 anos, a coorte relativa tinha valor 0,096 em 1981 e valor 0,075 em 2009, uma variação negativa de 0,02 implicando em um efeito marginal de 1,6% na probabilidade de matrícula. Um ano a mais de

[61] Ao se controlar pela coorte de 7 a 14 anos, os efeitos imediatos da diminuição do tamanho (absoluto ou relativo) da coorte não são capturados. Além disto, o efeito para uma coorte específica é prolongado ao se agregar o grupo etário. Assim, para cada regressão em análise, o tamanho da coorte relativa terá como base o grupo etário a ser analisado.

escolaridade do pai aumenta a probabilidade de matrícula em 1,1% na regressão 1, 1,2% na regressão 2 e 2,4% na regressão 3[62]. Estes valores para a escolaridade da mãe aumentam a probabilidade de matrícula, respectivamente, em 1,8%, 1,8% e 2,8%.

TABELA 2: EFEITOS MARGINAIS PARA PROBABILIDADE DE MATRÍCULA

	Regressão 1 8 - 9 anos	**Regressão 2 10 - 13 anos**	**Regressão 3 14 - 17 anos**
Coorte Relativa	-0.8578	0.4139	-0.7750
	(0.014)	(0.008)	(0.013)
Mulher	0.0150	0.0138	0.0572
	(0.000)	(0.000)	(0.000)
Urbana	0.0316	0.0463	0.1051
	(0.000)	(0.000)	(0.000)
Nº Irmãos 0 a 6 anos	-0.0165	-0.0183	-0.0241
	(0.000)	(0.000)	(0.000)
Nº Irmãos 7 a 17 anos	-0.0030	-0.0010	-0.0054
	(0.000)	(0.000)	(0.000)
Educação Pai	0.0145	0.0144	0.0263
	(0.000)	(0.000)	(0.000)
Educação Mãe	0.0234	0.0225	0.0326
	(0.000)	(0.000)	(0.000)
Educação Pai quadrado	-0.0006	-0.0005	-0.0005
	(0.000)	(0.000)	(0.000)
Educação Mãe quadrado	-0.0010	-0.0010	-0.0010
	(0.000)	(0.000)	(0.000)

OBS.: D*ummies* de 8, 10 e 14 anos omitidas.

[62] As médias de escolaridade do pai para as amostras de Regressão 1, Regressão 2 e Regressão 3 são, respectivamente 4,9, 4,6 e 4,4 anos de estudo.

4. REGRESSÃO POR SÉRIES DO ENSINO FUNDAMENTAL E ENSINO MÉDIO

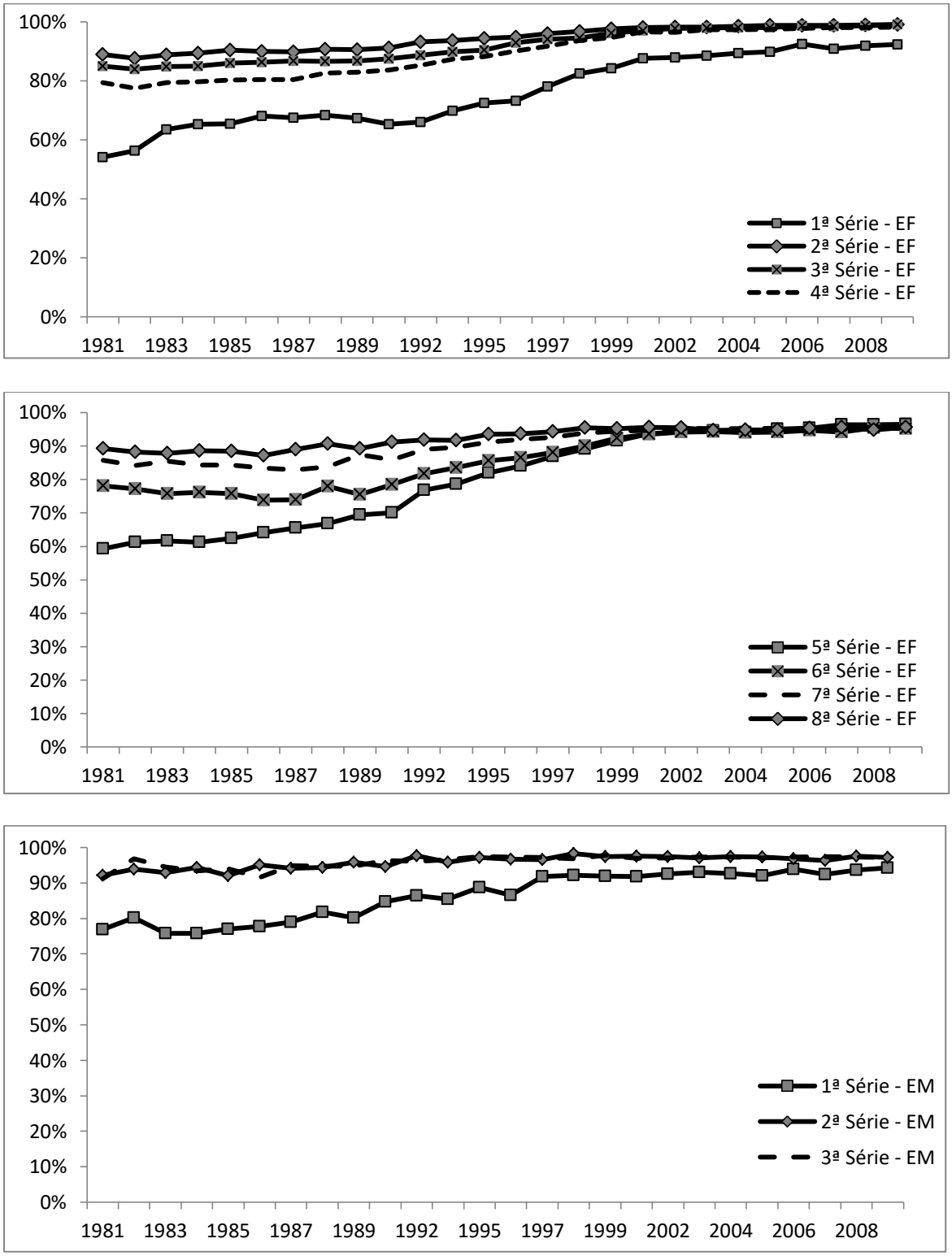
Ao se estabelecer a regressão por grupos etários, como apresentado na seção anterior na qual a metodologia de Lam & Marteleto guiou a estratégia empírica, o efeito de tamanho da coorte pode estar capturando outras informações. No Apêndice C são apresentadas diversas estatísticas sobre defasagem idade-série e a documentação dos dados da PNAD mostra, por exemplo, que na década de 1980 os indivíduos com 7 anos constituíam a menor taxa de matricula entre os indivíduos do grupo etário de 7 a 10 anos e a sua taxa de matrícula cresceu *pari passu* a diminuição do número de matriculados na 1ª série do Ensino Fundamental, ao passo que a defasagem na 1ª séria diminui ao longo das décadas[63]. Assim, faremos a decomposição da estratégia empírica por série. Para separar a base de dados por séries, usaremos o seguinte critério:

- Indivíduo i de 7 a 17 anos estuda ou não na primeira série: serão classificados os indivíduos que declararam estudar na primeira série do Ensino Fundamental e aqueles que não estudam e declararam que nunca concluíram alguma série com aprovação.
- Indivíduo i de 7 a 17 anos estuda ou não na segunda série: serão classificados os indivíduos que declaram estudar na segunda série do Ensino Fundamental e aqueles que não estudam e declararam ter concluído com aprovação a primeira série.
- Indivíduo i de 7 a 17 anos estuda ou não na terceira série: serão classificados os indivíduos que declaram estudar na terceira série do Ensino Fundamental e aqueles que não estudam e declararam ter concluído com aprovação a segunda série.
- Indivíduo i de 7 a 17 anos estuda ou não na quarta série: serão classificados os indivíduos que declaram estudar na quarta série do Ensino Fundamental e aqueles que não estudam e declararam ter concluído com aprovação a terceira série.

E assim sucessivamente. O Gráfico 2.6 apresenta a decomposição da taxa de matrícula para cada uma das séries e, segundo o gráfico, a 1ª e a 5ª séries do Ensino Fundamental assim como a 1ª série do Ensino Médio apresentam as menores taxas de matrícula na década de 1980 e apresentam uma importante tendência de aumento do percentual de matriculados.

[63] No Apêndice 3 são apresentadas estatísticas agregadas de defasagem idade-série além de uma referência metodológica para o cálculo da defasagem. O Brasil fez uma profunda transformação no ensino básico com diminuição da defasagem escolar para as coortes mais jovens. Este processo possibilita melhores indicadores de matrícula.

GRÁFICO 2.6: PERCENTUAL QUE ESTUDA POR SÉRIE DOS ENSINOS FUNDAMENTAL E MÉDIO



Fonte: Microdados da PNAD de 1981 a 2009
OBS. EF equivale a Ensino Fundamental e EM equivale a Ensino Médio

A estratégia empírica segue a equação (2.4) porém decompondo a amostra por séries j segunda a especificação:

$$P(E_{i,j,c,s,t} = 1) = \Phi(\beta_0 + \beta_1 Coorte_Potencial_{j,t,s} + \beta_2 X_{i,s,t} + \beta_3 Urbano + \beta_4 Idade + \beta_5 Ano + \beta_6 S) \quad (2.5)$$

Onde:

- é a probabilidade de o indivíduo de i da coorte c na localidade s e ano t estudar na série j, sendo $j = 1, \ldots, 11$
- Para cada série j, a $Coorte_Potencial_{j,t,s}$ é a coorte potencial demandante não defasada. Por exemplo, para a 1ª série, $j = 1$, a coorte potencial constitui o tamanho relativo da população de 7 anos no total da população. Para a 2ª série, $j = 2$, a coorte potencial constitui o tamanho relativo da população de 8 anos no total da população e assim sucessivamente.
- As outras variáveis são explicadas na especificação inicial.

Os resultados para as onze séries, oito séries referentes ao ensino fundamental e três séries referentes ao ensino médio, são apresentados nas Tabelas 3, 4 e 5. Os efeitos marginais destas regressões são apresentados no Apêndice 4. Em termos agregados, uma queda do tamanho relativo da coorte potencial aumenta a probabilidade de matrícula da 1ª a 4ª séries, da 5ª a 8ª séries, ambos do ensino fundamental, e da 1ª a 3ª séries do ensino médio. Ao desagregarmos por séries, o efeito mostrou-se negativo e com maior magnitude para a 1ª série, à 5ª série, ambos do ensino fundamental, e à 1ª série do ensino médio. Todas estas séries apresentavam as menores taxas de matrículas conforme verificado no Gráfico 2.6. A 3ª série do Ensino Médio apresentou parâmetro negativo, mas é importante ressaltar que a base de dados é constituída por indivíduos de 7 a 17 anos, os potenciais demandantes da 3ª série do ensino médio. Dois exercícios contrafactuais foram realizados: 1) simulação da taxa de matrícula caso a coorte relativa ficasse constante a valores de 1981 e 2) simulação da taxa de matrícula caso a educação dos pais ficasse constante a valores de 1981. Para a 1ª e 5ª séries do ensino fundamental e 1ª série do ensino médio, onde o parâmetro da coorte potencial relativa apresentou correlação negativa com a probabilidade de matrícula, as simulações contrafactuais

com coorte relativa a valores de 1981 não explicaram os ganhos na expansão do acesso ao ensino básico.

TABELA 3: COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 4ª SÉRIES DO ENSINO FUNDAMENTAL

VARIÁVEIS	(1) 1ª-4º series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie	(5) 4ª serie
Coorte Relativa	-3.694***	-8.869***	2.352***	0.345	3.694***
	(0.0329)	(0.111)	(0.234)	(0.239)	(0.258)
Mulher	0.117***	0.0640***	0.112***	0.0600***	0.00763***
	(0.000236)	(0.000312)	(0.000663)	(0.000698)	(0.000728)
Urbana	0.183***	0.118***	0.0353***	0.0697***	0.281***
	(0.000264)	(0.000356)	(0.000731)	(0.000771)	(0.000793)
Irmãos 0-6 anos	-0.103***	-0.0718***	-0.0919***	-0.0880***	-0.0739***
	(0.000119)	(0.000153)	(0.000334)	(0.000371)	(0.000412)
Irmãos 7-17 anos	-0.00314***	0.00449***	0.00612***	-0.00995***	-0.00247***
	(7.96e-05)	(0.000102)	(0.000228)	(0.000248)	(0.000260)
Educação Pai	0.0951***	0.0810***	0.0580***	0.0482***	0.0565***
	(0.000109)	(0.000148)	(0.000314)	(0.000328)	(0.000335)
Educação Mãe	0.142***	0.134***	0.0658***	0.0719***	0.0832***
	(0.000110)	(0.000148)	(0.000320)	(0.000334)	(0.000334)
Educação Pai ^{2}	-0.00348***	-0.00288***	-0.00130***	-0.000286***	-0.000419***
	(1.05e-05)	(1.40e-05)	(3.27e-05)	(3.44e-05)	(3.49e-05)
Educação Mãe ^{2}	-0.00535***	-0.00476***	-0.000525***	-0.00137***	-0.00346***
	(1.13e-05)	(1.51e-05)	(3.62e-05)	(3.76e-05)	(3.58e-05)
Constante	0.178***	-0.0680***	1.093***	1.015***	-1.094***
	(0.00412)	(0.00420)	(0.00885)	(0.0100)	(0.00874)
Nº Observações	207,900,324	86,415,076	48,953,255	40,173,779	32,358,214
Dummies de Ano	SIM	SIM	SIM	SIM	SIM
Dummies de UF	SIM	SIM	SIM	SIM	SIM

Erro Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

TABELA 4: COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 5ª A 8ª SÉRIES DO ENSINO FUNDAMENTAL

VARIÁVEIS	(1) 5ª-8º series	(2) 5ª serie	(3) 6ª serie	(4) 7ª serie	(5) 8ª serie
Coorte Relativa	-0.720***	-5.867***	16.42***	39.10***	8.142***
	(0.0615)	(0.247)	(0.331)	(0.437)	(0.514)
Mulher	0.0612***	-0.0666***	0.112***	0.109***	0.103***
	(0.000440)	(0.000686)	(0.000867)	(0.00115)	(0.00136)
Urbana	0.581***	0.801***	0.284***	0.102***	-0.0235***
	(0.000500)	(0.000745)	(0.00104)	(0.00144)	(0.00181)
Irmãos 0-6 anos	-0.106***	-0.0815***	-0.0970***	-0.0771***	-0.0242***
	(0.000295)	(0.000427)	(0.000587)	(0.000851)	(0.00110)
Irmãos 7-17 anos	-0.0439***	-0.0307***	-0.0155***	-0.0191***	-0.0297***
	(0.000165)	(0.000250)	(0.000327)	(0.000439)	(0.000543)
Educação Pai	0.0650***	0.0552***	0.0458***	0.0260***	0.01000***
	(0.000199)	(0.000323)	(0.000392)	(0.000530)	(0.000617)
Educação Mãe	0.0879***	0.0781***	0.0528***	0.0501***	0.0390***
	(0.000203)	(0.000326)	(0.000402)	(0.000545)	(0.000626)
Educação Pai ^{2}	-0.000186***	0.000652***	0.000704***	0.00181***	0.00290***
	(1.83e-05)	(3.41e-05)	(3.58e-05)	(4.62e-05)	(5.00e-05)
Educação Mãe ^{2}	-0.00207***	-0.00167***	-2.58e-05	0.000748***	6.84e-05
	(1.95e-05)	(3.55e-05)	(3.85e-05)	(5.06e-05)	(5.13e-05)
Constante	0.196***	0.0739***	-0.0321***	0.640***	0.748***
	(0.00926)	(0.0103)	(0.0124)	(0.0190)	(0.0188)
Nº Observações	67,636,686	29,343,293	18,098,311	11,837,806	8,357,276
Dummies de Ano	SIM	SIM	SIM	SIM	SIM
Dummies de UF	SIM	SIM	SIM	SIM	SIM

Erro Padrão entre parênteses

TABELA 5: COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 3ª SÉRIES DO ENSINO MÉDIO

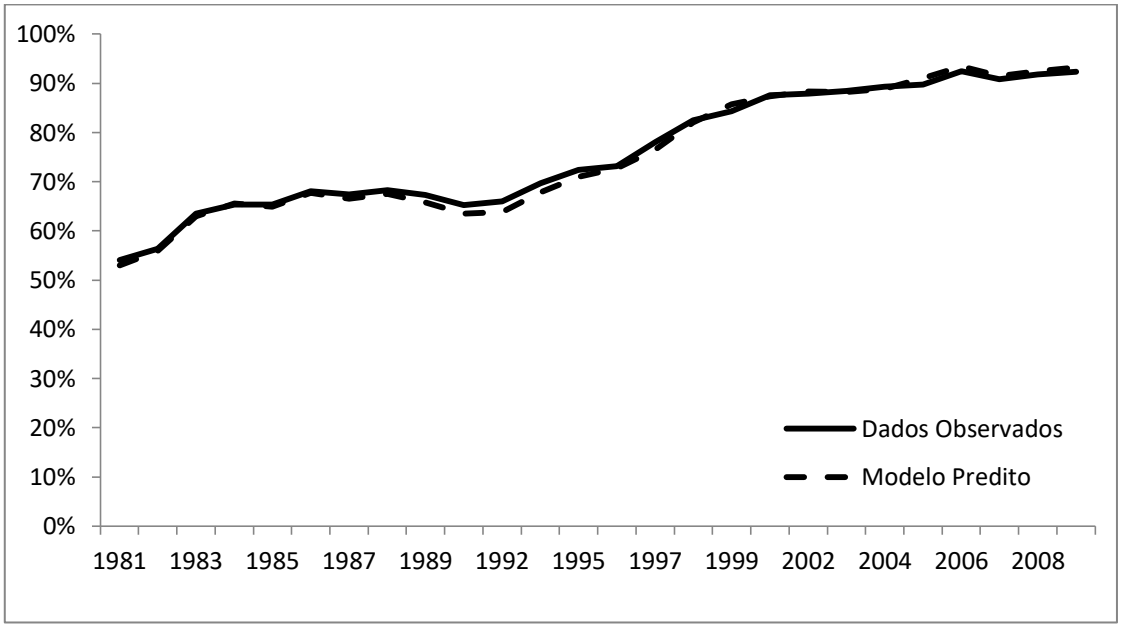
VARIÁVEIS	(1) 1ª-3ª series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie
Coorte Relativa	-9.417***	-16.69***	32.69***	-7.46***
	(0.206)	(0.556)	(1.113)	-1.517
Mulher	0.205***	0.147***	0.217***	0.227***
	(0.00119)	(0.00150)	(0.00266)	(0.00403)
Urbana	0.188***	0.248***	0.0198***	-0.104***
	(0.00159)	(0.00196)	(0.00374)	(0.00595)
Irmãos 0-6 anos	-0.0537***	-0.0542***	-0.0162***	-0.0796***
	(0.00102)	(0.00124)	(0.00237)	(0.00411)
Irmãos 7-17 anos	-0.0809***	-0.0687***	-0.0229***	-0.104***
	(0.000492)	(0.000604)	(0.00119)	(0.00176)
Educação Pai	0.0573***	0.0602***	0.0291***	0.0801***
	(0.000526)	(0.000658)	(0.00119)	(0.00172)
Educação Mãe	0.0698***	0.0583***	0.0111***	0.105***
	(0.000555)	(0.000702)	(0.00127)	(0.00182)
Educação Pai ^{2}	-0.000460***	-0.00123***	0.00194***	-0.00267***
	(3.85e-05)	(4.88e-05)	(8.70e-05)	(0.000116)
Educação Mãe ^{2}	-0.000473***	0.000943***	0.00122***	-0.00330***
	(4.28e-05)	(5.70e-05)	(9.15e-05)	(0.000127)
Constante	0.755***	0.732***	-0.862***	-0.502***
	(0.0186)	(0.0215)	(0.0351)	(0.0484)
Nº Observações	9,420,388	5,442,901	2,722,593	1,206,043
Dummies de Ano	SIM	SIM	SIM	SIM
Dummies de UF	SIM	SIM	SIM	SIM

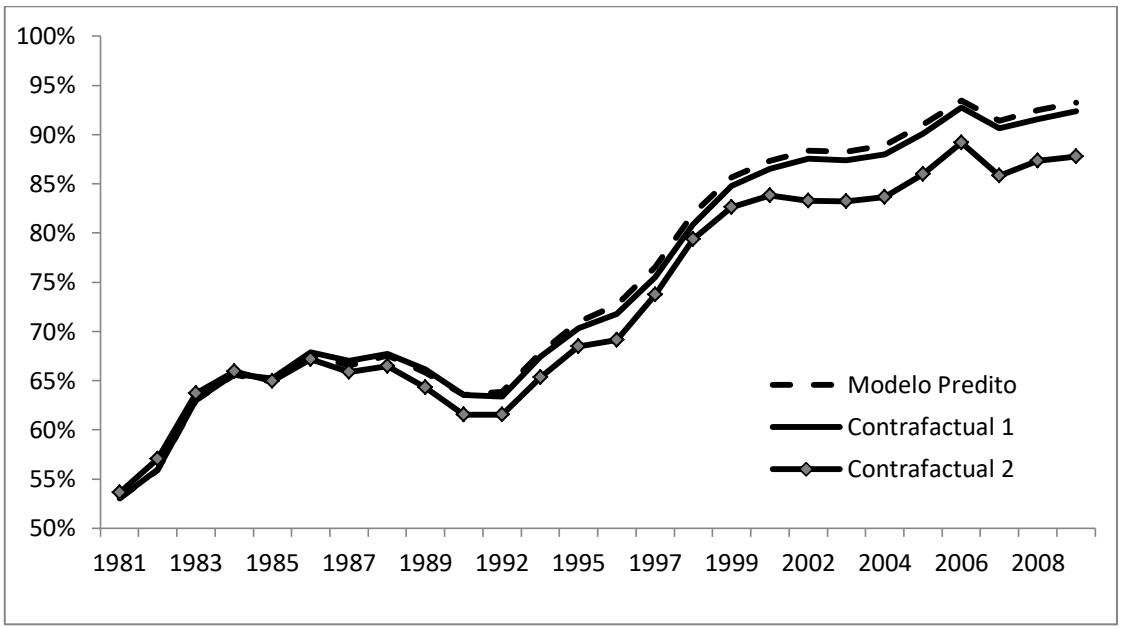
Erro Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

A simulação contrafactual da escolaridade dos pais mostrou que para a 1ª série do ensino fundamental, o maior problema para o acesso a educação ao longo do século XX como documenta Viégas & Souza (2006), mostrou efeito conforme é verificado no Gráfico 2.7[64]. Com valores constantes de 1981, ou seja, se a escolaridade dos pais permanecesse na média e 3 anos de estudo, o percentual de indivíduos matriculados na 1ª série, do total de demandantes da 1ª série, seria 88%, contra 93% observado nos dados.

[64] As estatísticas descritivas da educação dos pais é apresentada no Apêndice A. Para indivíduos de 7 a 10 anos, a escolaridade do pai e da mãe cresce de 3 anos de estudo, em 1981, para 7 anos de estudo em 2009, na média.

GRÁFICO 2.7: SIMULAÇÃO CONTRAFACTUAL PARA A PRIMEIRA SÉRIE DO ENSINO FUNDAMENTAL





Fonte: Microdados da PNAD de 1981 a 2009
OBS. Contrafactual 1: Coorte relativa a valores de 1981.
OBS2. Contrafactual 2: Educação dos pais a valores de 1981

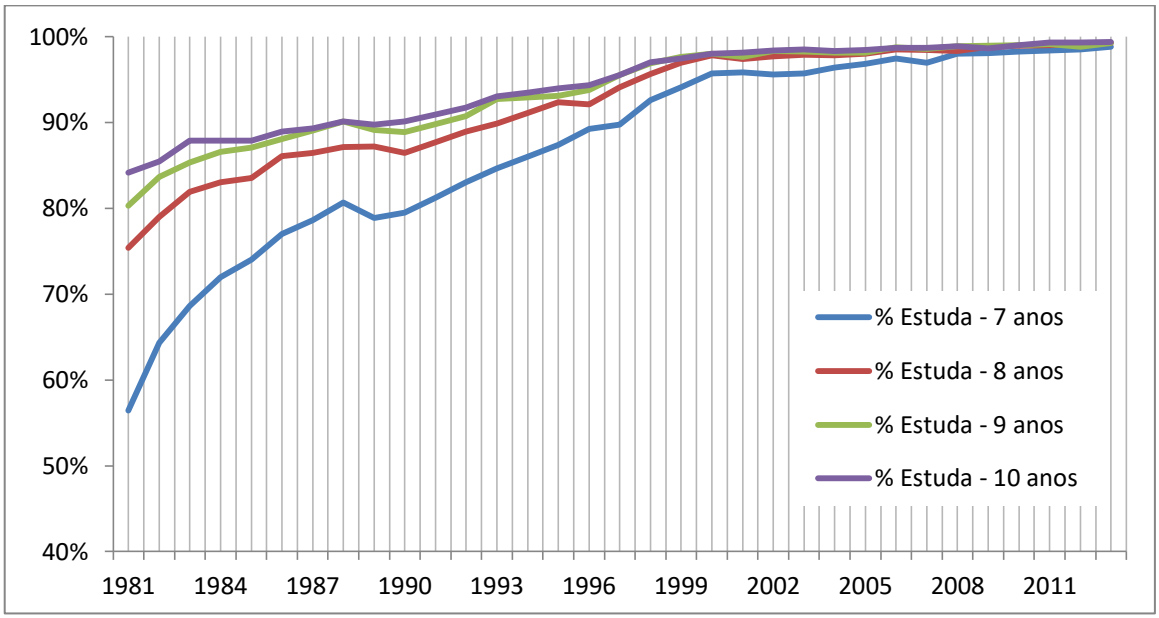
5. CONCLUSÃO

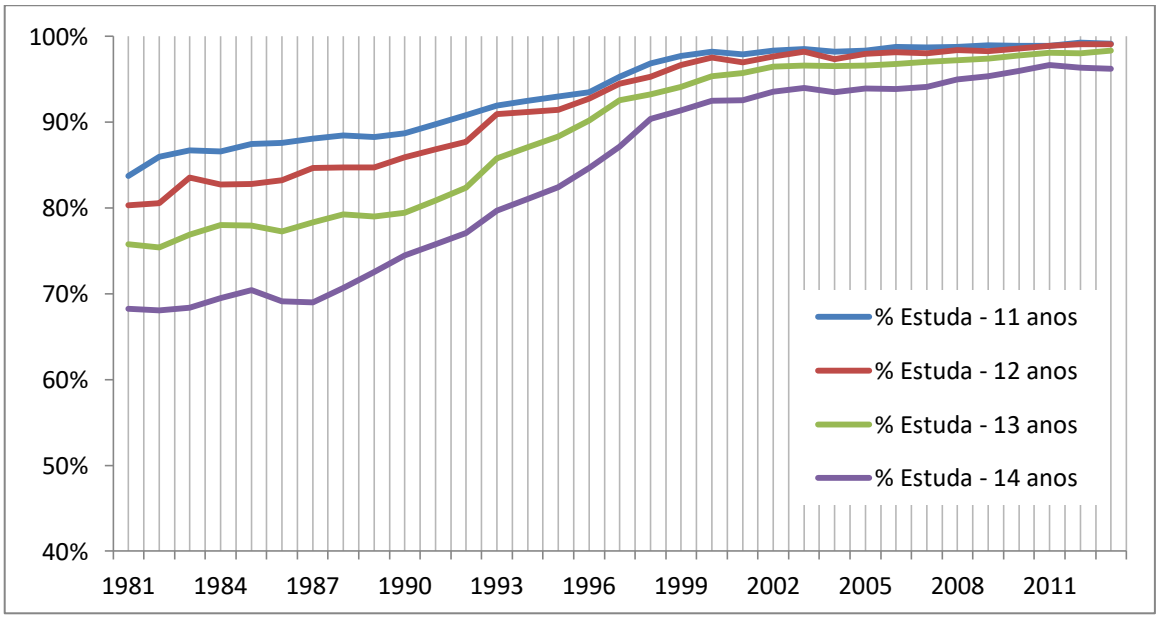
O presente artigo investiga o efeito demográfico sobre indicadores educacionais para o Brasil ao longo de três décadas, período de profundas transformações em termos educacionais. Segunda tradição da literatura teórica sobre educação e demografia, Becker e Lewis (1973) mostram que há um *trade-off* entre quantidade de filhos e qualidade. Parte da literatura empírica investigou o efeito causal do tamanho da família sobre a educação dos filhos, com experimentos com gêmeos, como por exemplo, Ponczek e Souza (2012) para o Brasil. Os resultados apontavam para um aumento da escolaridade dos filhos quando pais decidem ter menos filhos. No entanto, há ainda um segundo fator demográfico que seria o tamanho da coorte ao longo do tempo e como sua variação afeta a probabilidade de um indivíduo estar matriculado.

Os resultados encontrados neste artigo sugerem que tamanho da família é uma variável importante para a probabilidade de estar estudando e este efeito é heterogêneo, dependendo da ordem de nascimento do indivíduo. Quanto maior a quantidade de filhos entre 0 e 6 anos, menor a probabilidade de um indivíduo de 7 a 10 anos estar matriculado. Este mesmo efeito negativo é apresentado com menor magnitude para o número de irmão de 7 a 14 anos. A demografia, representada pela coorte relativa, contribuiu positivamente na probabilidade de matrícula ao longo das décadas, mas na margem devido à baixa magnitude. Quando decomposto por séries, a 1ª e a 5ª séries do ensino fundamental assim como a 1ª série do ensino médio, todas estas séries com importante aumento da probabilidade de matrícula devido à expansão do ensino básico, apresentaram correlação negativa entre probabilidade de matrícula e tamanho relativo da coorte potencial da série. As simulações contrafactuais mostram que mesmo com o efeito esperado pela hipótese gerada pelo modelo que racionaliza a oferta de vagas, a magnitude dele não explica os melhores indicadores de probabilidade de matrícula. Na segunda simulação contrafactual, educação dos pais foi importante, em magnitude, para o aumento da probabilidade de matrícula na primeira série do ensino fundamental, a série que constituía um gargalo pela defasagem escolar.

APÊNDICE A: OUTRAS ESTATÍSTICAS E RESULTADOS

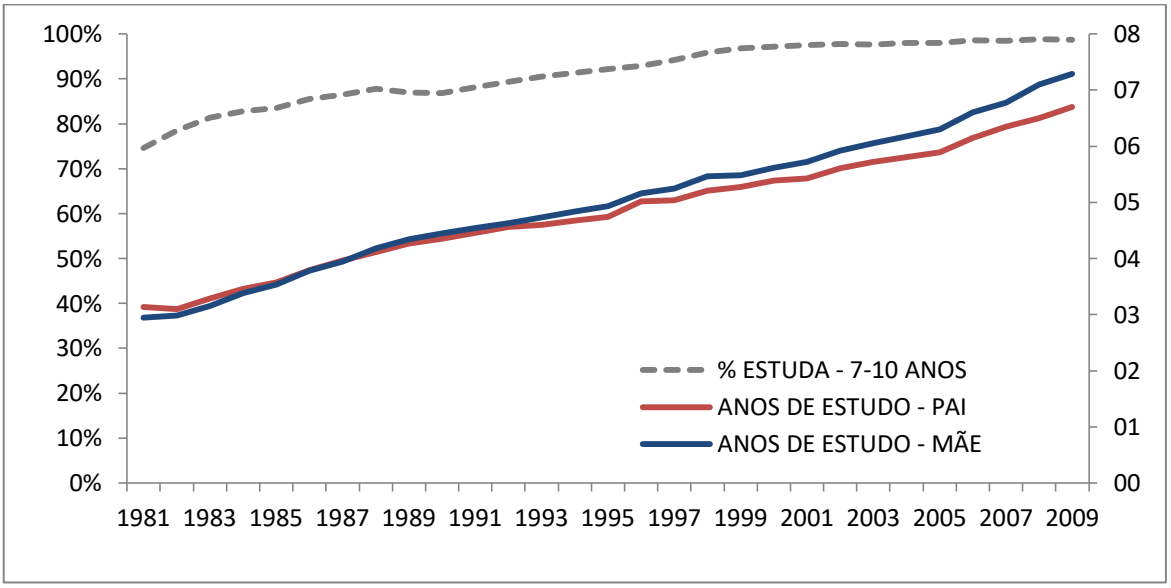
GRÁFICO A1.1: EVOLUÇÃO DA TAXA DE MATRÍCULA E COORTE RELATIVAPOR GRUPO ETÁRIO DESAGREGADO

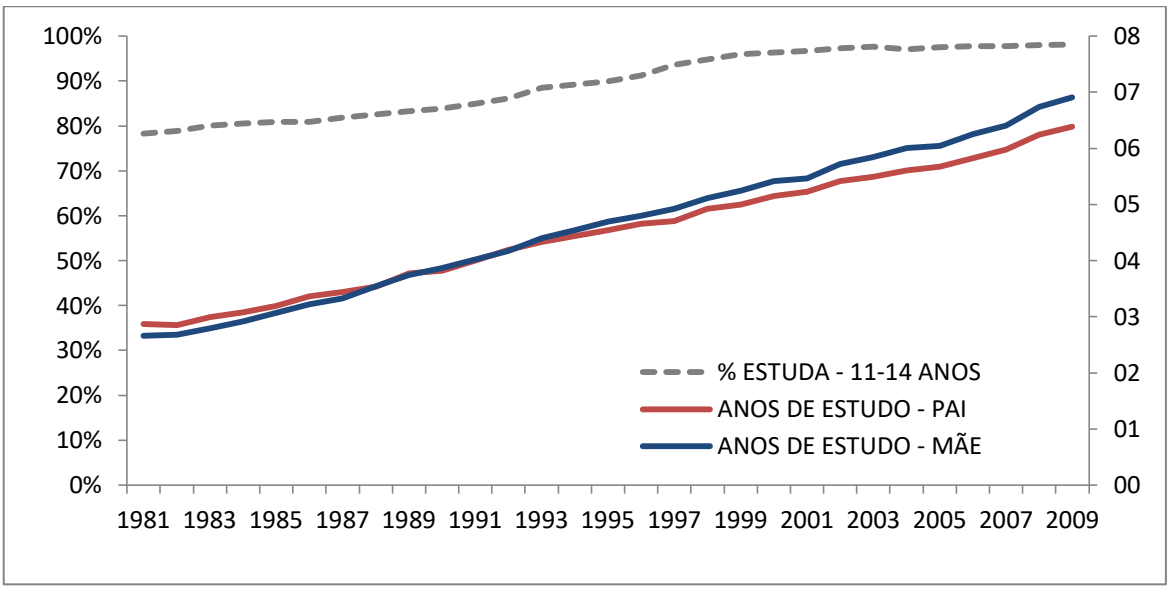




Fonte: Microdados da PNAD de 1981 a 2009
OBS.: Interpolação linear para os anos de 1991, 1994 e 2000.

GRÁFICO A1.2: EVOLUÇÃO DA TAXA DE MATRÍCULA E ESCOLARIDADE MÉDIA DOS PAIS POR GRUPOS ETÁRIOS





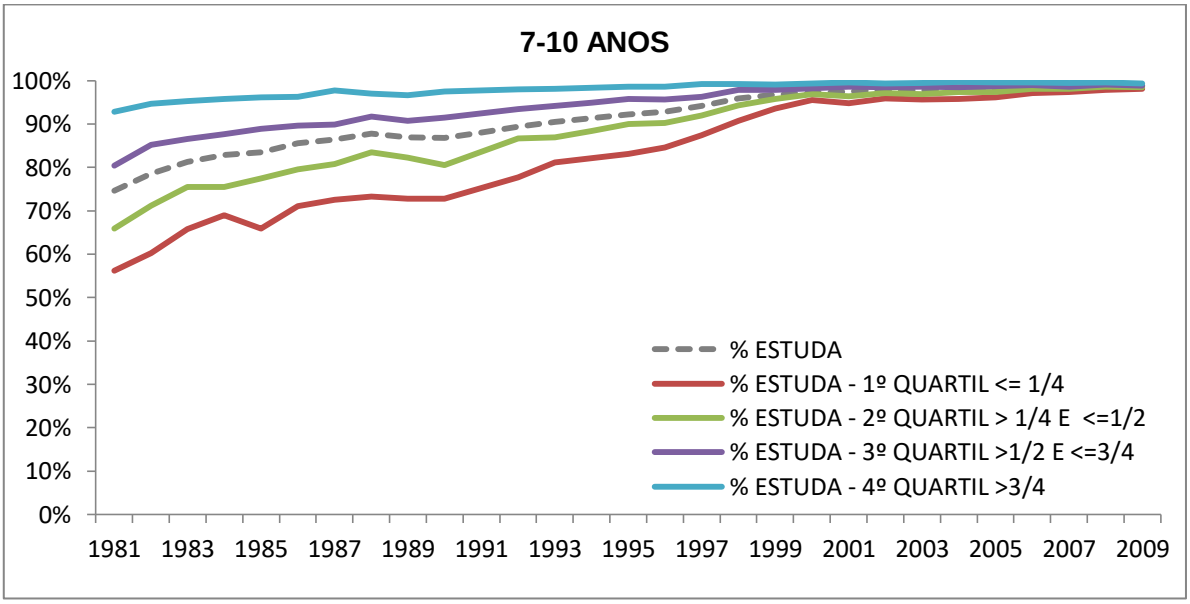
Fonte: Microdados da PNAD de 1981 a 2009
OBS.: Interpolação linear para os anos de 1991, 1994 e 2000.
OBS 2: % Estuda no eixo da esquerda. Anos médios de estudo no eixo da direita.

