

Max Cardoso de Resende

**Análises econométricas sobre  
a permanência dos alunos do  
Ensino Médio da rede pública  
catarinense**

Florianópolis / SC

Março / 2018



Max Cardoso de Resende

**Análises econométricas sobre a  
permanência dos alunos do Ensino  
Médio da rede pública catarinense**

Esta Tese está sendo submetida para obtenção do Título de “Doutor em Economia”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Economia.

Universidade Federal de Santa Catarina - UFSC

Centro Sócio Econômico - CSE

Programa de Pós-Graduação em Economia - PPGECO

Orientador: Dr. Francis Carlo Petterini Lourenço

Florianópolis / SC

Março / 2018

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Resende, Max Cardoso de  
Análises econométricas sobre a permanência dos  
alunos do Ensino Médio da rede pública catarinense  
/ Max Cardoso de Resende ; orientador, Francis  
Carlo Petterini Lourenço, 2018.  
243 p.

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro Sócio-Econômico, Programa de Pós  
Graduação em Economia, Florianópolis, 2018.

Inclui referências.

1. Economia. 2. Economia da Educação. 3. Ensino  
Médio. 4. Evasão escolar. 5. Modelos  
probabilísticos. I. Lourenço, Francis Carlo  
Petterini. II. Universidade Federal de Santa  
Catarina. Programa de Pós-Graduação em Economia. III.  
Título.

MAX CARDOSO DE RESENDE

**Análises econométricas sobre a permanência dos alunos do Ensino Médio na rede pública catarinense**

Tese submetida para obtenção do Título de "Doutor em Economia", e aprovada em sua forma final pelo Programa de Pós-Graduação em Economia.

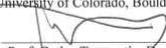
Florianópolis, 26 de março de 2018.

  
Prof. Adylson Jair da Silveira, Dr.  
Coordenador do Curso

**Banca Examinadora:**

  
Prof. Francis Carlo Petterini Lourenço, Dr.  
Orientador  
PPGECO/UFSC

  
Prof. Carlos Martins-Filho, Dr.  
University of Colorado, Boulder

  
Prof. Carlos Tramontin, Dr.  
ESAG/UDESC

  
Prof. Pedro Barbeta, Dr.  
INE/UFSC

  
Prof. Fernando Seabra, Dr.  
PPGECO/UFSC



# Agradecimentos

Gostaria de agradecer e dedicar esta tese a meus pais e irmã, José Francisco, Ludmilla e Heloá, pelo apoio incondicional que me prestaram, para que eu pudesse cumprir com os meus deveres. Sempre apoiaram as minhas escolhas e acreditam no meu potencial.

As minhas gurias, Ana Paula e Valentina, que me acompanharam ao longo de 3 anos deste doutorado, com muito amor, fazendo com que me sentisse uma pessoa especial.

Ao meu orientador, Professor Francis Petterini por sua dedicação, confiança e paciência. Por apoiar e acreditar na minha capacidade em realizar o trabalho.

Ao Professor Carlos Martins-Filho por me receber de maneira acolhedora, pelos ensinamentos e por me mostrar que fazer ciência não é fácil.

Agradeço aos comentários e sugestões oferecidos pelos professores componentes da banca examinadora: Prof. Carlos Tramontin, Prof. Fernando Seabra e Prof. Pedro Barbeta.

Agradeço aos demais professores, do programa de pós-graduação da Universidade Federal de Santa Catarina, em especial a Evelise, secretária do curso, pelos seus conselhos, carinho e paciência.

Agradeço aos meus colegas de doutorado, Dinorá, Adilson, Aulo, Evandro, Diogo, Mamadu e Júlio, por terem me ajudado nesta longa caminhada e, ao amigo e companheiro de casa por cerca de 6 anos, Gabriel.

Sou grato, também, a todos os meus amigos que me acompanharam até aqui, sempre apoiando e incentivando, cujos nomes não poderia de deixar de expor aqui: Marcelo, Gilmar, Paulo e Fernando Henrique, este companheiro desde os tempos de PUCSP.

A CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão de bolsa de estudos para o curso de doutorado.

# Resumo

O que leva um adolescente a abandonar os estudos ao longo do Ensino Médio? O entendimento desta questão é crucial, por exemplo, para formuladores de políticas que pretendam lançar um novo desenho de política pública, ou para pais e professores que precisam agir para evitarem essa decisão. Esta pesquisa trata fundamentalmente dessa situação. Mais especificamente, busca identificar as variáveis individuais, familiares e escolares que influenciam o comportamento acadêmico de um jovem de 15 a 17 anos. E para conduzir uma sequência de três ensaios sobre o tema, recorre-se a uma revisão bibliográfica e aos modelos probabilísticos univariados e bivaridos e de correção amostral de Heckman. Para a realização dos exercícios empíricos, foi utilizada uma base de dados fornecida pela Secretaria de Educação (SEDSC), que permite o acompanhamento dos alunos matriculados na rede pública catarinense de ensino. O primeiro ensaio apresenta uma revisão da literatura da Economia da Educação, no qual são discutidos os motivos da importância da educação, nos contextos macro e microeconômicos, além da contextualização do tema no cenário nacional. O segundo ensaio faz uso de modelos probabilísticos bivariados para compreender se e como a reprovação afeta a probabilidade de abandono e evasão escolar, com foco nos alunos de 1º ano do Ensino Médio. Por fim, o terceiro ensaio trata da possibilidade de viés de estimação amostral e realiza um exercício empírico que utiliza o modelo de correção amostral de Heckman. Por consequência do apresentado nesta pesquisa, será

possível verificar o quanto uma reprovação é capaz de influenciar a decisão do aluno em abandonar os estudos, além de seu perfil. Com base na revisão da literatura e nos exercícios empíricos desta pesquisa, verifica-se forte evidência de que a situação de aprovação/reprovação é uma função de uma miríade de elementos, intra e extraescolar; no entanto, encontrou-se uma evidência robusta de que a permanência na escola está, sobretudo, relacionada com a situação de reprovação, o que seria um argumento a favor de políticas de progressão continuada.

**Palavras-chave:** Educação. Ensino Médio. Modelos probabilísticos bivariados. Modelo de seleção de Heckman.

# Abstract

What leads a teenager to dropout of high school? Understanding this issue is crucial, for example, for policy makers wishing to launch a new public policy design, or for parents and teachers who need to act to avoid this decision. This research deals fundamentally with this situation. More specifically, it seeks to identify the individual, family and school variables that influence the academic behavior of a young person who aged from 15 to 17 years. In order to conduct a sequence of three essays in this matter, we call upon a bibliographical review and the use of univariate and bivariate probabilistic models and Heckman's sample selection model. For the accomplishment of the empirical exercises, a database was provided by the Secretary of Education (SEDSC) that allows the monitoring throughout the years of students enrolled in the Santa Catarina Public School system. The first essay presents a literature review in Economics of Education, where the reasons for the importance of education in the macro and microeconomic contexts are discussed, as well as the contextualization of the theme in the national scenario. The second essay makes use of bivariate probabilistic models to understand if and how an academic failure affects the probability of school dropout and, focusing on first graders students of High School. Finally, the third essay deals with the possibility of sample estimation bias, which leads to the use of Heckman sample correction model. Based on the literature review and the empirical exercises of this research, ensure that there is strong evidence that the ap-

proved/fail situation is a function of a myriad of elements, intra and extra school; however, robust evidence has emerged that the permanence in school is mainly related to the fail situation, which has implications in favor of policies of continuous transition.

**Keywords:** Education. High school. Bivariate probabilistic models. Heckman's sample selection model

# Lista de ilustrações

Figura 1 – Indicadores de Rendimento - Ensino Médio Brasil 1999/2014 . . . . .	92
Figura 2 – Ocupação de Jovens: 15 a 17 anos . . . . .	94
Figura 3 – Árvore de acompanhamento estudantil . . . . .	127
Figura 4 – Histograma do número de faltas de alunos aprovados e reprovados . . . . .	133
Figura 5 – Matriz de correlação para variáveis explicativas	134
Figura 6 – Modelo de fluxo escolar . . . . .	158



# Lista de tabelas

Tabela 1 – Probabilidades dos eventos no Modelo <i>biprobit</i>	120
Tabela 2 – Desempenho acadêmico: Quantidade total de alunos. 2008 a 2012 . . . . .	128
Tabela 3 – Estatísticas descritivas: Base de dados SEDSC	131
Tabela 4 – Descrição das variáveis extraídas da SEDSC .	139
Tabela 5 – Estimações <i>probit</i> e <i>biprobit</i> para a transição acadêmica da 1ª série 2ª série do Ensino Médio	143
Tabela 6 – Probabilidades marginais preditas: modelos <i>probit</i> . . . . .	145
Tabela 7 – Probabilidades marginais preditas: modelos <i>biprobit</i> . . . . .	146
Tabela 8 – Probabilidades conjuntas preditas: modelos <i>biprobit</i> . . . . .	147
Tabela 9 – Efeitos marginais médios: M1_B . . . . .	149
Tabela 10 – Efeitos marginais médios: Modelo M2_B . . .	150
Tabela 11 – Efeitos marginais médios: M3_B . . . . .	151
Tabela 12 – Estatísticas descritivas: Base de dados SEDSC	183
Tabela 13 – Matriz de correlação: Equação de interesse . .	184
Tabela 14 – Descrição das variáveis extraídas da SEDSC .	186
Tabela 15 – Especificações para a função de seleção . . .	190
Tabela 16 – Funções de Seleção - Coeficientes estimados usando Modelo <i>Probit</i> . . . . .	191
Tabela 17 – Funções de Seleção - Probabilidades previstas	193