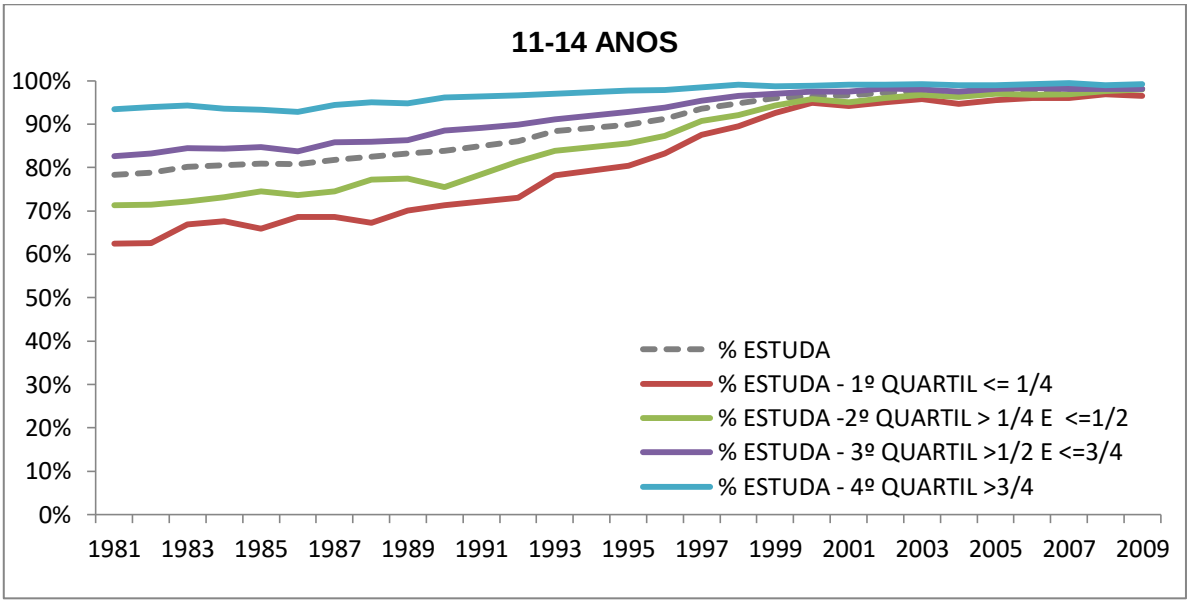
APÊNDICE A.2: DECOMPOSIÇÃO POR QUARTIL DE RENDA FAMILIAR.

Com base nas informações familiares incorporadas para cada indivíduo através de um pareamento para se identificar o pai e a mãe na base de dados, foram criadas informações sobre renda familiar e a posição da renda familiar na distribuição da renda total. Para cada família, a renda familiar foi criada somando-se a renda de todas as fontes, ou seja, a renda de todos os trabalhados e a renda não-trabalho, do pai e da mãe[65]. Esta renda familiar agregada foi incorporada para cada membro da família. Com base nesta informação, foram geradas quatro medidas de distribuição da renda familiar. 1ª) Foi gerada uma distribuição da renda familiar para cada unidade da federação e para cada ano. 2ª) Foi gerada uma distribuição da renda familiar para cada unidade da federação porém todos os anos foram analisados conjuntamente. Neste caso, a renda familiar foi deflacionada com o IPCA/INPC e analisada em valores reais. 3ª) Foi gerada uma distribuição da renda familiar para cada ano porém as unidades da federação foram analisadas conjuntamente. 4º) Foi gerada uma distribuição da renda familiar agregada em valores reais sem decompor por ano ou unidade da federação. Uma vez gerada a distribuição da renda familiar, analisou-se a posição da renda familiar de um indivíduo em relação à distribuição tal que se este indivíduo tivesse uma renda familiar entre as famílias 25% mais pobres, tal indivíduo estaria no primeiro quartil da distribuição de renda. Se o indivíduo tivesse uma renda familiar superior os 25% mais pobres, mas inferior à mediana da distribuição de renda, tal indivíduo estaria no segundo quartil. Se o indivíduo tivesse uma renda familiar acima da mediana da distribuição, porém inferior à renda dos 25% mais ricos, tal indivíduo estaria no terceiro quartil da distribuição. Por fim, se o indivíduo tivesse uma renda entre os 25% mais ricos, ele estaria no quarto quartil da distribuição de renda familiar. Este processo foi realizado para cada uma das quatro distribuições de renda e as estatísticas descritivas criadas foram equivalentes entre si. Assim, somente o resultado da distribuição da renda familiar gerada por ano e por unidade da federação será apresentado.

[65] Não foi considerada a renda do filho, quando este trabalhava, na renda familiar assim como qualquer outro parente.

GRÁFICO A2.1: PERCENTUAL MATRICULADO POR QUARTIL DE RENDA FAMILIAR – ANÁLISE POR GRUPO ETÁRIO





Fonte: Microdados da PNAD de 1981 a 2009
OBS.: Interpolação linear para os anos de 1991, 1994 e 2000.

O Gráfico A2.1 apresenta o percentual de indivíduos de 7 a 10 anos e o percentual de indivíduos de 11 a 14 anos que estudavam decompostos por quartil de renda familiar[66]. Indivíduos pertencentes ao quartil dos 25% mais ricos apresentam alta taxa de matrícula no ensino regular e o processo de universalização do acesso à escola básico favoreceu principalmente os indivíduos pertencentes aos quartis mais pobres da distribuição da renda familiar. As Tabelas A3.1 e A3.2 apresentam o percentual de alunos matriculados no ensino regular público por quartil de renda familiar[67]. Entre 2002 e 2013, o percentual de alunos matriculados no ensino público é estável para todos os quartis e para os grupos etários 7-9 anos e 11-14 anos. Na média, 86% dos alunos matriculados no ensino regular frequentavam a rede pública. Se selecionarmos os indivíduos pertencentes aos primeiro, segundo e terceiro quartis (excluindo os 25% mais ricos), este percentual aumenta para 96%.

TABELA A3.1: PERCENTUAL QUE ESTUDA EM ESCOLA PÚBLICA POR QUARTIL RENDA FAMILIAR PARA 7 A 10 ANOS

	1º quartil	2º quartil	3º quartil	4º quartil
2002	97%	96%	89%	51%
2003	97%	95%	89%	52%
2004	97%	95%	89%	50%
2005	97%	95%	88%	50%
2006	97%	95%	87%	48%
2007	97%	94%	87%	49%
2008	96%	94%	86%	50%
2009	96%	93%	85%	46%
2010	95%	94%	85%	47%
2011	93%	94%	86%	48%
2012	94%	91%	84%	51%
2013	93%	92%	84%	48%

[66] Para cada ano e para cada Unidade da Federação é gerada uma distribuição da renda familiar onde apenas uma única informação por família é utilizada, para não se gerar efeito de composição. Ao se definir os valores críticos para os quartis da renda, é possível caracterizar em qual posição da distribuição da renda o indivíduo pertence, ou seja, em qual quartil da renda ele pertence.

[67] Esta informação está disponível na PNAD somente a partir do ano de 2002, quando a pesquisa domiciliar sofre algumas alterações. Para anos anteriores é possível encontrar algumas informações agregadas nas Estatísticas do Século XX do IBGE.

TABELA A3.2: PERCENTUAL QUE ESTUDA EM ESCOLA PÚBLICA POR QUARTIL RENDA FAMILIAR PARA 11 A 14 ANOS

	1º quartil	2º quartil	3º quartil	4º quartil
2002	98%	98%	93%	58%
2003	98%	98%	93%	59%
2004	97%	97%	92%	58%
2005	98%	98%	93%	59%
2006	98%	97%	92%	56%
2007	97%	97%	92%	58%
2008	97%	97%	92%	57%
2009	97%	97%	90%	55%
2010	96%	96%	91%	57%
2011	95%	96%	91%	60%
2012	96%	95%	90%	57%
2013	95%	95%	89%	57%

APÊNDICE 2: ESTIMAÇÕES DE LAM & MARTELETO (1981-2009)

O artigo de Lam & Mateleto, principal referência para investigar o efeito da tamanho absoluto da coorte na probabilidade de matrícula, segue a seguinte especificação

$$P(E_{i,c,s,t} = 1) = \Phi(\beta_0 + \beta_1 TxCresc_{c,t,s} + \beta_2 X_{i,s,t} + \beta_3 Urbano + \beta_4 Idade + \beta_5 Ano + \beta_6 S)$$

Onde:

- $P(E_{i,c,s,t} = 1)$ é a probabilidade de o indivíduo de i da coorte c na localidade s e ano t estudar
- $TxCresc_{c,t,s}$ é a taxa de crescimento do grupo etário de 7 a14 anos no ano t e na localidade s.
- $X_{i,s,t}$ é o vetor de características domiciliares definidos como Educação da Mãe, Educação do Pai e seus respectivos valores ao quadrado, Número de irmãos de 0 a 6 anos e Número de irmãos de 7 a 14 anos
- $Urbano$ é uma *dummy* com valor 1 se o domicílio que o indivíduo i habita é localizado na zona urbana
- $Idade$ é o vetor que representam quatro *dummies* de idade
- Ano é capturado por uma tendência temporal linear
- S é um vetor de dummies de localidade (uma *dummy* para cada uma das 5 Regiões Geográficas.)

A base de dados construída na seção 2 é aplicada à especificação econométrica acima e a taxa de crescimento da coorte é calculada com base na projeção populacional imputada, conforme documentado ao longo do texto na seção 2. A Tabela A2.1 atualiza o exercício empírico de Lam & Marteleto que, originalmente, investigou a probabilidade de matrícula entre 1977 e 1999. Os resultados da tabela referem-se ao período de 1982 a 2009, sendo que o ano de 1981 foi excluído pois ele foi utilizado para iniciar a série de taxa de crescimento do tamanho da coorte. Em seu artigo original, os autores encontraram os valores de 0,68, 0-,42 e -0,74 para a taxa de crescimento de 7-14 anos respectivamente para os grupos etários 8-9 anos, 10-13 anos e 14-17 anos. Os resultados apresentados na Tabela A2.1 evidenciam, a priori, uma relação

negativa entre a taxa de crescimento da coorte 7-14 anos para o grupo etário de 8-9 anos e uma magnitude maior para o grupo etário de 10-13 anos.

TABELA A2.1 COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA POR GRUPO ETÁRIO

	Regressão 1 8 - 9 anos	**Regressão 2 10 - 13 anos**	**Regressão 3 14 - 17 anos**
Taxa Crescimento Coorte 7-14 anos	-0.877***	-0.864***	-0.634***
	(0.0119)	(0.00907)	(0.00935)
Crescimento x Educação Mãe	0.217***	0.195***	0.186***
	(0.00239)	(0.00200)	(0.00181)
Crescimento x Homem	0.0493***	-0.213***	-0.293***
	(0.0140)	(0.0107)	(0.0106)
Mulher	0.106***	0.0905***	0.203***
	(0.000508)	(0.000385)	(0.000377)
Urbana	0.208***	0.283***	0.350***
	(0.000528)	(0.000398)	(0.000411)
Nº Irmãos 0 a 6 anos	-0.111***	-0.110***	-0.0776***
	(0.000242)	(0.000189)	(0.000216)
Nº Irmãos 7 a 17 anos	-0.0224***	-0.00866***	-0.0221***
	(0.000162)	(0.000129)	(0.000129)
Educação Pai	0.0982***	0.0879***	0.0898***
	(0.000214)	(0.000168)	(0.000165)
Educação Mãe	0.163***	0.137***	0.106***
	(0.000216)	(0.000170)	(0.000167)
Educação Pai^2	-0.00407***	-0.00299***	-0.00179***
	(1.97e-05)	(1.63e-05)	(1.50e-05)
Educação Mãe^2	-0.00724***	-0.00592***	-0.00285***
	(2.09e-05)	(1.73e-05)	(1.59e-05)
Tendência temporal	0.0360***	0.0345***	0.0323***
	(3.63e-05)	(2.79e-05)	(2.57e-05)
Região Nordeste	-0.0340***	-0.000413	0.0752***
	(0.00122)	(0.00102)	(0.000970)
Região Sudeste	0.370***	0.0258***	-0.332***
	(0.00124)	(0.00102)	(0.000961)
Região Sul	0.385***	-0.0969***	-0.578***
	(0.00139)	(0.00110)	(0.00104)
Região Centro-Oeste	0.0776***	-0.0167***	-0.146***

	(0.00144)	(0.00120)	(0.00117)
Idade 9 anos	0.191***		
	(0.000482)		
Idade 11 anos		-0.0592***	
		(0.000521)	
Idade 12 anos		-0.201***	
		(0.000514)	
Idade 13 anos		-0.430***	
		(0.000509)	
Idade 15 anos			-0.261***
			(0.000491)
Idade 16 anos			-0.524***
			(0.000502)
Idade 17 anos			-0.786***
			(0.000530)
Constante	0.128***	0.551***	-0.0589***
	(0.00159)	(0.00129)	(0.00124)
Pseudo R^2	0.24	0.19	0.20
Nº Observações	69,162,087	106,585,024	62,186,316

Erro Padrão entre parênteses (*** p<0.01, ** p<0.05, * p<0.1)

OBS.: *Dummy* da Região Norte e *dummies* de 8, 10 e 14 anos omitidas

APÊNDICE 3: DEFASAGEM IDADE-SÉRIE

A taxa de defasagem escolar da série j na localidade s e ano t é definida por[68]:

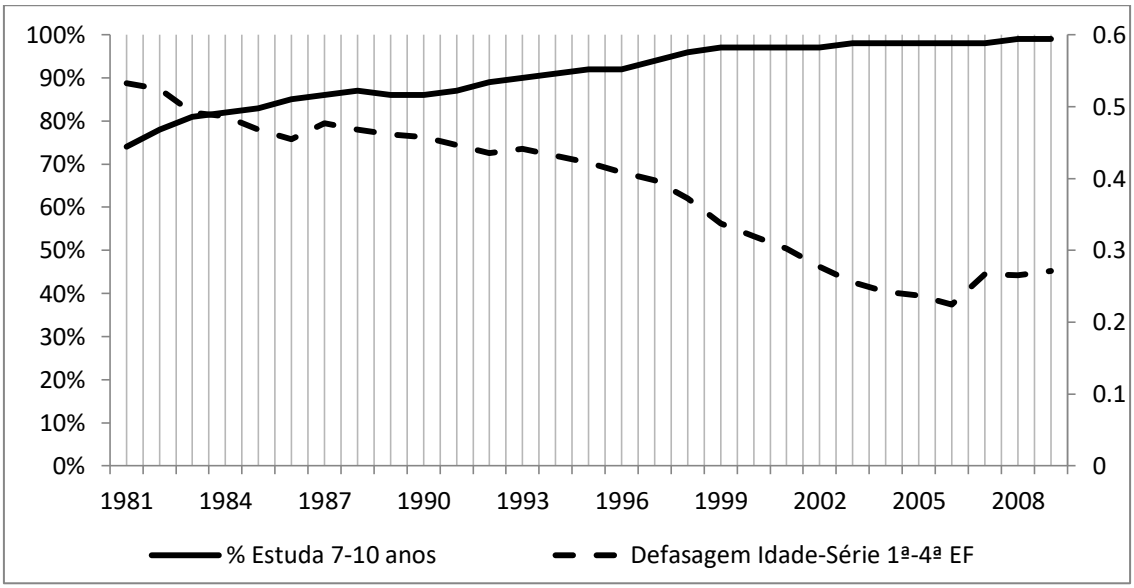
$$Defasagem\ Idade - Série_{j,s,t} = \frac{Matrícula\ total_{j,s,t} - Matrícula\ Idade\ Própria_{j,s,t}}{Matrícula\ total_{j,s,t}}$$

Para cada série do Ensino Fundamental, usou-se a "Idade Própria" com desvio-padrão de 1 ano. Assim, indivíduos de 6 a 8 anos foram utilizados para a primeira série do Ensino Fundamental, 7 a 9 anos para a segunda série e assim sucessivamente. O Gráfico A3.1 apresenta a evolução da taxa de matrícula para indivíduos de 7 a 10 anos, os potenciais demandantes da 1ª a 4ª séries do Ensino Fundamental, no eixo esquerdo, e a evolução do índice de defasagem escolar da 1ª a 4º séries para o Brasil entre 1981 e 2009, analisado no eixo da direita. O gráfico mostra uma correlação negativa entre as séries sendo que a quebra da tendência da defasagem escolar em 2006 deve-se a adoção gradual do Ensino Fundamental com a nona série. Em 1981, 50% dos alunos matriculados entre a 1ª e a 4ª séries estavam com idade defasada e este valor cai para 22% em 2006. O ano de 1997 apresenta uma quebra na tendência intensificando o processo de queda da defasagem escolar[69]. O Gráfico A3.2 apresenta o percentual de alunos de 7 a 10 anos defasados com 26% dos alunos de 7 a 10 anos defasados em 1981 e este percentual cai para 7% em 2006. O Gráfico A3.2 apresenta também uma quebra de tendência em 1997 intensificando o processo de queda da defasagem dos alunos de 7 a 10 anos. Os Gráficos A3.3 e A3.4 realizam a mesma análise para o grupo etário de 11 a 14 anos e para a 5ª a 8ª séries do Ensino Fundamental.

[68] Curi, Menezes-Filho (2007) apresentam um trabalho com dados da PNAD com o cálculo da defasagem idade-série.

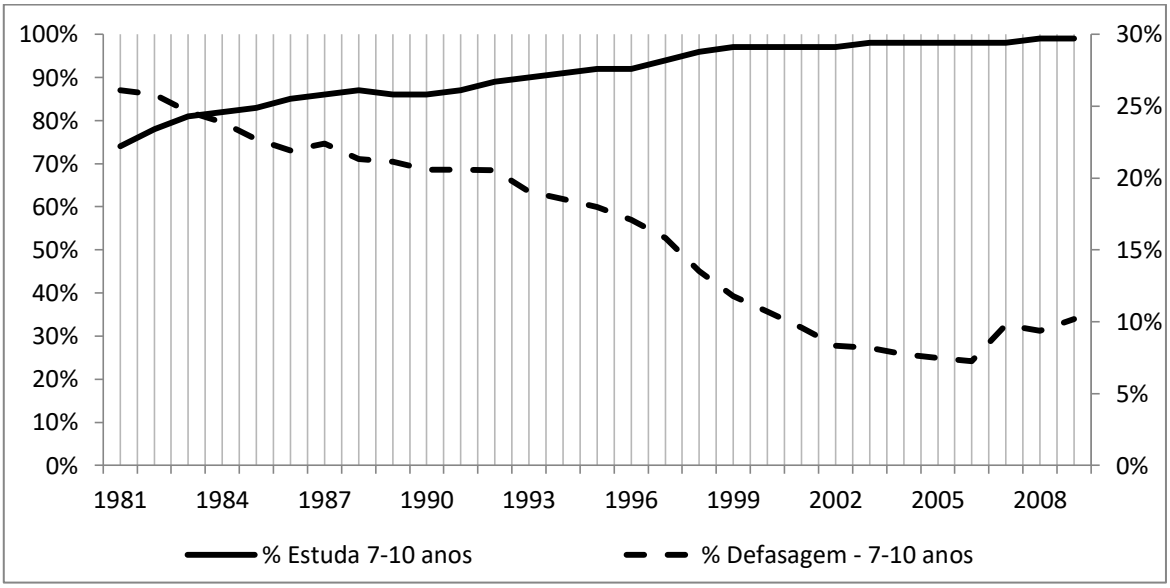
[69] Em 1996 foi redefinida a Lei de Diretrizes e Bases da Educação Nacional

GRÁFICO A3.1: PERCENTUAL DA POPULAÇÃO DE 7 A 10 ANOS MATRICULADA NO ENSINO REGULAR E SUA TAXA DE DEFASAGEM IDADE-SÉRIE



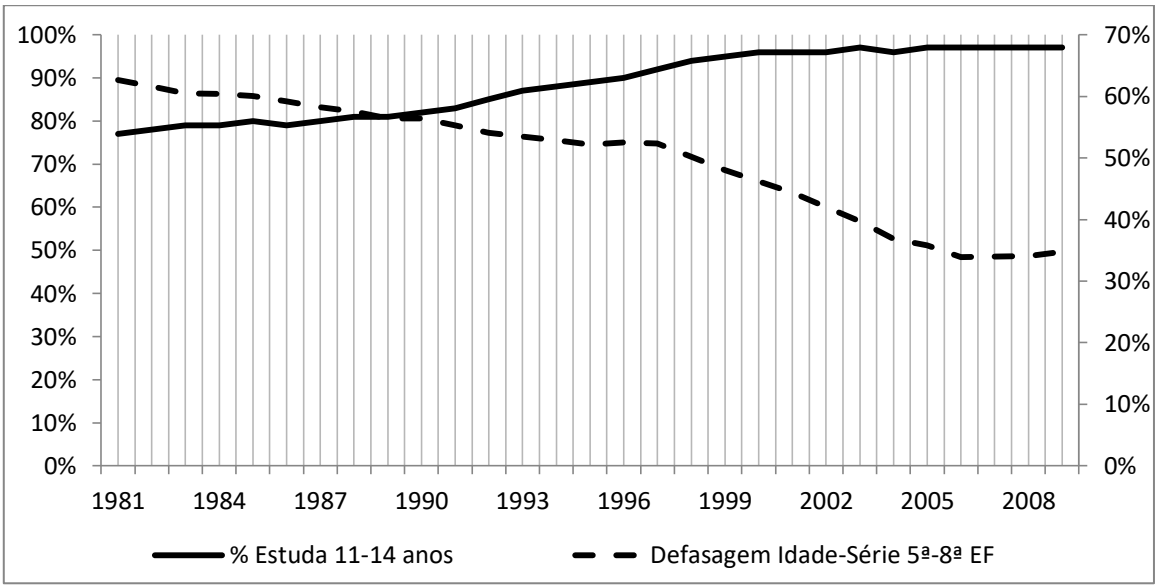
Fonte: Microdados da PNAD 1981 – 2009
OBS. % População 7–10 anos é analisada no eixo da esquerda. Defasagem idade-série é analisada no eixo da direita

GRÁFICO A3.2: PERCENTUAL DA POPULAÇÃO DE 7 A 10 ANOS MATRICULADA NO ENSINO REGULAR E SUA TAXA DE DEFASAGEM DO GRUPO ETÁRIO



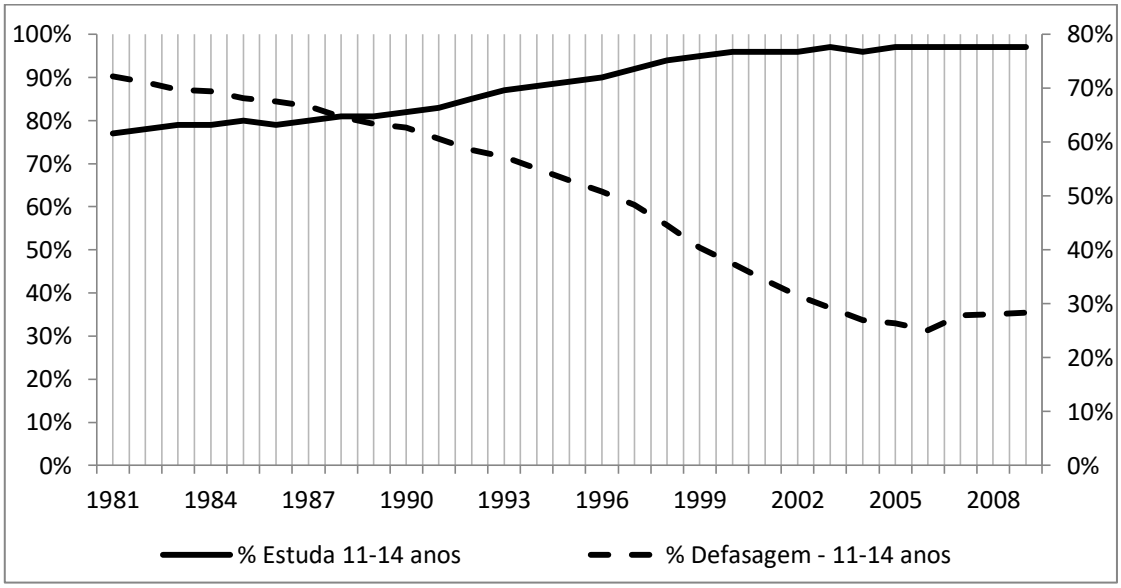
Fonte: Microdados da PNAD 1981 – 2009
OBS. % População 7–10 anos é analisada no eixo da esquerda. Defasagem é analisada no eixo da direita

GRÁFICO A3.3: PERCENTUAL DA POPULAÇÃO DE 11 A 14 ANOS MATRICULADA NO ENSINO REGULAR E A TAXA DE DEFASAGEM IDADE-SÉRIE



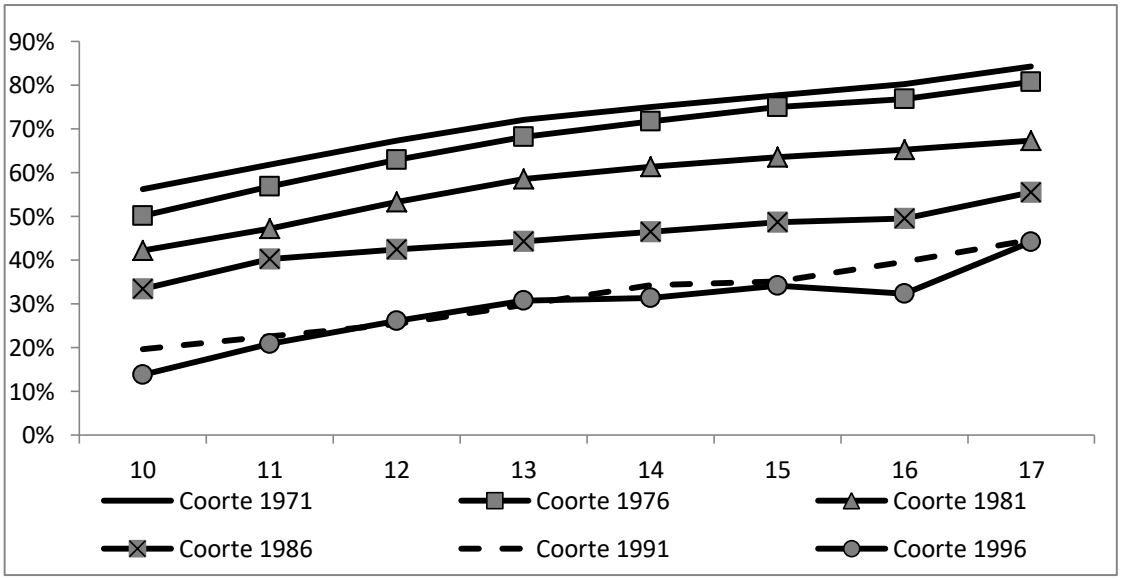
Fonte: Microdados da PNAD 1981 – 2009
OBS. % População 11–14 anos é analisada no eixo da esquerda. Defasagem idade-série é analisada no eixo da direita

GRÁFICO A3.4: PERCENTUAL DA POPULAÇÃO DE 11 A 14 ANOS MATRICULADA NO ENSINO REGULAR E SUA TAXA DE DEFASAGEM DO GRUPO ETÁRIO



Fonte: Microdados da PNAD 1981 – 2009
OBS. % População 11–14 anos é analisada no eixo da esquerda. Defasagem é analisada no eixo da direita

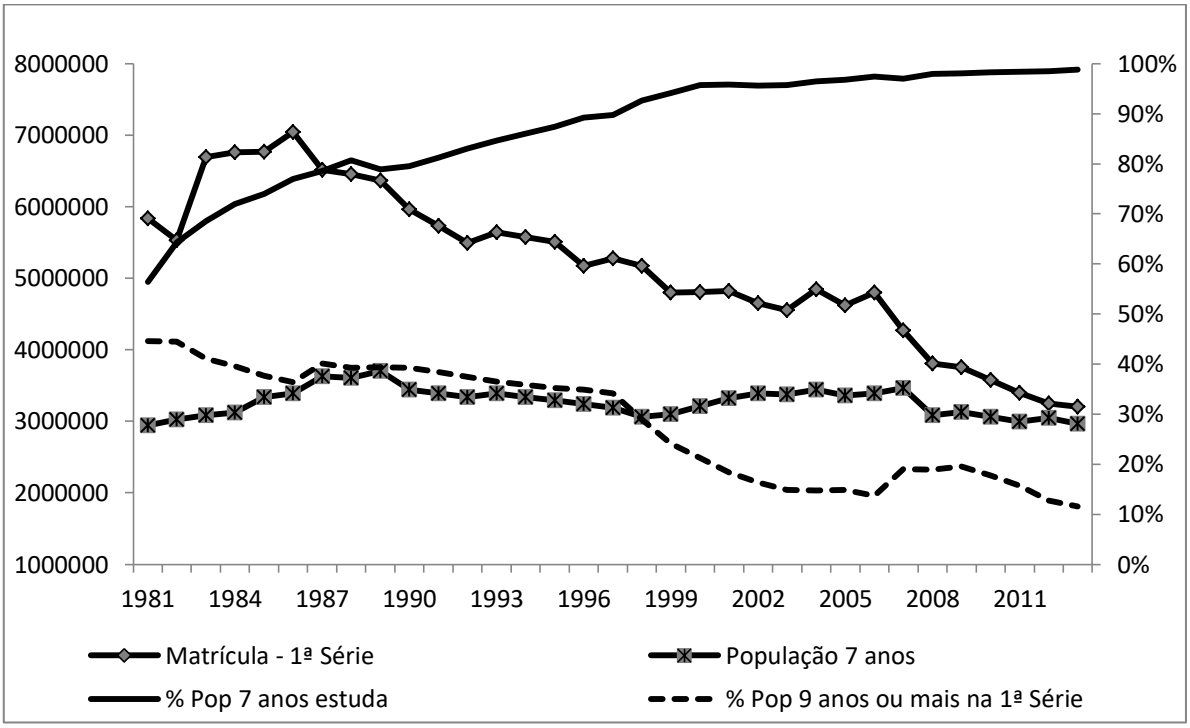
GRÁFICO A3.5: DEFASAGEM ESCOLAR POR COORTE – CICLO ESCOLAR



Fonte: Microdados da PNAD 1981 – 2009
OBS. % Matriculados com defasagem escolar por coorte de nascimento entre 10 e 17 anos

O Gráfico A3.5 apresenta a defasagem por idade para diferentes coortes. Com 10 anos de idade, a coorte nascida em 1971 tinha uma 56% dos seus alunos matriculados em séries defasadas do ensino regular, a coorte nascida em 1986 tinha 33% dos seus alunos matriculados em séries defasadas enquanto que a coorte nascida em 1996 tinha 14% dos seus alunos defasados quando estes tinham 10 anos. Com 17 anos de idade, a coorte nascida em 1971 tinha 84% de seus alunos defasados enquanto que para a mesma idade, a coorte nascida em 1996 tinha 44% de seus alunos defasados. Esta profunda mudança na defasagem escolar produz dois efeitos para o desempenho educacional brasileiro. O primeiro relaciona-se com a demanda por educação e uma melhor acomodação dos alunos em sua série potencial ajuda no processo de universalização do acesso a educação. O segundo relaciona-se com a progressão escolar, fazendo o estudante acumular mais capital humano pois uma menor defasagem idade-série possibilita menor evasão escolar.

GRÁFICO A3.6: MATRÍCULA, EM VALORES ABSOLUTAS, PARA A 1ª SÉRIE ENSINO FUNDAMENTAL



Fonte: Microdados da PNAD 1981 – 2009
OBS. *Matrícula 1ª Série* e *População 7 anos* são analisados no eixo da esquerda. *% Pop 7 anos estuda* e *% Pop 9 anos ou mais na 1ª Série* são analisados no eixo da direita.

O Gráfico A3.6 apresenta o número absoluto de matriculados na 1ª série do Ensino Fundamental, o tamanho absoluto da população com 7 anos que é a potencial demandante por vagas na 1ª série, o percentual da população com 7 anos que estuda e o percentual de alunos com nove anos ou mais matriculados na 1ª série. Entre 1981 e 2013, o número absoluto de matriculados na primeira série apresentou tendencia negativa, principalmente a partir de 1987 e, ao longo da série, o número de vagas na primeira série excedeu o valor absoluto da população de 7 anos, com tendência de convergência para estas duas séries. O aumento do percentual de indivíduos com 7 anos que estudam é acompanhado pela queda da defasagem na primeira série. Em 1981, 56% dos indivíduos com 7 anos estudavam e este percentual aumenta para 81% em 1991, 96% em 2001 e 98% em 2011. Em 1981, 45% dos indivíduos matriculados na primeira série tinham 9 anos ou mais e estes indivíduos potencialmente deveriam estar matriculados em séries mais avançadas. Este percentual diminui para 14% em 2006.

APÊNDICE 4: RESULTADOS DOS EFEITOS MARGINAIS

TABELA A4.1: EFEITOS MARGINAIS DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 4ª SÉRIES DO ENSINO FUNDAMENTAL

VARIÁVEIS	(1) 1ª-4º series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie	(5) 4ª serie
Coorte Relativa	-0.7284	-2.5297	0.2284	0.0381	0.4820
	(0.0065)	(0.0317)	(0.0228)	(0.0263)	(0.0337)
Mulher	0.0230	0.0183	0.0109	0.0066	0.0010
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Urbana	0.0361	0.0336	0.0034	0.0077	0.0366
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Irmãos 0-6 anos	-0.0204	-0.0205	-0.0089	-0.0097	-0.0096
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Irmãos 7-17 anos	-0.0006	0.0013	0.0006	-0.0011	-0.0003
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Pai	0.0187	0.0231	0.0056	0.0053	0.0074
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe	0.0281	0.0381	0.0064	0.0079	0.0109
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Pai ^{2}	-0.0007	-0.0008	-0.0001	0.0000	-0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe ^{2}	-0.0011	-0.0014	-0.0001	-0.0002	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Erro Padrão entre parênteses

TABELA A4.2: EFEITOS MARGINAIS DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 5ª A 8ª SÉRIES DO ENSINO FUNDAMENTAL

VARIÁVEIS	(1) 5ª-8º series	(2) 5ª serie	(3) 6ª serie	(4) 7ª serie	(5) 8ª serie
Coorte Relativa	-0.1249	-0.9845	2.7689	5.5095	1.1335
	(0.0107)	(0.0414)	(0.0559)	(0.0616)	(0.0716)
Mulher	0.0106	-0.0112	0.0188	0.0153	0.0144
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Urbana	0.1008	0.1345	0.0479	0.0143	-0.0033
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)
Irmãos 0-6 anos	-0.0184	-0.0137	-0.0164	-0.0109	-0.0034
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Irmãos 7-17 anos	-0.0076	-0.0051	-0.0026	-0.0027	-0.0041
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Educação Pai	0.0113	0.0093	0.0077	0.0037	0.0014
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Educação Mãe	0.0153	0.0131	0.0089	0.0071	0.0054
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Educação Pai ^{2}	0.0000	0.0001	0.0001	0.0003	0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe ^{2}	-0.0004	-0.0003	0.0000	0.0001	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Erro Padrão entre parênteses

TABELA A4.3: EFEITOS MARGINAIS DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 3ª SÉRIES DO ENSINO MÉDIO

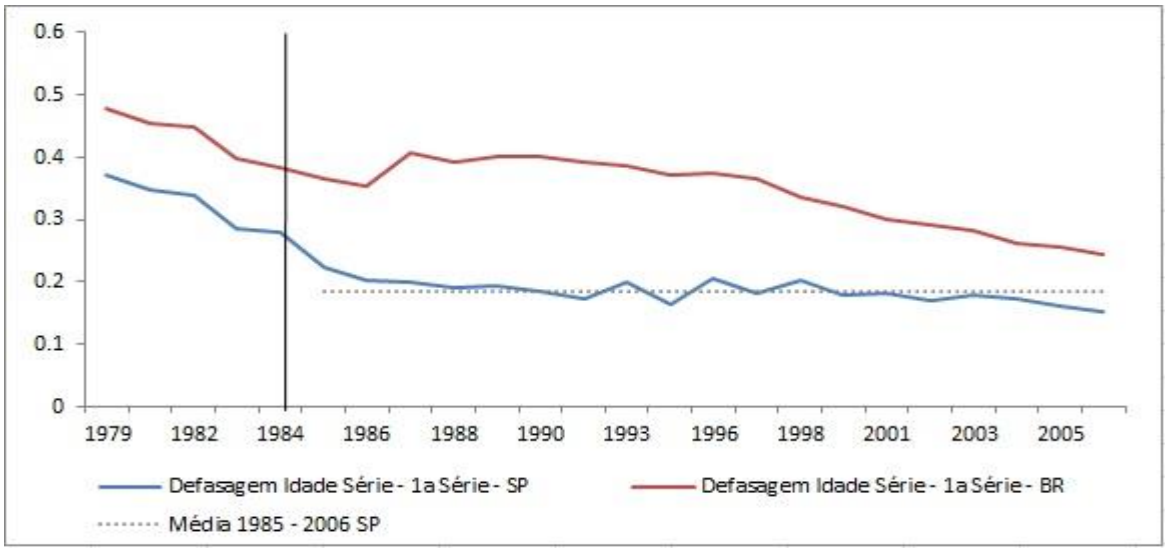
VARIÁVEIS	(1) 1ª-3º series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie
Coorte Relativa	-1.5317	-3.0949	3.4175	-0.7952
	(0.0335)	(0.1031)	(0.1164)	(0.1616)
Mulher	0.0334	0.0273	0.0226	0.0242
	(0.0002)	(0.0003)	(0.0003)	(0.0004)
Urbana	0.0305	0.0460	0.0021	-0.0111
	(0.0003)	(0.0004)	(0.0004)	(0.0006)
Irmãos 0-6 anos	-0.0087	-0.0100	-0.0017	-0.0085
	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Irmãos 7-17 anos	-0.0132	-0.0127	-0.0024	-0.0111
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Educação Pai	0.0093	0.0112	0.0030	0.0085
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Educação Mãe	0.0114	0.0108	0.0012	0.0112
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Educação Pai ^{2}	-0.0001	-0.0002	0.0002	-0.0003
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe ^{2}	-0.0001	0.0002	0.0001	-0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Erro Padrão entre parênteses

APÊNDICE 5: PROGRESSÃO CONTINUADA NO ESTADO DE SÃO PAULO

As altas taxas de defasagem idade-série sempre constituiram um problema no ensino básico brasileiro e a dificuldade de se capturar o efeito do tamanho (relativo ou absoluto) da coorte na probabilidade de matrícula deve-se, em parte, por movimentos demográficos e movimentos de ajuste idade-série caminharem *pari-passu* ao longo da década de 1990. Para isto serão apresentadas nesta seção os resultados da regressão por série, seguindo a especificação da seção 4 e dos resultados apresentados na Tabela 3, mas separando a amostra com o Brasil exclusive São Paulo e uma amostra apenas com o estado de São Paulo. Viegas & Souza (2006) documentam que o Estado de São Paulo implementou no ano de 1984 a progressão continuada da 1ª à 2ª série do ensino fundamental e este processo pode ser visualizado no Gráfico A5.1 onde até 1984, a defasagem na 1ª Série em São Paulo seguia a tendência da defasagem no Brasil (excluído São Paulo). Após a implementação do Ciclo com progressão da 1ª à 2ª série, a defasagem em São Paulo manteve-se, na média, invariante no tempo com média próxima de 20%, enquanto o Brasil apresentou uma tendência de ajuste após 1996.