Tabela 18 – Modelo 1: Estimacões <i>MQO</i> e <i>Heckman MLE</i> para a 1ª série, turno da manhã . . . . .	194
Tabela 19 – Modelo 2: Estimacões <i>MQO</i> e <i>Heckman MLE</i> para a 1ª série, turno da manhã . . . . .	196
Tabela 20 – Modelo 3: Estimacões <i>MQO</i> e <i>Heckman MLE</i> para a 1ª série, turno da manhã . . . . .	198
Tabela 21 – Modelo 4: Estimacões <i>MQO</i> e <i>Heckman MLE</i> para a 1ª série, turno da manhã . . . . .	200
Tabela 22 – Modelo 1: Alunos da manhã × Alunos do noturno . . . . .	204
Tabela 23 – Modelo 2: Alunos da manhã × Alunos do noturno . . . . .	206
Tabela 24 – Modelo 3: Alunos da manhã × Alunos do noturno . . . . .	208
Tabela 25 – Modelo 4: Alunos da manhã × Alunos do noturno . . . . .	210
Tabela 26 – Modelo 1: Estimacões <i>Heckman MLE</i> : 2008 e 2009 . . . . .	234
Tabela 27 – Modelo 1: Estimacões <i>Heckman MLE</i> : 2010 e 2011 . . . . .	235
Tabela 28 – Modelo 2: Estimacões <i>Heckman MLE</i> : 2008 e 2009 . . . . .	236
Tabela 29 – Modelo 2: Estimacões <i>Heckman MLE</i> : 2010 e 2011 . . . . .	237
Tabela 30 – Modelo 3: Estimacões <i>Heckman MLE</i> : 2008 e 2009 . . . . .	238
Tabela 31 – Modelo 3: Estimacões <i>Heckman MLE</i> : 2010 e 2011 . . . . .	239

Tabela 32 – Modelo 4: Estimacões <i>Heckman MLE</i> : 2008 e 2009 . . . . .	240
Tabela 33 – Modelo 4: Estimacões <i>Heckman MLE</i> : 2010 e 2011 . . . . .	241



# Sumário

	<b>Introdução</b> . . . . .	<b>21</b>
<b>1</b>	<b>O ESTADO DA ARTE DA ECONOMIA DA EDUCAÇÃO</b> . . . . .	<b>27</b>
1.1	<b>Introdução</b> . . . . .	<b>27</b>
1.2	<b>A perspectiva macroeconômica da educação</b> . . . . .	<b>32</b>
1.2.1	A especificação Mincer . . . . .	33
1.2.2	A teoria da sinalização . . . . .	37
1.2.3	Os modelos macroeconômicos de crescimento endógeno . . . . .	39
1.3	<b>Em direção a uma Economia da Educação</b> . . . . .	<b>42</b>
1.3.1	A perspectiva microeconômica . . . . .	42
1.3.2	O Relatório Coleman . . . . .	46
1.4	<b>Determinantes do desempenho acadêmico</b> . . . . .	<b>50</b>
1.4.1	Tempo de instrução . . . . .	50
1.4.2	A escola . . . . .	52
1.4.3	Professores . . . . .	55
1.4.4	A estrutura curricular . . . . .	63
1.4.5	A família . . . . .	78
1.5	<b>A educação no Brasil sob a perspectiva da Economia da Educação</b> . . . . .	<b>80</b>

1.6	Considerações finais . . . . .	89
2	<b>UM MODELO BIVARIADO PARA A PROGRESSÃO ACADÊMICA NO ENSINO MÉDIO . . . . .</b>	<b>91</b>
2.1	Introdução . . . . .	91
2.2	Modelos teóricos de evasão e persistência escolar . . . . .	97
2.3	Metodologia Econométrica . . . . .	117
2.3.1	A modelagem bivariada <i>biprobit</i> . . . . .	117
2.3.2	A correlação $\rho$ . . . . .	122
2.3.3	Os efeitos marginais para a modelagem <i>birpobit</i> . . . . .	123
2.4	A base de microdados da SEDSC . . . . .	125
2.5	O modelo econométrico . . . . .	134
2.6	Resultados empíricos . . . . .	142
2.6.1	Efeitos marginais para o modelo <i>biprobit</i> . . . . .	147
2.7	Considerações finais . . . . .	152
3	<b>MODELO DE INFERÊNCIA DE HECKMAN APLICADO AO ABANDONO ESCOLAR NO ENSINO MÉDIO . . . . .</b>	<b>155</b>
3.1	Introdução . . . . .	155
3.2	A questão do viés de seleção amostral . . . . .	162
3.3	Revisão teórica . . . . .	164
3.4	O Modelo de Heckman . . . . .	167
3.4.1	A estimação de dois estágios de Heckman . . . . .	174
3.4.2	O método de máxima verossimilhança . . . . .	178

<b>3.5</b>	<b>A base de microdados</b> . . . . .	<b>181</b>
<b>3.6</b>	<b>O modelo econométrico</b> . . . . .	<b>184</b>
3.6.1	Especificando a função de seleção . . . . .	189
<b>3.7</b>	<b>Resultados empíricos</b> . . . . .	<b>194</b>
3.7.1	Comparação entre turnos . . . . .	202
<b>3.8</b>	<b>Considerações finais</b> . . . . .	<b>211</b>
	<b>Conclusão</b> . . . . .	<b>215</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>217</b>
	<b>ANEXO A – INFORMAÇÕES COMPLE-</b> <b>MENTARES: ENSAIO 3</b> . . . . .	<b>233</b>



# Introdução

O que leva um adolescente a abandonar a escola pública? O que diferencia um aluno aprovado de um reprovado? O perfil de um estudante que reprova é semelhante ao do que abandona os estudos? O entendimento dessas questões, historicamente, faz parte de debates e reflexões no âmbito da educação pública no Brasil, as quais têm por objetivo quantificar, qualificar e aprimorar tanto o comportamento quanto a capacidade do estudante e, por consequência, fornecer subsídios para a formulação de políticas públicas voltadas ao combate da crise de audiência e dos altos índices de reprovação do Ensino Médio.

A história do Ensino Médio é marcada pelo descaso do Estado brasileiro em torná-lo mais acessível as camadas mais populares da população e em formular políticas públicas que motivem esses grupos sociais a ingressar e permanecer nesse nível de educação. Parcialmente, essa perspectiva começa a se alterar a partir da década de 1990, com a promulgação de marcos legislativos, programas específicos e reformas curriculares, como, por exemplo, a Lei de Diretrizes e Bases (LDB) (1996), a instituição do Exame Nacional do Ensino Médio (ENEM) (1998), a criação do Fundo de Manutenção e Desenvolvimento da Educação Básica (FUNDEB) (2006) e, mais recentemente, a reforma curricular intitulada como Novo Ensino Médio (2017), entre outros. Para [Krawczyk et al. \(2009\)](#), merece destaque a mudança de

tratamento dada a esse nível de ensino desde a homologação da LDB, ao expressar que “a inclusão do Ensino Médio no âmbito da educação básica e o seu caráter progressivamente obrigatório demonstram o reconhecimento da importância política e social que ele possui.” Essa mudança reproduz tanto o anseio das camadas mais pobres por mais escolarização quanto a necessidade de atendimento de maiores exigências apresentadas no contexto econômico e no âmbito da concorrência do mercado de trabalho e do comércio internacional.

A justificativa para esta pesquisa surgiu da crise de audiência do Ensino Médio, evidenciada pela manutenção em patamares elevados das taxas de abandono e reprovação nos últimos vinte anos. Então, quais são os problemas da não completude do ciclo básico e da má formação acadêmica do jovem brasileiro? Se a pessoa abandona o Ensino Médio, ela sequer terá um curso técnico ou ingressará no Ensino Superior, o que, por consequência, compromete o nível futuro de capital humano e produtividade no país, por isso é um objeto de estudo da economia. Portanto, precisa-se entender as razões da não permanência do aluno na escola. Nesse sentido, é importante observar que a permanência nada mais é do que uma sequência de aprovações, em última instância, a variável-chave que precisa ser estudada são as características de sequência de aprovação.

O presente trabalho pretende, através de uma perspectiva da Economia da Educação, investigar as causas da progressão acadêmica dos estudantes de Ensino Médio do estado de Santa Catarina, para o período de 2008 a 2012, a partir de uma base

---

de dados fornecida pela Secretaria de Educação (SEDSC), que contempla todos os alunos matriculados na rede pública catarinense. O uso desses dados traz um caráter singular e original para os resultados que serão estimados a partir desta pesquisa, assim será possível obter um retrato mais acurado da educação de Ensino Médio no estado e, dessa maneira, contribuir com o poder público para a eficácia e efetividade na implementação de políticas e ações educacionais.

Trata-se de analisar questões de caráter individual, familiar, econômico, social e de infraestrutura acadêmica. A intenção é compreender a integração entre as diversas variáveis que constituem o processo de educação, como estas influenciam o desempenho acadêmico de estudantes, na sua maioria, de 15 a 17 anos, e de que forma será possível realizar mudanças de políticas para que os alunos com maior risco de abandonar a rede pública de ensino permaneçam.

Esta pesquisa é composta por três ensaios, os quais estão interligados pelo objetivo geral da análise das questões referentes à permanência e transição serial na rede pública de ensino catarinense. Para tal fim, cada ensaio explora a questão da educação no estado sob diferentes pontos de vista. Destarte, cada um tem a capacidade de ser interpretado por si só, podendo ser caracterizado como um artigo.

O primeiro ensaio traz uma revisão da literatura da Economia da Educação, no qual se discute os motivos da importância da educação, nos contextos macro e microeconômicos, além de se apresentar fatos estilizados em relação às questões escolares, familiares e pedagógicas para a interpretação do comportamento estudantil. Apresenta, também, os principais trabalhos empíricos que tratam do tema no Brasil, buscando examinar quais são os fatores que determinam tanto o desempenho acadêmico quanto a audiência escolar, com o objetivo de elucidar a discussão sobre essas temáticas no âmbito do Ensino Médio.

O segundo ensaio tem por objetivo identificar as características individuais, familiares e escolares mais relevantes para a aprovação acadêmica, com foco na transição da 1ª para a 2ª Série, em virtude deste ser o momento de maior probabilidade para a repetência ao longo dessa etapa do ciclo básico. Diante disso, para conduzir uma sequência de exercícios empíricos sobre o tema, são utilizados modelos probabilísticos univariados e bivariados. Por consequência desse entendimento, será possível compreender se e como a reprovação afeta a probabilidade de abandono e evasão escolar.

O terceiro e último ensaio trata da possibilidade de viés de estimação amostral contida nas estimações do segundo ensaio, o que pode ocorrer em função do atrito na base de dados. Neste contexto, a estratégia empírica deste ensaio consiste em utilizar a técnica de correção de Heckman (1977), que expressa-se em duas equações: a equação de participação (ou seleção), que trata da probabilidade do aluno permanecer na rede pública

de ensino; e a equação de regressão (ou interesse), que trata do desempenho acadêmico, o qual inclui o termo de correção derivado da distribuição do erro da equação de participação. Por consequência desse entendimento, tem-se como objetivo identificar as características individuais e escolares que mais afetam o rendimento acadêmico e a permanência do aluno na escola pública e, assim, contribuir para a análise da crise de audiência do Ensino Médio no Brasil.



# 1 O estado da arte da Economia da Educação

## 1.1 Introdução

Em um contexto de sistema capitalista, o desenvolvimento e a organização da atividade humana se tornaram questões-chaves no que tange o crescimento econômico e a distribuição de renda das nações, alçando as questões sobre política educacional a uma posição de destaque dentro dessa perspectiva. Por consequência, atribuiu-se à educação os papéis de disseminar, formar e qualificar a mão de obra para o mercado de trabalho e, principalmente, em países subdesenvolvidos, a inclusão social.

Segundo [Psacharopoulos e Patrinos \(2004\)](#), a difusão da mentalidade desenvolvimentista para diversos países permitiu a formação de uma extensa literatura a partir dos anos de 1950, não apenas focada no papel central da educação na economia *per se*, mas também na fusão, transbordamentos e externalidades que esta tem sobre outros campos da ciência, como economia do trabalho, pedagogia, sociologia e psicologia. Em âmbito individual, a educação afeta os ganhos, a empregabilidade e a qualidade de vida, gerando efeitos sobre a saúde, a estrutura familiar, a realização intelectual, entre outros. Por sua vez, sob uma perspectiva de país, o estoque de capital humano está bastante correlacionado com sua taxa de prosperidade e crescimento, além de ser um

fator determinante de desigualdade tanto entre os seus habitantes, quanto em relação a outros países. Nesse contexto multidisciplinar, a análise econômica busca estabelecer, a partir de abordagens quantitativas, inferências causais entre as diversas variáveis que possam vir a afetar os resultados acadêmicos e, assim, fornecer informações para que as políticas públicas educacionais sejam melhor desenhadas e implementadas.

A década de 1960 foi marcada por uma forte expansão da economia mundial, com taxas médias anuais de crescimento de 4,0% do PIB, conforme dados do Banco Mundial, impulsionada, sobretudo, pelas transformações tecnológicas que passaram a fazer parte do cotidiano das nações. Logo, em virtude das necessidades de mercado em dar prosseguimento ao processo contínuo de inovação, com vistas à manutenção da competitividade para empresas e países que desencadeassem uma série de mudanças em termos de custo, qualidade, produtividade e formação de capital, tornou-se cada vez mais necessário capacitar a força de trabalho. Assim, a educação foi apresentada como a principal responsável por formar a nova mão de obra, impondo à escola a incumbência de tal tarefa.

Diante dessa nova perspectiva para a mão de obra, os anos de 1960 se caracterizam pelas tentativas de se estimar a taxa de retorno do investimento em capital humano ou a contribuição da educação no crescimento econômico. [Schultz \(1961\)](#), [Baker \(1964\)](#) e [Mincer \(1974\)](#) trataram a educação como um investimento, em que as pessoas acumulam capital humano a fim de receberem, no futuro próximo, uma remuneração maior, em que

a demanda por educação será determinada pela igualdade entre o seu custo marginal, o qual consiste de custos diretos (mensalidade) e indiretos (custos de oportunidade decorrentes da redução da renda corrente pela opção por estudar) e os benefícios marginais, que são obtidos pelo valor presente da renda de toda a vida. Ao ser tratada como um investimento, a demanda por educação está negativamente correlacionada tanto com a taxa de juros quanto com ambos os custos, mas caso o retorno sobre capital humano seja incerto, o valor esperado do retorno poderá ser tanto maior quanto menor que a taxa de juros no nível ótimo.

Os anos de 1970 e 1980 foram marcados pela implementação de testes que focaram na existência de triagens para a contribuição individual da educação no processo de alocação profissional. Essa perspectiva surge da premissa de que as empresas não podem observar a produtividade de seus trabalhadores diretamente. Dado que o custo do aprendizado em escolas é menor para pessoas mais talentosas, os indivíduos com alta produtividade poderão aumentar a sua demanda por educação até o ponto em que indivíduos com baixa produtividade possam se recusar a imitar a decisão por mais anos de educação dos primeiros. Dessa maneira, a demanda por estudos se torna um indicador de expectativas de produtividade do trabalhador por parte dos empregadores, e que será levada em conta no processo de seleção.

A partir dos anos 1990 até hoje, o foco está na identificação de externalidades e complementariedades promovidas pela educação tanto para o crescimento econômico quanto na quantificação de efeitos não mercadológicos (crime, doenças, ca-

racterísticas cognitivas, etc.) gerados por ela, além de assuntos relacionados à gestão. Portanto, com o passar dos anos, ocorreu uma mudança de perspectiva de análise, a qual deixou de ser voltada para níveis agregados de educação no crescimento econômico e passou a ser centrada na microeconomia.

Diante das contribuições de [Schultz \(1961\)](#), [Blaug \(1985\)](#) e [Psacharopoulos e Patrinos \(2004\)](#), pode-se classificar a trajetória histórica da teoria econômica da educação, de modo que a perspectiva macroeconômica vem cedendo espaço para a microeconômica. Temos, assim, a seguinte perspectiva:

- As abordagens iniciais de [Schultz \(1961\)](#), [Baker \(1964\)](#), [Mincer \(1974\)](#)<sup>1</sup> e [Nelson e Phelps \(1966\)](#) que formalizaram o tratamento da educação como investimento ao invés de uma atividade de consumo, bem como a possibilidade de mensurá-la. Isto é, sob uma perspectiva macroeconômica, a importância de medidas de educação para o crescimento de um país;
- O tratamento de [Arrow et al. \(1973\)](#), [Spence \(1978\)](#) e [Stiglitz \(1975\)](#) sobre a possibilidade da educação agir como um sinalizador particular para o empregador e não mais somente como uma contribuição social;
- Os aportes de [Romer \(1986\)](#), [Lucas \(1988\)](#) e [Acemoglu e Angrist \(2000\)](#) sobre o papel da educação na teoria do crescimento econômico endógeno e em relação às exter-

---

<sup>1</sup> Grupo de estudos da Universidade de Chicago.

nalidades geradas no contexto geral do sistema educativo. Entre os temas tratados, incluem-se análises de custos e benefícios privados e sociais (por nível de ensino, por tipo de currículo, por tipo de ensino), análises de eficiência, eficácia, equidade e alocação de recursos, entre outros;

- Em paralelo ao desenvolvimento de modelos de crescimento endógeno, tivemos a expansão da abordagem microeconômica da educação, através da qual contribuições importantes foram trazidas por [Hanushek \(1986\)](#), [Correa \(1995\)](#), [Greenwald, Hedges e Laine \(1996\)](#), [Rivkin, Hanushek e Kain \(2005\)](#), [Allen e Fraser \(2007\)](#), entre outros.

Diante da relevância do tema, também é crescente o número de trabalhos empíricos no Brasil, que buscam examinar quais são os fatores que determinam tanto o desempenho acadêmico quanto a audiência escolar como, por exemplo, os trabalhos de [Barros e Mendonça \(2008\)](#) e [Neri et al. \(2015\)](#). Assim, buscase trazer informações consistentes que embasem as discussões sobre ambas as temáticas no âmbito do Ensino Médio, a fim de contribuir com o desenvolvimento de políticas públicas mais eficazes.