GRÁFICO A5.1: DEFASAGEM NA 1ª SÉRIE DO ENSINO FUNDAMENTAL PARA O BRASIL EXCLUSIVE SÃO PAULO E PARA APENAS SÃO PAULO



Fonte: Microdados da PNAD 1977 – 2009

OBS. *Em 1984 é introduzida a progressão continuada para escolas públicas estaduais no Estado de São Paulo apenas da 1ª à 2ª séries do Ensino Fundamental.*

TABELA A5.1: COEFICIENTES DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 4ª SÉRIES DO ENSINO FUNDAMENTAL

	(1)		(2)		(3)		(4)		(5)	
VARIÁVEIS	1-4 series BR menos SP	1-4 series apenas SP	1 serie BR-menos SP	1 serie apenas SP	2 serie BR-menos SP	2 serie apenas SP	3 serie BR-menos SP	3 serie apenas SP	4 serie BR-menos SP	4 serie apenas SP
Coorte Reltiva	-3.370***	-37.21***	-16.48***	-100.6***	7.310***	-21.53***	34.46***	-161.4***	60.83***	-231.2***
	(0.0127)	(0.219)	(0.0611)	(0.936)	(0.120)	(3.042)	(0.133)	(4.064)	(0.144)	(3.898)
Mulher	0.133***	-0.0216***	0.0751***	-0.0362***	0.130***	-0.0259***	0.0803***	-0.0459***	0.0438***	-0.122***
	(0.000250)	(0.000720)	(0.000328)	(0.00101)	(0.000700)	(0.00207)	(0.000744)	(0.00202)	(0.000780)	(0.00195)
Urbana	0.164***	0.275***	0.110***	0.105***	0.0161***	0.191***	0.0417***	0.373***	0.240***	0.679***
	(0.000265)	(0.000917)	(0.000355)	(0.00135)	(0.000731)	(0.00259)	(0.000779)	(0.00239)	(0.000807)	(0.00224)
Irmãos 0-6 anos	-0.0970***	-0.176***	-0.0662***	-0.154***	-0.0836***	-0.197***	-0.0878***	-0.0973***	-0.0635***	-0.156***
	(0.000124)	(0.000431)	(0.000158)	(0.000605)	(0.000347)	(0.00115)	(0.000388)	(0.00120)	(0.000433)	(0.00122)
Irmãos 7-17 anos	0.00176***	-0.0482***	0.00934***	-0.0478***	0.00459***	-0.0106***	0.00201***	-0.0739***	-0.00260***	-0.00884***
	(8.31e-05)	(0.000268)	(0.000106)	(0.000374)	(0.000239)	(0.000745)	(0.000261)	(0.000744)	(0.000275)	(0.000725)
Educação Pai	0.0978***	0.0789***	0.0829***	0.0651***	0.0608***	0.0627***	0.0475***	-0.00134	0.0365***	0.0902***
	(0.000117)	(0.000311)	(0.000158)	(0.000436)	(0.000329)	(0.000967)	(0.000345)	(0.00102)	(0.000359)	(0.000898)
Educação Mãe	0.150***	0.0912***	0.143***	0.0610***	0.0661***	0.0720***	0.0650***	0.0942***	0.0837***	0.0691***
	(0.000117)	(0.000320)	(0.000157)	(0.000451)	(0.000341)	(0.000941)	(0.000352)	(0.00112)	(0.000358)	(0.000917)
Educação Pai 2	-0.00374***	-0.00275***	-0.00294***	-0.00234***	-0.00193***	0.000556***	-0.000564***	0.00453***	0.000781***	-0.00189***
	(1.16e-05)	(2.58e-05)	(1.56e-05)	(3.44e-05)	(3.46e-05)	(0.000101)	(3.65e-05)	(0.000118)	(3.85e-05)	(9.08e-05)
Educação Mãe 2	-0.00576***	-0.00241***	-0.00544***	4.81e-05	-4.72e-05	-0.00338***	-0.000958***	0.00173***	-0.00299***	-0.00288***
	(1.24e-05)	(2.94e-05)	(1.65e-05)	(3.93e-05)	(3.94e-05)	(9.68e-05)	(3.94e-05)	(0.000155)	(3.88e-05)	(9.76e-05)
Constante	0.125***	3.420***	0.134***	2.515***	0.973***	1.923***	0.270***	4.571***	-2.168***	3.402***
	(0.00147)	(0.0182)	(0.00184)	(0.0192)	(0.00385)	(0.0612)	(0.00668)	(0.0878)	(0.00425)	(0.0811)
Nº Observações	174,543,565	33,356,759	76,054,071	10,361,005	40,181,993	7,871,661	32,443,558	7,482,711	25,863,943	6,052,777
Controle Ano	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM
Controle UF	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM	SIM

Erro Padrão entre parênteses
*** p<0.01, ** p<0.05, * p<0.1

TABELA A5.2: EFEITOS MARGINAIS DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 4ª SÉRIES DO ENSINO FUNDAMENTAL – BRASIL EXCLUSVE SÃO PAULO

VARIÁVEIS	(1) 1ª-4º series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie	(5) 4ª serie
Coorte Relativa	-0.7125	-4.8750	0.7817	4.1522	8.6764
	(0.0027)	(0.0180)	(0.0129)	(0.0160)	(0.0205)
Mulher	0.0282	0.0222	0.0139	0.0097	0.0062
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Urbana	0.0346	0.0324	0.0017	0.0050	0.0342
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Irmãos 0-6 anos	-0.0205	-0.0196	-0.0089	-0.0106	-0.0091
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Irmãos 7-17 anos	0.0004	0.0028	0.0005	0.0002	0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Pai	0.0207	0.0245	0.0065	0.0057	0.0052
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Educação Mãe	0.0316	0.0424	0.0071	0.0078	0.0119
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Educação Pai ^{2}	-0.0008	-0.0009	-0.0002	-0.0001	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe ^{2}	-0.0012	-0.0016	0.0000	-0.0001	-0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Erro Padrão entre parênteses

TABELA A5.3: EFEITOS MARGINAIS DAS REGRESSÕES PARA PROBABILIDADE DE MATRÍCULA DA 1ª A 4ª SÉRIES DO ENSINO FUNDAMENTAL – APENAS SÃO PAULO

VARIÁVEIS	(1) 1ª-4º series	(2) 1ª serie	(3) 2ª serie	(4) 3ª serie	(5) 4ª serie
Coorte Relativa	-4.4986	-21.0872	-1.2523	-11.0460	-21.9960
	(0.0264)	(0.1961)	(0.1770)	(0.2780)	(0.3706)
Mulher	-0.0026	-0.0076	-0.0015	-0.0031	-0.0116
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)
Urbana	0.0332	0.0220	0.0111	0.0255	0.0646
	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Irmãos 0-6 anos	-0.0213	-0.0323	-0.0115	-0.0067	-0.0149
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Irmãos 7-17 anos	-0.0058	-0.0100	-0.0006	-0.0051	-0.0008
	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Educação Pai	0.0095	0.0136	0.0036	-0.0001	0.0086
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Educação Mãe	0.0110	0.0128	0.0042	0.0064	0.0066
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Educação Pai ^{2}	-0.0003	-0.0005	0.0000	0.0003	-0.0002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Educação Mãe ^{2}	-0.0003	0.0000	-0.0002	0.0001	-0.0003
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Erro Padrão entre parênteses

As regressões comparando a amostra com apenas São Paulo e a amostra com o Brasil exclusive São Paulo sugerem um efeito negativo para todas as quatro primeiras séries do Ensino Fundamental, sinal não encontrado para a amostra do Brasil. Além disto, a Tabela A5.3 mostra que o efeito marginal na variação do tamanho relativo da coorte é muio superior em magnitude, fazendo com que o tamanho da coorte tenha favorecido à universalização do acesso ao ensino básico em São Paulo, estado com baixa defasagem idade-série desde a década de 1980.

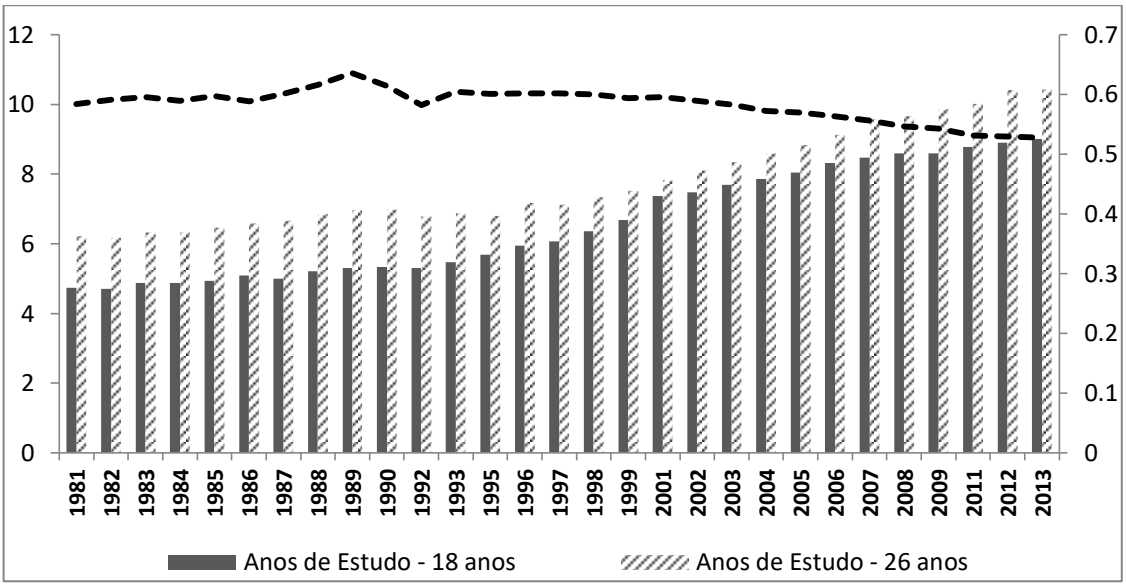
A ESCOLARIDADE DAS COORTES E A EVOLUÇÃO SALARIAL BRASILEIRA ENTRE 1981 E 2013

1. INTRODUÇÃO

O mercado de trabalho brasileiro sofre profundas transformações principalmente na mudança da composição de sua mão-de-obra por grau de escolaridade. Na média, 83% dos trabalhadores de 18 a 60 anos em 1981 tinham até 8 anos de estudo, o que equivale a ter frequentado no máximo o Ensino Fundamental, com ou sem conclusão. Para este mesmo ano, 11% dos trabalhadores tinham de 9 a 11 anos de estudo, equivalente a ter frequentado o Ensino Médio e 6% destes trabalhadores tinham de 12 a 17 anos de estudo, equivalente a ter frequentado, com conclusão ou não, o Ensino Superior. Em 2013, 48% dos trabalhadores tinham até 8 anos de estudo, 38% dos trabalhadores tinham de 9 a 11 anos de estudo e 16% dos trabalhadores tinham de 12 a 17 anos de estudo. Estas mudanças estão associadas, primeiro, às mudanças educacionais ocorridas no Brasil principalmente nas décadas de 1980 e 1990, com a queda da defasagem idade-série, a expansão do ensino básico, universalização do acesso à escola e, principalmente, ao avanço nas taxas de matrículas do ensino médio[70]. O Gráfico 3.1 apresenta a evolução dos anos médios de estudo para indivíduos com 18 anos e indivíduos com 26 anos entre 1981 e 2013. Na média, indivíduos com 18 anos tinham 4,7 anos de estudo e indivíduos com 26 anos tinham 6,2 anos de estudo em 1981. Potencialmente, um indivíduo com 18 anos teria seu Ensino Médio concluído e apresentaria 11 anos de estudo, implicando em uma defasagem média de 6,3 anos de estudo. Para o ano de 2013, na média, um indivíduo com 18 anos tinha 9 anos de estudo, tendo diminuído sua defasagem média para 2 anos e este processo de melhora nos anos médios de estudo foi intensificado a partir da segunda metade da década de 1990, como pode-se ver no gráfico. Ainda para o ano de 2013, indivíduos com 26 anos tinham, na média, 10,4 anos de estudo.

[70] No Capítulo 2, *Tamanho da coorte, tamanho da família e matrícula escolar: evidências do Brasil*, as estatísticas educacionais brasileiras para o período 1981-2009 mostraram que, à década de 1980, a defasagem idade-série concentrava indivíduos com idade superior nas primeiras séries do Ensino Fundamental. Reacomodar indivíduos em sua série potencial dada sua idade possibilitou organizar a demanda por ensino e a sua continuidade para que houvesse a expansão do ensino médio no Brasil.

GRÁFICO 3.1: COEFICIENTE DE GINI E ANOS MÉDIOS DE ESTUDO PARA INDIVÍDUOS COM 18 ANOS E PARA INDIVÍDUOS COM 26 ANOS ENTRE 1981 E 2013



Fonte: Microdados PNAD 1981-2013; **Coef. Gini:** PNAD/IPEA

OBS: Coeficiene de Gini é representado no eixo da esquerda e seu valor varia de 0, quando não há desigualdade (indivíduos teriam a mesma renda) e 1 quando o grau de desigualdade é máximo (toda a renda estaria concentrada em apenas um indivíduo e os outros não teriam renda)

O processo de aumento da escolaridade média dos indivíduos ao longo dos anos produziu mudanças na composição da oferta de trabalho que era menos qualificada, na média, para a década de 1980 e passou a ter um aumento da oferta de indivíduos com qualificação intermediária (Ensino Médio), a partir da segunda metade da década de 1990, e indivíduos com alta qualificação (Ensino Superior), principalmente a partir da segunda metade da década de 2000. Este processo de mudança da composição por educação mudou a oferta relativa da economia brasileira e gerou efeitos para os salários relativos por conta de um movimento geracional, onde os indivíduos mais velhos e com baixa qualificação, ao se aposentarem, têm a sua reposição no mercado de trabalho por coortes entrantes com melhores indicadores educacionais. O aumento da oferta de trabalhadores qualificados e intermediários, impulsionado pelas coortes mais novas, e a queda da oferta de trabalhadores pouco qualificados diminuem o salário relativo ao trabalhador pouco qualificado, que frequentou no máximo o Ensino Fundamental. Este processo está

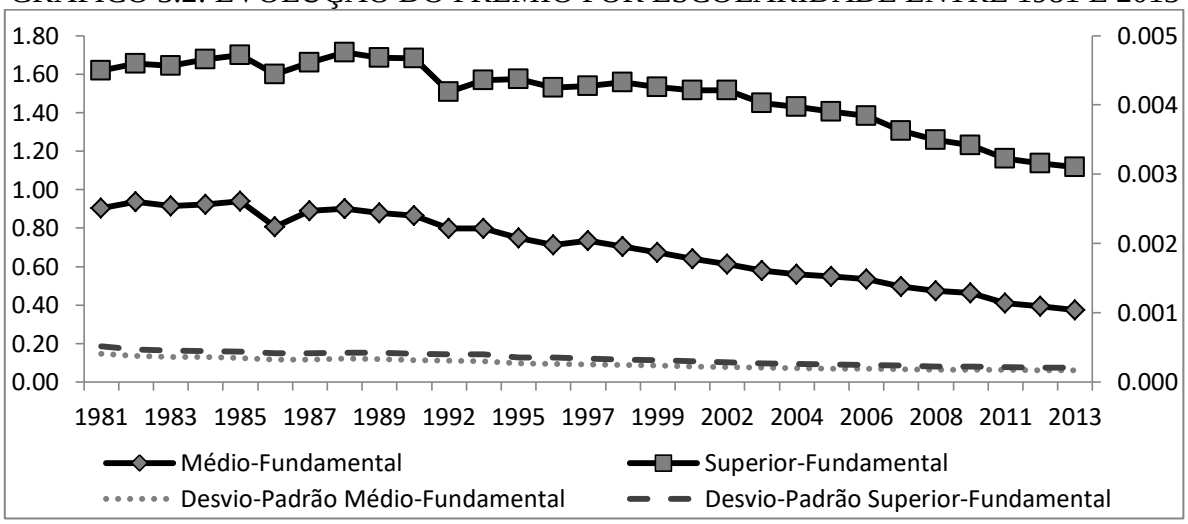
correlacionado com a queda da desigualdade representada pelo coeficiente de Gini, apresentado no Gráfico 3.1[71].

O prêmio por escolaridade, representado pelo percentual de ganho salarial por uma determinada qualificação, é apresentado no Gráfico 3.2 como resultado de 33 regressões, uma para cada ano ao longo das séries, onde o logaritmo do salário hora é regredido contra dummies de qualificação (Ensino Médio e Ensino Superior), dummies de grupos etários (18-25 anos, 26-40 anos e 41-60 anos), *dummy* de sexo e controles de localidades representados por Unidades da Federação[72]. Segundo os dados, em 1981, um indivíduo com Ensino Médio, completo ou incompleto, ganharia 90% a mais que um indivíduo com o Ensino Fundamental, completo ou incompleto, controlando para o grupo etário. Em relação ao Ensino Superior, o prêmio é de 165% a mais no salário em relação a indivíduos com Ensino Fundamental, controlando para o grupo etário. Estes prêmios diminuem para 37% a mais no salário para o Ensino Médio e 112% para o Ensino Superior, ambos em relação ao Ensino Fundamental, em 2013. A queda maior no prêmio do Ensino Médio é explicada pelo aumento relativo de indivíduos com 9 a 11 anos de estudo ao longo destes anos e este evento apresenta correlação com a queda da desigualdade sintetizada pelo coeficiente de Gini.

[71] O coeficiente de Gini é representado por um intervalo entre 0 e 1 onde 0 seria uma situação sem desigualdade (todos os indivíduos teriam a mesma renda) e 1 uma situação com desigualdade máxima (apenas um indivíduo concentraria toda a renda). No Brasil, na média, o coeficiente de Gini era 0,60 na década de 1980, 0,60 na década de 1990 e 0,56 entre 2001 e 2013. Estas mudanças com queda da desigualdade na renda a partir da década de 2000 estão correlacionadas com as mudanças nos salários relativos. Ao se regredir a série do coeficiente de Gini contra o logaritmo do salário relativo Ensino Médio/ Ensino Fundamental, o logaritmo do salário relativo Ensino Superior/Ensino Fundamental e uma tendência temporal, o modelo ajusta-se com um R^2ajustado de 0,99 e uma queda de 1% no salário relativo Ensino Médio/ Ensino Fundamental explica uma queda de 0,18 no coeficiente de Gini. Uma queda de 1% no salário relativo Ensino Superior/ Ensino Fundamental explica uma queda de 0,08 no coeficiente.

[72] O mesmo exercício de 33 regressões, uma para cada ano ao longo da série, foi feito para outras cinco especificações: 1) apenas homens e *dummies* de qualificação, 2) Apenas mulheres e *dummies* de qualificação, 3) Homens, mulheres, *dummy* de sexo e *dummies* de qualificação, 4) Especificação (3) adicionado controle de localidade (UF) e 5) Adicionando controles para tipo de emprego (conta própria, empregado, empregador etc.). Todos os resultados mostraram-se robustos e similares entre si e similares às tendências apresentadas pelo prêmio de qualificação calculado na Seção 3 que utiliza a metodologia de Acemoglu & Autor (2012).

GRÁFICO 3.2: EVOLUÇÃO DO PRÊMIO POR ESCOLARIDADE ENTRE 1981 E 2013



Fonte: Microdados PNAD 1981-2013 para indivíduos de 18 a 60 anos que trabalhavam 20h ou mais por semana

OBS 1: Para cada ano, o prêmio por escolaridade é calculado por uma regressão do logaritmo do salário contra uma *dummy* para ensino médio, uma *dummy* para ensino superior, *dummies* de idade (18-25 anos, 26-40 anos e 41-60 anos), *dummy* de sexo e controla-se por localidade no nível da Unidade da Federação.

OBS 2: Os resultados das dummies de educação são relativos ao Ensino Fundamental. Considerou-se Ensino Fundamental indivíduos de 0 a 8 anos de estudo, Ensino Médio de 9 a 11 anos de estudo e Ensino Superior indivíduos de 12 a 17 anos de estudo.

OBS 3: Prêmio por escolaridade é o resultado relativo das dummies e são apresentados no eixo da esquerda. O desvio-padrão de cada parâmetro é apresentado no eixo da direita e representado pelas linhas tracejadas e pontilhadas.

Este artigo tem por finalidade investigar como as mudanças nas ofertas relativas, introduzidas pelo aumento da escolaridade das novas coortes entrantes no mercado de trabalho, conduziram a uma mudança nos salários relativos e como estes eventos estão correlacionados com a queda da desigualdade recente no Brasil. Um modelo de oferta relativa e salário relativo, seguindo a tradição de Katz & Marphy (1992) e Acemoglu & Autor (2012) é utilizado para se entender às transformações ocorridas no mercado de trabalho. Em termos agregados, o modelo demonstra que as mudanças na oferta relativa, com um aumento do número de trabalhadores com Ensino Médio em relação ao Ensino Fundamental e aumento do número de trabalhadores com Ensino Superior em relação ao Ensino Fundamental, conduziram a uma queda em seus respectivos salários relativos. No entanto, ao desagregar por grupos etários, os ajustes entre oferta relativa e salário relativo são menos claros e um modelo com substituição entre grupos etários para um mesmo grupo de qualificação é utilizado, seguindo a tradição de Card & Lemieux (2001).

2. BASE DE DADOS E ESTATÍSTICAS

O conjunto de dados utilizados na investigação empírica foi obtido das Pesquisas Nacionais por Amostra de Domicílios (PNAD) de 1981 a 2013. Os anos de 1991, 1994, 2000 e 2010 não disponibilizam dados por serem anos censitários (1991, 2000 e 2010) ou pela não realização da pesquisa anual (1994). As séries de mercado de trabalho foram agregadas a partir de um conjunto de microdados respeitando os critérios de compatibilização para que o conjunto de dados ficasse harmônico ao longo dos anos. A necessidade de compatibilização da série deve-se tanto por mudanças metodológicas, como na investigação da ocupação semanal que mudou da PNAD de 1990 para a PNAD de 1992, assim como as mudanças no ciclo educacional que foram absorvidas pela pesquisa domiciliar. Para a construção da base de dados foram coletadas informações sobre mercado de trabalho e educação para cada ano.

Foram considerados ocupados os indivíduos que declararam ter trabalhado na semana de referência ou possuírem trabalho, porém estavam afastados por algum motivo específico. A partir da PNAD de 1992 a investigação da ocupação da semana de referência também considerou, para definição de ocupados, os trabalhos de cultivo e construção no domicílio. Estas declarações não foram consideradas para que a série fosse compatibilizada e as informações a partir de 1992 fossem metodologicamente compatíveis com as informações da década de 1980. Para cada indivíduo foi considerado apenas as declarações da ocupação principal e suas respectivas renda do trabalho principal e número de horas semanais[73]. A renda do trabalho foi deflacionada utilizando o Índice de Preços ao Consumidor IPCA/INPC divulgado pelo IBGE sendo setembro de 2012 o período de referência.

A componente educação é medida em anos completos de estudo e esta variável é construída com a declaração da última série concluída com aprovação. A PNAD disponibiliza uma variável de anos de estudo, mas os intervalos dos valores atribuídos a

[73] Como exercício de robustez, as estatísticas descritivas e as regressões foram feitas também com a declaração de todos os trabalhados, renda de todos os trabalhos e número de horas semanais em todos os trabalhos. Os resultados foram praticamente idênticos.

esta variável diferem principalmente a partir de 1992, quando houve uma mudança metodológica na pesquisa. Para a construção de uma variável "anos de estudo" que fosse compatível e comparável ao longo dos anos, utilizou-se a última série concluída com aprovação tanto em cursos seriados regulares quanto em supletivos. Os ciclos escolares Elementar, Médio Primeiro Ciclo, Médio Segundo Ciclo, 1º Grau e 2º Grau foram ajustados ao seu equivalente do Ensino Fundamental e Ensino Médio. A mudança do ciclo básico de 8 séries para 9 anos do Ensino Fundamental, realizada a partir de 2007, também foi compatibilizada ao seu equivalente do Ensino Fundamental padrão com 8 séries . Para maiores detalhes na confecção de dados educacionais a partir dos microdados da PNAD, RIANI E GOLGHER (2004) e RIGOTTI (2004) são referências importantes na literatura nacional.

A construção da série agregada de dados considerou 3 grupos de escolaridade (pouco qualificado, médio qualificado e muito qualificado) e três grupos etários (18 a 25 anos, 26 a 40 anos e 41 a 60 anos). Definiu-se pouco qualificado o indivíduo que tinha de 0 a 8 anos de estudo e, portanto, frequentou no máximo o ensino fundamental do ciclo básico, médio qualificado o indivíduo que tinha de 9 a 11 anos de estudo, equivalente a ter frequentado o ensino médio e, por fim, muito qualificado o indivíduo com 12 anos ou mais de estudo, equivalente a ter frequentado o ensino superior[74]. A oferta considerou o total de horas do trabalho principal para cada indivíduo pertencente ao seu grupo específico de escolaridade e etário. Para as estatísticas descritivas, o salário foi ajustado para uma jornada de 44 horas semanais e o salário relativo foi calculado com base na média do salário ajustado para cada ano, grupo etário e grupo educacional. Para as regressões, considerou-se apenas a média do salário-hora para cada ano, grupo etário e grupo educacional. Em ambas estatísticas de oferta de trabalho e salário médio, a amostra selecionada considerou apenas trabalhadores com uma jornada no trabalho principal de 20 horas semanais ou mais, representando 94% da amostra de trabalhadores. Além disto, consideraram-se apenas trabalhadores entre 18 anos e 60 anos, representando 87% do total de trabalhadores. Para todas as séries de dados,

[74] Os dados de mercado de trabalho mostram que o Brasil fez uma transição com queda do percentual de trabalhadores com o ensino fundamental incompleto e teve, como contrapartida, um aumento do número de trabalhadores com ensino médio completo e ensino superior completo. Dentre estes últimos, ensino médio completo passou a ser o grupo com maior trajetória de crescimento. No Apêndice A são apresentadas estas estatísticas desagregadas.

o fator de expansão amostral (peso amostral) foi utilizado para que a agregação fornecesse informações representativas do universo socioeconômico brasileiro[75].

O Gráfico 3.3 apresenta a evolução do salário médio real para indivíduos que trabalharam pelo menos 20 horas na semana e tinham de 18 a 60 anos de idade. Durante a década de 1980 a economia brasileira conviveu com sucessivos planos de estabilização de sua moeda e a natureza destes planos, além de suas consequências, explicam as flutuações do salário real durante esta década[76]. Ao regredir os dados por décadas contra uma tendência temporal, verifica-se que as décadas de 1980 e de 1990 apresentam ausência de tendência (positiva ou negativa) enquanto que a década de 2000 apresenta uma tendência positiva e estatisticamente significativa do salário real[77]. O Gráfico 3.4 decompõe os dados dos salários médios reais por grupos de qualificação onde Ensino Fundamental representa indivíduos com 0 a 8 anos de estudo, Ensino Médio indivíduos com 9 a 11 anos de estudo e, por fim, Ensino Superior indivíduos com 12 anos ou mais de estudo. Ao se regredir cada uma das três séries contra uma tendência temporal, verificamos que nas décadas de 1980 e 1990 todas são estatisticamente não significativas, isso representa que na média da década, os salários reais dos trabalhadores pouco, médio e muito qualificados não variaram. Para a década de 2000, o salário real dos trabalhadores pouco qualificados apresenta tendência positiva, o salário real dos trabalhadores médio qualificados apresentou-se invariante e, por fim, o salário real do trabalhador muito qualificado apresentou forte tendência negativa ao longo da década de 2000.

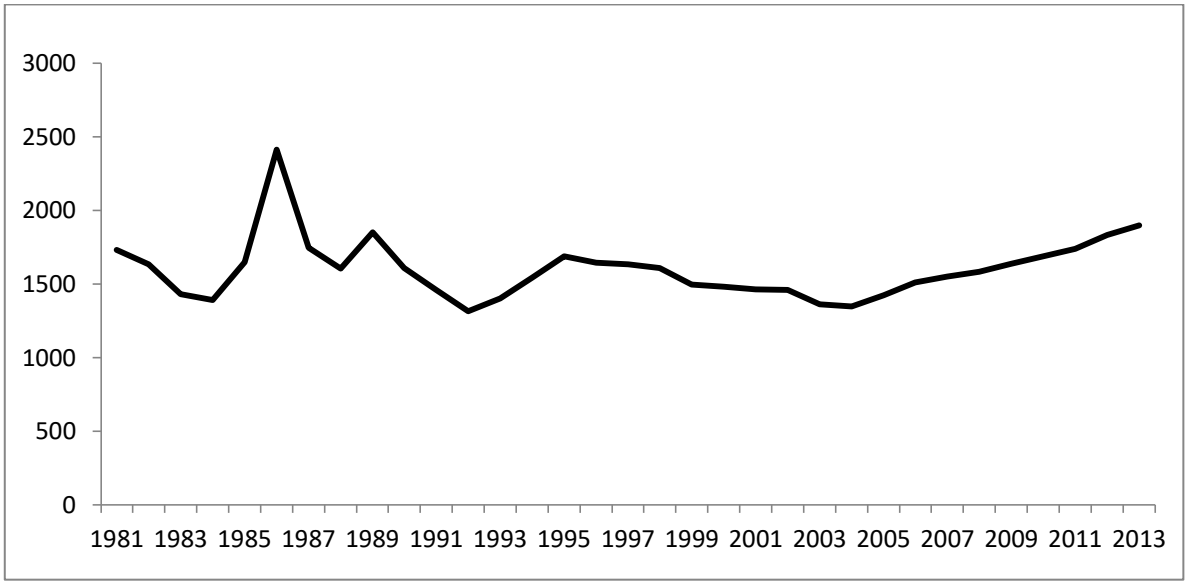
---

[75] Da amostra total de indivíduos que trabalhavam 20 horas ou mais na semana de referência e tinham de 18 a 60 anos, apenas 5% declararam ter outro trabalho além do trabalho principal. Para todas as estatísticas descritivas, os resultados apresentados usando trabalho principal são idênticos aos resultados apresentados usando todos os trabalhos semanais.

[76] O Plano Cruzado, primeiro plano de estabilização, tinha como núcleo de ação o congelamento dos preços e uma política de gatilho salarial que incorporava ganhos reais aos salários dos trabalhadores. Estas políticas de combate à inflação inercial além de suas consequências explicam as flutuações do salário real ao longo da década de 1980. Para uma análise aprofundada dos planos de estabilização do processo inflacionário brasileiro, Castro (2005) e Franco (2005) são importantes referências.

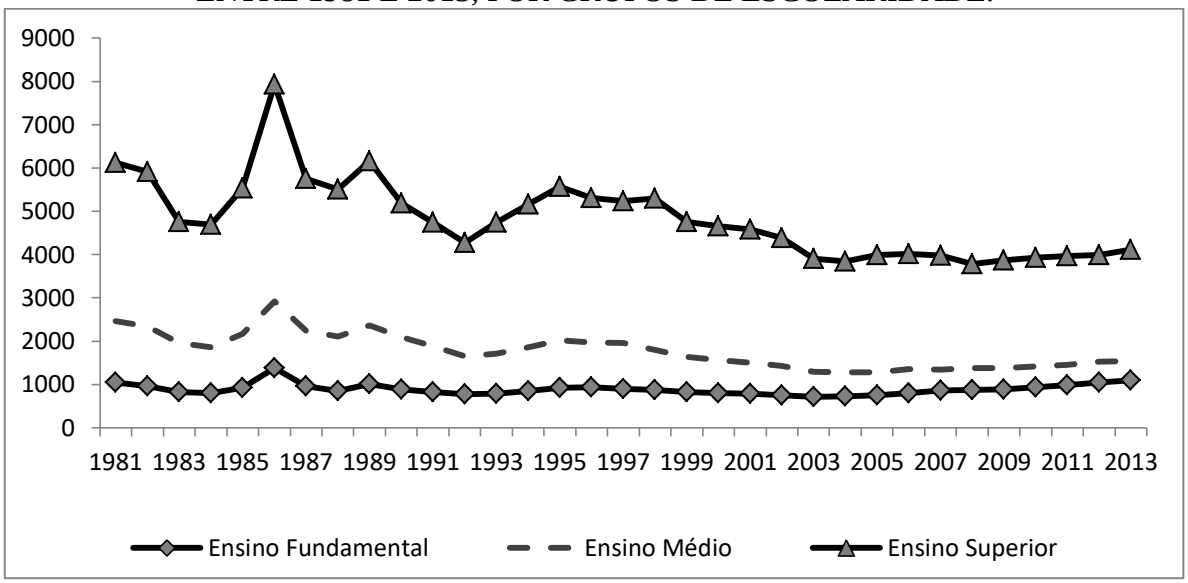
[77] A regressão da série de dados do GRÁFICO 3.3 apenas para a década de 1980 apresenta uma tendência com parâmetro 18,74 e um p-valor de 0,58, mostrando-se estatisticamente não significativo. A década de 1990 apresenta um parâmetro de tendência temporal de 16,41 com p-valor de 0,23 e, por fim, a década de 2000 apresenta um parâmetro de 30,93 com p-valor de 0,003. Em todas as três regressões, as séries foram estimadas com uma constante e uma tendência temporal como variável explicativa. O R^2da primeira regressão (década de 1980) foi 0,04, o da segunda regressão (década de 1990) foi 0,17 e, por fim, o da terceira regressão (década de 2000) foi 0,68.

GRÁFICO 3.3: EVOLUÇÃO DO SALÁRIO REAL MÉDIO DOS TRABALHADORES ENTRE 1981 E 2013



Fonte: Microdados PNAD 1981-2013
OBS.: IPCA/IPCA Setembro 2012 = 100

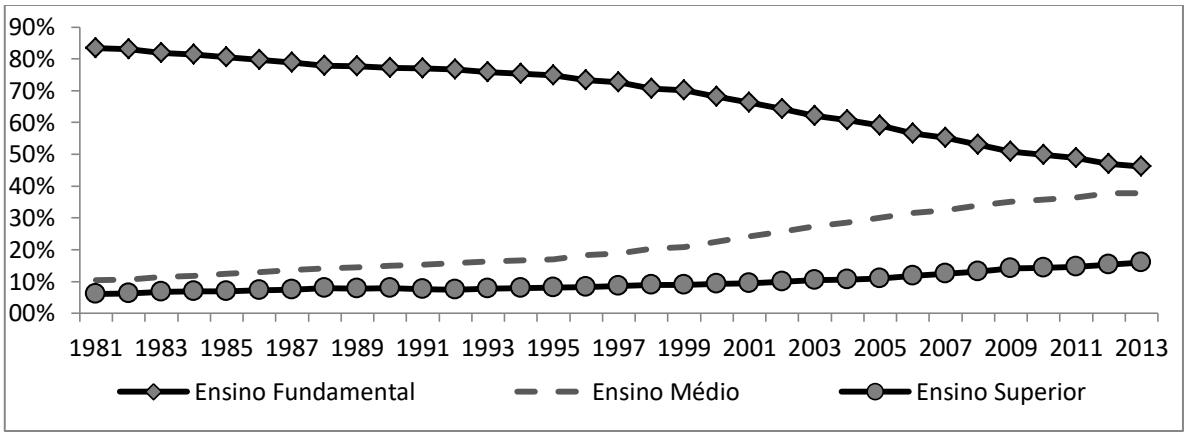
GRÁFICO 3.4: EVOLUÇÃO DO SALÁRIO REAL MÉDIO DOS TRABALHADORES, ENTRE 1981 E 2013, POR GRUPOS DE ESCOLARIDADE.



Fonte: Microdados PNAD 1981-2013
OBS 1.: IPCA/IPCA Setembro 2012 = 100
OBS 2: Ensino Fundamental equivale de 0 a 8 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Superior** equivale de 12 a 17 anos de estudo.