O objetivo do presente ensaio é apresentar uma revisão da literatura da Economia da Educação, onde são discutidos os motivos da importância da educação, nos contextos macro e microeconômicos, além de apresentar o que é amplamente acordado em relação às questões escolares, familiares e pedagógicas para a interpretação do comportamento estudantil.

O ensaio está organizado em 6 seções. Além desta introdução, a Seção 2 traz uma breve descrição dos principais modelos macroeconômicos de crescimento e sua relação com a educação. A Seção 3, trata do surgimento da Economia da Educação, além de apresentar a tentativa clássica mais conhecida de se estimar o desempenho acadêmico baseado em princípios microeconômico, o Relatório Coleman. A Seção 4, apresenta as principais variáveis que guiam o comportamento estudantil. Por sua vez, a Seção 5, mostra como os fundamentos da Economia da Educação podem colaborar na compreensão da crise de audiência do Ensino Médio da rede pública no Brasil. Por fim, a Seção 6, conclui o ensaio.

## 1.2 A perspectiva macroeconômica da educação

Antes do século XIX, as políticas governamentais adotadas tinham como objetivo a redistribuição de riqueza dos ricos para os pobres, principalmente, através da adoção de práticas protecionistas a alguns grupos em situações particulares, como penúria e crises bancárias. Nesse contexto, o investimento sistemático em capital humano não era visto como essencial para o processo de desenvolvimento econômico sustentável de um país, e os gastos com escolas, treinamento e outras formas de investimentos similares eram insignificantes (TANZI; SCHUKNECHT, 2000; OZTURK, 2001).

Esse contexto se alterou no decorrer do século XIX, quando os governos dos países, hoje considerados industrializados, começaram a fornecer educação primária, em vistas à prerrogativa de que a educação poderia contribuir para o crescimento econômico e na estabilidade social. Em 1900, o ensino público primário universal já era regra entre esses países e os investimentos em educação representavam mais de 1% do produto interno bruto (PIB) de países como França, Alemanha, Bélgica e Japão. Antes da Segunda Guerra Mundial, essas inversões já haviam aumentado consideravelmente, assim como o tempo médio legal que uma pessoa deveria frequentar a escola (TANZI; SCHUKNECHT, 2000).

Assim, no decorrer do século XX, educação, habilidades, talento e capacidade de aquisição e incorporação de conhecimento se tornaram um determinante essencial para a produtividade tanto de um indivíduo quanto de uma nação. Para Ozturk (2001), esse século pode ser tratado como a "Era do Capital Humano", no sentido de que o determinante básico acerca dos padrões de vida depende de quão eficiente um país é em desenvolver e utilizar habilidades e conhecimentos, e, em última instância, na abrangência populacional do seu sistema educacional.

### 1.2.1 A especificação Mincer

Com o conceito de capital humano formalizado por Schultz (1961) e Baker (1964), o investimento em educação passou a ser um condicionante da dinâmica econômica de longo prazo, uma vez que o trabalho humano quando qualificado atra-

vés da educação, apresenta melhora nos rendimentos individuais e amplia a produtividade econômica. Por consequência, o aumento do estoque de capital humano elevam as taxas de lucro das empresas, além de gerar externalidades positivas sobre a formação das famílias, saúde e criminalidade. As teorias de capital humano e sinalização deslocam para o contexto individual as responsabilidades da inserção social, do emprego e do desempenho profissional e tratam a educação como um insumo de produção qualquer ao equiparar capital físico e trabalho, além de ter como premissa a ideia de que os investimentos em educação devem ser determinados por regras de mercado, logo a decisão por uma demanda maior de educação é na essência uma decisão de investimento, no qual se comparam custos e retornos. Sob essa perspectiva, [Harmon, Oosterbeek e Walker \(2003\)](#) enumera um conjunto de implicações:

- A taxa interna de retorno da educação (TIR) é a taxa de desconto que iguala o valor presente dos benefícios aos custos. Nesse caso, se a TIR for maior que a taxa de juros do mercado, mais vantajoso será para o indivíduo investir em educação. Ao tomar uma decisão de investimento qualquer, a pessoa que dá mais (menos) valor a renda corrente do que a futura, terá um valor maior (menor) para a taxa de desconto, o que faz com que ela tenha um menor interesse pelos estudos. [Barbosa-Filho e Pessoa \(2010\)](#) ainda acrescentam que, quando uma pessoa toma como premissa que o investimento em educação é uma decisão econômica, pode-se compreender melhor tanto as irregularidades do

mercado de trabalho quanto as causas de desigualdade de renda;

- Custos diretos da educação reduzem o benefício líquido da escolaridade;
- Se a probabilidade de estar empregado é alta quando se tem maior escolaridade, então, um aumento no seguro desemprego pode reduzir o prêmio por uma maior educação;
- A maior escolaridade implica em uma probabilidade maior de se realizar treinamentos enquanto se está empregado;
- Os salários das pessoas com maior escolaridade sofrem menos variações do que os daquelas com menos anos de estudo quando o governo implementa algum tipo de política que venha a ter alguma incidência sobre o lado da demanda do mercado de trabalho.

Em relação a primeira implicação, pode-se aferir diversas características, tais como a heterogeneidade na evolução temporal dos salários ao longo dos ciclos econômicos entre as pessoas mais e menos educadas, o tempo de permanência e de rotatividade nos postos de trabalho entre as diversas faixas etárias da população economicamente ativa, o maior investimento em educação por parte de pessoas tecnicamente mais eficientes, a maior taxa de retorno de pessoas mais instruídas, entre outras.

Quanto a segunda, uma explicação para a heterogeneidade da renda do trabalho era a de que as pessoas são diferentes por natureza, de modo que os salários também seriam diferentes. Conforme [Mincer \(1958\)](#) *apud* [Barbosa-Filho e Pessôa \(2010\)](#), em razão da premissa da teoria dos diferenciais compensatórios de salário, que diz que a divergência entre salários se origina em função do tempo maior de escolaridade suprimir completamente os custos privados em educar-se, a distribuição de salários deveria ser assimétrica. Nesse contexto, a aproximação empírica da estrutura teórica do capital humano é a forma funcional da equação de salários ([MINCER, 1974](#)):

$$\ln w = \ln w + \beta S + \delta X + \gamma_1 E + \gamma_2 E^2 + u_i, \quad (1.1)$$

onde o subscrito  $i$  refere-se ao indivíduo,  $\ln w$  é o logaritmo do salário do indivíduo com  $S$  anos de estudos,  $\ln w$ , o logaritmo do salário do trabalhador sem escolaridade alguma,  $X$ , é um vetor de variáveis de controle observáveis,  $E$ , o tempo de experiência no mercado de trabalho,  $\beta$  a semielasticidade da educação no salário do indivíduo para cada ano a mais de estudo e  $u_i$  um termo aleatório que representa outros fatores que não especificados ou observados, assumidos não correlacionados assintoticamente a  $X$  e  $S$ .

A função salário do capital humano (Equação 1.1) está ajustada para o logaritmo natural dos salários, para permitir relações não lineares entre as variáveis dependentes e independentes, reduzir problemas de assimetria de distribuição dos salários e gerar estimativas mais robustas devido a redução da amplitude dos valores dos salários e, assim, facilitar na comparação relativa

entre estes. Note, também, que a experiência é inserida como um termo quadrático, dado o interesse em se capturar a concavidade do perfil dos salários por experiência<sup>2</sup>. A partir das derivações empíricas do Modelo de Mincer, a Economia da Educação se consolidou no arcabouço da economia do trabalho.

Dentro dessa perspectiva, para [Waltenberg \(2006\)](#), as escolas seriam tratadas como firmas especializadas na produção de educação e o sistema educativo seria uma indústria, porém com diferenças em relação às indústrias capitalistas em geral, no sentido de não ter como objeto explícito, a obtenção de lucro, e sim a instrução cognitiva, moral, intelectual, divulgação de conhecimento, pesquisa e desenvolvimento, etc., além de que os seus consumidores (estudantes) não arcam com todos os custos de aquisição do serviço ofertado.

### 1.2.2 A teoria da sinalização

A teoria da sinalização, apresentada por [Arrow et al. \(1973\)](#), [Stiglitz \(1975\)](#) e [Spence \(1978\)](#), corrobora com as conclusões advindas da teoria do capital humano, de que indivíduos com uma escolaridade média maior tem um aumento na probabilidade de receber maiores salários. Todavia, discordam do papel da escola no trato das habilidades individuais, uma vez que essas são inerentes a pessoa, e a escola apenas as realçaria, tornando o sistema educacional apenas um sinalizador para o mercado de trabalho.

---

<sup>2</sup> Aquele que investir mais em educação, terá rendas mais elevadas e duradouras, além de ser mais produtivo que aqueles sem escolaridade alguma.

Por sua vez, os salários serão definidos pelo grau de produtividade associado à maior habilidade sinalizada pela maior escolaridade e não pela acumulação de capital humano. [Spence \(1978\)](#) explica que a pessoa menos habilidosa não tem a capacidade de apresentar uma aptidão maior do que a que ela realmente tem pois, para estas pessoas, é muito custoso se educar mais. Assim, em uma situação extrema, o dispêndio de recursos públicos e privados em educação seria ineficiente, o que levaria a uma forte retração nos investimentos neste setor.