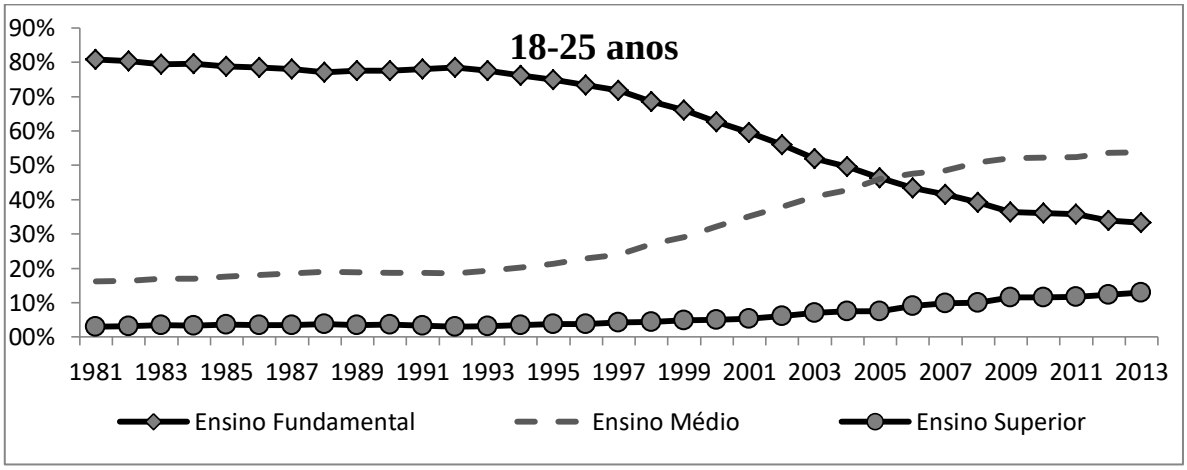
O Gráfico 3.5 decompõe, para cada ano, o percentual de trabalhadores entre Ensino Fundamental, Ensino Médio e Ensino Superior. Em 1981, 83% do total de trabalhadores entre 18 e 60 anos tinham qualificação compatível com o Ensino Fundamental, 11% compatível com o Ensino Médio e 6% compatível com o Ensino Superior. Ao longo das décadas houve uma tendência de aumento dos trabalhadores com Ensino Fundamental representando 35% do total de trabalhadores na década de 2000. A expansão do Ensino Superior ocorrida no Brasil na década de 2000 aumenta a participação de trabalhadores com esta qualificação, principalmente na segunda metade da década de 2000. O Gráfico 3.6 decompõe os resultados apresentados no Gráfico 3.5 para três grupos etários, 18-25 anos, 26-40 anos e 41- 60 anos. O processo de diminuição da participação de indivíduos pouco qualificados no total de trabalhadores é visto nos três grupos etários, mas com muita intensidade para trabalhadores entre 18 e 25 anos. No entanto, uma coorte de 18 a 25 anos em um ano específico pertencerá, por completo, à coorte de 26 a 40 anos após 8 anos e, com isto, podemos interpretar este efeito entre coortes apresentados na comparação dos três gráficos. No Apêndice A são apresentadas outras estatísticas descritivas além de uma decomposição para um número maior de grupos de escolaridade e, pelo Gráfico 3.A3 e 3.A4 verificamos que a escolaridade dos grupos etários cresce *pari passu* quando analisamos em anos médios de estudo. O percentual de trabalhadores que ainda estudam aumenta ao longo da década de 1990 para os três grupos de escolaridade, mas com muita intensidade para os grupos mais jovens.

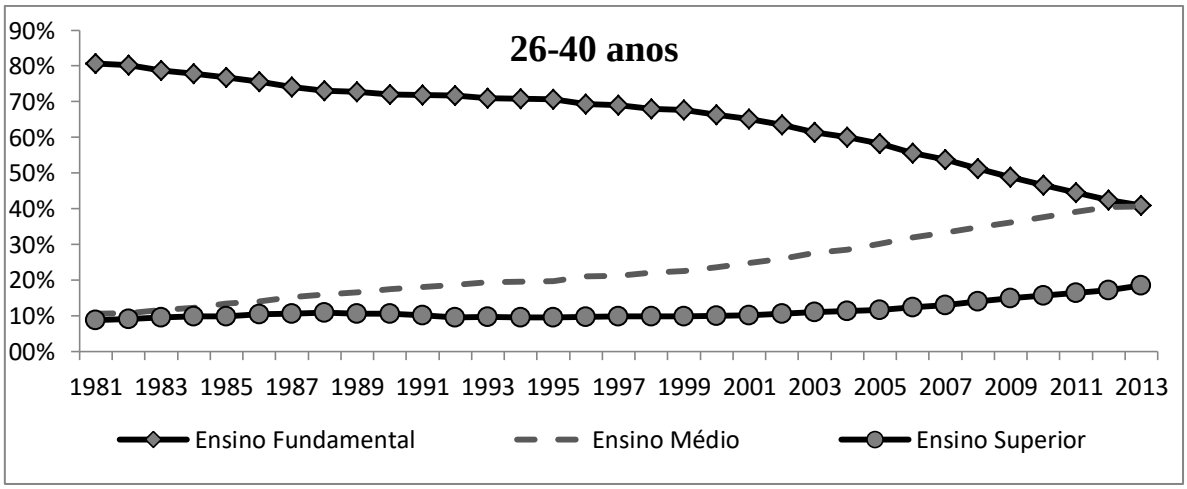
GRÁFICO 3.5: PERCENTUAL DE TRABALHADORES POR GRUPOS DE ESCOLARIDADE

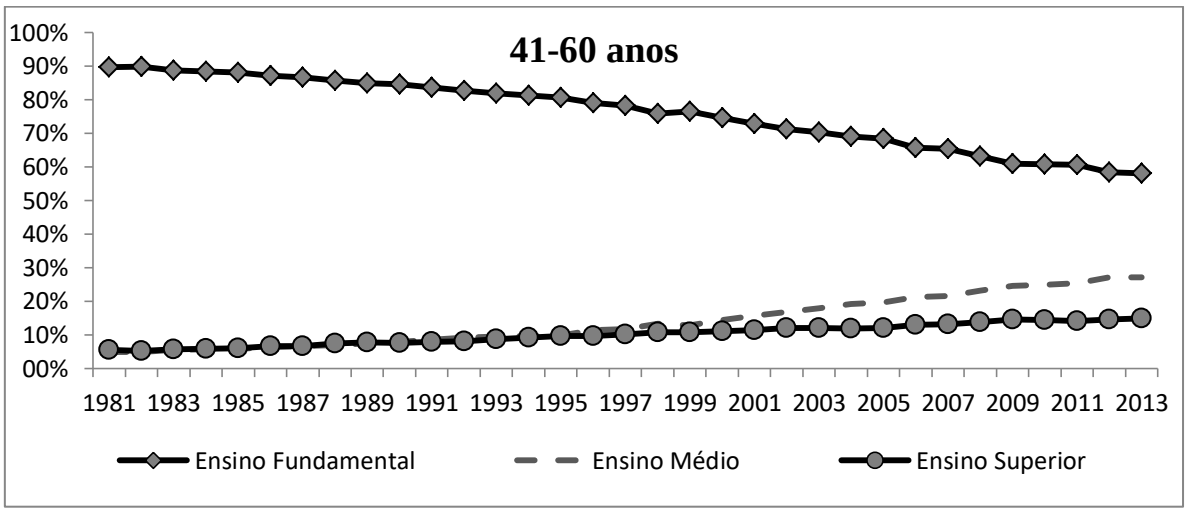


Fonte: Microdados PNAD 1981-2013
OBS: Ensino Fundamental equivale de 0 a 8 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Superior** equivale de 12 a 17 anos de estudo.

GRÁFICO 3.6: PERCENTUAL DE TRABALHADORES POR GRUPOS DE ESCOLARIDADE E GRUPO ETÁRIO







Fonte: Microdados PNAD 1981-2013
OBS: Ensino Fundamental equivale de 0 a 8 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Superior** equivale de 12 a 17 anos de estudo.

3. OFERTA RELATIVA E SALÁRIO RELATIVO

Na literatura de mercado de trabalho, o prêmio salarial por qualificação foi analisado por diversos autores relacionando-o com a oferta relativa de mão-de-obra, assim como em Katz & Murphy (1992), Acemoglu & Autor (2012) e Card & Lemieux (2001). Uma proposta alternativa para se construir a série de salário relativo seria apresentar o prêmio salarial medido por regressões Mincerianas[78]. Uma diferença deste presente trabalho para artigos clássicos de oferta relativa e salário relativo consiste na determinação da coorte como ano de nascimento diferentemente de Katz & Murphy (1992) e Acemoglu & Autor (2012) onde coorte é definida por experiência potencial[79]. Já no trabalho de Card & Lemieux (2001), a coorte é definida pelo ano de nascimento e assume-se que grupos etários diferentes não são perfeitamente substitutos e esta elasticidade de substituição pode ser estimada empiricamente. No trabalho seminal de Katz & Murphy, os trabalhadores de diferentes grupos etários, mas com mesmo capital humano, são perfeitamente substitutos.

Seguindo o modelo canônico teórico de Acemoglu & Autor, a produção é descrita por uma tecnologia CES (*Constant Elasticity of Substitution*) que combina dois tipos de trabalhadores, qualificado (H) e pouco qualificado (L) tal que estes grupos educacionais não são perfeitamente substitutos e a substituição entre estes grupos se dá uma elasticidade de substituição σ. Há um fator de aumento tecnológico para o trabalho qualificado (A_H) e um para o trabalho pouco qualificado (A_L). Assume-se que θ é um parâmetro de importância relativa do trabalho pouco qualificado na produção total[80].

$$Y = \left[\theta(A_L L)^{\frac{\sigma-1}{\sigma}} + (1-\theta)(A_H H)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{3.1}$$

O salário recebido pelo trabalhador qualificado, w_H, é dado pela sua produtividade marginal do trabalho, $\partial Y/\partial H$, assim como o salário recebido pelo trabalhador pouco qualificado, w_L, é dado pela sua produtividade marginal, $\partial Y/\partial L$. Assim, o salário relativo

[78] O Gráfico 3.2 apresenta resultados de estimações por uma tradição Minceriana.

[79] Segundo Katz & Murphy, experiência potencial é definida pela função: $min\{idade - anos\ de\ estudo - 7, idade - 17\}$

[80] A Elasticidade de substituição entre os grupos de qualificação é dada, na CES, por $\frac{1}{1-\frac{\sigma-1}{1}} = \sigma$, onde $\sigma \in [0, \infty)$

do trabalhador qualificado em relação o trabalhador pouco qualificado ($^{w_H}/_{w_L}$) pode ser descrito por

$$w = \frac{w_H}{w_L} = \frac{1-\theta}{\theta}\left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{H}{L}\right)^{-\frac{1}{\sigma}} \tag{3.2}$$

Aplicando o logaritmo natural na equação (3.2), o logaritmo do salário relativo $\ln\left(^{w_H}/_{w_L}\right)$ pode ser expresso como função de uma constante, $ln\left(\frac{1-\theta}{\theta}\right)$, um termo que captura o que a literatura denomina por "*skill-bias of technology*", $ln\left(\frac{A_H}{A_L}\right)$, e a oferta relativa por grupos de qualificação, $ln\left(\frac{H}{L}\right)$

$$\ln(w) = ln\left(\frac{1-\theta}{\theta}\right) + \frac{\sigma-1}{\sigma}ln\left(\frac{A_H}{A_L}\right) - \frac{1}{\sigma}ln\left(\frac{H}{L}\right) \tag{3.3}$$

Acemoglu & Autor assumem a hipótese da forma específica de Tinbergen onde o logaritmo da tecnologia relativa pode ser descrito por uma tendência linear, ou seja, $ln\left(^{A_H}/_{A_L}\right) = \gamma_0 + \gamma_1 t$. Substituindo na equação (3.3), mudanças no salário relativo podem ser explicadas por uma tendência temporal linear, que captura a demanda por tecnologia, e por mudanças na oferta relativa por grupos de qualificação:

$$\ln(w_t) = ln\left(\frac{1-\theta}{\theta}\right) + \frac{\sigma-1}{\sigma}\gamma_0 + \frac{\sigma-1}{\sigma}\gamma_1 t - \frac{1}{\sigma}ln\left(\frac{H_t}{L_t}\right) \tag{3.4}$$

Para a economia brasileira, Fernandes e Menezes-Filho (2012) estimaram as elasticidades de substituição adicionando um terceiro grupo de qualificação, denominado trabalhadores intermediários. A justificativa para a introdução de um grupo intermediário aos modelos tradicionais de "*skill-biased technological change*", como Katz & Murphy (1992), Acemoglu & Autor (2012), e para modelos de decomposição com elasticidade de substituição *à la* Card & Lemieux (2001), reside na pouca participação de trabalhadores

com nível superior no total de trabalhadores além de trabalhadores com ensino médio, considerados pouco qualificados em modelos tradicionais, são mais próximos de uma qualificação intermediária no caso de países em desenvolvimento. O Gráfico 3.5 explicitou que historicamente o Brasil tem uma grande parcela de sua mão-de-obra com qualificação compatível com o ensino fundamental, no entanto, as mudanças direcionaram para um aumento da participação de trabalhadores com ensino médio completo e ensino superior completo[81]. Assim, apresentaremos os resultados relativos a três grupos de qualificação denominados *Ensino Fundamental*, representando o grupo dos pouco qualificados c com até 8 anos de estudo, o que representa ter frequentado no máximo o ensino fundamental. A oferta do grupo *Ensino Fundamental* será denominada por L e o salário por w_L. O segundo grupo denominado *Ensino Médio* corresponde aos indivíduos com 9 a 11 anos de estudo e, portanto, frequentaram o *Ensino Médio*. A oferta do grupo médio será denominada por M e o salário por w_M. Por fim, o terceiro grupo será denominado por Ensino Superior e corresponderá a indivíduos com 12 a 17 anos de estudo, equivalente a ter frequentado o ensino superior. A oferta do grupo Ensino Superior será denotada por H e o salário por w_H.

$$\ln\left(\frac{w_M}{w_L}\right) = \underset{(0.1865)}{-0{,}2942} + \underset{(0.0051)}{0.0174}\, t \; \underset{(0.080)}{-0.5447} \ln\left(\frac{M}{L}\right) \tag{3.5}$$

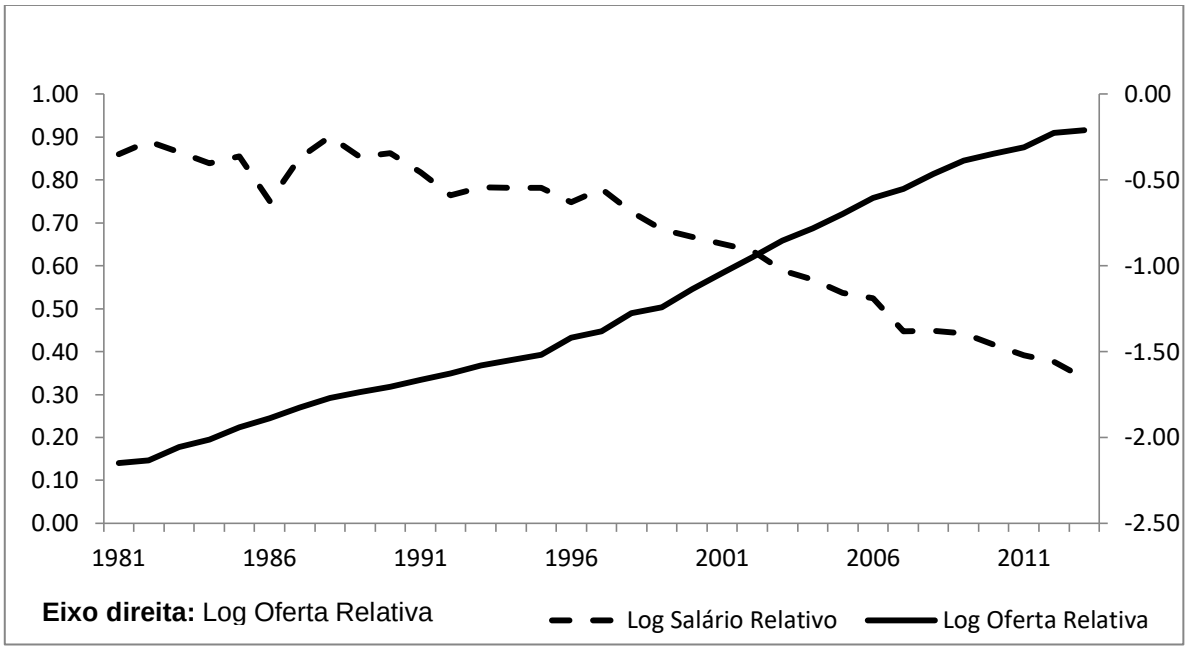
$R^2 = 0{,}96$ Nº obs. = 33

A Equação (3.5) apresenta o resultado da regressão do salário relativo do grupo médio qualificado em relação ao grupo pouco qualificado como função de uma tendência temporal e da oferta relativa destes trabalhadores. A tendência positiva mostra que, entre 1981 e 2013, um ano a mais aumenta o salário relativo em 1,74 % e este valor é significativo a 1%. A oferta relativa é negativamente relacionada com o salário relativo e uma queda de 1% na oferta relativa de trabalhadores médio qualificados em relação a trabalhadores pouco qualificados provoca um aumento de 0,55% no salário relativos destes grupos de educação. Da forma funcional (3.4) temos que $\sigma_{L,M} = {}^{-1}/_{0.5447} = 1{,}83$,

[81] No Apêndice A, o Gráfico 3.A2 apresenta uma decomposição da mão-de-obra brasileira por um número maior de grupos de qualificação. Entre 1981 e 2013, a perda de participação de trabalhadores com fundamental incompleto foi acompanhada por um aumento da participação de trabalhadores com ensino médio completo e ensino superior completo.

portanto a elasticidade de substituição entre trabalhadores médio qualificados e trabalhadores pouco qualificados equivale em magnitude a 1,83, corroborando a hipótese de Acemoglu & Autor de que os diferentes grupos de escolaridade não são substitutos perfeitos[82]. O Gráfico 3.7 apresenta a evolução do salário relativo em logaritmo (linha vermelha) e a evolução da oferta relativa em logaritmo (linha azul) comprovando que as séries são negativamente correlacionadas. Ao longo de todo o período houve um aumento relativo de trabalhadores médio qualificados em relação a trabalhadores pouco qualificados, algo já sinalizado pelas estatísticas descritivas. Esta mudança relativa foi acompanhada pela queda do prêmio de qualificação do ensino médio em relação ao ensino fundamental.

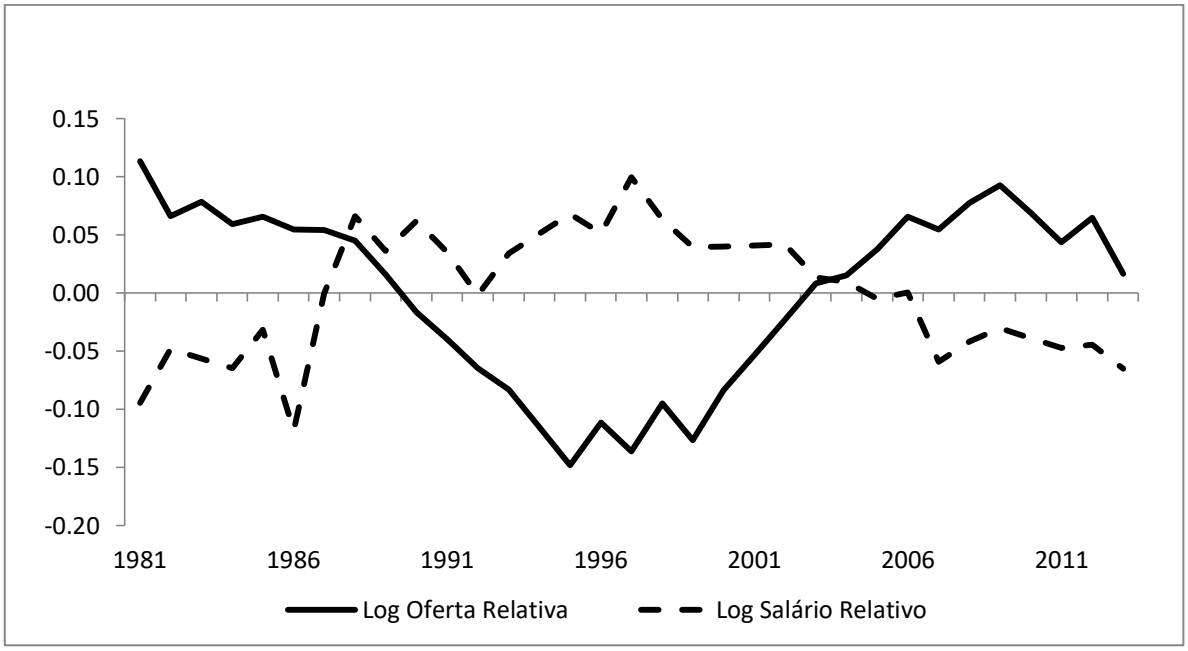
GRÁFICO 3.7: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO MÉDIO – ENSINO FUNDAMENTAL



Fonte: Microdados PNAD 1981-2013
OBS: Ensino Fundamental equivale de 0 a 8 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.

GRÁFICO 3.8: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*

[82] Para que os grupos de escolaridade fossem substitutos perfeitos na função de produção CES, a elasticidade de substituição entre ambos deveria ser infinita, fato não comprovado empiricamente.



Fonte: Microdados PNAD 1981-2013
OBS: Ensino Fundamental equivale de 0 a 8 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

O Gráfico 3.8 apresenta o logaritmo da oferta relativa e o logaritmo do salário relativo ambos sem a tendência temporal. Este resultado é obtido regredindo cada uma destas duas séries contra uma constante e uma tendência temporal e os resíduos destas regressões foram utilizados para na construção do gráfico. Essas séries mostram como se ajustam o logaritmo do salário relativo e o logaritmo da oferta relativa uma vez controlada a tendência temporal[83]. O grau de ajuste das séries é bastante alto e períodos com aumento da oferta relativa de médio qualificados em relação a pouco qualificados são acompanhados com a queda do salário relativo destes grupos de qualificação, corroborado pelo R^2 da série de 0,96.

[83] Ao regredir o resíduo do salário relativo contra uma constante mais o resíduo da oferta relativa, a constante apresenta parâmetro nulo e o parâmetro do resíduo da oferta relativa é -0,5447, o mesmo apresentado na regressão anterior (Eq. 3.5). Portanto, o gráfico com as séries sem tendência temporal demonstra como $\ln\left(\frac{w_M}{w_L}\right)$ e $\ln\left(\frac{M}{L}\right)$ se ajustam na regressão.

$$\ln\left(\frac{w_H}{w_M}\right) = \underset{(0.0510)}{0.5436} + \underset{(0.0016)}{-0.0068\,t} + \underset{(0.094)}{-0.7259}\ln\left(\frac{H}{M}\right) \tag{3.6}$$

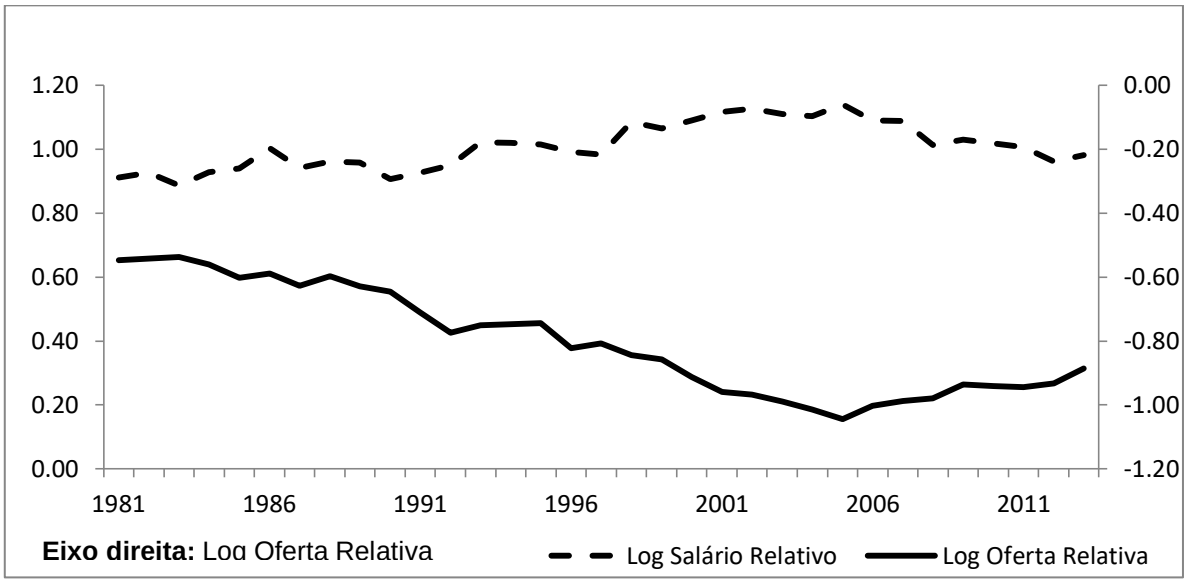
$R^2 = 0{,}78$ Nº obs. = 33

A Equação (3.6) apresenta a regressão do salário relativo do grupo de qualificados em relação ao grupo médio qualificado contra uma constante, uma tendência temporal e a oferta relativa destes grupos de qualificação. Diferente da regressão apresenta em (3.5) onde a tendência temporal fora positiva, o salário relativo dos qualificados em relação aos trabalhadores médio qualificados apresentou tendência negativa e estatisticamente significativa a 1% onde para cada ano entre 1981 e 2013 o salário relativo diminuiu 0,68%. A oferta relativa relaciona-se negativamente com o salário relativo e, para uma queda de 1% na oferta relativa, o salário relativo aumenta em 0,73%. A elasticidade de substituição entre trabalhadores qualificados e trabalhadores médio qualificados é $\sigma_{H,M} = {-1}/{0.7259} = 1{,}37$. Acemoglu & Autor (2012) calculam esta elasticidade para a economia americana entre 1963 e 2008 relacionando a média de indivíduos que foram para a universidade ("*some college*", "*college graduate*" e "*greater than college*"), equivalentes ao que se foi denominado qualificados (H) neste presente trabalho e a média de indivíduos que foram para o ensino médio ("*high school dropout*" e "*high school graduate*"), equivalentes ao que se foi denominado médio qualificados (M). O parâmetro estimado pelos autores foi – 0,612 com desvio-padrão de 0.128 e a elasticidade de substituição entre estes grupos de qualificação foi 1,63[84]. O Gráfico 3.9 apresenta as séries do logaritmo do salário relativo e do logaritmo da oferta relativa que se apresentam negativamente relacionados, como evidenciados pela regressão na equação (3.6). A série de oferta relativa mostra que entre 1981 e 2005 houve um aumento relativo de trabalhadores médio qualificados em relação a trabalhadores qualificados e, a partir de 2006, trabalhadores qualificados passam a crescer proporcionalmente mais do que trabalhadores médio

[84] Fernandes & Menezes-Filho (2012) replicam o trabalho de Katz & Murphy para o Brasil assumindo três grupos de qualificação com possibilidade de o grupo intermediário fazer tarefas simples e tarefas complexas. Os resultados dependem de uma calibração e a elasticidade de substituição entre trabalhadores intermediários e trabalhadores não qualificados variou de 1,441 a 2,976. A elasticidade de substituição entre trabalhadores qualificados e intermediários variou entre 1,60 e 2,04. Como resultado geral da pesquisa, as simulações indicaram que entre 1981 e 2009 o Brasil aumentou a demanda por trabalhadores intermediários em atividades simples e aumentou a demanda por trabalhadores qualificados em atividades complexas. Para maiores detalhes sobre os resultados, calibrações e simulações, Fernandes & Menezes-Filho (2012).

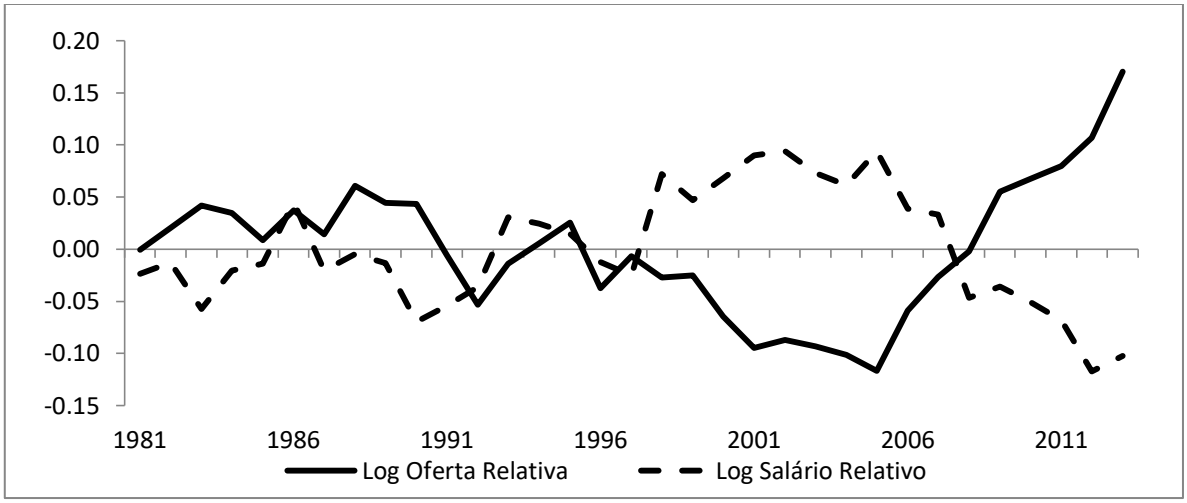
qualificados. Esta reversão na oferta relativa é acompanhada pela reversão no salário relativo em 2006.

GRÁFICO 3.9: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO SUPERIOR – ENSINO MÉDIO



Fonte: Microdados PNAD 1981-2013
OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.

GRÁFICO 3.10: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



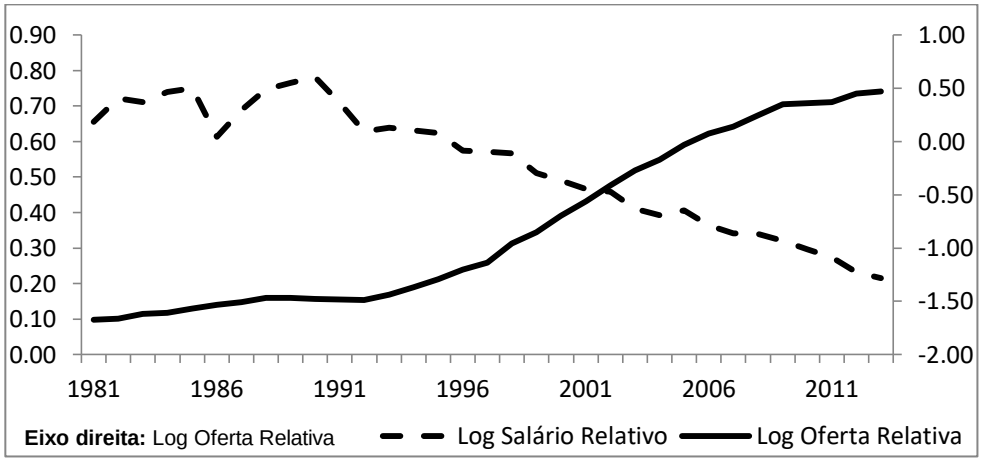
Fonte: Microdados PNAD 1981-2013
OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

3.1 DECOMPOSIÇÕES POR GRUPOS ETÁRIOS

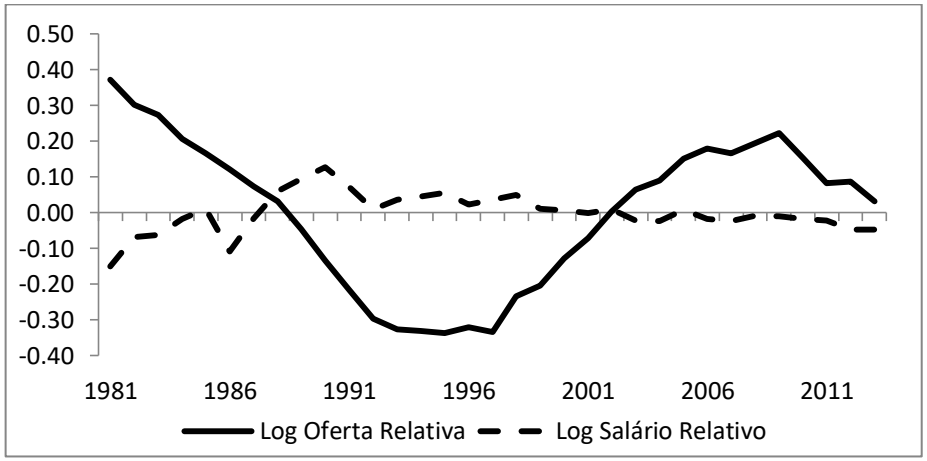
Os exercícios relacionando o logaritmo da oferta relativa e o logaritmo do salário relativo, assim como suas séries sem tendência temporal são apresentados nos Gráficos 3.11 a 3.13 para os grupos etários 18-25 anos, 26-40 anos e 41-60 anos respectivamente. No Apêndice A são apresentadas as regressões da equação (3.4) para cada um dos três grupos etários, assim como o R^2 para se verificar o quanto o modelo ajusta o salário relativo com a tendência temporal e a oferta relativa. Comparando o Ensino Médio com o Ensino Fundamental, os três grupos etários apresentaram a mesma tendência de aumento da oferta relativa e a mesma tendência de queda do salário relativo. Em especial, o grupo etário de 18 a 25 anos apresentou uma quebra da tendência aproximadamente em 1996, intensificando o aumento da oferta de trabalhadores com Ensino Médio (completo ou incompleto) comparado com número de trabalhadores com Ensino Fundamental (completo ou incompleto). Os outros grupos etários apresentaram uma tendência de aumento da oferta relativa de trabalhadores com Ensino Médio em relação a trabalhadores com Ensino Fundamental mais suave. O ajuste do salário relativo com a oferta relativa para Ensino Médio em relação ao Ensino Fundamental é feito pelos gráficos da direita, com ausência da tendência temporal que fora expurgada regredindo cada série contra uma constante e uma tendência. As séries apresentadas são os resíduos destas regressões. Para o grupo etário 18-25 anos, um aumento de 1% na oferta relativo do Ensino Médio em relação ao Ensino Fundamental diminui o salário relativo deste grupo etário em 0,18% e este resultado é estatisticamente significativo a 1%. Para o grupo etário 26-40 anos, um aumento de 1% na oferta relativa diminui o salário relativo deste grupo etário em 0,34% e este resultado também é estatisticamente significativo a 1%. Para o grupo etário 41-60 anos, a oferta relativa não se apresentou estatisticamente significativa (*p-valor* igual a 0,83) e esta ausência de correlação entre as séries pode ser comprovada pelo Gráfico 3.13. A relação entre Ensino Superior (completo e incompleto) com Ensino Médio, estabelecidas nos gráficos inferiores de 3.11 a 3.13 é condizente com a previsão do modelo de Acemoglu & Autor para os grupos etários 18-25 anos e 26-40 anos, porém, para trabalhadores de 41 a 60 anos, o modelo prevê que a oferta relativa está positivamente relacionada com o salário negativo, mas esta relação não é estatisticamente significativa (*p-valor* de 0,37). Na próxima seção, avançaremos no trabalho de Card & Lemieux para ajustarmos as lacunas empíricas deixadas pelo modelo no ajuste com os grupos etários.