Para [Barbosa-Filho e Pessôa \(2010\)](#), “uma evidência comumente alegada em favor da teoria da sinalização para os diferenciais de salários associados aos diferenciais de educação é o prêmio de diploma.” Assim, quando um ano adicional de estudo representa o ano de conclusão de algum ciclo escolar, a taxa de retorno desse ano é muito maior do que a de outro qualquer que não esteja associado a um diploma. Subentende-se que esse ganho extra venha dessa sinalização, o que contraria a visão da acumulação de capital humano sobre os anos de estudo, que não distingue o ano de conclusão dos demais, isto é, todos apresentam retornos idênticos.

Muitos desses efeitos refletem sobre o crescimento econômico agregado, entretanto, esses canais não os únicos. A educação pode gerar externalidades para a sociedade que façam com que o seu impacto sobre o produto nacional seja ainda maior do que o obtido nos estudos da educação, como uma decisão de investimento e na equação de salários. Como efeitos externos positivos, a educação facilita a comunicação entre as pessoas e a imple-

mentação de normas sociais desejáveis, reduz a criminalidade, conscientiza as pessoas quanto a vícios e cuidados de higiene, dentre outros benefícios que podem fazer com que o retorno social da educação seja tão importante ou maior que o privado para o crescimento econômico do país.

Assim, com o avanço da literatura, a educação passou a ser vista não apenas como um determinante do crescimento e progresso econômico, mas também como um dinamizador de externalidades positivas para a sociedade, pois alavanca maior consciência política, ética, moral e social do indivíduo, ocasionando a redução de fatores de risco à população e maior bem estar coletivo.

### 1.2.3 Os modelos macroeconômicos de crescimento endógeno

O modo como a educação era interpretada e inserida, nos modelos macroeconômicos neoclássicos de crescimento econômico, foi bastante alterada a partir do final da década de 1980 com o trabalho de [Lucas \(1988\)](#) em alusão ao crescimento econômico endógeno. A função de produção é agora um processo simultâneo de dois passos, em que recursos são utilizados para produzir educação, a qual entra no processo produtivo de modo que permita retornos crescentes de escala a nível global de economia. Assim, para uma economia qualquer, a educação pode aumentar o capital humano inerente à força de trabalho, que aumentaria a produtividade do trabalhador e que levaria a economia a um nível

de equilíbrio de produto maior.

A educação promove também o aumento na capacidade inovativa de uma economia - conhecimento de novas tecnologias, produtos e métodos - e facilita a difusão e transmissão do conhecimento necessário para entender, processar e implementar novas técnicas idealizadas por terceiros, as quais produzem aumentos na produtividade agregada da economia. Nesses tipos de modelos de crescimento econômico, o setor educacional é o responsável por gerar desenvolvimento contínuo e sustentado a longo prazo, em que as duas fontes principais de acumulação de capital humano (ou aquisição de habilidades) são: a educação e o *learning by doing*. Por fim, a participação do governo via políticas públicas faz-se necessária, uma vez que estas teriam a capacidade de afetar o crescimento através de mudanças na eficiência da escolaridade e na atratividade da educação.

Para analisar os efeitos da educação, os modelos de crescimento endógeno empregam uma medida quantitativa para escolaridade (a mais comum, a média de anos matriculados na escola para toda a força laboral) e uma relação linear positiva significativa entre desenvolvimento econômico e nível de educação. Todavia, para [Hanushek e Wößmann \(2007\)](#), o uso desse tipo de variável como medida do nível educacional implica assumir que qualquer ano a mais de escolaridade gera o mesmo incremento de conhecimento e técnica, não fazendo distinção entre os níveis de ensino (Infantil, Fundamental, Médio ou Superior) e dependência administrativa, além de assumir que, independente do país, a relação dela com o nível de desenvolvimento será constante.

Nesse sentido, é insensato assumir que o estudante médio em Gana ou no Brasil possa vir a ganhar a mesma quantidade de conhecimento por ano de estudo que um aluno médio finlandês ou japonês. Logo, esses problemas podem vir a gerar graves inconvenientes para a estimação de modelos de crescimento econômico que utilizam a educação como indutor de capital humano.

As estimativas utilizadas nos modelos econométricos têm como hipótese básica a relação positiva entre anos de estudo e taxas de crescimento, mas os questionamentos sobre essa causalidade estão presentes na literatura. Uma controvérsia significativa surge em relação ao que é mais importante para estimular o desenvolvimento econômico: a quantidade média de anos de escolaridade (Modelos Clássicos de Crescimento Endógeno) ou a variação nos anos de escolaridade média (Modelos Neoclássicos de Crescimento Endógeno). Não existe consenso sobre qual das duas medidas é preferível para se justificar o avanço econômico. Outro ponto de discussão deriva da comparação entre países. [Vandenbussche, Aghion e Meghir \(2006\)](#) confirmam que a educação é importante para facilitar a pesquisa, o desenvolvimento e a difusão de tecnologias, mas que países em desenvolvimento devem focar em habilidades básicas, uma vez que em locais onde o processo educacional ainda não está consolidado, a imitação é a melhor estratégia para se buscar o avanço estrutural, enquanto que a inovação deve ficar a cargo de países com um sistema de educação mais robusto.

Todos esses modelos propõem um conjunto de correlações no que tange a educação e a economia, de modo que alterações no primeiro afetam o segundo e vice-versa, mas com foco no mercado de trabalho, em especial, no seu rendimento. Porém, a explicação da relação entre ambas mudou com o passar dos anos, ao deixar de ser pautada no efeito do incremento da produtividade do trabalho, para agora atuar em conjunto com fatores estruturais, qualitativos e cognitivos, o que torna a questão da qualidade do sistema educacional primordial para o desenvolvimento econômico de um país, e que coloca a escola numa posição de destaque dentro desse cenário.

## 1.3 Em direção a uma Economia da Educação

### 1.3.1 A perspectiva microeconômica

A literatura microeconômica da educação é caracterizada como o estudo do comportamento racional de agentes individuais que desempenham papéis distantes dentro do sistema educacional. Assim, o modo de agir, de tomar decisões e de interagir de estudantes regulares, professores, administradores, formuladores de políticas públicas e do *modus operandi* das instituições na busca pela situação máxima de bem estar são os objetos de estudo deste campo do conhecimento. Essa mudança de perspectiva caminhou junto com o aprimoramento e a expansão das bases de dados em educação, as quais permitiram análises

em termos de aluno e escola. Para [Correa \(1995\)](#), a avaliação do desempenho desses agentes pode ser feita a partir de dois pontos distintos: o primeiro, que é considerado o método tradicional, investiga o comportamento de cada agente individualmente, isto é, sem permitir a interação com outros agentes; enquanto o segundo, sob um perspectiva mais moderna e dinâmica, permite a interação entre os diferentes agentes racionais.

A ênfase da literatura em Economia da Educação se dá sob a ótica empírica da análise de políticas públicas, onde os trabalhos desenvolvidos por [Coleman \(1968\)](#), [Blaug \(1985\)](#), [Hanushek \(1986\)](#), [Levin \(1989\)](#), [Klees \(1991\)](#), [Psacharopoulos \(1994\)](#), [Akerlof e Kranton \(2002\)](#) e [Psacharopoulos e Patrinos \(2004\)](#) são alguns dos que mais se destacam. Esses trabalhos adicionaram as preocupações de longo prazo relacionadas a custos, efetividade e falhas do modelo corrente de escola, além de interações entre as características sociais, habilidades cognitivas e psicológicas nas funções de produção da educação, alcançando um novo grau de complexidade de análise.

Um dos objetivos da Economia da Educação é explicar o comportamento de indivíduos que desempenham papéis considerados relevantes no processo educacional e, para tanto, diversas funções de produção foram propostas. Sob o ponto de vista histórico, [Hanushek \(1986\)](#) mostra que o estudo das funções de produção educacional<sup>3</sup> examinam a relação existente entre as diferentes combinações de covariadas que geram o melhor resultado possível, os quais são investigações sistemáticas norteadas

---

<sup>3</sup> Também conhecidas como análise de insumo-resposta ou custo-qualidade.

pelo arcabouço econométrico, em oposição a estudos experimentais, que buscam separar os vários fatores que influenciam na performance do estudante.

Nos modelos clássicos, os estudantes tomam a decisão sobre o tempo que irão se dedicar à escola (ou aprendizagem), de modo a equilibrarem as suas respectivas taxas de desconto com custos de oportunidade e qualidade da escola. Assim, o tratamento básico das funções de produção permite uma solução direta para o “custo mínimo”, isto é, a combinação de insumos que podem gerar qualquer resultado ao menor custo possível.

Nesse cenário, transformando o valor da dedicação à escola em uma medida simples de habilidades individuais, como apresentado por [Mincer \(1970\)](#) e, posteriormente por [Psacharopoulos e Patrinos \(2004\)](#), a dificuldade em relação a esta variável de resposta é que assume-se que um ano de escolaridade produz a mesma quantidade de desempenho acadêmico ou de melhoria de habilidade ao longo do tempo em qualquer país. Assim, essa medida apenas conta o tempo gasto na escola sem julgar o que acontece dentro dela, portanto, não fornece uma imagem completa e precisa dos resultados.

Os formuladores de política, os educadores, os estudantes e os familiares estão interessados no sucesso das políticas de educação pública implementadas no país, em função da conexão existente entre esta e a produtividade futura, no qual se faz necessário que o desempenho acadêmico dos estudantes seja o mais alto possível. O processo de aprendizagem pode ser mais eficiente e melhorado através do uso de pesquisas empíricas que

colaborem com a identificação de fatores que levem a melhoria da performance dos estudantes.