GRÁFICO 3.11: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **18 A 25 ANOS**

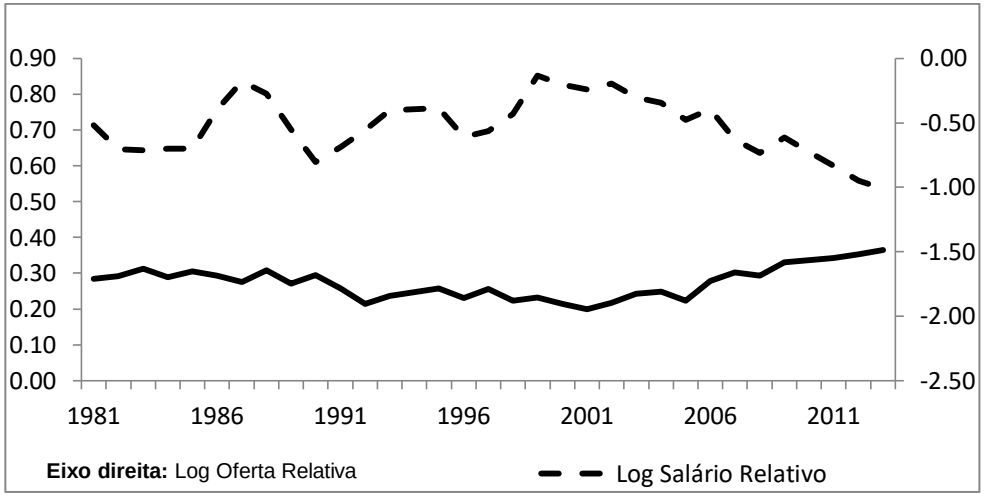
ENSINO MÉDIO – ENSINO FUNDAMENTAL



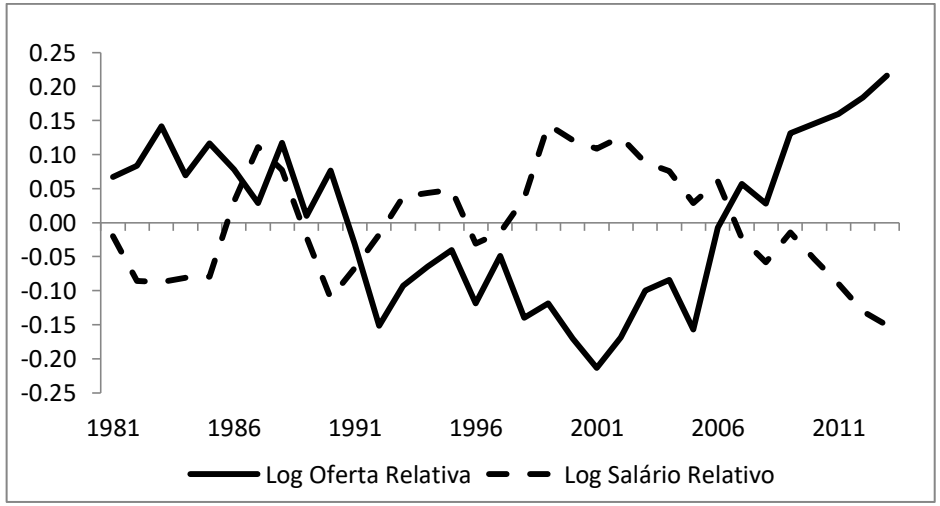
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



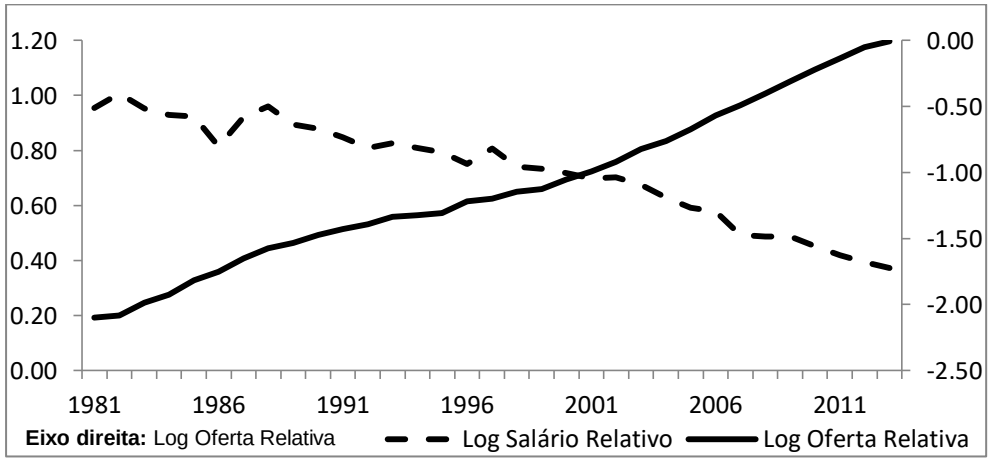
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

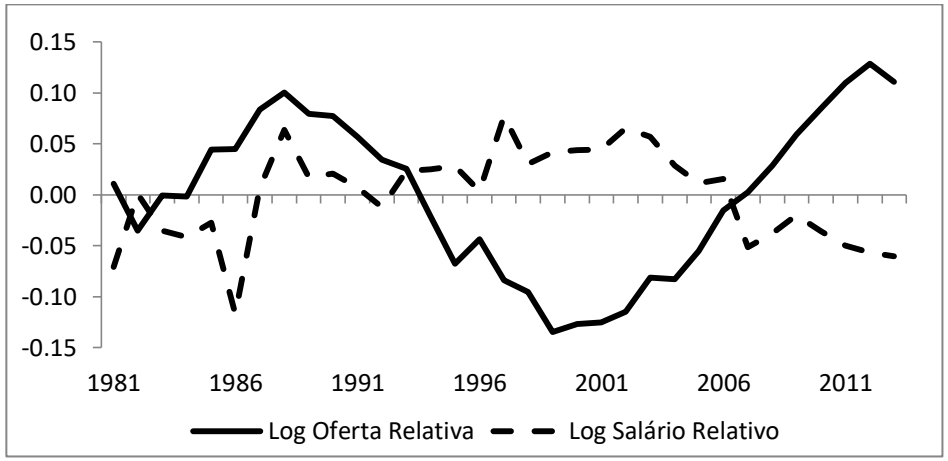
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3.12: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **26 A 40 ANOS**

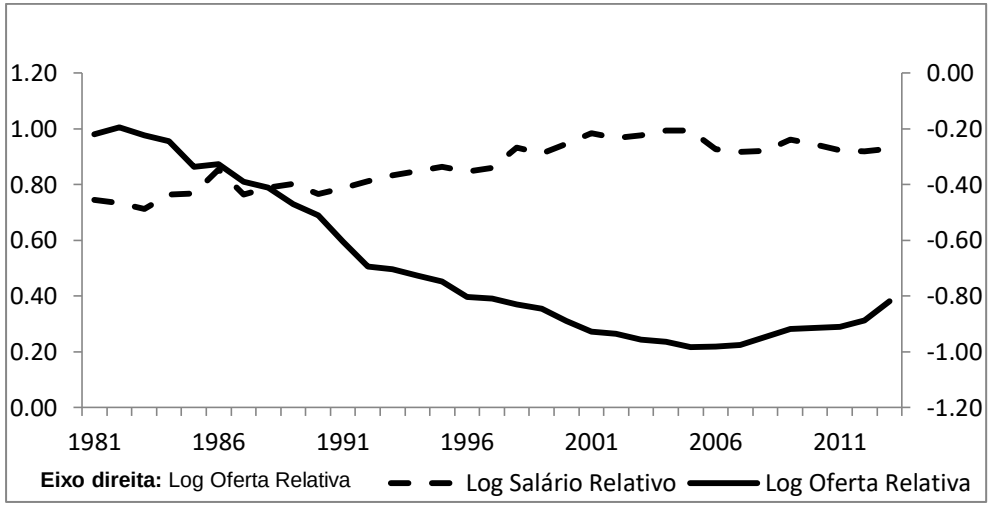
ENSINO MÉDIO – ENSINO FUNDAMENTAL



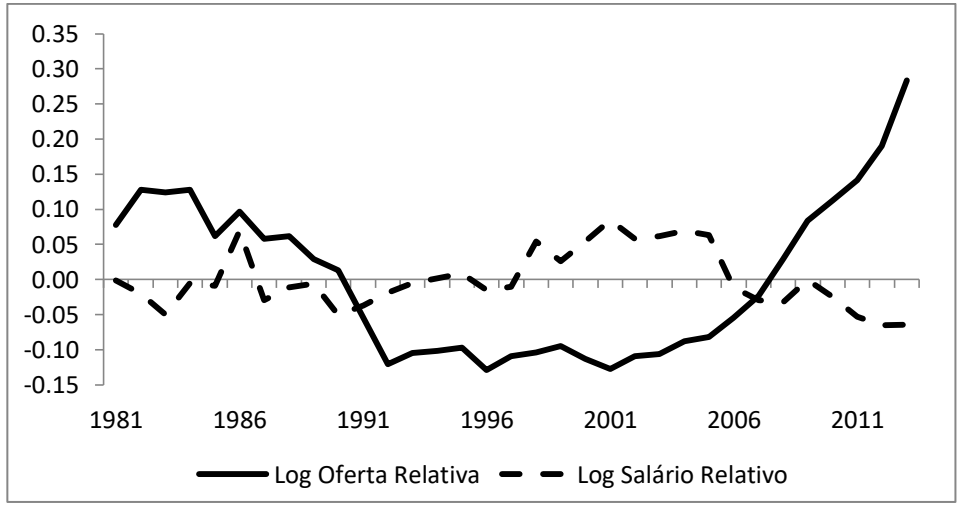
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL



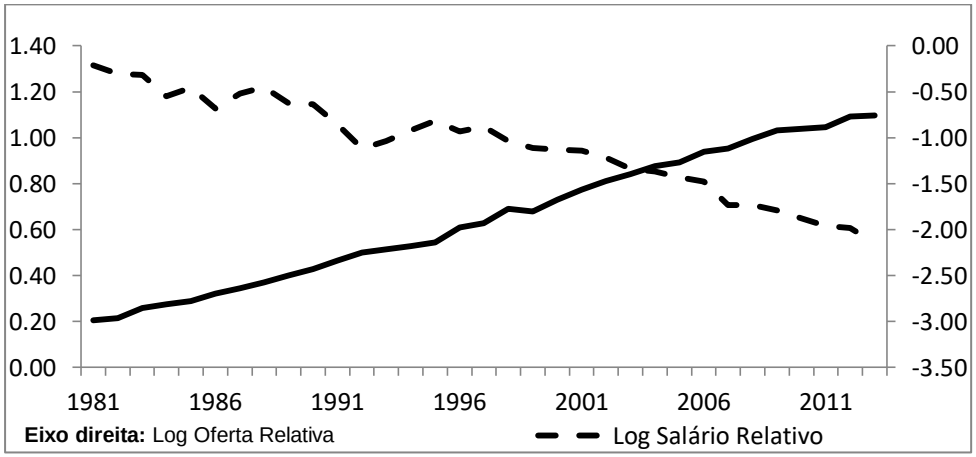
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

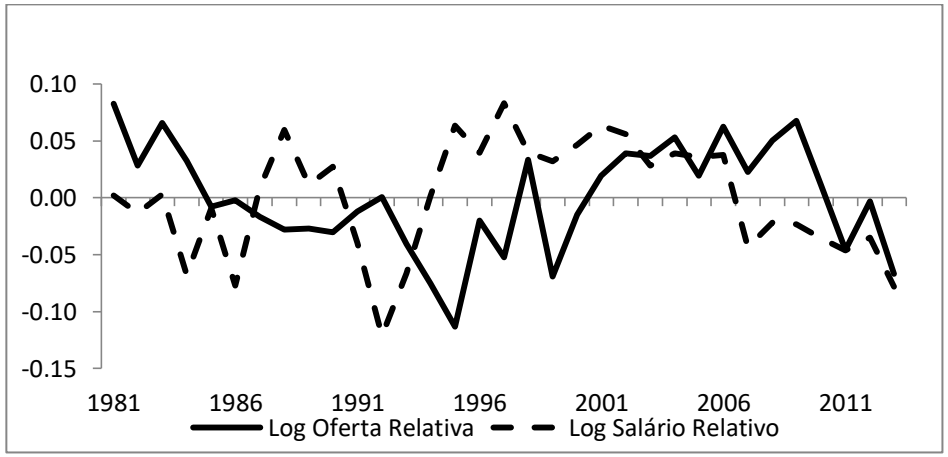
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3.13: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **41 A 60 ANOS**

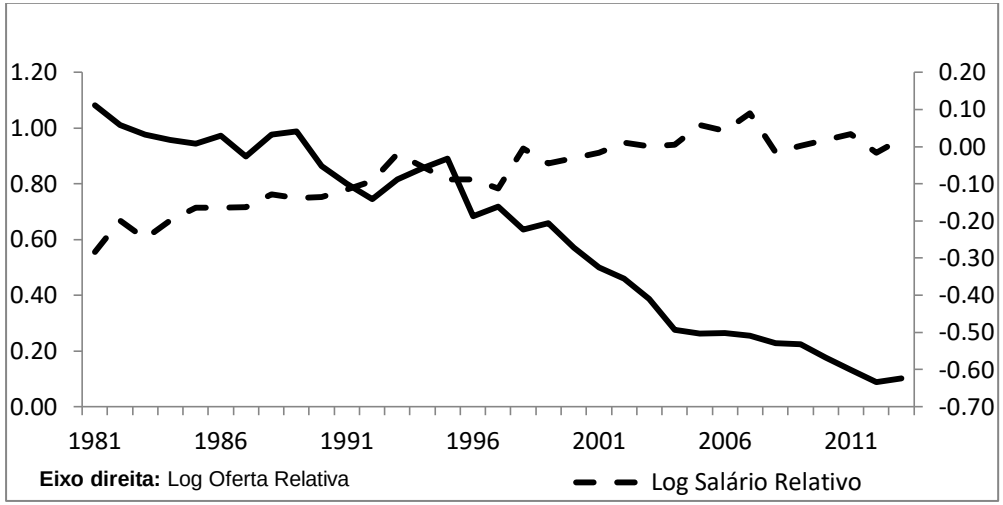
ENSINO MÉDIO – ENSINO FUNDAMENTAL



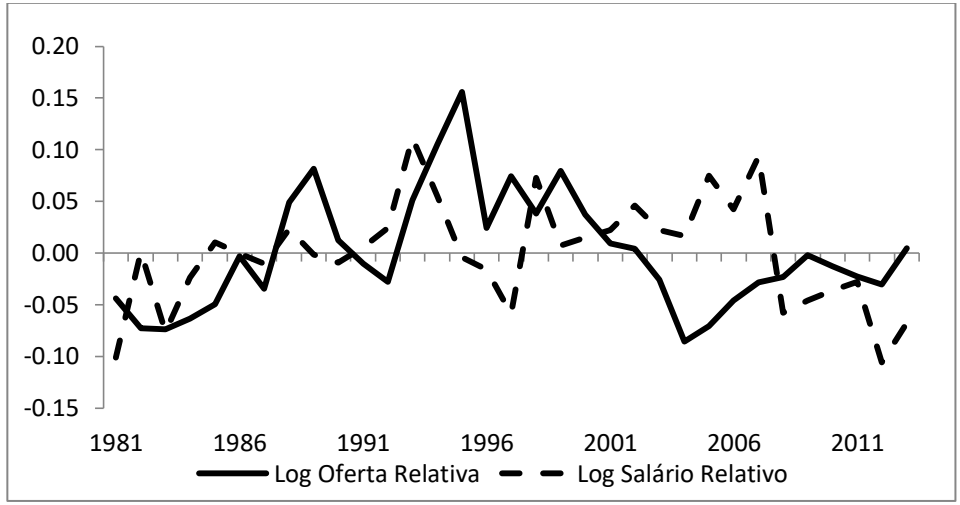
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL



Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

3.2. ELASTICIDADES DE SUBSTITUIÇÕES

O trabalho seminal de Card & Lemieux (2001) avança em relação ao trabalho de Katz & Murphy ao assumir que trabalhadores de um mesmo grupo de escolaridade, mas de diferente grupo etário não são substitutos perfeitos na função de produção agregada da economia[85]. Esta contribuição será utilizada para se entender a evolução do salário relativo por grupos etários avançando a partir dos resultados da seção anterior, mas incorporando três grupos de qualificação ao trabalho e ajustando-o à realidade de economias em desenvolvimento com uma parcela considerável de sua mão-de-obra com qualificação relativa ao ensino fundamental incompleto. Pecora & Menezes-Filho (2014) estimaram as elasticidades de substituição para a economia brasileira adicionando ao modelo de Card & Lemieux um terceiro grupo de qualificação denominado médio qualificado. A função de produção agregada da economia combina, assim como Acemoglu & Autor, um termo de oferta para trabalhadores qualificados, H_t, um termo de oferta para trabalhadores não qualificados, C_t, um parâmetro de produtividade do trabalho qualificado, θ_{ht}, um parâmetro de produtividade do trabalho não qualificado, θ_{ct}

$$y_t = \left(\theta_{ht} H_t^{\rho} + \theta_{ct} C_t^{\rho}\right)^{1/\rho} \tag{3.7}$$

A elasticidade de substituição[86] entre trabalhadores qualificados (H_t) e trabalhadores não qualificados (C_t), denominado σ_E e dada por $\sigma_E = \frac{1}{1-\rho}$.

---

[85] Assim como Acemoglu & Autor definiram que trabalhadores de diferentes grupos de escolaridade não são substitutos perfeitos e que a substituição entre estes grupos de escolaridade pode ser capturada empiricamente ao se estimar a função de produção CES em sua forma de salário relativo e oferta relativa, o trabalho de Card & Lemieux também o faz. No entanto, segundo estes autores, "*Theses studies analyze the evolution of the return to schooling under the assumption that diferente age groups with the same level of education are perfect substitutes in production*" (Card & Lemieux, 2001, p. 708).

[86] A taxa marginal de substituição entre trabalhadores qualificados (H) e trabalhadores pouco não qualificados (C), $TMS_{H,C}$, é dada por:

$$TMS_{H,C} = \frac{\theta_{ht}}{\theta_{ct}} \left[\frac{C_t}{H_t}\right]^{1-\rho}$$

A elasticidade de substituição entre trabalhadores qualificados (H) e trabalhadores não qualificados (C) é dada por:

$$\sigma_A = \frac{dLn\left({}^{C}/_{H}\right)}{dLnTMS_{H,C}} = \frac{1}{1-\rho}$$

Para se assumir a hipótese de não substituição perfeita entre os grupos etários para um dado grupo educacional, utiliza-se um agregador CES[87] para o trabalho qualificado, H_t, e um agregador CES para o trabalho não qualificado tal que em ambos existem j grupos etários. Para o agregador de trabalho qualificado, $\propto_j$ representa um parâmetro de produtividade média para o grupo etário j e, para o agregador de trabalho não qualificado, β_j representa um parâmetro de produtividade média para o grupo etário j. Em ambas as funções, a elasticidade de substituição[88] entre indivíduos de grupos etários diferentes é dada por $\sigma_A = \frac{1}{1-\eta}$.

$$H_t = \left[\sum_j \left(\propto_j H_{jt}^{\eta}\right)\right]^{1/\eta} \tag{3.8}$$

$$C_t = \left[\sum_j \left(\beta_j C_{jt}^{\eta}\right)\right]^{1/\eta} \tag{3.9}$$

Pecora & Menezes-Filho (2014) adicionam uma terceira equação ao modelo de Card & Lemieux tal que o trabalho não qualificado pode ser agregado pelo trabalho médio qualificado, M_{jt}, pelo trabalho pouco qualificado, L_{jt}, por um parâmetro de produtividade do trabalho médio qualificado, δ_{mj} e um parâmetro de produtividade do trabalho pouco qualificado, δ_{lj}. Em todos os casos, j representa o grupo etário e t representa o ano em que os dados serão agregados. A elasticidade de substituição entre o trabalho médio qualificado e o trabalho pouco qualificado é dada por $\sigma_U = \frac{1}{1-\gamma}$.

---

[87] "*Constant Elasticity of Substitution*" ou Elasticidade de Substituição Constante em português.

[88] Para o grupo educacional H, a taxa marginal de substituição entre trabalhadores do grupo etário j e trabalhadores do grupo etário j', $TMS_{j,j'}$, é dada por:

$$TMS_{j,j'} = \frac{\theta_{ht}}{\theta_{ct}}\left[\frac{H_{j',t}}{H_{j,t}}\right]^{1-\rho}$$

A elasticidade de substituição entre trabalhadores qualificados do grupo etário j (H_j) e trabalhadores qualificados do grupo etário *j'* ($H_{j'}$) é dada por:

$$\sigma_A = \frac{dLn\left(\frac{H_{j'}}{H_j}\right)}{dLnTMS_{j,j'}} = \frac{1}{1-\eta}$$

$$C_{jt} = \left(\delta_{mj} M_{jt}^{\gamma} + \delta_{lj} L_{jt}^{\gamma}\right)^{1/\gamma} \tag{3.10}$$

Um ponto importante a ser destacado é que o parâmetro de produtividade na equação (3.10) é constante no tempo variando apenas entre grupos etários enquanto o parâmetro de produtividade na equação (3.7) varia entre grupos etários e varia no tempo. Além disto, uma vez fixado δ_{mj} igual a 1, o parâmetro de produtividade do grupo etário δ_{lj} será interpretado como a produtividade relativa do indivíduo pouco qualificado em relação ao indivíduo médio qualificado, ambos com idade j. O salário para cada um dos três grupos de educação e para cada um dos *j* grupos etários pode ser calculado pela sua respectiva produtividade marginal. Assim, para o grupo educacional qualificado com idade *j*, o salário é a produtividade marginal representado pela derivada da equação (3.7) em relação a H_{jt}

$$w_{jt}^{H} = \frac{\partial y_t}{\partial H_{jt}} = \left(y_t^{1-\rho}\right)\theta_{ht}\left(H_t^{\rho-1}\right)\left(H_t^{1-\eta}\right)\alpha_j\left(H_{jt}^{\eta-1}\right) \tag{3.11}$$

Aplicando o logaritmo natural no salário dos indivíduos qualificados com idade j no ano t e utilizando $1-\eta = {}^{1}/_{\sigma_A}$, $1-\rho = {}^{1}/_{\sigma_E}$, tem-se

$$\begin{aligned} ln\left(w_{jt}^{H}\right) = (1-\rho)ln(y_t) + ln(\theta_{ht}) + \ln\left(\alpha_j\right) - \frac{1}{\sigma_E} ln(H_t) \\ - \frac{1}{\sigma_A}\left\{ln\left(H_{jt}\right) - ln(H_t)\right\} \end{aligned} \tag{3.12}$$

Os salários do grupo médio qualificado (w_{jt}^{M}) e do grupo pouco qualificado (w_{jt}^{L}) com idade *j* no ano *t* são dados pelas suas respectivas produtividades marginais, ou seja, a derivada da equação (3.7) em relação a M_{jt} e a derivada da equação (3.7) em relação a L_{jt}

$$w_{jt}^{M} = \frac{\partial y_t}{\partial M_{jt}} = \left(y_t^{1-\rho}\right)\theta_{ct}\left(C_t^{\rho-1}\right)\left(C_t^{1-\eta}\right)\beta_j\left(C_{jt}^{\eta-1}\right)\left(C_{jt}^{1-\gamma}\right)\delta_{mj}\left(M_{jt}^{\gamma-1}\right) \tag{3.13}$$

$$w_{jt}^{L} = \frac{\partial y_t}{\partial L_{jt}} = \left(y_t^{1-\rho}\right)\theta_{ct}\left(C_t^{\rho-1}\right)\left(C_t^{1-\eta}\right)\beta_j\left(C_{jt}^{\eta-1}\right)\left(C_{jt}^{1-\gamma}\right)\delta_{lj}\left(L_{jt}^{\gamma-1}\right) \tag{3.14}$$

Aplicando o logaritmo natural em ambos os salários e utilizando $1-\eta = {}^{1}/_{\sigma_A}$, $1-\rho = {}^{1}/_{\sigma_E}$ e $1-\gamma = {}^{1}/_{\sigma_U}$, tem-se

$$\begin{aligned} ln\left(w_{jt}^{M}\right) = (1-\rho)ln(y_t) + ln(\theta_{ct}) + \ln\left(\beta_j\right) + \ln(\delta_{mj}) - \frac{1}{\sigma_E}ln(C_t) \\ - \frac{1}{\sigma_A}\left\{ln\left(C_{jt}\right) - ln(C_t)\right\} - \frac{1}{\sigma_U}\left\{ln\left(M_{jt}\right) - ln\left(C_{jt}\right)\right\} \end{aligned} \tag{3.15}$$

$$\begin{aligned} ln\left(w_{jt}^{L}\right) = (1-\rho)ln(y_t) + ln(\theta_{ct}) + \ln\left(\beta_j\right) + \ln(\delta_{lj}) - \frac{1}{\sigma_E}ln(C_t) \\ - \frac{1}{\sigma_A}\left\{ln\left(C_{jt}\right) - ln(C_t)\right\} - \frac{1}{\sigma_U}\left\{ln\left(L_{jt}\right) - ln\left(C_{jt}\right)\right\} \end{aligned} \tag{3.16}$$

Para se estimar as elasticidades de substituições σ_E, σ_A e σ_U é necessária a construção dos salários relativos como função das ofertas relativas. Primeiro, o salário relativo do médio qualificado de idade j em relação ao pouco qualificado de idade j é construído e sua estimação recupera os parâmetros σ_U e δ_{lj} com δ_{mj} igual a 1. Com estes parâmetros, a equação (3.10) é construída e uma segunda estimação recupera o valor de σ_A. Com esta última elasticidade, três novas regressões recuperam os valores de α_j e β_j. Por fim, com as equações (3.8) e (3.9) construída, a última etapa de regressão recupera o parâmetro σ_E.

A base de dados utilizada nas regressões segue a construção da seção 2 com todas as PNAD disponíveis entre 1981 e 2013. As pesquisas domiciliares foram compatibilizadas para que as informações sobre mercado de trabalho e educação fossem comparáveis ao longo dos anos. Além disto, o fator de expansão da amostra foi utilizado para se agregar a oferta de trabalho, ajustada por horas semanais, para cada grupo etário, grupo educacional ano por ano. Assim como na seção anterior, trabalhadores pouco qualificados têm de 0 a 8 anos de estudo, equivalentes a terem frequentado no máximo o ensino fundamental.

Trabalhadores médio qualificados têm de 9 a 11 anos de estudo, equivalente a terem frequentado o ensino médio e, por fim, trabalhadores qualificados têm de 12 a 17 anos de estudo, equivalente a terem frequentado o ensino superior. Os grupos etários utilizados são 18 a 25 anos, 26 a 40 anos e 41 a 60 anos. No apêndice deste presente trabalho são apresentadas estatísticas descritivas e estimações para um número maior de grupos etários e os resultados mostram-se robustos aos encontrados com 3 grupos etários. Além disto, dado que as séries são criadas para 29 anos multiplicado pelo número de grupos etários, exercícios de estimação com 6 ou 7 grupos etários dobram ou mais do que dobram o número de observações, evento importante para se fazer testes de hipóteses. No geral, todas as estimações mesmo com um menor número de observações mostraram-se estatisticamente significativas a 1% em geral.

3.2.1. Salário relativo médio qualificado em relação ao pouco qualificado

O salário relativo do grupo médio qualificado com idade j em relação ao grupo não qualificado com idade j, w_{jt}^{M}/w_{jt}^{L}, é construído dividindo-se (3.13) por (3.14)

$$\frac{w_{jt}^{M}}{w_{jt}^{L}} = \frac{\delta_{mj}}{\delta_{lj}}\left[\frac{M_{jt}}{L_{jt}}\right]^{\gamma-1} \tag{3.17}$$

Aplicando o logaritmo natural na série e fazendo o parâmetro de produtividade relativa em função do grupo médio qualificado com idade j, δ_{lj}/δ_{mj}, tem-se:

$$ln\left(\frac{w_{jt}^{M}}{w_{jt}^{L}}\right) = -\ln\left(\frac{\delta_{lj}}{\delta_{mj}}\right) - \frac{1}{\sigma_U}\, ln\left(\frac{M_{jt}}{L_{jt}}\right) \tag{3.18}$$

Por simplificação, assumiremos que $\delta_{mj} = 1$ e, para cada idade j, δ_{lj} representa um parâmetro de produtividade relativa que empiricamente pode ser estimado através de

dummies de idade ($dummy_j$). Assim, a equação (3.18) pode ser tratada empiricamente como

$$ln\left(\frac{w_{jt}^{M}}{w_{jt}^{L}}\right) = \beta^{j}{}_{1} dummy_j + \beta_2 ln\left(\frac{M_{jt}}{L_{jt}}\right) + e_{jt} \qquad (3.19)$$

Onde e_{jt} é um componente aleatório com média condicional zero, $\delta_{lj} = exp\left(-\beta^{j}{}_{1}\right)$ para todo j, $\sigma_U = {}^{-1}/_{\beta_2}$ e a forma funcional (3.19) é estimada sem constante e com todas as *dummies* de idade. A tabela abaixo apresenta os resultados da Regressão 1 com somente o β_2 da equação estocástica (3.18). O parâmetro é estatisticamente significativo a 1% e o R^2 da Regressão 1 mostra que $ln\left({w_{jt}^{M}}/{w_{jt}^{L}}\right)$ e $ln\left({M_{jt}}/{L_{jt}}\right)$ se ajustam bem com correlação negativa entre estas séries[89].

REGRESSÃO 1: ELASTICIDADE DE SUBSTITUIÇÃO ENTRE NÃO QUALIFICADOS

${}^{-1}/_{\sigma_U}$	-0.285*** (0.011)
Nº Observações	87
R^2	0.993

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

OBS.: Regressão controla por 3 dummies de idade e não se utiliza constante

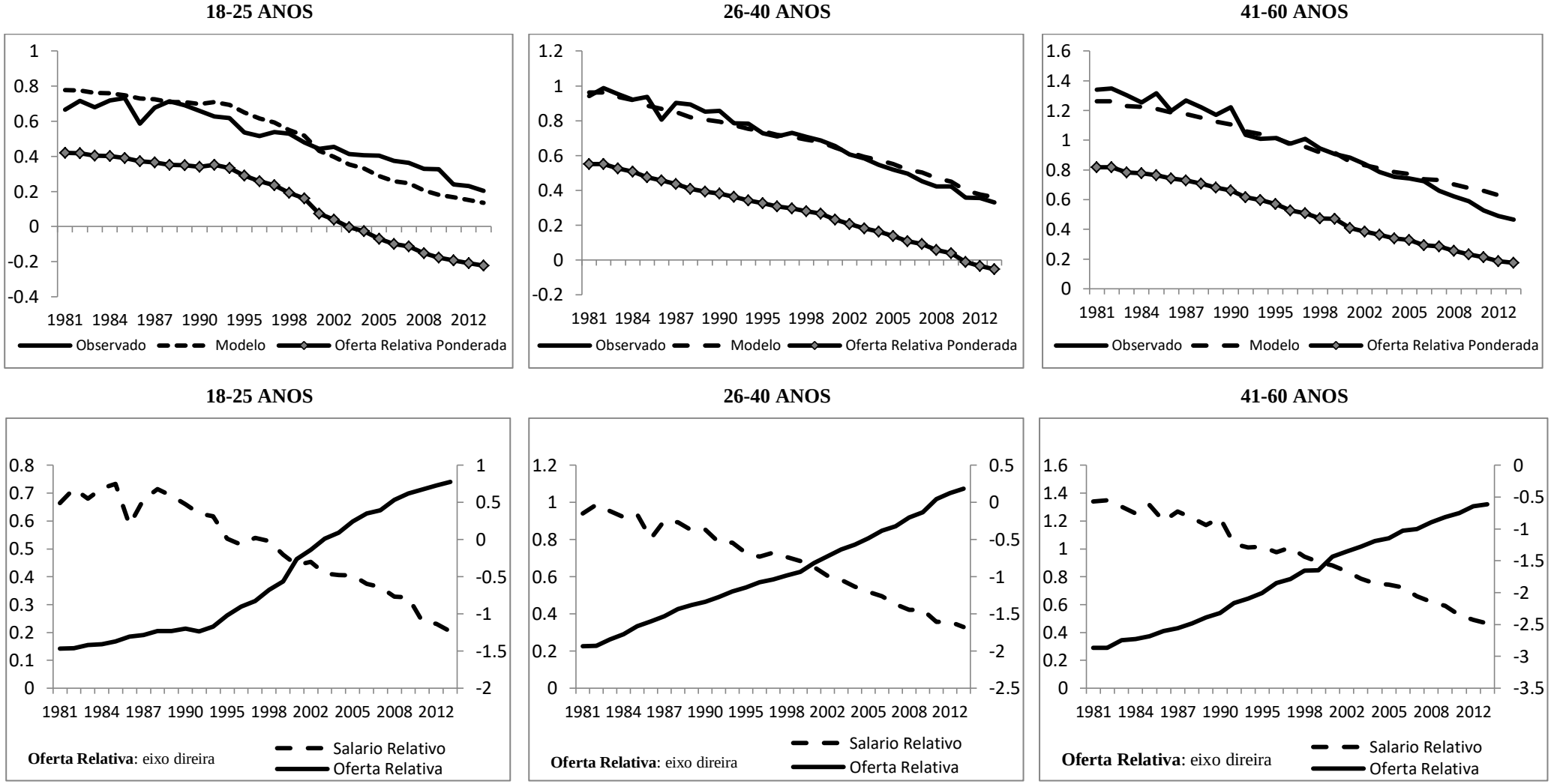
A Regressão 1 demonstra que um aumento de 1% na oferta relativa de médio qualificados em relação a pouco qualificados, ambos pertencentes ao grupo etário *j*, diminui seu salário relativo em 0,29%. A elasticidade de substituição entre o grupo médio qualificado e o grupo pouco qualificado, σ_U, equivale a 3,51 e este parâmetro é maior em magnitude do

[89] Os parâmetros de produtividade estimados para o grupo não qualificado foram: $\delta_{18-25} = 0{,}699$, $\delta_{26-40} = 0{,}663$, $\delta_{41-60} = 0{,}641$. Para o grupo médio qualificados, foi assumido $\delta_{18-25} = 1$, $\delta_{26-40} = 1$ e $\delta_{41-60} = 1$.

parâmetro encontrado na seção passada[90] que fora de 1,83. Para a Regressão 1, foi assumido que os diferentes grupos etários em um mesmo grupo de qualificação não são substitutos perfeitos, hipótese esta assumida na seção passada ao se estimar a elasticidade de substituição pela metodologia de Acemoglu & autor. O Gráfico 3.14 apresenta para cada grupo etário j (18-25 anos, 26-40 anos e 41-60 anos) dois gráficos. No gráfico superior, o salário relativo observado nos dados na PNAD (linha continua) são apresentados com o salário relativo predito pelo modelo descrito na equação (3.19) com os parâmetros estimados na Regressão 1 (linha tracejada). Para todos os grupos etários, o modelo predito ajusta-se aos dados observados, algo já sinalizado pela estimação pelo R^2. A linha com indicação de um losango representa a oferta relativa ponderada, ou seja, $-0{,}285.\, ln\left(\frac{M_{jt}}{L_{jt}}\right)$ com a oferta relativa variando para cada um dos três grupos etários ao longo dos anos. Portanto, a trajetória de queda do salário relativo, para cada grupo etário, é explicada pela trajetória de queda da oferta relativa ponderada e a diferença entre o modelo predito e a oferta relativa ponderada representa a produtividade relativa do grupo pouco qualificado em relação ao grupo médio qualificado. No gráfico inferior, o salário relativo observado e utilizado na regressão é contrastado com a oferta relativa observada. O salário relativo em logaritmo é analisado no eixo da esquerda enquanto a oferta relativa em logaritmo é analisada pelo eixo da esquerda.

[90] Equação (3.5). Na seção anterior, a série de dados era agregada enquanto que nesta seção, agrega-se por grupos etários.

GRÁFICO 3.14: REGRESSÃO 1 - OFERTA RELATIVA E SALÁRIO RELATIVO PARA CADA GRUPO ETÁRIO



Fonte: Resultados da Regressão 1 com microdados PNAD 1981-2013

3.2.2. Salário relativo qualificado em relação ao médio qualificado e ao pouco qualificado

Com os parâmetros estimados na Regressão 1 é possível recuperar os valões de δ_{mj}, δ_{lj} e γ e construir a função $C_{jt} = \left(\delta_{mj} M_{jt}^{\gamma} + \delta_{lj} L_{jt}^{\gamma}\right)^{1/\gamma}$. Esta função será importante na segunda etapa da regressão para se recuperar a elasticidade de substituição entre grupos etários, σ_A. A partir das equações (3.11), (3.13) e (3.14) é possível construir o salário relativo para qualificados em relação a trabalhadores médio qualificados, ambos pertencentes ao grupo etário j (w_{jt}^{H}/w_{jt}^{M}) e o salário relativo para qualificados em relação ao pouco qualificado, com ambos pertencentes ao grupo etário j, (w_{jt}^{H}/w_{jt}^{L}). Ambos os salários relativos são avaliados no período t e, por simplificação, indexaremos por l como $I = L$ quando a referência for o grupo pouco qualificado e $I = M$ quando a referência for o grupo médio qualificado.

$$\frac{w_{jt}^{H}}{w_{jt}^{I}} = \frac{\theta_{Ht}}{\theta_{Ct}} \left[\frac{H_t}{C_t}\right]^{\rho-1} \left[\frac{H_t}{C_t}\right]^{1-\eta} \frac{\alpha_j}{\beta_j} \left[\frac{H_{jt}}{C_{jt}}\right]^{\eta-1} \left[\frac{C_{jt}}{I_{jt}}\right]^{\eta-1} \delta_{Ij}^{-1} \qquad I = L, M \qquad (3.20)$$

Aplicando o logaritmo natural no salário relativo e utilizando $1 - \eta = 1/\sigma_A$, $1 - \rho = 1/\sigma_E$ e $1 - \gamma = 1/\sigma_U$, tem-se

$$\ln\left(\frac{w_{jt}^{H}}{w_{jt}^{I}}\right) = ln\left(\frac{\theta_{Ht}}{\theta_{Ct}}\right) + ln\left(\frac{\alpha_j}{\beta_j}\right) - \ln(\delta_{Ij}) - \left[\frac{1}{\sigma_E} - \frac{1}{\sigma_A}\right] ln\left(\frac{H_t}{C_t}\right) - \frac{1}{\sigma_A} ln\left(\frac{H_{jt}}{C_{jt}}\right) - \frac{1}{\sigma_U} ln\left(\frac{C_{jt}}{I_{jt}}\right) \qquad I = L, M \qquad (3.21)$$

Na equação (3.21), o termo $ln\left(\frac{\theta_{Ht}}{\theta_{Ct}}\right) - \left[\frac{1}{\sigma_E} - \frac{1}{\sigma_A}\right] ln\left(\frac{H_t}{C_t}\right)$ varia apenas no tempo, porém não varia entre os grupos de educação e, empiricamente, pode ser representado por um

conjunto de *dummies* de ano ($dummy_t$). O termo $ln\left(\frac{\alpha_j}{\beta_j}\right) - \ln(\delta_{Ij})$ varia entre os grupos etários e grupos de educação, porém não varia no tempo e pode, empiricamente, ser representado por um conjunto de *dummies* de idade e educação ($dummy_{I,j}$) tal que $dummy_{M,j} = 1$ se pertencer ao grupo de média qualificação com idade j e $dummy_{L,j} = 1$ se pertencer ao grupo de pouco qualificação e idade j. Assim, assumindo que o termo estocástico $e_{I,j,t}$ possui média condicional zero, a equação (3.21) pode ser estimada por mínimos quadrados ordinários, sem constante aditiva e representada por[91]

$$\ln\left(\frac{w_{jt}^{H}}{w_{jt}^{I}}\right) = \beta_1 dummy_t + \beta_2 dummy_{I,j} + \beta_3 ln\left(\frac{H_{jt}}{C_{jt}}\right) + \beta_4 ln\left(\frac{C_{jt}}{I_{jt}}\right) + e_{I,j,t} \qquad I = L, M \qquad (3.22)$$

Os resultados de (3,22) são sintetizados na Regressão 2 com os parâmetros β_3 e β_4, ambos estatisticamente significativos a 1%. Nota-se que ${}^{-1}/_{\sigma_U}$ apresentou parâmetro -0,285 na regressão 1 e -0,280 na regressão 2 e estes valores similares são validados pela construção teórica do modelo. A elasticidade de substituição entre grupos etários (para o mesmo grupo educacional), σ_A, é 10,29.

[91] Os parâmetros da equação (3.21) representam:
$\beta_1 = ln\left(\frac{\theta_{Ht}}{\theta_{Ct}}\right) - \left[\frac{1}{\sigma_E} - \frac{1}{\sigma_A}\right] ln\left(\frac{H_t}{C_t}\right)$; $\beta_2 = ln\left(\frac{\alpha_j}{\beta_j}\right) - \ln(\delta_{Ij})$; $\sigma_A = {}^{-1}/_{\beta_3}$; $\sigma_U = {}^{-1}/_{\beta_4}$

REGRESSÃO 2: ELASTICIDADE DE SUBSTITUIÇÃO ENTRE GRUPOS ETÁRIOS	
$-1/\sigma_U$	-0.280*** (0.009)
$-1/\sigma_A$	-0.089*** (0.024)
Nº Observações	174
R^2	0.999

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

OBS.: Regressão controla por 5 dummies de idade-educação, 29 dummies de ano e não se utiliza constante

A regressão 2 apresenta a elasticidade de substituição entre grupos etários, σ_A e, consequentemente o valor de η. Assim, a próxima etapa será a estimação dos valores de α_j e β_j para se construir H_t e C_t representados pelas equações (3.8) e (3.9). Da equação (3.12) tem-se

$$ln\left(w_{jt}^{H}\right) + \frac{1}{\sigma_A}\left\{ln\left(H_{jt}\right)\right\} = (1-\rho)ln(y_t) + ln(\theta_{ht}) + \ln\left(\alpha_j\right) - \left[\frac{1}{\sigma_E} + \frac{1}{\sigma_A}\right] ln(H_t) \quad (3.23)$$

Onde $\frac{1}{\sigma_A}\left\{ln\left(H_{jt}\right)\right\}$ é construído usando o resultado da regressão 2 para σ_A e usando os valores observados e agregados pela PNAD para H_{jt} . O termo $(1-\rho)ln(y_t) + ln(\theta_{ht}) - \left[\frac{1}{\sigma_E} + \frac{1}{\sigma_A}\right] ln(H_t)$ varia apenas no tempo e, empiricamente, pode ser representado por um conjunto de *dummies* de ano ($dummy_t$). O termo $\ln\left(\alpha_j\right)$ varia apenas entre os grupos etários e, empiricamente, pode ser representado por um conjunto de *dummies* de idade ($dummy_j$). Assim, assumindo que o termo estocástico e_t apresenta média condicional zero, a equação (3.23) pode ser estimada por MQO sem constante aditiva na forma funcional.

$$ln\left(w_{jt}^{H}\right) + \frac{1}{\sigma_A}\left\{ln\left(H_{jt}\right)\right\} = \beta_1 dummy_t + \beta_2 dummy_j + e_t \quad (3.24)$$

Onde

$$\alpha_j = \begin{cases} \exp(\beta_{2j}) & se\ a\ dummy\ for\ usada \\ 1 & se\ a\ dummy\ for\ excluída \end{cases}$$

A estimação por MQO da equação (3.24) usando o grupo etário 18-25 anos como referência, ou seja, excluído da estimação para não produzir multicolinearidade perfeita, produziu os resultados α_{18-25} igual a 1, α_{26-40} igual a 0,47 significativo a 1% e, por fim, α_{41-60} igual a 0,35 significativo a 1%. Assim, com os valores de α_j para os três grupos etários e sendo $\eta = 1 - \frac{1}{\sigma_A} = 0{,}9$, é possível construir o agregador CES H_t descrito em (3.8). A próxima etapa consiste em estimar os valores de β_j para se construir o agregador CES C_t. De (3.15) e (3.16) e utilizando $I = L$ para denominar pouco qualificado e $I = M$ para denominar médio qualificado, tem-se:

$$\begin{aligned} ln(w_{jt}^I) + & -\frac{1}{\sigma_A} ln(C_{jt}) + \frac{1}{\sigma_U}\{ln(I_{jt}) - ln(C_{jt})\} \\ & = (1-\rho)ln(y_t) + ln(\theta_{It}) + \ln(\beta_j) + \ln(\delta_{Ij}) \qquad I = L, M \qquad (3.25) \\ & - \left[\frac{1}{\sigma_E} + \frac{1}{\sigma_A}\right] \ln(C_t) \end{aligned}$$

Onde $-\frac{1}{\sigma_A} ln(C_{jt}) + \frac{1}{\sigma_U}\{ln(I_{jt}) - ln(C_{jt})\}$ pode ser calculado usando os resultados[92] da regressão 2 para σ_A e σ_U. O termo $(1-\rho)ln(y_t) + ln(\theta_{It}) - \left[\frac{1}{\sigma_E} + \frac{1}{\sigma_A}\right] \ln(C_t)$ varia apenas no tempo e pode ser representado por um conjunto de *dummies* de ano ($dummy_t$). O termo $\ln(\beta_j) + \ln(\delta_{Ij})$ varia com o grupo etário j e com o grupo educacional I e pode ser representado, empiricamente, por um conjunto de *dummies* de idade e educação ($dummy_{I,j}$). Assumindo-se que o termo estocástico $e_{I,t}$ apresenta média condicional zero, a equação (3.25) pode ser estimada por MQO seguindo a forma funcional:

[92] σ_A e σ_U foram estimados conjuntamente na regressão 2. O termo C_{jt} foi construído na segunda etapa usando os resultados da regressão 1. E, por fim, I_{jt} é agregado a partir dos dados da PNAD tanto para $I = L$ quanto para $I = M$

$$ln(w_{jt}^{I}) + -\frac{1}{\sigma_A} ln(C_{jt}) + \frac{1}{\sigma_U}\{ln(I_{jt}) - ln(C_{jt})\} = \beta_1 dummy_t + \beta_2 dummy_j + e_{I,t} \qquad I = L, M \qquad (3.26)$$

Onde

$$\beta_j = \begin{cases} \exp(\beta_{2j} - ln\delta_{Lj}) & se\ I = L \\ \exp(\beta_{2j}) & se\ I = M \end{cases}$$

A estimação por MQO da equação (3.26) para o grupo o grupo pouco educado, $I = L$, sendo o grupo etário 18-25 anos a referência, ou seja, β_{18-25} igual a 1, gerou como resultados β_{26-40} igual a 0,63 significativo a 1% e β_{41-60} igual a 0,54 também significativo a 1%. Com os valores de β_j para os três grupos etários e o valor de η é possível construir a série agregada de C_t, representada pela equação (3.9), para obter na próxima etapa a estimação da elasticidade de substituição entre o grupo qualificado e o grupo não qualificado, σ_E. Com as séries H_t e C_t construídas utilizando a elasticidade de substituição e os parâmetros de produtividade, o próximo passo é estimar a equação (3.21) e capturar a elasticidade de substituição σ_E. Desta equação, o termo $ln\left(\frac{\alpha_j}{\beta_j}\right) - \ln(\delta_{Ij})$ varia com o grupo etário *j* e com o grupo educacional *l* e pode, empiricamente, ser representado por um conjunto de *dummies* de idade e educação ($dummy_{I,j}$). Assim como na seção anterior, a hipótese da forma específica de Tinbergen é utilizada para se estabelecer uma forma funcional ao parâmetro de produtividade relativa ${\theta_{Ht}}/{\theta_{Ct}}$ tal que $ln\left(\frac{\theta_{Ht}}{\theta_{Ct}}\right) = \beta_0 + \beta_1 t$. Substituindo estes termos na equação (3.21) e assumindo que o termo estocástico $e_{I,j,t}$ apresenta média condicional zero, a equação pode ser estimada por MQO as seguinte forma funcional[93]

[93] As elasticidades de substituição são recuperadas por: $\sigma_E = {-1}/{\beta_3}$, $\sigma_A = {-1}/{\beta_4}$ e $\sigma_U = {-1}/{\beta_5}$

$$\ln\left(\frac{w_{jt}^{H}}{w_{jt}^{I}}\right) = \beta_0 + \beta_1 t + \beta_2 dummy_{I,j} + \beta_3 ln\left(\frac{H_t}{C_t}\right) + \beta_4\left\{ln\left(\frac{C_t}{C_{jt}}\right) - ln\left(\frac{H_t}{H_{jt}}\right)\right\} + \beta_5 ln\left(\frac{C_{jt}}{I_{jt}}\right) + e_{I,j,t} \qquad I = L, M \qquad (3.27)$$

A tabela Regressão 3 apresenta os resultados dos parâmetros β_1, β_3, β_4 e β_5 com o desvio-padrão de cada parâmetro, a significância calculada a partir do p-valor além do número de observações e do R^2 ajustado. Segundo Card & Lemieux, a diferença salarial entre qualificados, H_{jt}, e pouco qualificados, C_{jt}, ambos pertencentes ao mesmo grupo etário *j* e avaliados no ano *t* depende da oferta relativa agregada entre estes grupos de qualificação, ${H_t}/{C_t}$ e depende da oferta relativa do grupo etário j. O termo $ln\left(\frac{C_t}{C_{jt}}\right) - ln\left(\frac{H_t}{H_{jt}}\right)$ pode ser rescrito por $ln\left(\frac{H_{jt}}{C_{jt}}\right) - ln\left(\frac{H_t}{C_t}\right)$ onde ${H_{jt}}/{C_{jt}}$ representa a oferta relativa do grupo etário entre qualificados e não qualificados. Segundo os autores, se $ln\left(\frac{H_{jt}}{C_{jt}}\right) - ln\left(\frac{H_t}{C_t}\right)$ for constante e igual a zero ao longo do tempo, isto significa que a oferta relativa para cada grupo etário j, (H_{jt}/C_{jt}), apresenta taxa de variação igual à oferta relativa agregada, (H_t/C_t) . Este fato é observado empiricamente para o Brasil entre 1981 e 2013 segundo os resultados que serão apresentados nos Gráficos 3.15 a 3.17. Para qualquer grupo etário *j*, um aumento de 1% na oferta relativa agregada de qualificados em relação aos não qualificados[94], ${H_t}/{C_t}$, diminui o salário relativo deste grupo etário em 0,67% e a elasticidade de substituição entre qualificados e não qualificados é $\sigma_E = 1{,}5$.

[94] Note que não qualificados de um grupo etário *j*, C_{jt}, é uma combinação entre trabalhadores médio qualificados, ou que fizeram o ensino médio com ou sem conclusão, e os trabalhadores pouco qualificados, ou que fizeram o ensino fundamental com ou sem conclusão

REGRESSÃO 3: ELASTICIDADE DE SUBSTITUIÇÃO ENTRE QUALIFICADOS E NAO QUALIFICADOS

$-1/\sigma_U$	-0.285*** (0.013)
$-1/\sigma_A$	-0.097*** (0.033)
$-1/\sigma_E$	-0.665*** (0.087)
Tendência	0.009*** (0.003)
Nº Observações	174
R^2	0.981

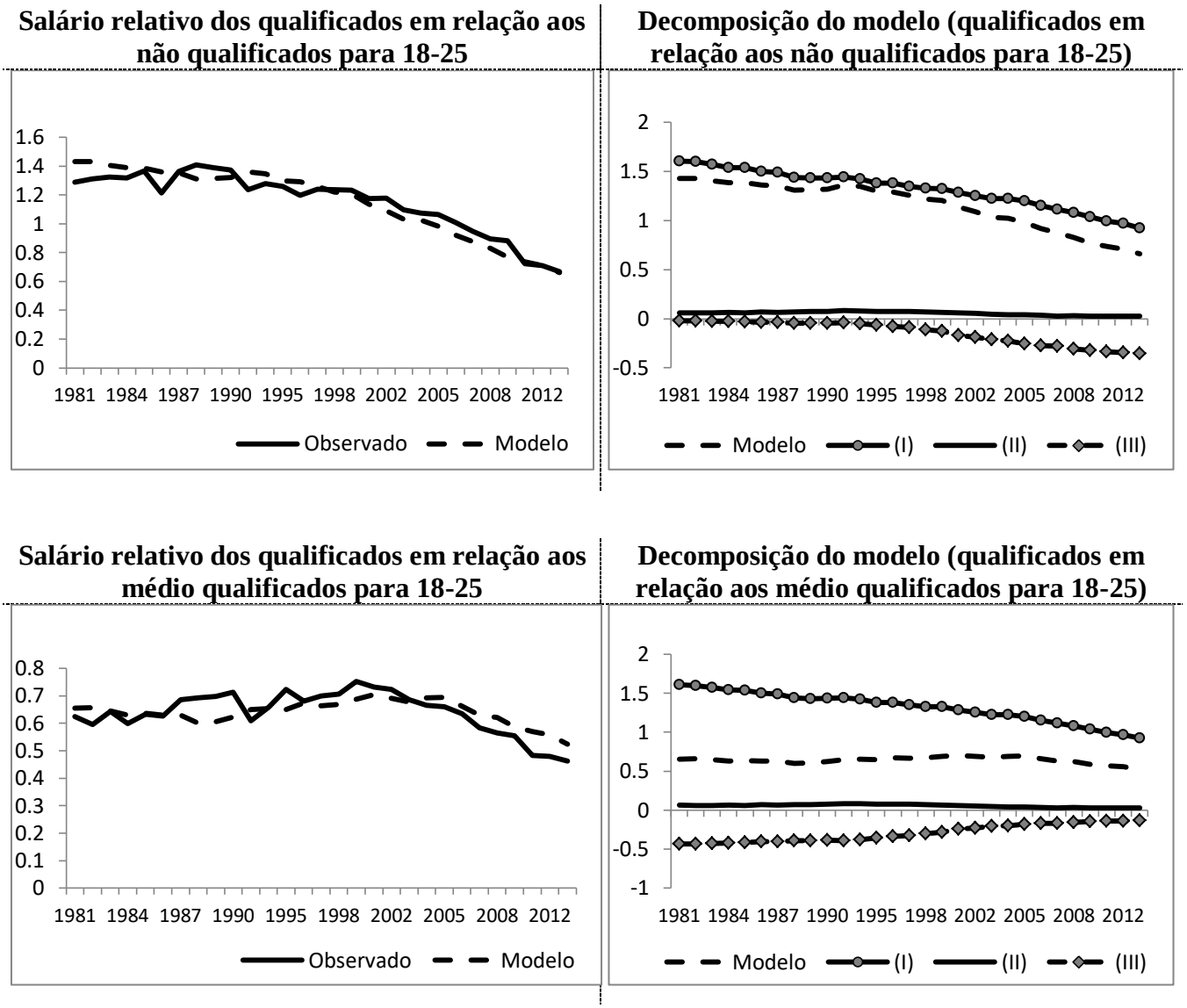
* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

OBS.: Regressão controla por 5 dummies de idade-educação e uma constante

Os Gráficos 3.15 a 3.17 apresentam os resultados da regressão 3 para cada um dos grupos etários 18-25 anos, 26-40 anos e 41-60 anos respectivamente. Em cada gráfico que representa seu grupo etário, a primeira linha apresenta o salário relativo de qualificados em relação a pouco qualificados (w_{jt}^H/w_{jt}^L) de seu grupo etário específico e a segunda linha apresenta o salário relativo de qualificados em relação a médio qualificados (w_{jt}^H/w_{jt}^M) de seu grupo etário. Para todos os grupos etários e todos os grupos de qualificação, o salário relativo observado dos dados da PNAD (linha contínua) e o modelo predito (linha tracejada) são semelhantes, corroborando o alto R^2 ajustado apresentado na regressão 3. Estes resultados são apresentados nos gráficos da esquerda onde o modelo predito é comparado com os dados observados para cada grupo etário e grupo educacional (pouco qualificado e médio qualificado). Os gráficos da direita decompõem o modelo predito em três componentes, enunciados pela equação (3.28) onde os valores de β_3, β_4 e β_5 são utilizados da tabela Regressão 3.

$$\ln\left(\frac{w_{jt}^H}{w_{jt}^I}\right) = \cdots + \underbrace{\beta_3 ln\left(\frac{H_t}{C_t}\right)}_{(I)} + \underbrace{\beta_4\left\{ln\left(\frac{C_t}{C_{jt}}\right) - ln\left(\frac{H_t}{H_{jt}}\right)\right\}}_{(II)} + \underbrace{\beta_5 ln\left(\frac{C_{jt}}{I_{jt}}\right)}_{(III)} + e_{I,j,t} \qquad (3.28)$$

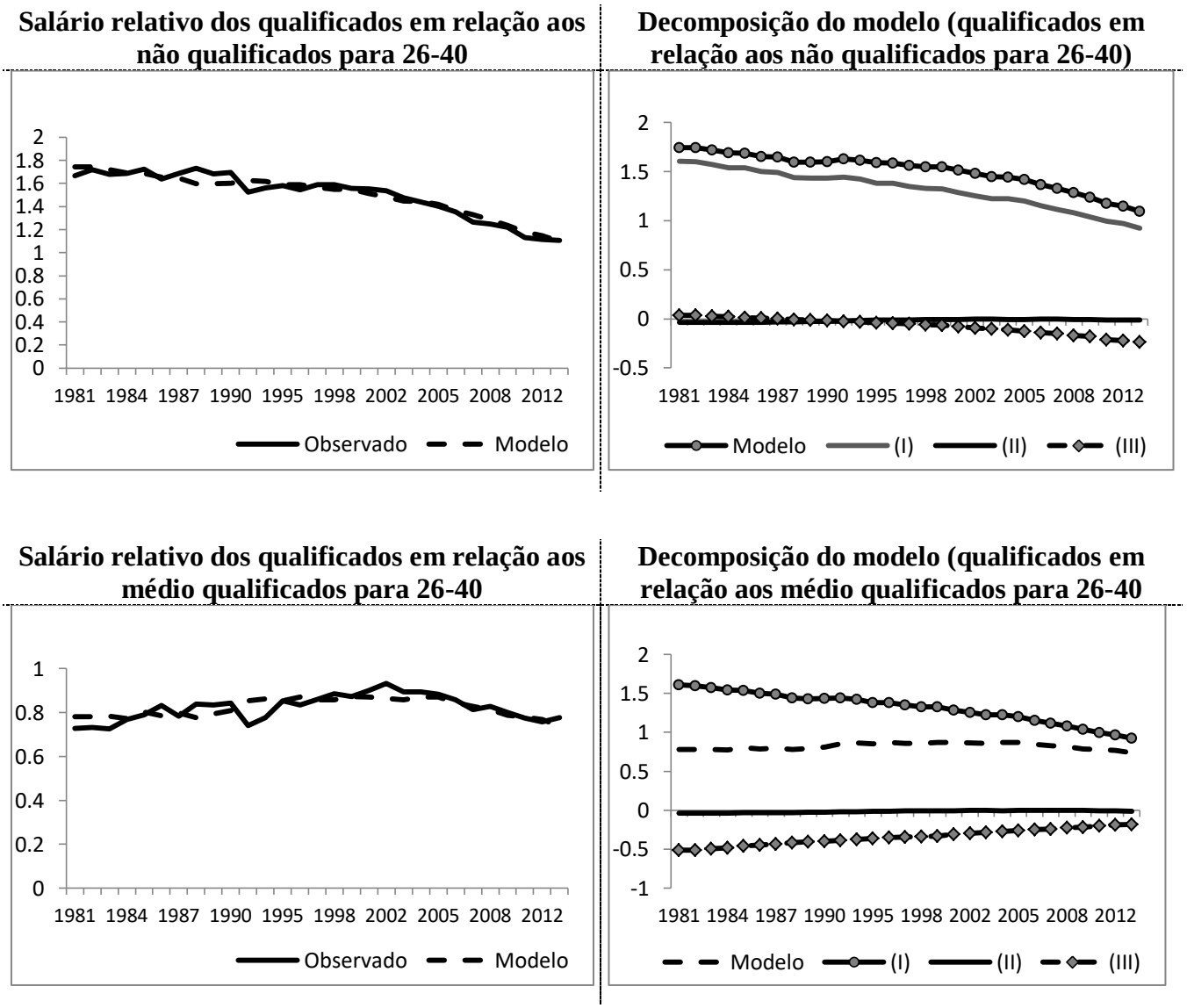
GRÁFICO 3.15: REGRESSÃO 3 - DECOMPOSIÇÃO DA OFERTA RELATIVA E SALÁRIO RELATIVO PARA CADA GRUPO DE ESCOLARIDADE E GRUPO ETÁRIO 18-25



Fonte: Resultados da Regressão 3 com microdados PNAD 1981-2013

Para todos os resultados apresentados nos gráficos (3.15) a (3.17), $\{ln(C_t/C_{jt}) - ln(H_t/H_{jt})\}$ mostrou-se invariante no tempo e por isto ela não explica tendências de queda do salário relativo para cada grupo etário. Em geral, as evoluções dos salários relativos são explicadas pelos componentes (I) e (III). Para o grupo etário 18-25 anos, a diminuição do salário relativo de qualificados em relação a pouco qualificados é explicada por mudanças na oferta relativa do mesmo grupo etário, o componente (III). Este mesmo componente explica a evolução do salário relativo de qualificados em relação a médio qualificados.

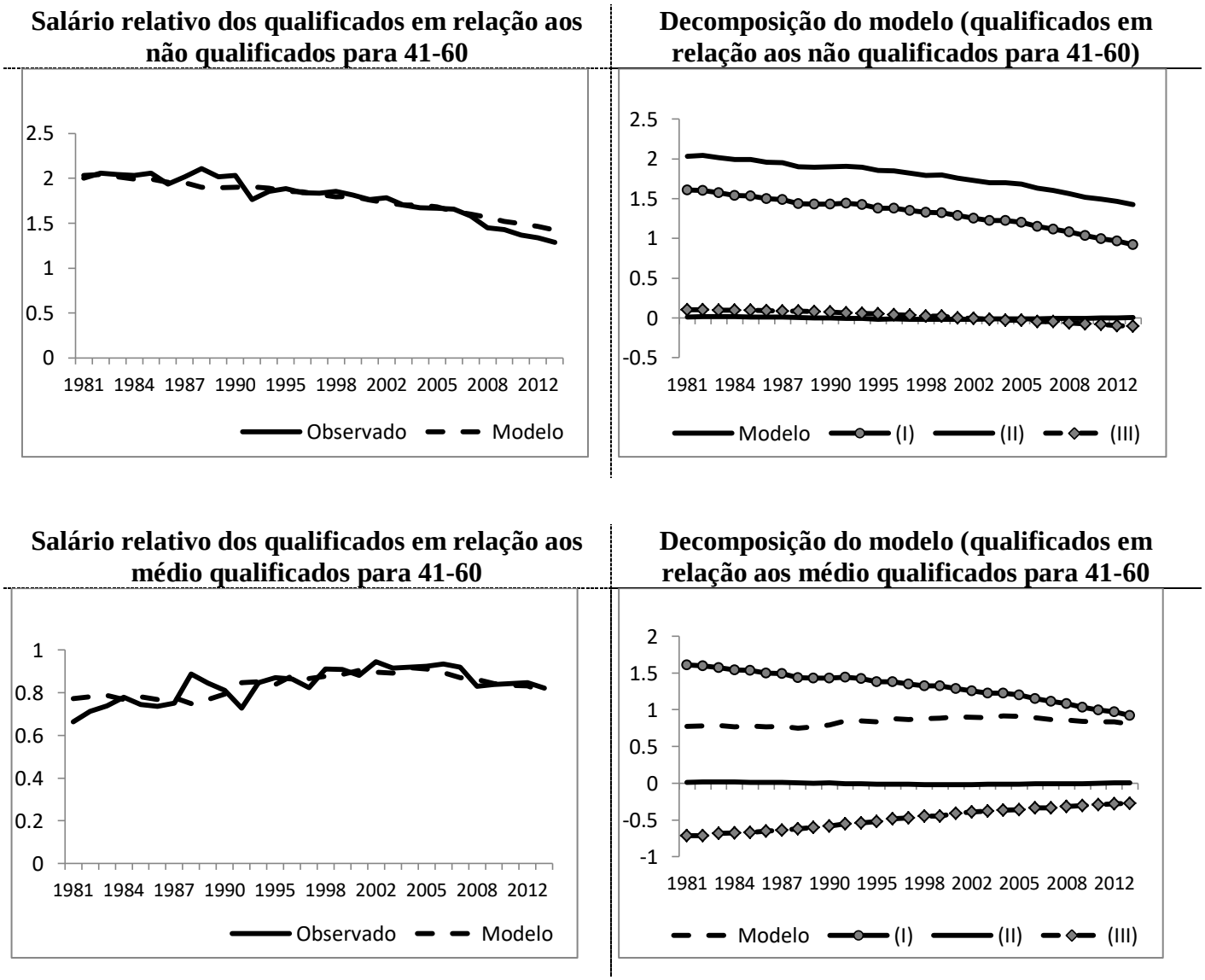
GRÁFICO 3.16: REGRESSÃO 3 - DECOMPOSIÇÃO DA OFERTA RELATIVA E SALÁRIO RELATIVO PARA CADA GRUPO DE ESCOLARIDADE E GRUPO ETÁRIO 26-40



Fonte: Resultados da Regressão 3 com microdados PNAD 1981-2013

Para o grupo etário 26-40 anos, a diminuição do salário relativo de qualificados em relação a pouco qualificados é explicada pelos componentes (I) e (III). Em relação a médio qualificados, gráficos inferiores, a evolução com o aumento do salário relativo é explicada pela evolução do componente (III), ou seja, mudanças na oferta relativa no mesmo grupo etário.

GRÁFICO 3.17: REGRESSÃO 3 - DECOMPOSIÇÃO DA OFERTA RELATIVA E SALÁRIO RELATIVO PARA CADA GRUPO DE ESCOLARIDADE E GRUPO ETÁRIO 41-60



Fonte: Resultados da Regressão 3 com microdados PNAD 1981-2013

A queda do salário relativo dos qualificados em relação aos pouco qualificados para o grupo etário 41-60 anos, como apresentado no Gráfico 3.17, é explicada pela pelo aumento da oferta relativa agregada. Ou seja, mudanças na oferta relativa agregada, muito impactada por pela entrada de coortes jovens com melhores indicadores de capital humano explicam o aumento do salário real de trabalhadores pouco qualificados com idade entre 41 e 60 anos em relação a trabalhadores qualificados pertencentes ao mesmo grupo etário.

Uma vez que as coortes mais jovens entram no mercado de trabalho mais qualificadas, o salário relativo destas coortes diminui. Este processo faz com que a oferta de mão-de-obra pouco qualificada fica mais concentrada com as coortes mais velhas e seu salário real aumenta. Ao mesmo tempo, o Gráfico 3.15 demonstra que a forte queda do salário relativo de trabalhadores qualificados em relação a trabalhadores não qualificados para o grupo etário 18-25 anos foi fortemente impactada pelas mudanças na oferta relativa de C_{jt}/L_{jt} ponderada pela elasticidade de substituição entre trabalhadores médio qualificados e trabalhadores pouco qualificados. O artigo de Fernandes & Menezes-Filho (2012), através de suas simulações, encontraram que, para o período 1981 a 2009, houve um aumento da participação de indivíduos médio qualificados (qualificação intermediária) em atividades simples que, em sua essência, são destinadas a trabalhadores não qualificados.

4. CONTRAFACTUAL

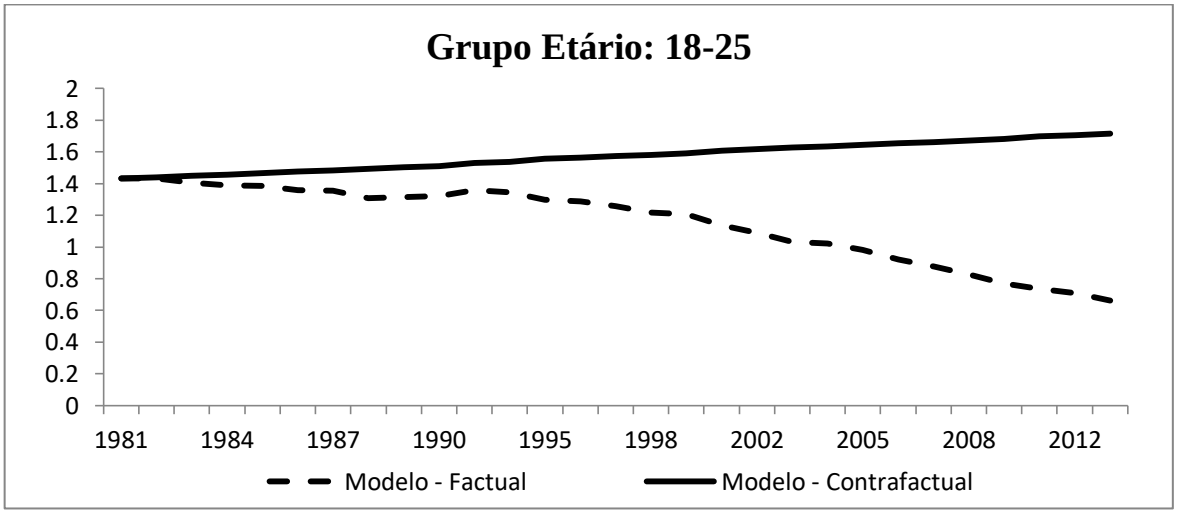
A análise contrafactual procura responder qual seriam os salários relativos, por grupo etário, se ao longo das últimas décadas a oferta relativa não mudasse e permanecesse a valores de 1981, quando 83% dos trabalhadores eram não qualificados no Brasil. Para fazer tal exercício, serão utilizados os modelos preditos apresentados nos Gráficos 3.15, 3.16 e 3.17 e tais modelos ajustam-se aos dados observados e tais resultados serão denominados "factuais", preditos a partir do modelo empírico[95]. O contrafactual será representado pela equação (3.28) com as ofertas relativas a valores de 1981, o que implica que o único termo que varia no tempo é a tendência, t.

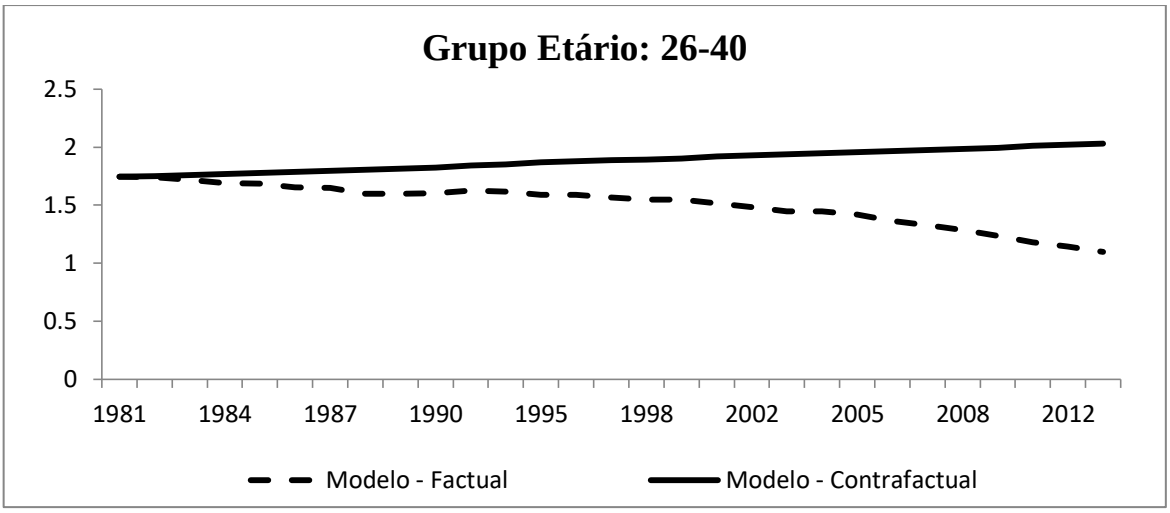
CONTRAFACTUAL

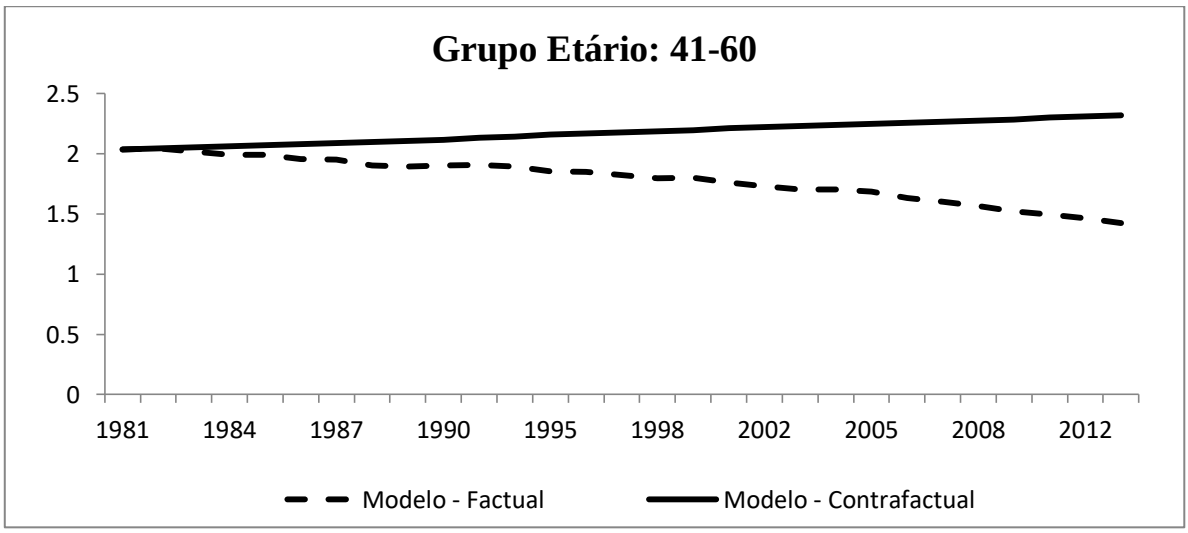
$$\ln\left(\frac{\widetilde{w}_{jt}^{H}}{\widetilde{w}_{jt}^{M}}\right) \approx \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 dummy_{I,j} + \hat{\beta}_3 ln\left(\frac{H_{1981}}{C_{1981}}\right) + \hat{\beta}_4\left\{ln\left(\frac{C_{1981}}{C_{j1981}}\right) - ln\left(\frac{H_{1981}}{H_{j1981}}\right)\right\} + \hat{\beta}_5 ln\left(\frac{C_{j1981}}{M_{j1981}}\right) = \ln\left(\widetilde{\frac{w_j^H}{w_{jt}^M}}\right)_c \quad (3.28)$$

[95] Os modelos preditos são resultados da aplicação da equação (3.27) com os parâmetros estimados na Regressão 3 e as ofertas relativas por grupo etário observadas nos dados.

GRÁFICO 3.18: ANALISE CONTRAFACTUAL PARA O SALÁRIO RELATIVO SUPERIOR-FUNDAMENTAL POR GRUPOS ETÁRIOS



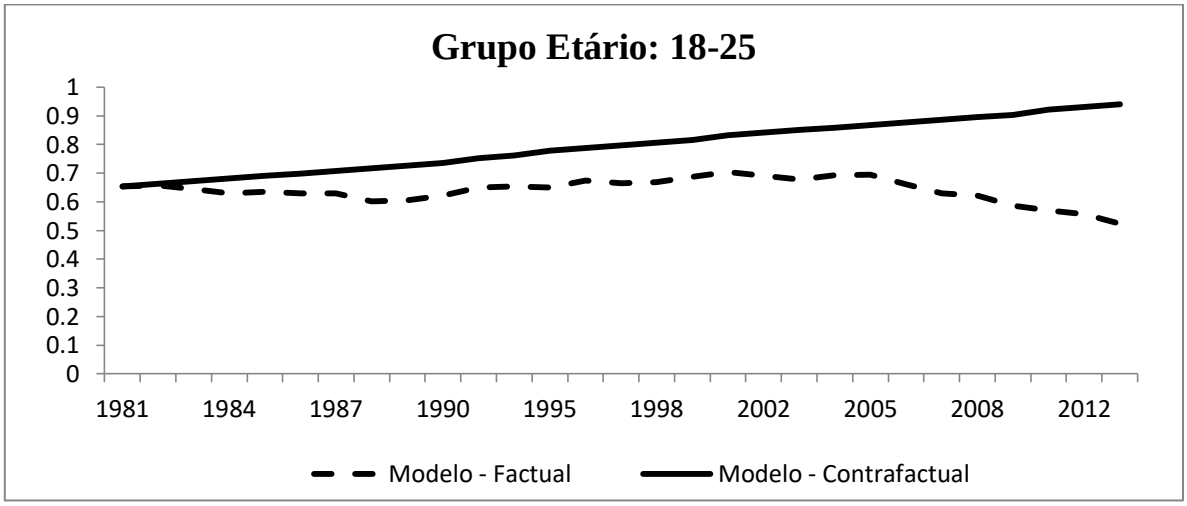


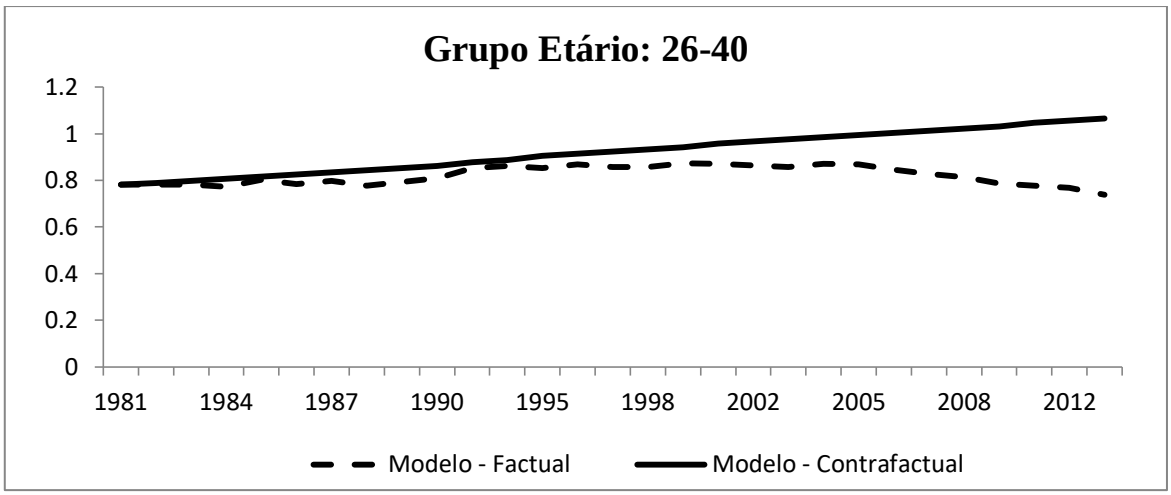


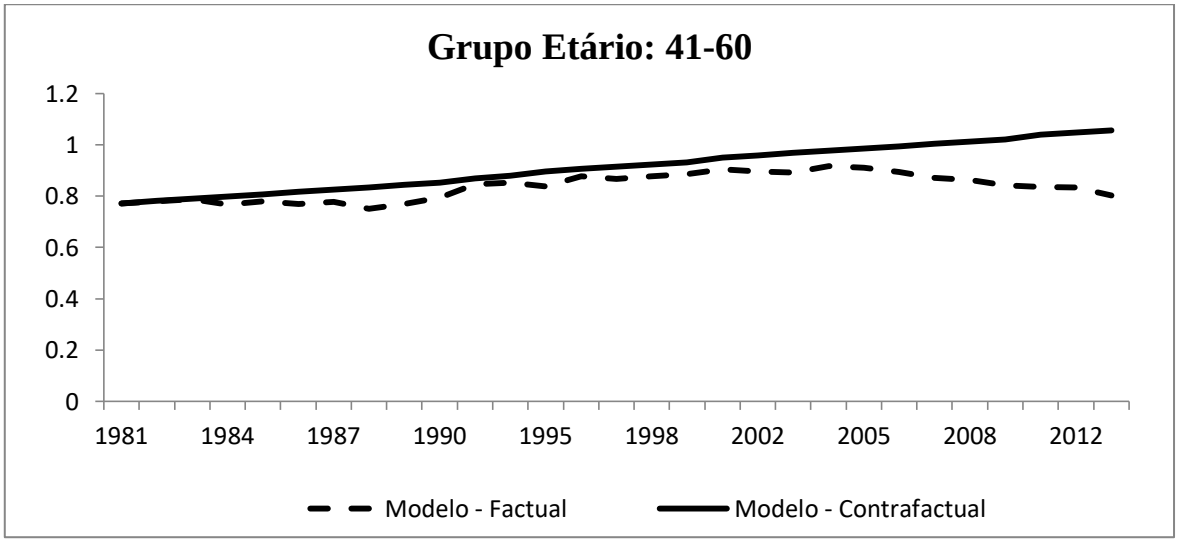
OBS. Resultados gerados a partir dos parâmetros estimados na Regressão 3

Os Gráficos 3.18 e 3.19 apresentam os resultados comparativos entre o factual, ajuste do modelo aos dados observados pela PNAD, e o contrafactual, previsão do modelo caso as ofertas relativas por grupo etário se mantivessem constantes a valores de 1981. Os resultados apresentados no Gráfico 3.18 referem-se ao logaritmo do salário relativo de qualificados (Ensino Superior) em relação a não qualificados (Ensino Fundamental) por grupo etário (18-25 anos, 26-40 anos e 41-60 anos). Os salários relativos observados caem para os grupos etários 18-25 e 26-40 por conta do aumento da oferta relativa de qualificados e por conta do aumento da oferta relativa de trabalhadores com Ensino Médio em relação a trabalhadores com ensino Fundamental. Para o grupo etário 41-60 anos, apenas mudanças na oferta relativa de qualificados (Superior) em relação a não qualificados (Ensino Médio e Ensino Fundamental agregados) explicam mudanças no salário relativo. Caso não ocorressem mudanças nas ofertas relativas, o modelo contrafactual projetaria um aumento do salário relativo de ensino superior em relação a ensino médio. Comparando os grupos etários tanto no salário relativo do ensino superior em relação ao ensino fundamental quanto no salário relativo de ensino superior em relação a ensino médio, a diferença entre o modelo factual e o modelo contrafactual é maior para o grupo etário de 18 a 25 anos e menor para o grupo etário de 41 a 60 anos.