Os trabalhos desenvolvidos na concepção da microeconômica da educação mostram três preocupações relevantes no que tange o processo educacional. O primeiro trata do tempo de duração deste, o segundo se refere a qualidade dos recursos empregados nele e, o último apresenta a independência entre o efeito da qualidade dos insumos e o seu tempo de uso.

A primeira pesquisa relevante baseada em técnicas estatísticas de regressão para análise da performance acadêmica foi realizada por [Thomas \(1962\)](#) *apud* [Collier \(1994\)](#). O autor identificou que apenas três variáveis apresentavam uma correlação positiva com o desempenho acadêmico do estudante (medido por notas em avaliações nacionais), a saber: remuneração inicial dos professores, a sua experiência profissional e o número de livros disponíveis na biblioteca da escola. Esse estudo assumia apenas relações diretas entre as variáveis insumo e a variável resultado e formulava uma equação única para calcular a magnitude, a significância e a direção dos efeitos. Apesar da simplicidade, este trabalho foi capaz de questionar as hipóteses assumidas por educadores na definição dos fatores que influenciavam o desempenho estudantil.

### 1.3.2 O Relatório Coleman

A tentativa clássica mais conhecida de se estimar o desempenho acadêmico baseada em princípios microeconômicos foi produzido por [Coleman \(1968\)](#), trabalho este que ficou popularmente conhecido como Relatório Coleman (*Coleman Report*) e que foi concebido como um estudo acerca da distribuição de recursos educacionais nos Estados Unidos por raça e origem étnica. Coleman foi muito além de suas proposições básicas e acabou por criar uma grande base de dados, contendo informações de mais de meio milhão de estudantes, de mais de 3000 escolas, corroborando para a definição das principais variáveis sociais e econômicas que influenciavam o desempenho acadêmico naquele momento.

Todavia, alguns resultados do Relatório Coleman são bastante controversos como, por exemplo, a de que diferenças no desempenho acadêmico de estudantes de diferentes escolas está pouco relacionado as características físicas de cada instituição, mas que está bastante correlacionada com o perfil familiar, no modo como esses estudantes interagem entre si e na habilidade de se expressar do professor, ou seja, o meio social em que o aluno está inserido é mais relevante para o seu sucesso escolar do que as características inerentes à estrutura física da escola. O relatório concluiu também que os recursos destinados ao setor são ineficientemente utilizados provocando reações pouco favoráveis à aceitação da abordagem da função de produção por parte dos formadores de política.

O procedimento estatístico também foi alvo de críticas. O uso de modelos com variância amostral particionada em diversos componentes, devido a diferentes fatores (variáveis), apesar de reportarem uma elevada intercorrelação entre as variáveis explicativas, mostraram-se pouco útil para os formuladores de política, uma vez que esta nada dizia a respeito da direção e da magnitude do efeito. E, por último, uma preocupação também recaiu sobre o uso de dados em corte transversal ao invés de longitudinais, o que comprometia consideravelmente as inferências sobre o passado e o futuro das variáveis de resultado.

Para [Hanushek \(1986\)](#), as críticas refletem tanto uma visão geral contrária a realização de qualquer avaliação empírica evolucionista de todo o sistema educacional quanto a má interpretação de resultados específicos e de problemas analíticos que sempre foram inerentes ao sistema. Hoje, apesar do relatório ser uma das análises mais citadas no campo da Economia da Educação, este é muito mais lembrado por sua importância histórica na questão da investigação das relações entre estudantes, professores e escola nos resultados educacionais associados do que por produzir introspecções a respeito do sistema educacional.

A partir da base de dados produzida pelo Relatório Coleman e para mitigar as críticas em relação a metodologia estatística recebidas por ele, [Jencks et al. \(1972\)](#) reinterpretou tanto a base quanto reestruturou os procedimentos estatísticos, com o auxílio de análises de tendência e do método de mínimos quadrados ordinários (*MQO*). Em concordância com Coleman (1968), mostrou que o ambiente sócioeconômico e as atividades extraclasse nos

quais o estudante está envolvido provavelmente influenciam mais no seu desempenho acadêmico do que as características escolares, como tamanho do orçamento ou grau de experiência dos professores. Entretanto, discordavam sobre qual seria o resultado mais apropriado para a escolaridade, ao invés deste ser demonstrado pela realização acadêmica, como previa os estudos anteriores, deveria ser visto como uma igualdade de oportunidades, sendo definida pelo que é acessível ao aluno em termos de oportunidades educacionais. Os autores concluíram que a visão tradicional sobre a ampliação do investimento direto em escolas estava equivocada, uma vez que não foi percebida correlação entre o que é investido em uma escola e o desempenho acadêmico, e muito menos entre a composição social da escola de Ensino Médio e o sucesso acadêmico.

Outros estudos sucederam a temática de [Coleman \(1968\)](#), [Jencks et al. \(1972\)](#) e [Boardman, Davis e Sanday \(1977\)](#) utilizaram a estratégia de equações simultâneas e análise de regressão de dois estágios (*2SLS*), a qual envolveu a presença de seis variáveis endógenas, a saber: motivação, expectativas de estudantes, pais e professores, eficiência e desempenho acadêmico. Os autores concluíram que o processo educacional resulta de uma grande quantidade de fatores interdependentes e que as atitudes dos estudantes é a variável de maior significância para o seu sucesso, além de que a habilidade de comunicação dos professores possui uma correlação positiva com o desempenho escolar.

[Cohn et al. \(1975\)](#) fizeram o uso de 12 variáveis endógenas, como: nacionalidade, hábitos saudáveis, potencial de

criatividade, desenvolvimento vocacional, conhecimentos básicos de matemática e verbal, entre outras. Uma estrutura de equação foi desenvolvida usando correlações de ordem zero e regressões graduais etapa acima<sup>4</sup>, além das técnicas tradicionais de *MCO* e *2SLS*, constatando que fatores escolares e professores têm uma relação positiva estatisticamente significativa com os resultados dos estudantes.

O modelo econômico que guia a abordagem tradicional da função de produção da educação para princípios microeconômicos é bastante direto. O resultado do processo educacional, isto é, o desempenho acadêmico individual, é diretamente relacionado a uma série de insumos, onde alguns são controlados pelos formadores de política – estrutura física de escolas, grade curricular, professores – enquanto outros não – características e estrutura familiar, capacidade intrínseca de aprendizado do estudante, habilidades cognitivas, entre outros. – Temos então que, enquanto o desempenho escolar pode ser mensurado em qualquer momento discreto do tempo, o processo educacional é cumulativo, onde algum elemento inserido no passado pode afetar os níveis correntes da performance acadêmica dos estudantes e muito mais complexo que uma simples análise de insumo-produto possa sugerir (HANUSHEK, 1986; COLLIER, 1994).

---

<sup>4</sup> Procedimentos estatísticos para comparações múltiplas que são realizados por etapas, em que os resultados da etapa atual influenciam, se houverem, comparações feitas na etapa seguinte, para um subconjunto de médias. Tais procedimentos podem ser de dois tipos: etapa abaixo (*step-down*) e etapa acima (*step-up*).

## 1.4 Determinantes do desempenho acadêmico

### 1.4.1 Tempo de instrução

A temática dos efeitos do tempo de instrução e dedicação aos estudos no desempenho escolar também é destaque na literatura. para [Walberg, Fraser e Welch \(1986\)](#), o tempo de instrução, medido pelo número de semestres em que o estudante cursou alguma disciplina de ciências no Ensino Médio e a quantidade de deveres de casa, apresentara uma correlação positiva com o sucesso acadêmico, assim como o ambiente familiar e escolar. [Betts \(1996\)](#) avaliou o efeito dos deveres de casa mais longos nos resultados das avaliações realizadas nas disciplinas de Matemática e notou que esse é um preditor muito mais significativo na melhora no desempenho em testes relacionados à matéria do que fatores tradicionais, como o tamanho de classe e a experiência de professores, além de apresentar uma produtividade marginal constante em relação às notas. Assim, a inferência lógica do autor é a de que mesmo um professor pouco treinado e que enfrente diariamente uma sala com grande quantidade de alunos consegue fazer com que estes obtenham altos índices de desempenho acadêmico, desde que passe uma grande quantidade de dever de casa.

A partir de informações fornecidas pelo *Illinois Goal Assessment Program (IGAP)*, entre os anos de 1994 a 1997, para mais de 2500 escolas, [Coates \(2003\)](#) propôs um modelo de painel

para verificar a importância do tempo de instrução. Formalmente:

$$S_{jtk} = \alpha + \gamma X_{jt}^d + \mu X_{jt}^s + \beta MPD_{jt} + \delta (MPD_{jt} \times X_{jt}) + \varepsilon_{jtk}, \quad (1.2)$$

em que  $S_{jtk}$  é a nota média no teste *IGAP* da disciplina  $k$  da escola  $j$  no ano  $t$ ,  $X_{jt}^d$  é um vetor de características específicas da localidade da escola  $j$  no ano  $t$ ,  $X_{jt}^s$  é um vetor de características específicas de uma escola  $j$  no ano  $t$ ,  $MPD_{jt}$  representa o vetor de variáveis do tempo de instrução,  $X_{jt} = X_{jt}^d, X_{jt}^s$ ,  $\gamma$ ,  $\mu$ ,  $\beta$  e  $\delta$  (captura os efeitos interativos) são os vetores de parâmetros a serem estimados. Algumas das variáveis utilizadas no modelo foram: o tamanho médio das classes; renda; raça; variável *dummy*<sup>5</sup> para algum tipo de benefício social fornecido pelo governo à família do aluno; experiência e qualificação de professores, entre outras. Quanto às que se referem ao tempo de instrução, valeu-se do percentual de alunos que foram à escola todos os dias e do tempo médio semanal de instrução para as disciplinas de Ciências, Matemática, Estudos Sociais e Inglês.