GRÁFICO 3.19: ANALISE CONTRAFACTUAL PARA O SALÁRIO RELATIVO SUPERIOR-MÉDIO POR GRUPOS ETÁRIOS







OBS. Resultados gerados a partir dos parâmetros estimados na Regressão 3

5. CONCLUSÃO

Este artigo investigou como as mudanças educacionais no mercado de trabalho brasileiro, entre 1981 e 2013, impactaram os salários relativos e como este evento guarda uma relação com a desigualdade de renda analisada a partir do coeficiente de Gini. No começo da década de 1980, os trabalhadores tinham, em sua maioria, pouca qualificação, com uma média de 4,7 anos de estudo e isto se deveu às dificuldades encontradas tanto no acesso à educação quanto para a permanência, dado a alta defasagem escolar apresentada no Brasil em décadas passadas. Com a melhora nos indicadores de acesso e na queda da defasagem escolar, as coortes mais jovens acumularam mais capital humano e ampliaram tanto a oferta de trabalho médio qualificado, ou seja, com ensino médio, quanto muito qualificado, ou seja, com ensino superior.

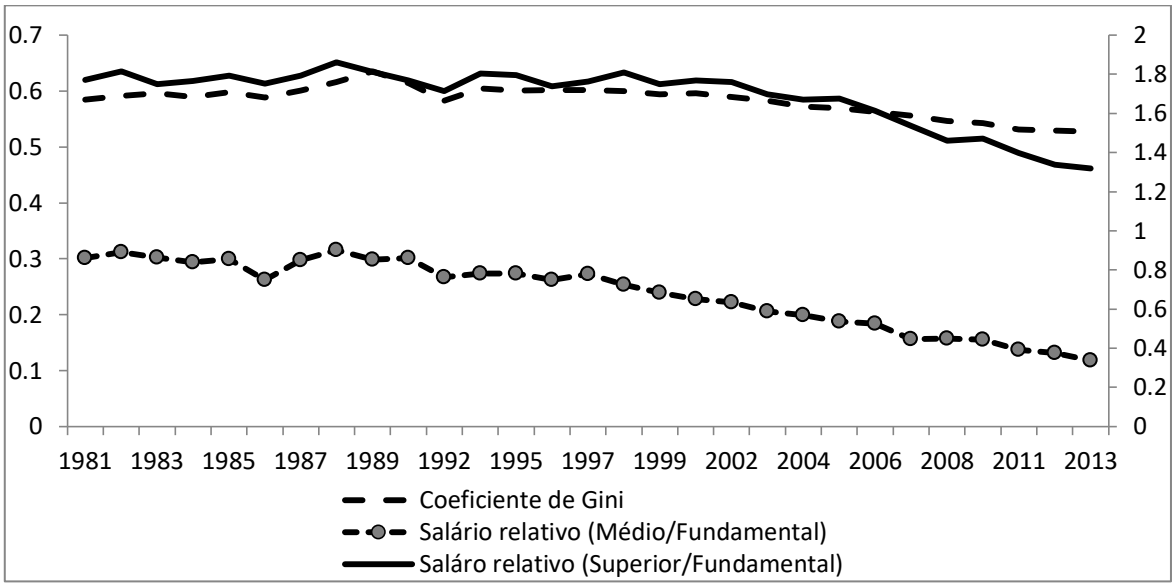
A aplicação da metodologia de Katz & Murphy e Acemoglu & Autor aos dados brasileiros de mercado de trabalho mostraram que, em termos agregados, mudanças na oferta relativa conduziram a mudanças no salário relativo. Entre 1981 e 2013 houve um aumento da oferta relativa de trabalhadores médio qualificados em relação a trabalhadores pouco qualificados e isto se deve pela ampliação da participação do número de trabalhadores com Ensino Médio completo e pela queda da participação do número de trabalhadores com Ensino Fundamental incompleto. Este aumento da oferta relativa impactou numa queda do salário relativo destes grupos de qualificação e, uma vez eliminada a tendência da série, o modelo demostra que a oferta relativa é negativamente relacionada com o salário relativo e uma queda de 1% na oferta relativa de trabalhadores médio qualificados em relação a trabalhadores pouco qualificados provoca um aumento de 0,55% no salário relativos destes. Entre 1981 e 2005 houve uma queda da oferta relativa de trabalhadores qualificados em relação a trabalhadores médio qualificados e este fato deve-se à maior expansão do Ensino Médio no período, uma vez que tanto a proporção de trabalhadores com ensino médio quanto a proporção de trabalhadores com ensino superior cresceram ao longo das três décadas. A partir de 2006, a proporção de trabalhadores com superior cresce mais que a proporção de trabalhadores com ensino médio e a oferta relativa de qualificados em relação a médio qualificados aumenta, explicado pela expansão do ensino superior ao

longo da década de 2000. A reversão com o aumento da oferta relativa de qualificados em relação a médio qualificados é acompanhada por uma queda do salário relativo destes. A aplicação deste exercício empírico por grupos etários gerou resultados condizentes com as hipóteses apresentadas por Acemoglu &Autor para grupos etários mais jovens, mas o modelo não previu o comportamento para o grupo etário de 41 a 60 anos por exemplo.

Um modelo com substituição entre grupos etários com mesma qualificação, seguindo a tradição da literatura de Card & Lemieux, foi estimado para o Brasil. Segundo os resultados, o salário relativo de qualificados em relação a não qualificados (não qualificados definidos por ensino médio e ensino fundamental) é explicado pelas mudanças relativas da oferta do grupo etário para as coortes mais jovens. Para a coorte de 41 a 60 anos, mudanças no salário relativo de qualificados em relação a pouco qualificados (ensino fundamental) são explicados por mudanças na oferta relativa agregada e não por mudanças na oferta relativa da própria coorte. Com o aumento da oferta qualificada das coortes mais jovens, para demanda por trabalho pouco qualificado, relativo ao ensino fundamental, terá agora uma oferta de trabalhadores pouco qualificados menor, em termos relativos, composta por trabalhadores de coortes mais velhas e representados em cada ponto do tempo pelo grupo etário 41 a 60 anos. As simulações contrafactuais do modelo, assumindo que a oferta relativa era invariante e assumia os valores de 1981 para toda a série, demonstrou que o salário relativo de qualificados em relação a não qualificados (ensino médio e ensino fundamental) apresentaria trajetória crescente.

APÊNDICE A: OUTRAS ESTATÍSTICAS DESCRITIVAS E RESULTADOS EMPÍRICOS

GRÁFICO 3A.1: EVOLUÇÃO DO COEFICIENTE DE GINI E DOS SALÁRIOS RELATIVOS ENTRE 1981 E 2013



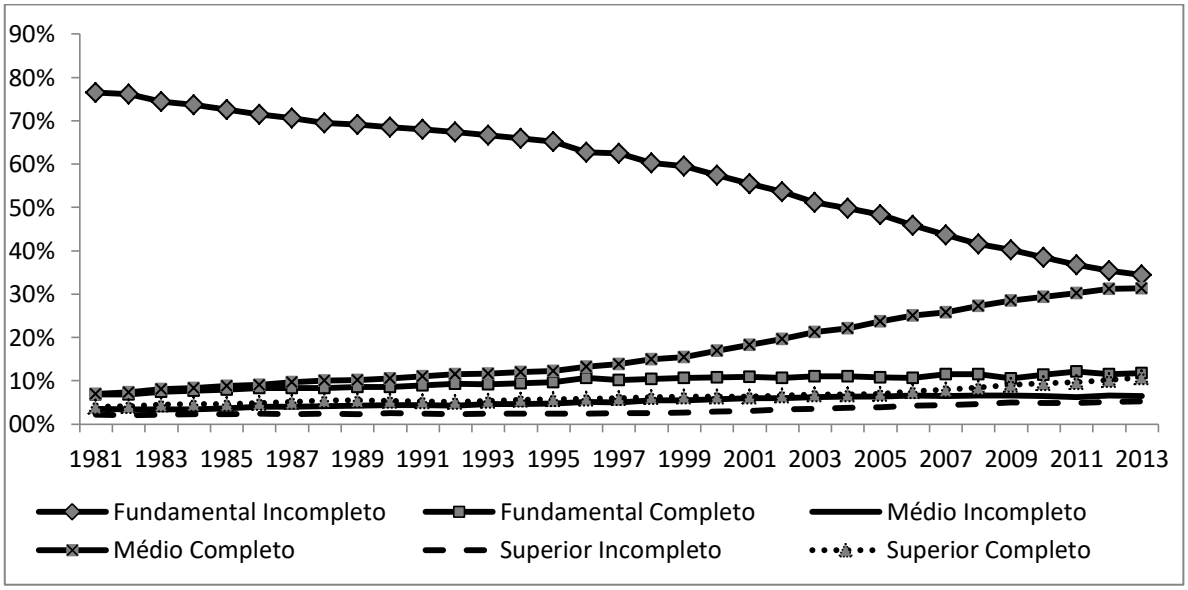
Fonte Salários Relativos: Microdados PNAD 1981-2013 para indivíduos de 18 a 60 anos que trabalhavam 20h ou mais por semana

Fonte Coef. Gini: PNAD/IPEA

OBS: Coeficiene de Gini é representado no eixo da esquerda e seu valor varia de 0, quando não há desigualdade (indivíduos teriam a mesma renda) e 1 quando o grau de desigualdade é máximo (toda a renda estaria concentrada em apenas um indivíduo e os outros não teriam renda)

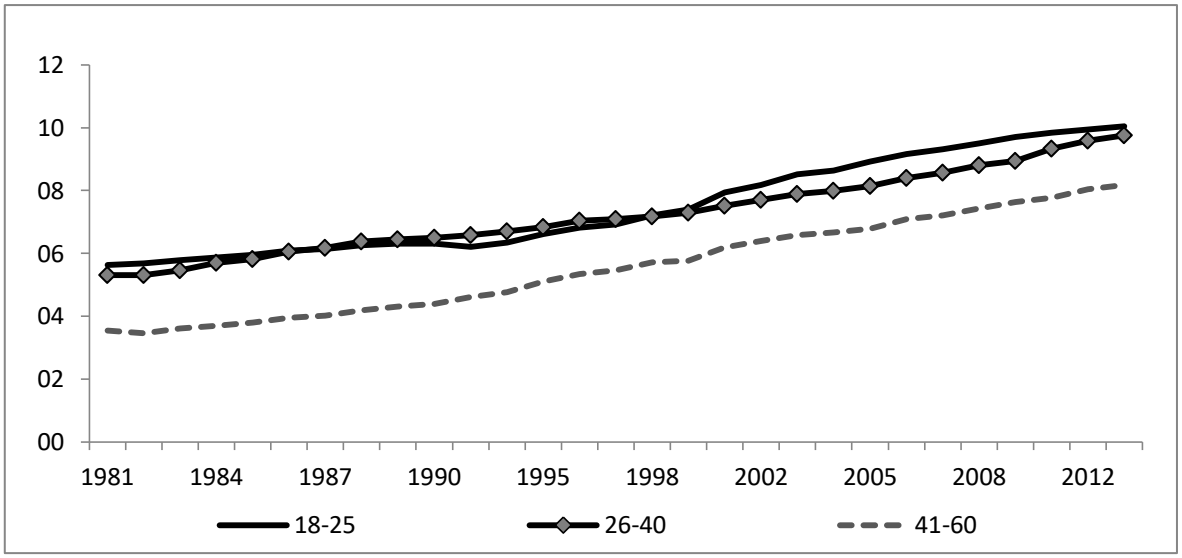
OBS 2: O Salário relativo é apresentado no eixo da direita e representa o logaritmo da razão salário médio agregado de um grupo de escolaridade (Ensino Médio ou Ensino Superior) pelo salário médio agregado de indivíduos com apenas o Ensino Fundamental (0 a 8 anos de estudo)

GRÁFICO 3A2: PERCENTUAL DE TRABALHADORES POR GRUPOS DE ESCOLARIDADE



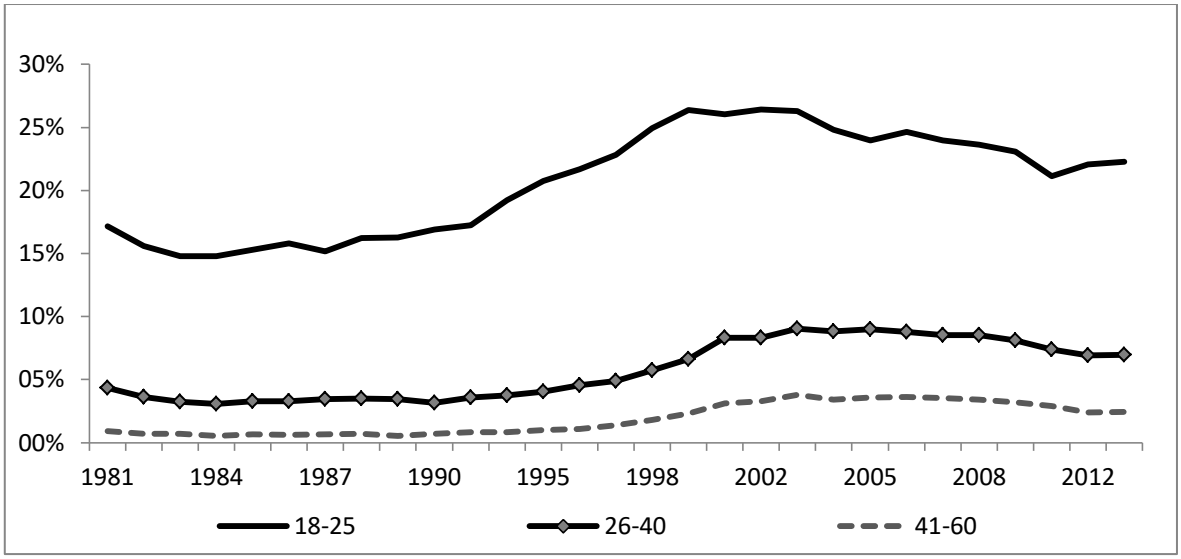
Fonte: Microdados PNAD 1981-2013

GRÁFICO 3A3: ANOS DE ESTUDO POR GRUPOS ETÁRIOS



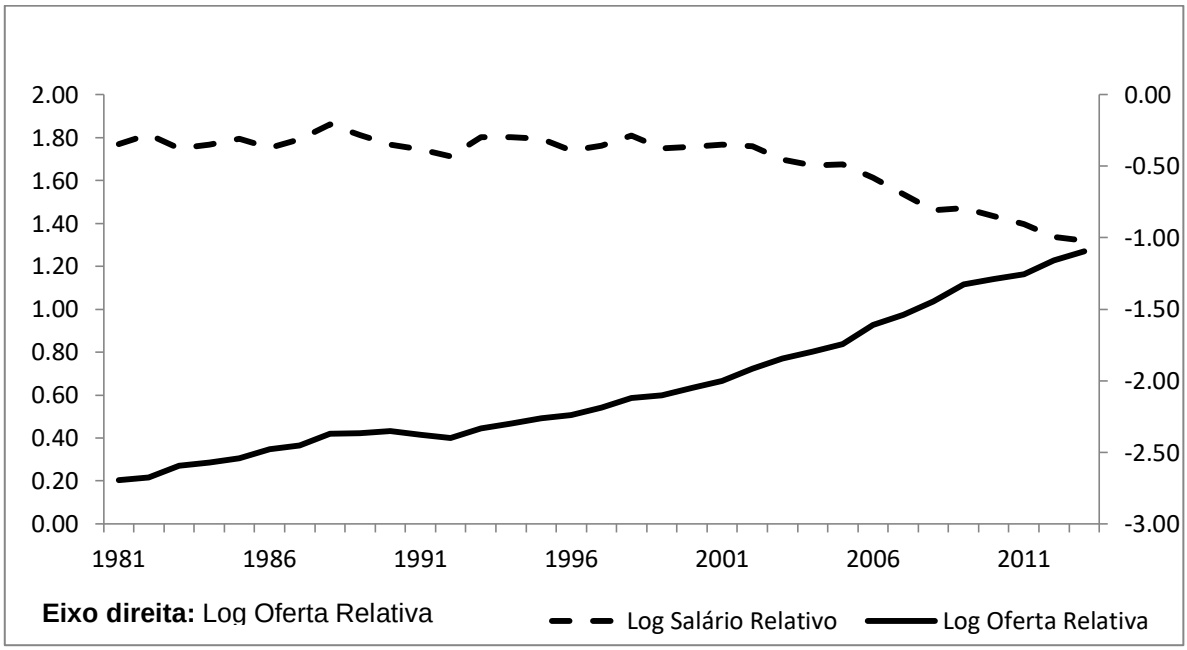
Fonte: Microdados PNAD 1981-2013

GRÁFICO 3A4: PERCENTUAL QUE TRABALHA E ESTUDA POR GRUPO ETÁRIO



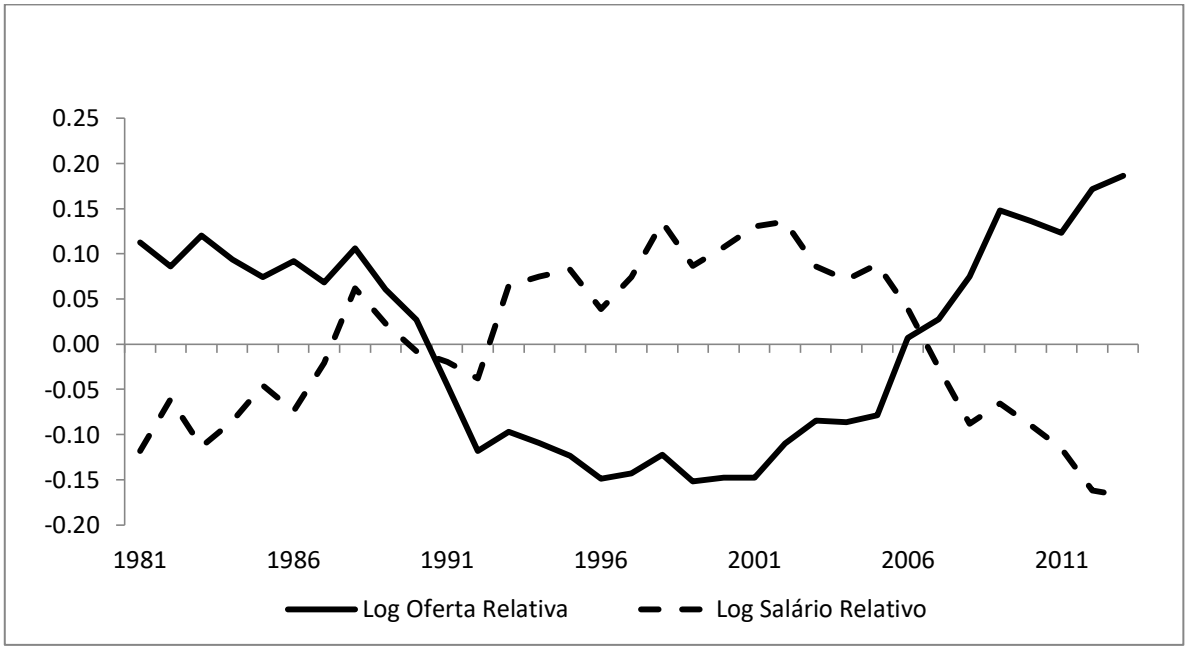
Fonte: Microdados PNAD 1981-2013

GRÁFICO 3A5: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO SUPERIOR – ENSINO FUNDAMENTAL



Fonte: Microdados PNAD 1981-2013
OBS: Ensino Superior equivale de 12 a17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.

GRÁFICO 3A6: OFERTA RELATIVA E SALÁRIO RELATIVO: ENSINO SUPERIOR – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



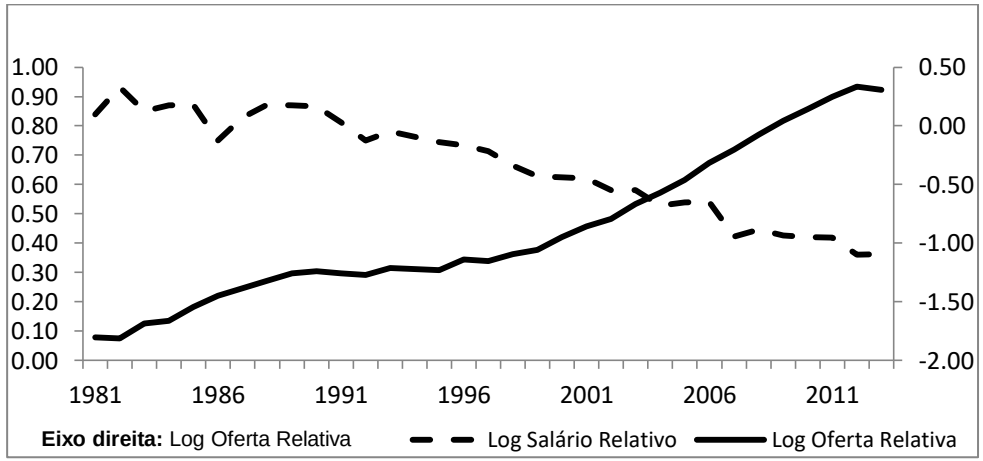
Fonte: Microdados PNAD 1981-2013
OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo.
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

$$\ln\left(\frac{w_H}{w_M}\right) = \underset{(0.2090)}{-\ 0.0467} + \underset{(0.0036)}{0.0199}\ xt + \underset{(0.073)}{-\ 0.6820}\ \ln\left(\frac{H}{M}\right) \qquad \sigma = 1{,}47$$

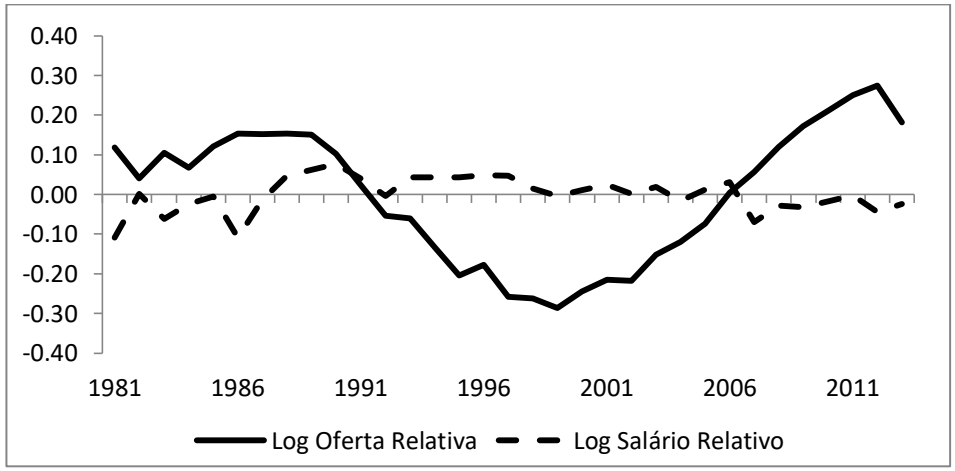
$R^2 = 0{,}90$ Nº obs. = 33

GRÁFICO 3A7: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **26 A 30 ANOS**

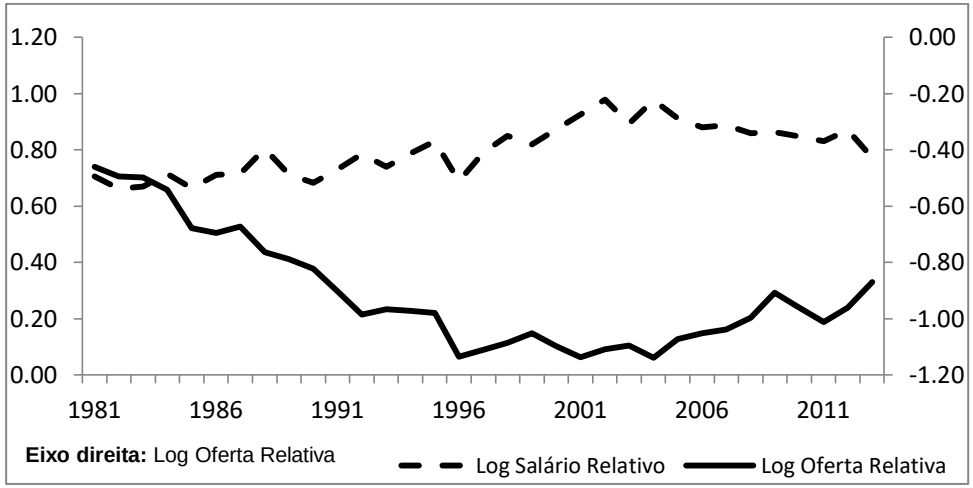
ENSINO MÉDIO – ENSINO FUNDAMENTAL



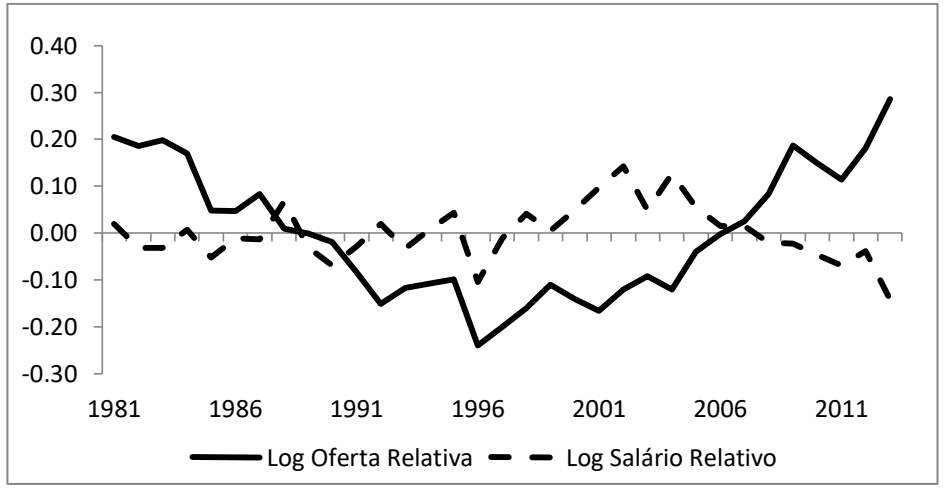
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



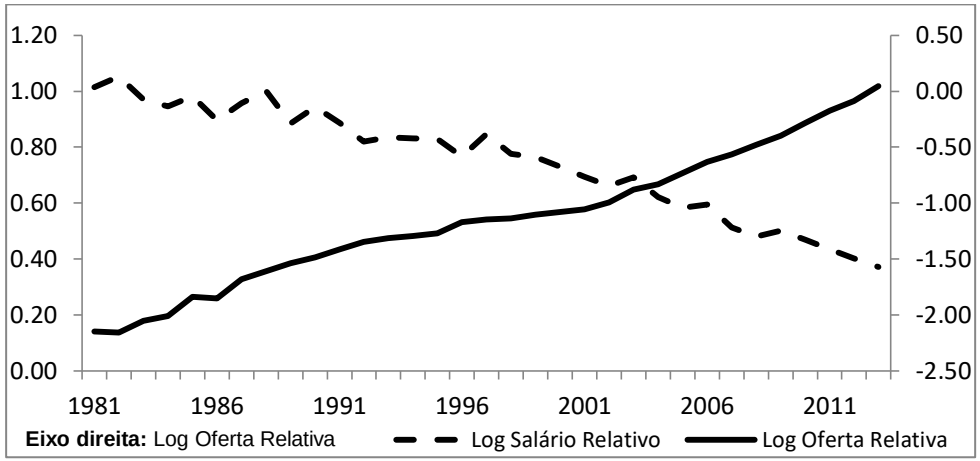
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

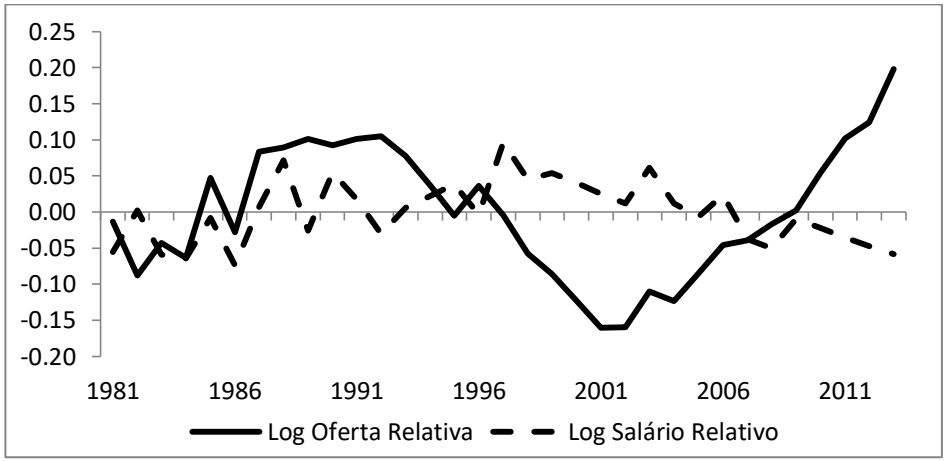
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A8: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **31 A 35 ANOS**

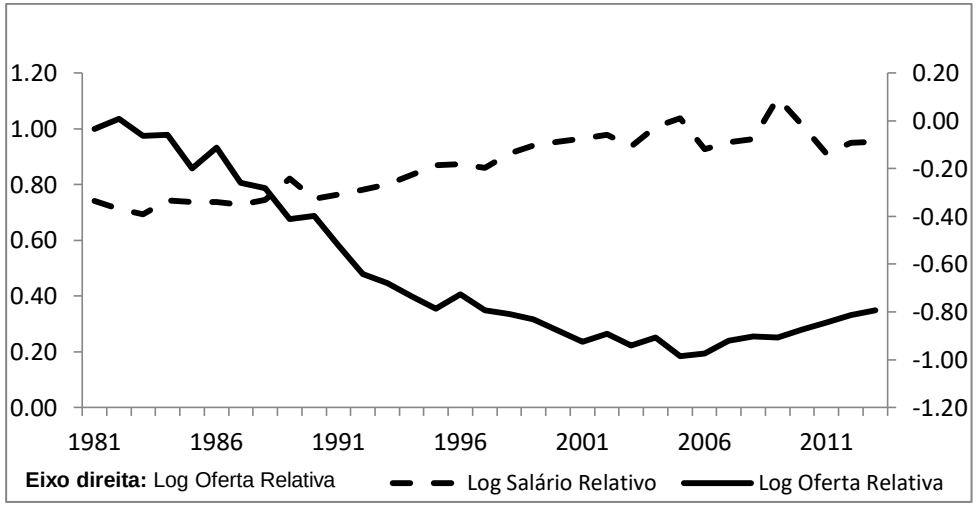
ENSINO MÉDIO – ENSINO FUNDAMENTAL



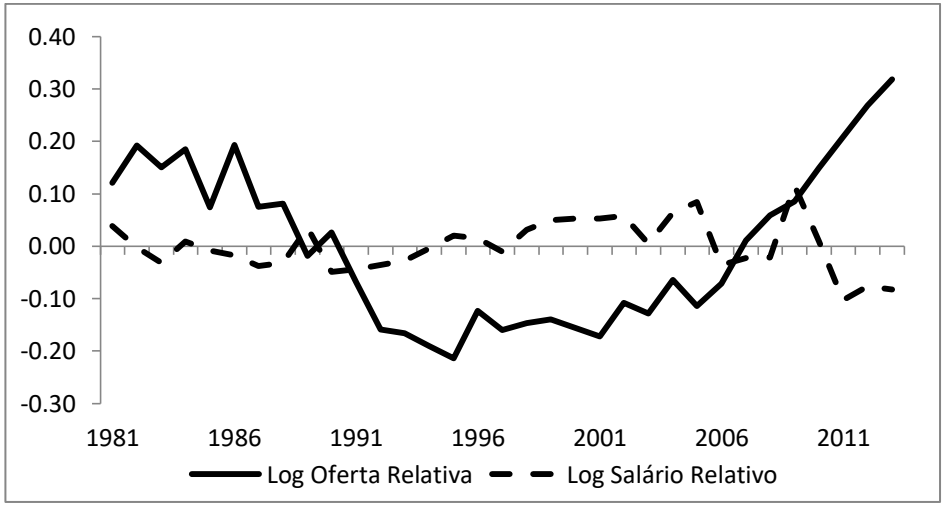
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



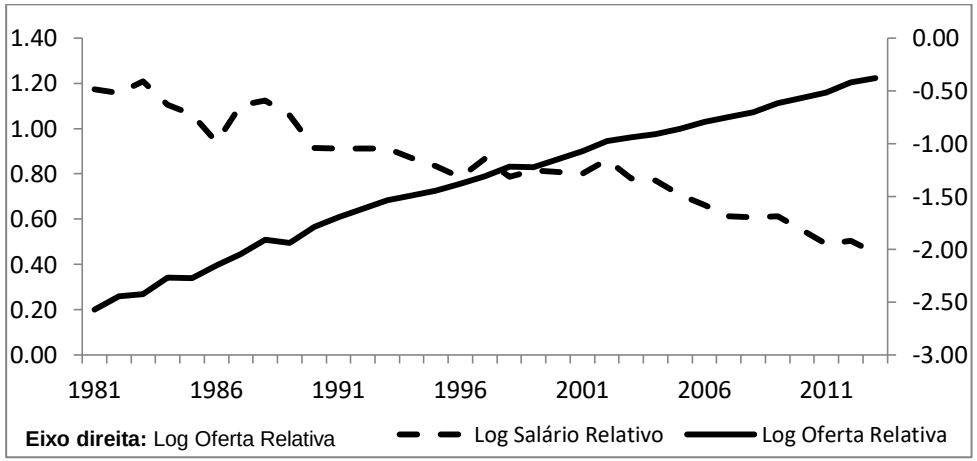
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

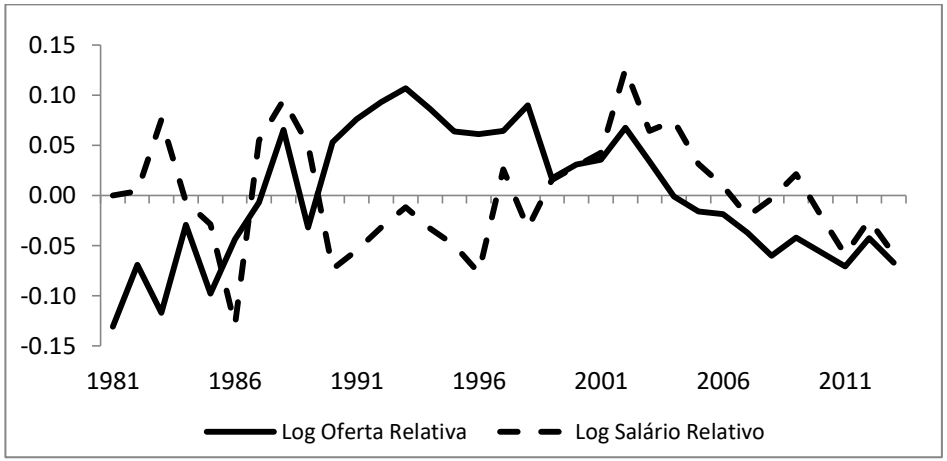
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A9: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **36 A 40 ANOS**

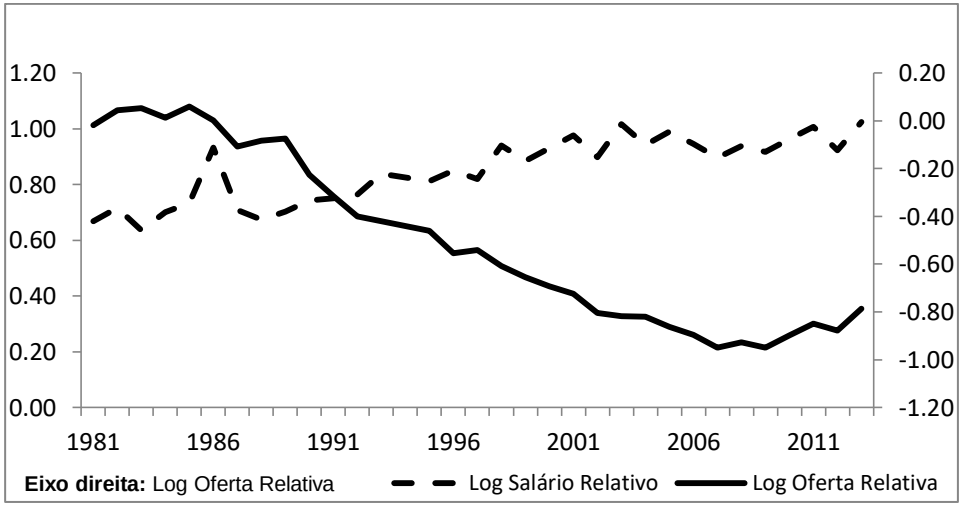
ENSINO MÉDIO – ENSINO FUNDAMENTAL



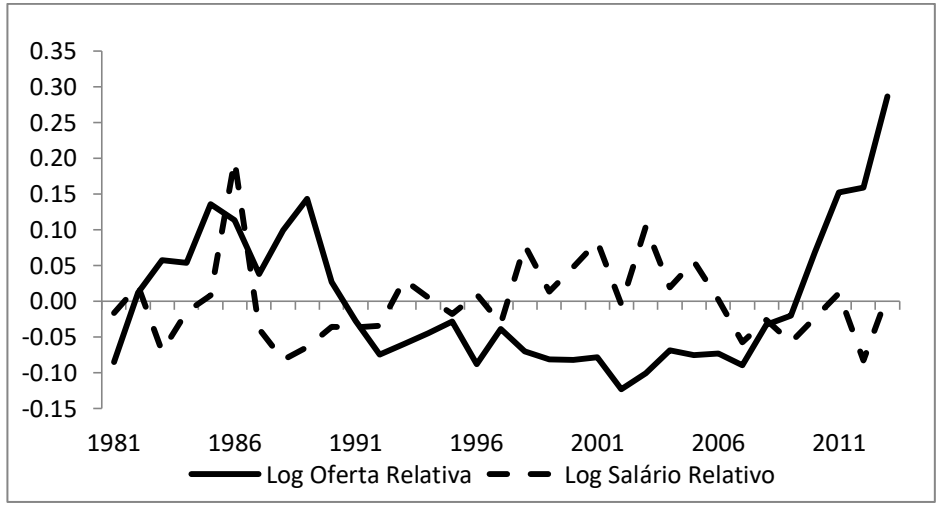
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



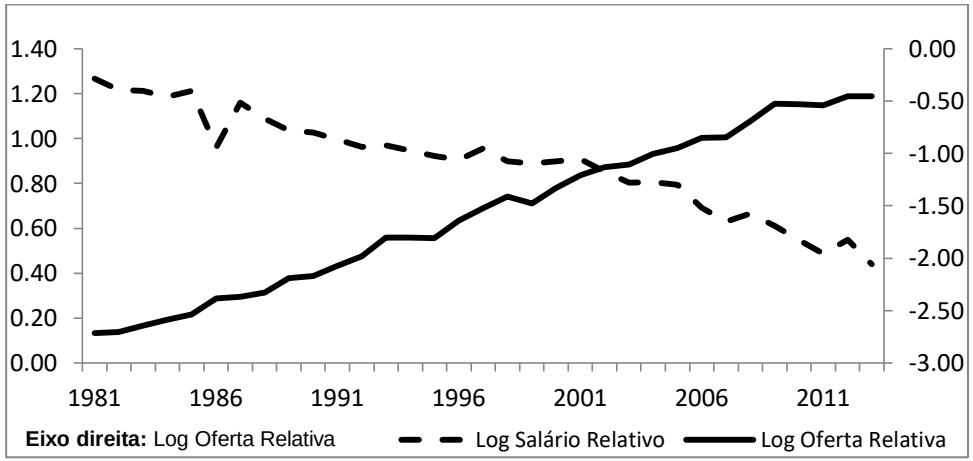
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

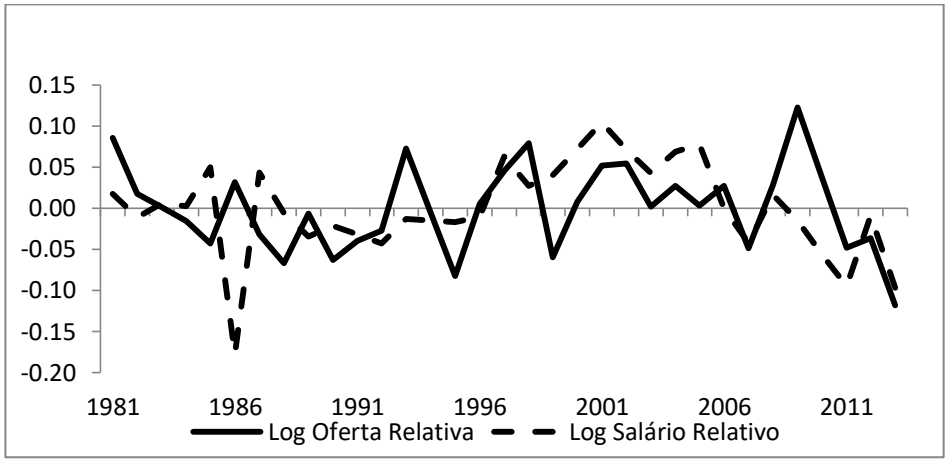
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A10: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **41 A 45 ANOS**

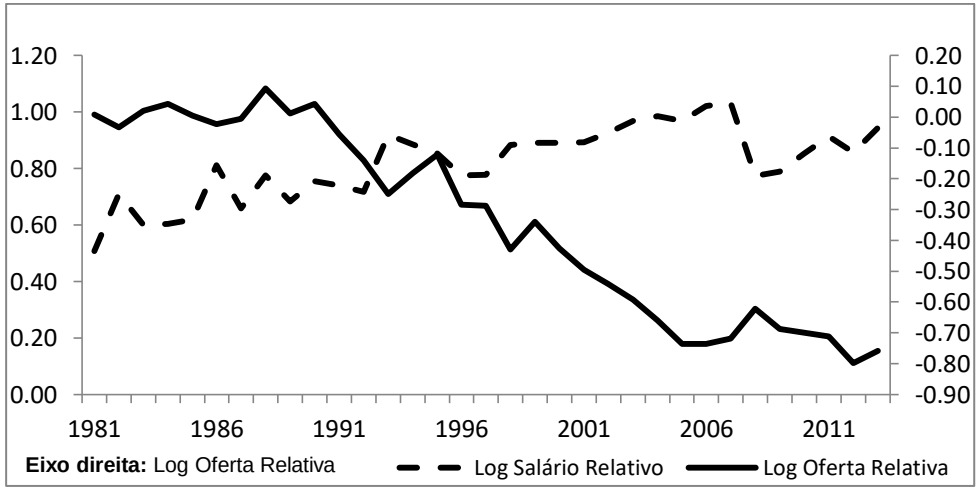
ENSINO MÉDIO – ENSINO FUNDAMENTAL



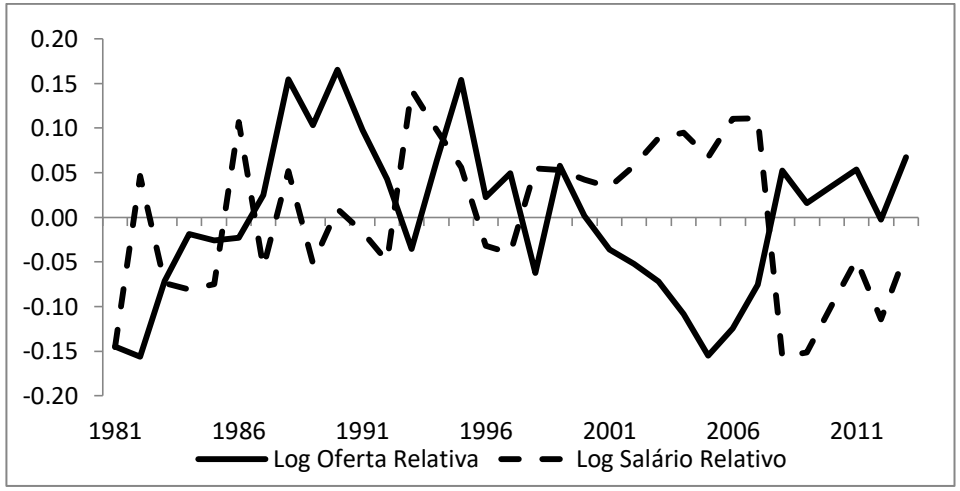
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



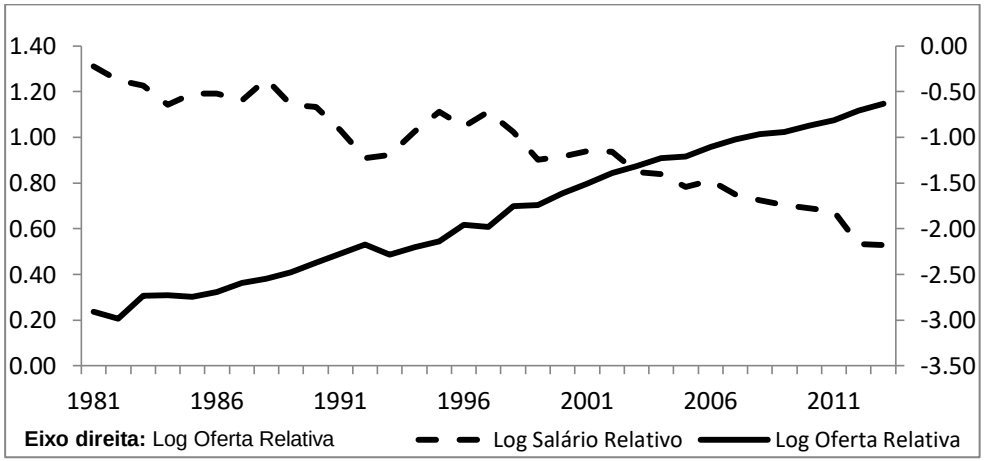
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

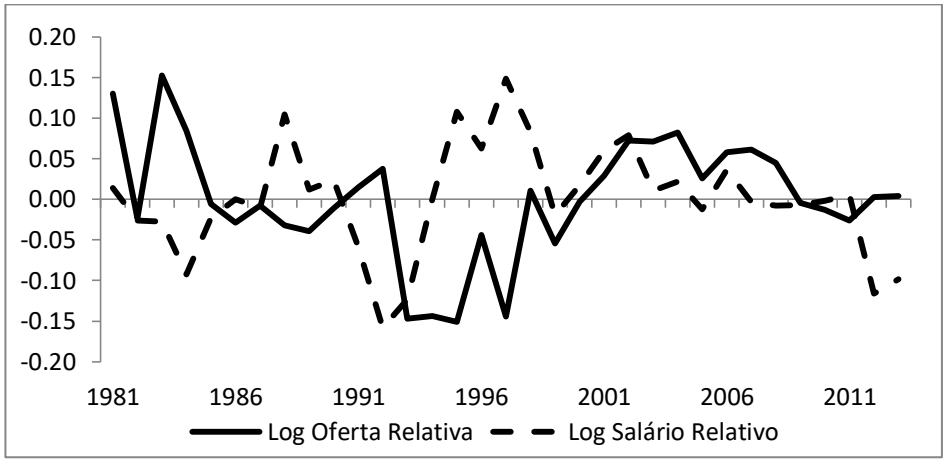
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A11: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **46 A 50 ANOS**

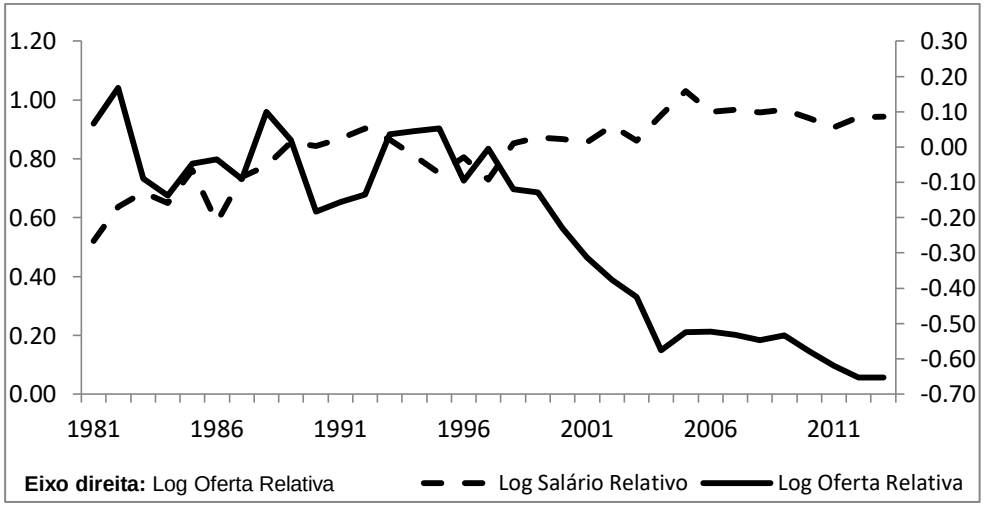
ENSINO MÉDIO – ENSINO FUNDAMENTAL



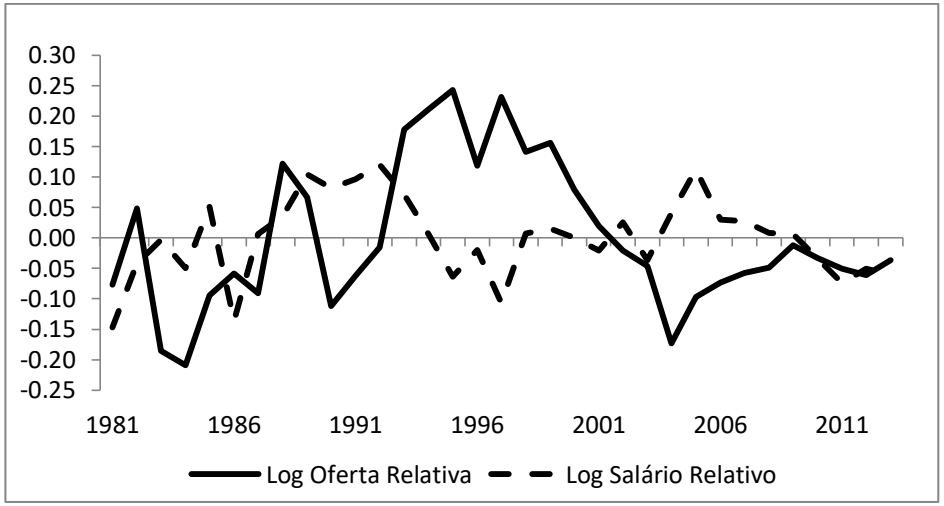
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



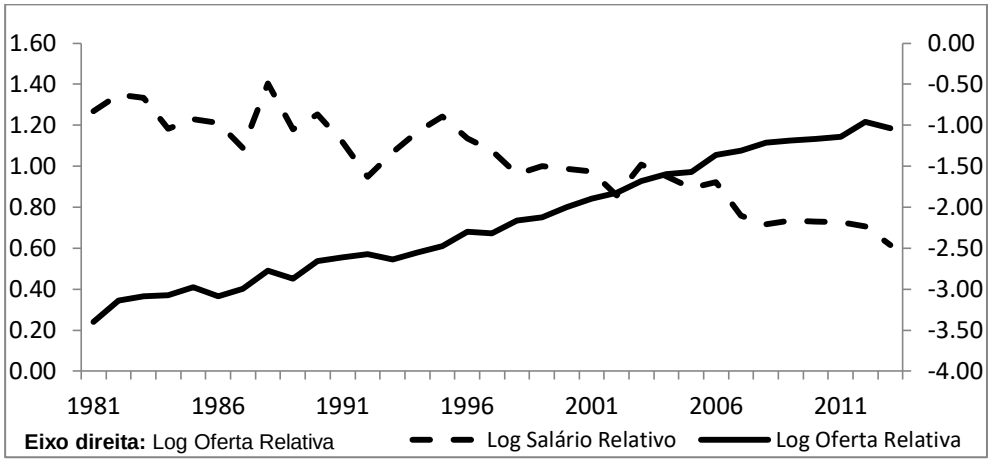
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

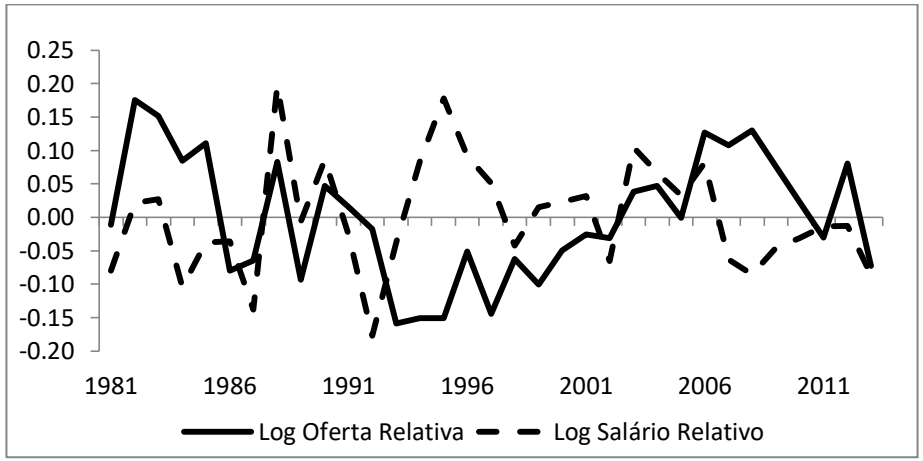
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A12: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **51 A 55 ANOS**

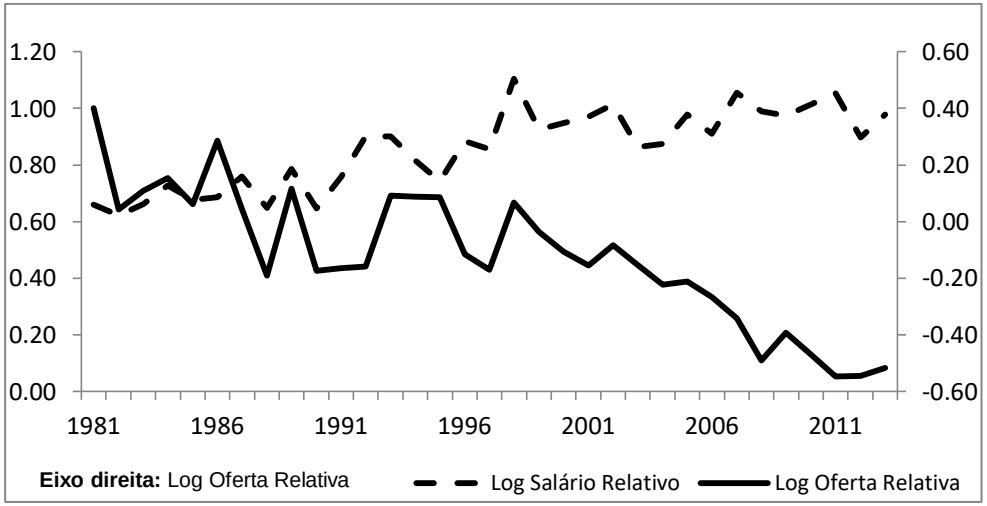
ENSINO MÉDIO – ENSINO FUNDAMENTAL



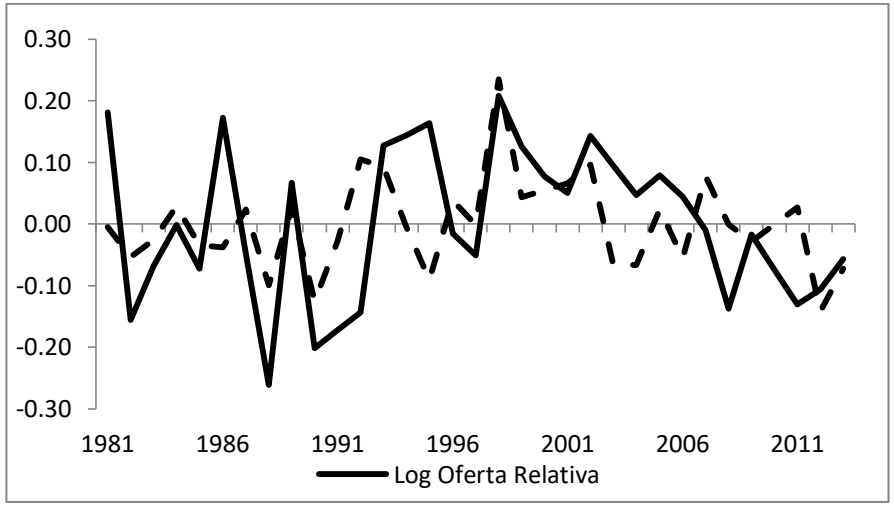
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



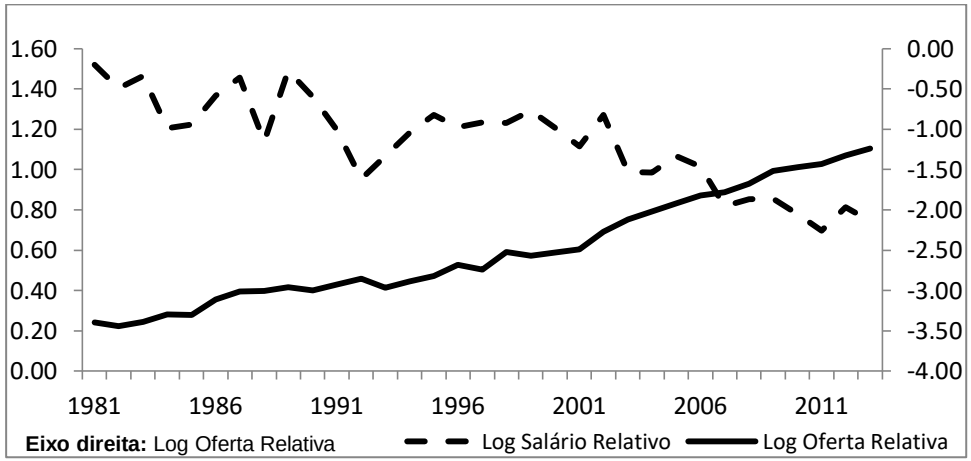
Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

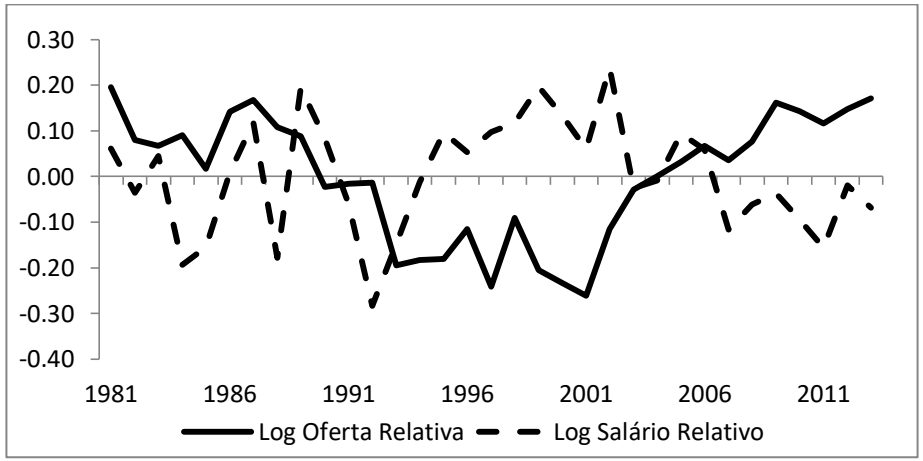
***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

GRÁFICO 3A13: OFERTA RELATIVA E SALÁRIO RELATIVO PARA O GRUPO ETÁRIO **56 A 60 ANOS**

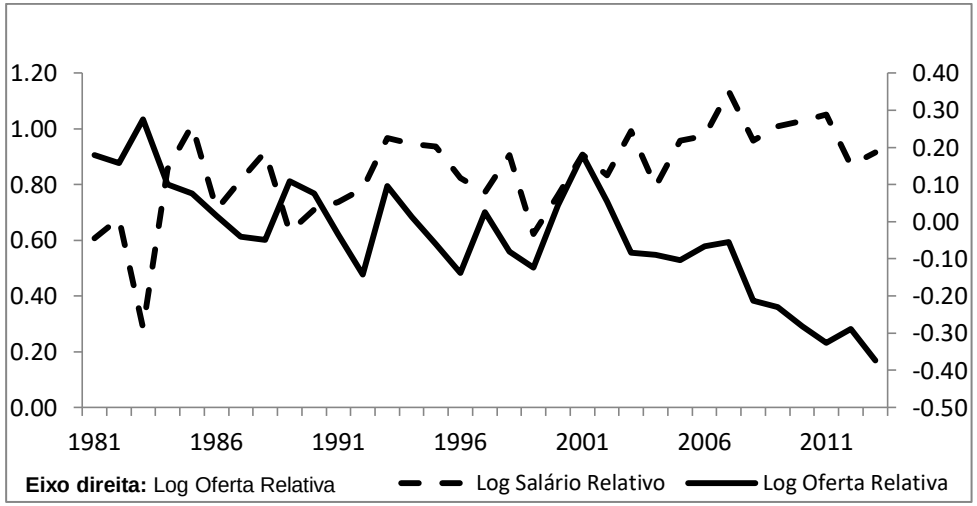
ENSINO MÉDIO – ENSINO FUNDAMENTAL



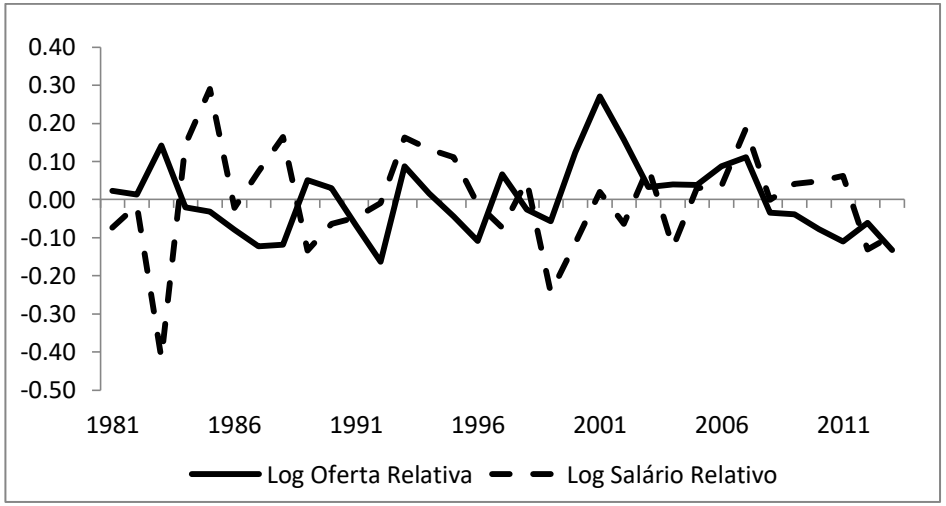
ENSINO MÉDIO – ENSINO FUNDAMENTAL SEM TENDÊNCIA TEMPORAL*



ENSINO SUPERIOR – ENSINO MÉDIO



ENSINO SUPERIOR – ENSINO MÉDIO SEM TENDÊNCIA TEMPORAL*



Fonte: Microdados PNAD 1981-2013

OBS: Ensino Superior equivale de 12 a 17 anos de estudo. **Ensino Médio** equivale de 9 a 11 anos de estudo. **Ensino Fundamental** equivale de 0 a 8 anos de estudo.

***:** As séries de oferta relativa e salário relativo foram regredidas, por MQO, usando uma constante e uma tendência temporal como especificação. Os resultados apresentados no gráfico são os resíduos destas duas regressões.

REGRESSÕES DA EQUAÇÃO (3.4) POR GRUPOS ETÁRIOS

	18-25		26-40		41-60	
	Ensino Médio-Ensino Fundamental	Ensino Superior-Ensino Médio	Ensino Médio-Ensino Fundamental	Ensino Superior-Ensino Médio	Ensino Médio-Ensino Fundamental	Ensino Superior-Ensino Médio
Tendência	-0.003	-0.000	0.002	0.002	-0.019	0.015***
	(0.003)	(0.001)	(0.005)	(0.002)	(0.014)	(0.004)
$-\frac{1}{\sigma}$	-0.175***	-0.480***	-0.336***	-0.225***	-0.041	0.152
	(0.035)	(0.087)	(0.086)	(0.056)	(0.194)	(0.165)
R^2	0.94	0.49	0.96	0.83	0.94	0.82

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

REGRESSÃO 3: ELASTICIDADE DE SUBSTITUIÇÃO ENTRE QUALIFICADOS E NÃO QUALIFICADOS

	6 GRUPOS ETÁRIOS 26-60 ANOS	**7 GRUPOS ETÁRIOS 18-60 ANOS**	**3 GRUPOS ETÁRIOS 18-60 ANOS**	**3 GRUPOS ETÁRIOS 18-60 ANOS APENAS HOMENS**
$-1/\sigma_U$	-0.340*** (0.010)	-0.313*** (0.009)	-0.285*** (0.013)	-0.284*** (0.014)
$-1/\sigma_A$	-0.094*** (0.024)	-0.082*** (0.021)	-0.097*** (0.033)	-0.082*** (0.014)
$-1/\sigma_E$	-0.518*** (0.071)	-0.583*** (0.071)	-0.665*** (0.087)	-0.757*** (0.099)
Tendência	0.002 (0.002)	0.006*** (0.002)	0.009*** (0.003)	0.008*** (0.002)
Nº Observações	348	406	174	174
R^2	0.9782	0.979	0.981	0.980

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

CONCLUSÃO

As mudanças ocorridas no Brasil durante o século XX foram estruturais e trouxeram consequências para a dinâmica populacional, acumulação de capital humano e efeitos nos diferenciais de salários no mercado de trabalho, assim como na desigualdade de renda. Os movimentos iniciados pelas coortes de mulheres, principalmente as nascidas a partir da década de 1930 do século XX, mostraram uma mudança na taxa de fecundidade. Na investigação proposta, uma vez controlando a coorte em seu município e interagindo os efeitos da coorte com os efeitos do município, educação explicou aproximadamente 30% na média a redução da taxa de fecundidade. Esta redução da taxa de fecundidade provocou, principalmente na década de 1990, uma mudança no tamanho relativo da população. Esta mudança foi usada como exógena para se investigar, no segundo capítulo da tese, se a queda da fecundidade favoreceu o processo de universalização do acesso à educação básica. Mesmo quando as estimações apresentaram um parâmetro negativo para o tamanho (relativo ou absoluto) da coorte, a magnitude apresentou contribuição na margem. Para o modelo com decomposição por séries, a educação dos pais apresentou-se mais relevante para o aumento da probabilidade de matrícula quando analisada a primeira série do Ensino Fundamental, o principal gargalo para o avanço no aumento da matrícula. As coortes mais educadas entraram no mercado de trabalho e alteraram a composição da oferta, aumentando a proporção de trabalhadores qualificados em relação a trabalhadores pouco qualificados principalmente nas décadas de 1990 e 2000. Para os grupos etários mais jovens, o aumento da oferta relativa de trabalhadores qualificados em relação a trabalhadores pouco qualificados explica a queda do salário relativo deste grupo. Para os grupos etários mais velhos, de 41 a 60 anos, aumento da oferta relativa agregada explica mudanças no salário relativo desse grupo etário. As simulações contrafactuais mostram que as mudanças relativas impactaram mais os trabalhadores jovens, de 18 a 25 anos e impactaram menos os grupos etários mais velhos, de 41 a 60 anos. Estas mudanças nos salários relativos estão relacionadas com desigualdade e coeficiente de Gini.

REFERÊNCIAS BIBLIOGRÁFICAS

AARONSON, D. et al. *Fertility transitions along the extensive and intensive margins*. The American Economic Review, 2014.

ACEMOGLU, D.; AUTOR, D. *What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology*. Journal of Economic Literature, 2012.

ANGRIST, J; LAVY, V.; SCHLOSSER, A. *New evidence on the causal link between the quantity and quality of children*. Working Paper 11835. NBER, 2005.

BECKER, G. S. *An economic analysis of fertility*, NBER Chapters, in: Demographic and Economic Change in Developed Countries, pages 209-240 National Bureau of Economic Research, Inc, 1960..

BECKER, G. *A theory of allocation of time*. The Economic Journal, 1965.

BECKER, G; LEWIS, G. *Interaction between quantity and quality of children*. In SCHULTZ T. (Org) Economics of the Family: Marriage, Children, and Human Capital. University of Chicago Press. 1974.

BECKER, G.; MURPHY, K.; TAMURA, R. *Human capital, fertility, and economic growth*. The Journal of Political Economy, 1990.

BLACK, D. et al. *Are children "normal"?* The Review of Economics and Statistics, 2013.

BLACK, S.; DEVEREUX, P.; SALVANES, K. *The more the merrier? the effect of family size and birth order on children's education*. The Quarterly Journal of Economics, 2005

CASTRO, L. B. *Esperança, frustração e aprendizado: A história da Nova República (1985-1989)***.** In: Giambiagi et at (org) *Economia Brasileira Contemporânea (1945-2004)*. Editora Campus, 2004.

CAMERON, S,.HECKMAN, J. *The Dynamics of Educational Attainment for Black, Hispanic, and White Males*. Journal of Political Economy, 2001, vol. 109, no. 3

CONNELLY, R. *A Framework for Analyzing the Impact of Cohort Size on Education and Labor Earning*. The Journal of Human Resources, 1986

CONNELLY, R.; GOTTSCHALK, P. *Effect of Cohort Composition on Human Capital Accumulation across Generations*. Journal of Labor Economics, 1995

CURI, A.; MENEZES FILHO, N. *Os Determinantes das Matrículas no Ensino Fundamental e Médio*. In Análise da Pesquisa Nacional por Amostra de Domicílios – PNAD: 2005. – Educação. MTE, CGEE, 2007

DEDECA, C., S. A *Pesquisa Nacional por Amostra de Domicílios – PNAD. Síntese Metodológica*. Revista Brasileira de Estudos Populacionais, Volume. 15, n. 2, 1998

DIÁRIO OFICIAL DA UNIÃO. Lei orgânica do ensino secundário, 1942. Disponível em: http://www2.camara.leg.br/legin/fed/declei/1940-1949/decreto-lei-4244-9-abril-1942-414155-publicacaooriginal-1-pe.html

FERNANDES, R.; MENEZES-FILHO, N. *Educação, salários e a alocação de trabalhadores entre tarefas: Teoria e evidência para o Bra*sil. Política e Planejamento Econômico, 2012.

FIENBERG, S.E.; MASON, W.M. Specification and implementation of age, period and cohort models. In MASON, W.M.; FIENBERG, S.E (Org.), *Cohort analysis in social research. Beyond the identification problem*.Springer-Verlag New York Inc., 1985

FRANCO, G. H. B. *Auge e declínio do inflacionismo no Brasil*. In: Giambiagi et at (org) *Economia Brasileira Contemporânea (1945-2004)*. Editora Campus, 2004.

GALOR, O. *From estagnation to growth: unified growth theory*. In: AGHION, Philippe; DURLAUF, Steven N. Handbook of economic growth; 2005.

GALOR, O; WEIL, D.N. *Population, technology, and growth: from Malthusian stagnation to the demographic transition and beyond*. The American Economic Review. V. 90, N.4; Setembro de 2000.

GOLDIN, C.; Katz, L. *The race between education and technology*, 2008.

HAYASHI, F. *Econometrics*. Princeton University Press, 2000.

IBGE. *Brasil, Grandes Regiões e Unidades da Federação: Estimativas da população com data de referência em 1º de julho de cada ano: 1980/2031*. Projeção da População por Sexo e Idade para o Período 1980 - 2050 - Revisão 2008

KAI, H. M. *Três ensaios em demografia econômica*. Tese de doutorado, PUC-RIO, 2009

KATZ, L; MURHY, K. *Changes in Relative Wages, 1963-1987: Supply and Demand Factors*. The Quarterly Journal of Economics, 1992.

LAM, D.; MARTELETO, L. *A escolaridade das crianças brasileiras durante a transição demográfica: aumento no tamanho da coorte versus diminuição no tamanho da família*. Pesquisa e Planejamento Econômico (PPE), v.36, n.2, 2006.

MASON, W.M.; FIENBERG, S.E. *Cohort analysis in social research. Beyond the identification problem*.Springer-Verlag New York Inc., 1985

PERCORA, A. ; MENEZES-FILHO, N. *O Papel da Oferta e da Demanda por Qualificação na Evolução do Diferencial de Salários por Nível Educacional no Brasil*. Estudos Econômicos, 2014.

PONCZEK, V.; SOUZA, P. *New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country*. Journal of Human Resources, 2012

RIANI, J.; GOLGHER, A. *Indicadores educacionais confeccionados a partir de bases de dados do IBGE* In. RIOS-NETO, E; RIANI (Org), J. Introdução à demografia da educação. – Campinas: Associação Brasileira de Estudos Populacionais - ABEP, 2004

RIGOTTI, J. *Variáveis de educação dos censos demográficos brasileiros de 1960 a 2000.* In. RIOS-NETO, E; RIANI (Org), J. Introdução à demografia da educação. – Campinas: Associação Brasileira de Estudos Populacionais - ABEP, 2004

ROSENZWEIG, M; WOLPIN, K. *Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment*. Econometrica, Vol. 48, No. 1, 1980.

SOARES, R. *Mortality reductions, educational attainment, and fertility choice*. The American Economic Review, 2005;

VIÉGAS, L.; SOUZA, M. *A progressão continuada no Estado de São Paulo: considerações a partir da perspectiva de educadores*. ABRAPEE, V.10, N.2, 2006.

WELCH, F. *Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust*. Journal of Political Economy, 1979.

WILLIS, R. *A new approach to the economic theory of fertility behavior*. Journal of Political Economy, 1973.

WOOLDRIDGE, J. *Econometrics analysis of cross section and panel data*. MIT Press, 2010