O exercício empírico promovido por Coates (2003) mostrou, entre vários resultados, que o tamanho médio da sala (número de alunos por classe) não é estatisticamente significativo para a melhora do desempenho acadêmico mas que deve ser inserido de forma não linear na função de produção; que aqueles alunos cuja a família recebe algum tipo de benefício governamental têm

<sup>5</sup> Modo de se introduzir características qualitativas em modelos econométricos, também chamadas de variáveis binárias ou dicotômicas, uma vez que assumem apenas um de dois valores – em geral 0 ou 1 – para indicar a presença ou ausência de determinada característica.

o desempenho acadêmico pouco satisfatório, enquanto que as notas dos alunos que frequentam as aulas diariamente e que tem um tempo de instrução maior, principalmente em Matemática, apresentam uma performance superior a média no exame; e que a interação de salas de aula menores com um maior tempo de exposição as disciplinas tem um efeito ainda mais positivo sobre o desempenho no teste *IGAP*.

A discussão apresentada acerca da inclusão do tempo de instrução na função de produção educacional mostra que é tão relevante quanto o perfil dos professores, dos estudantes, da escola e do ambiente sócioeconômico. E que adicionalmente, o tempo de instrução maior em uma determinada disciplina pode afetar positivamente outras e que esta se contrapõe aos efeitos de salas de aula muito grande. Nessa concepção, a presente análise pode servir como guia e apoio para a reformulação e melhoria de políticas públicas de educação implementadas em vários países.

### 1.4.2 A escola

A Economia da Educação também busca avaliar a relação entre as características das escolas e o sucesso acadêmico. Para tanto, [Hanushek \(1996\)](#) de posse da base de dados longitudinal *High School and Beyond (HS&B)*, propôs o seguinte modelo:

$$A_{ij} = g(T_{ij}, F_{ij}, C_{ij}, S_{ij}), \quad (1.3)$$

onde  $A_{ij}$  é o nível de realização acadêmica para o indivíduo  $i$  da escola  $j$ ,  $T_{ij}$  é um resultado padrão de teste anterior,  $F_{ij}$  é um

vetor de características individuais e familiares,  $C_{ij}$  é um vetor que traz o perfil do ambiente comunitário em que o aluno está, e  $S_{ij}$  é um vetor de características da escola. Faz-se necessário o uso de controles adequados para que as diferenças de perfis familiares, ambiente comunitário e experiência acadêmica anterior do aluno possam ser isolados dos efeitos das características da escola, evitando, assim, problemas de viés durante o processo de estimação.

Uma vez que as bases de dados disponíveis não conseguem fazer essa desvinculação, o viés pode surgir de sua própria estruturação, ou seja, o nível de agregação dos dados pode influenciar na estimação da relação entre desempenho acadêmico e características da escola. Com o intuito de superar esse problema, os autores reestruturam a Equação (1.3), que passa a apresentar a seguinte especificação:

$$A_{ijs} = \alpha_{js} + \beta_{ijs}T_{ijs} + \eta_{ijs}F_{ijs} + \theta_{ijs}C_{js} + \psi_{ijs}S_{js} + \varepsilon_{ijs}, \quad (1.4)$$

na qual o subscrito  $s$  indica o estado (ou município) de residência,  $\varepsilon_{ijs}$  é um erro de previsão do tipo i.i.d.<sup>6</sup>, enquanto que  $S_{js}$  representa uma medida única de qualidade da escola como, por exemplo, investimento por estudante, número de funcionários, etc.

Infelizmente, essa hipótese é uma condição suficiente para agregação perfeita apenas quando o modelo empírico é perfeitamente especificado, e como, na prática essa é uma situação

<sup>6</sup> O termo de erro,  $\varepsilon$ , é uma variável aleatória, supostamente com média zero e variância constante:  $\varepsilon_{ijs} \sim (0, \sigma^2)$ .

muito difícil de acontecer, problemas de viés podem vir a ocorrer. Portanto, a questão que deve tentar ser respondida é se o viés é derivado do nível de agregação dos dados da escola ou não, para que se possa evitar problemas de endogeneidade na seleção da escola.

Hanushek, Rivkin e Taylor (1996) utilizam para estimação o método de mínimos quadrados de dois estágios (2SLS), com foco nos aspectos-chave da agregação da base de dados e dos efeitos dos recursos escolares. Nesse sentido, concluíram que a expansão dos investimentos em escolas com vistas a reduzir a razão professor-estudante e o aumento dos salários dos professores têm pouco impacto sistemático sobre o sucesso acadêmico de estudantes do ciclo básico matriculados em escolas públicas e, ainda, que não há evidências de que a maior disponibilidade de recursos financeiros de uma escola em relação à outra seja mais importante do que os padrões cognitivos do aluno para o sucesso acadêmico.

Por fim, ressaltam que existem diferenças significativas de qualidade entre as escolas, mas que estas também não estão relacionadas às questões financeiras. Esse pressuposto adiciona complexidade ao desenvolvimento de políticas públicas de educação, uma vez que mostra que a simples expansão dos recursos destinados às escolas não garante a melhora do desempenho acadêmico do estudante, e que políticas que busquem alterar a estrutura de incentivos têm uma maior probabilidade de sucesso.

### 1.4.3 Professores

A respeito de políticas públicas educacionais, questões relacionadas ao impacto do professor no desempenho acadêmico também devem ser consideradas. Primeiro, em virtude das interpretações realizadas a partir do Relatório Coleman, realmente existe diferença significativa e sistemática entre escolas e professores quanto às suas capacidades em elevar o desempenho acadêmico? Segundo, quão importante são as diferenças em qualidade de professor na determinação dos resultados dos alunos? E, por último, uma questão também ligada ao estudo clássico de Coleman: se as características observáveis de escolas e professores são capazes de capturar as diferenças qualitativas que influenciam o resultado acadêmico dos alunos, qual será a magnitude desse efeito? Assim, com o objetivo de avançar nos estudos sobre o desempenho escolar e da efetividade de políticas públicas, a partir da década de 1970, foram realizadas diversas investigações estatísticas que buscaram identificar, no processo educacional, a relevância do professor, entre os quais se sobressaem os trabalhos de [Hanushek \(1971\)](#), [Summers e Wolfe \(1977\)](#) e [Rivkin, Hanushek e Kain \(2005\)](#).

A partir de uma amostra de 1061 estudantes, do terceiro ano do Ensino Fundamental, de uma escola da Califórnia, [Hanushek \(1971\)](#) desenvolveu um teste empírico capaz de verificar o nível de importância do professor para o processo educacional. Para tanto, subdividiu a amostra de estudantes em três partes, as quais foram definidas por etnia e ocupação laboral dos pais: a primeira foi composta por estudantes brancos cujos pais desen-

volviam atividades manuais; a segunda também composta por alunos brancos, mas com pais que se dedicavam à atividades em um ambiente de escritório e; o último grupo foi formado por estudantes com descendência mexicana, cujos pais realizavam trabalhos manuais.

O teste foi implementado a partir de um modelo de regressão com variável independente dicotômica,  $T_{ij}$ , para cada professor da amostra, de modo que, se o  $j$ -ésimo estudante tem o  $i$ -ésimo professor,  $T_{ij} = 1$  e  $T_{kj} = 0$ , em que  $k \neq i$ . O modelo completo apresenta a seguinte especificação:

$$A_{j3} = \sum_{i=1}^n t_i T_{ij} + \alpha S_j + b A_{j2} + u_j, \quad (1.5)$$

em que o desempenho do  $j$ -ésimo aluno da 3ª Série ( $A_{j3}$ ) é função do resultado obtido na 2ª Série ( $A_{j2}$ ), do gênero  $j$ -ésimo estudante e  $S_j$ , sendo 1 para feminino e 0, caso contrário e da capacidade de se expressar do professor  $T_{ij}$ .

Os resultados das estimações sugeriram que a performance dos estudantes brancos é influenciada tanto pelas qualidades específicas do professor, enquanto que, para os descendentes mexicanos, estas não são importantes. As características do professor mais relevantes para o processo educacional foram a sua qualidade verbal, o quão recente foi a sua graduação e a sua capacidade de manter a ordem, a disciplina e a interação dos alunos em sala de aula. Todavia, percebeu-se que a experiência profissional e títulos de mestrado ou doutorado não contribuem para a melhora da performance acadêmica dos estudantes e que, portanto, a política de contratação de professores está equivocada,

uma vez que ela dá importância às características que parecem não estar correlacionadas com a melhora acadêmica.

Sucintamente, as Equações (1.4) e (1.5) sugerem que o desempenho acadêmico (A) é uma função das qualidades genéticas do estudante associada a seu *status* socioeconômico, as qualidades do professor e da escola e as influências de grupo, sendo que a maioria das investigações empíricas realizadas nesse sentido as interpreta como uma função de produção baseada em hipóteses de minimização de custos.

Entretanto, essa perspectiva foi contestada por [Summers e Wolfe \(1977\)](#), que enumeraram uma série de razões que explicariam o porquê de não se encontrar com clareza uma relação entre insumo e produto em educação. Em primeiro lugar, ao se utilizar funções de produção tradicionais, em que o próprio consumidor (características inerentes do aluno) é insumo, torna-se difícil a separação dos efeitos endógenos dos de contexto. Em segundo lugar, no processo educacional nem todos os insumos podem ser controlados pelo formulador de política e, praticamente não há evidências de que todas as escolas alcancem tal eficiência produtiva. Em terceiro lugar, o uso em estimações de variáveis apresentadas em médias gerais, ao invés de dados per capita, promove uma desagregação menor dos dados, podendo levar a resultados menos plausíveis em relação aos insumos. Ademais, a segmentação das amostras e as técnicas de estimações condicionais não permitem que as características dos estudantes variem com o efeito das variáveis de controle das políticas.

Com o objetivo de se testar a hipótese de que a desagregação de dados traria informações mais significantes sobre as variáveis que afetam o desempenho acadêmico, como foco naquelas influenciadas diretamente pelo poder público (professor, estrutura da escola, currículo), [Summers e Wolfe \(1977\)](#) utilizaram uma base de dados que engloba mais de 2000 alunos de 103 escolas diferentes da cidade da Filadélfia (EUA), as quais foram selecionadas aleatoriamente, de diferentes séries. A variável dependente selecionada foi a alteração no índice de aproveitamento para os últimos três anos, enquanto as independentes foram de três tipos: características genéticas e socioeconômicas do aluno (QI, raça, renda familiar); características escola-aluno (tamanho da sala, qualidade do professor, etc.); e características de grupo (proporção de negros, de estudantes com notas superiores a média da escola, etc.).

Apesar da relevância de todas essas variáveis, os autores se dedicaram, em especial, a identificar quais características dos professores mais colaboram com a melhora da performance acadêmica. Notou-se que aqueles professores que se graduaram em universidades mais renomadas estavam associados com estudantes com taxas de desempenho melhores e, entre seus alunos, os de menor renda familiar foram os que apresentaram a maior taxa de crescimento nos resultados escolares. Corroborando com [Hanushek \(1971\)](#), a questão da experiência foi tida como irrelevante, apesar de não ser crível afirmar que esta característica impacte de forma homogênea<sup>7</sup> sobre os alunos nos diversos níveis de ensino.

---

<sup>7</sup> Foi percebido que estudantes com notas piores evoluíam mais com pro-

Por último, concluíram que a nota do *National Teacher Exam*<sup>8</sup>, utilizada pelo Estado para avaliar a qualidade do professor tinha um efeito negativo na questão do aprendizado e que outras medidas, e não apenas esta, deveriam ser usada em conjunto para se aferir o potencial do professor.

Através de um modelo que fosse capaz de estimar as diferenças nas contribuições dos professores no aprendizado estudantil e que eliminasse as fontes de contaminação ou viés oriundas do processo de seleção de estudantes pela escola e das atividades práticas dos professores, Rivkin, Hanushek e Kain (2005) realizaram uma análise empírica a partir de uma base de dados longitudinal com informações sobre o desempenho acadêmico de mais de meio milhão de estudantes, de 3000 escolas públicas do estado do Texas.

Para evitar os problemas de variáveis omitidas e de vício de agregação dos modelos lineares clássicos, esse trabalho focou nos determinantes da taxa de aprendizagem para um período específico e fez uso de modelos de valor adicionado, formulação esta que elimina a influência de uma grande quantidade de variáveis perturbadoras<sup>9</sup> e que controla as variações nas condições iniciais, para assim, elucidar como as escolas podem influenciar

---

fessores novos, enquanto que melhores estudantes tinham um melhor rendimento com professores mais experientes.

<sup>8</sup> Conjunto de provas que são realizadas por professores recém-graduados como parte de um processo de certificação exigido por vários estados e organizações profissionais norte-americanas.

<sup>9</sup> Também conhecidas como *confounding variables*. No modelo apresentado, se referem à informações passadas, não observadas e históricas de características familiares e escolares.

a performance acadêmica em um período pré-determinado.

O modelo paramétrico convencional de valor adicionado proposto por Rivkin, Hanushek e Kain (2005) é similar aos exercícios econométricos propostas por Hanushek (1971) e Summers e Wolfe (1977), e descreve a taxa do ganho em desempenho acadêmico ( $\Delta A_{ijgs}^c$ ) para o indivíduo  $i$  da classe  $c$  com o professor  $j$  na série  $g$  da escola  $s$  da seguinte maneira:

$$\Delta A_{ijgs}^c = A_{ijgs}^c - A_{ij'g-1s'}^c \quad (1.6)$$

$$\Delta A_{ijgs}^c = \beta_X X_{ig}^c + \beta_T T_{jgs}^c + \beta_S S_{gs}^c + f_i + \varepsilon_{ijgs}^c, \quad (1.7)$$

O ganho em desempenho, medido pela diferença de notas de testes obtidas pelos estudantes nas séries  $g$  e  $g - 1$ , é função do histórico familiar ( $X$ ), das características do professor ( $T$ ), da escola ( $S$ ) e habilidades intrínsecas do estudante ( $f$ ). As características inerentes aos estudantes referem-se a um conjunto de habilidades cognitivas, motivação e traços de personalidade que influenciam a taxa de desempenho, mas que não variaram ao longo dos anos pesquisados.

Em virtude de problemas de erros de medida e como os processos de seleção de escolas, turmas e alocação de professores, normalmente não estão disponíveis para os pesquisadores, as estimações do modelo podem estar sujeitas a problemas de viés. Assim, com o intuito de não definir estatisticamente as covariadas do processo educacional, os autores desenvolveram as variáveis em termos de efeito sistemático total de estudantes,

famílias, professores e escolas, transformando a abordagem paramétrica apresentada na Equação (1.7) em uma semiparamétrica, a qual decompõe a produção educacional durante a série  $g$  em um conjunto de fatores fixos e variantes no tempo:

$$\Delta A_{ijgs}^c = \gamma_i + \theta_j + \delta_S + u_{ijgs}^c, \quad (1.8)$$

onde a melhora no desempenho acadêmico ao longo da série  $g$  é tida como uma função aditiva de fatores fixos do aluno ( $\gamma$ ), do professor ( $\theta$ ), da escola ( $\delta$ ) e de um componente aleatório que varia no tempo,  $u$ . Os componentes fixos do estudante que afetam a sua taxa de aprendizagem retratam características familiares, como o nível de escolaridade e renda dos pais, fatores escolares, tais como o currículo escolar, dos recursos financeiros e estrutura física; e, por fim, ao perfil do professor, que busca captar a sua qualidade média.

Na estrutura semiparamétrica, a variância de  $\theta$  indica as alterações na qualidade do professor em termos de ganhos em desempenho acadêmico dos estudantes, a qual é obtida a partir de um método de análise comparativa que usa informações sobre a sua rotatividade<sup>10</sup> e da média de melhora nos resultados escolares, para gerar uma estimativa de limite inferior da variância intraescola da qualidade do professor.

Para a realização da técnica, deve-se primeiramente reescrever as variáveis da Equação (1.8) em termos de seus valores médios, lembrando que  $\delta$  não se altera por se tratar de uma análise

<sup>10</sup> Tem-se por hipótese que a alternância de professores deve aumentar a variância nos resultados escolares entre turmas e séries da mesma escola.

intraescolar:

$$\overline{\Delta A_{gs}^c} = \overline{\gamma_{gs}^c} + \overline{\theta_{gs}^c} + \delta_S + \overline{u_{gs}^c}. \quad (1.9)$$

A partir da equação acima, pode-se comparar o ganho médio em desempenho acadêmico para alunos da mesma série, mas de salas diferentes ( $c$  e  $c'$ ):

$$\overline{\Delta A_{gs}^c} - \overline{\Delta A_{gs}^{c'}} = \left( \overline{\gamma_{gs}^c} - \overline{\gamma_{gs}^{c'}} \right) + \left( \overline{\theta_{gs}^c} - \overline{\theta_{gs}^{c'}} \right) + \left( \overline{u_{gs}^c} - \overline{u_{gs}^{c'}} \right). \quad (1.10)$$

Perceba que essa diferença de médias elimina os fatores fixos das escolas que afetam a taxa de crescimento do desempenho acadêmico, mas a performance escolar continua sendo uma função das características dos estudantes, professores e de um termo de erro. Por fim, a partir de transformações matemáticas, da limitação da amostra, os estudantes que permaneceram na mesma escola entre as séries  $g$  e  $g'$  e da imposição de hipóteses que caracterizam formalmente o conceito de que os professores foram selecionados aleatoriamente para as classes durante o período de análise, os autores eliminaram os efeitos fixos dos estudantes e, assim, identificaram a magnitude da variância intraescolar da qualidade do professor no crescimento do desempenho acadêmico:

$$E \left[ \left( \overline{\Delta A_{gs}^c} - \overline{\Delta A_{g's}^c} \right) - \left( \overline{\Delta A_{gs}^{c'}} - \overline{\Delta A_{g's}^{c'}} \right) \right] = 4 \left( \sigma_{\theta_s}^2 - \sigma_{\theta_s \theta_{s'}} \right) + E(e_s), \quad (1.11)$$

em que  $\sigma_{\theta_s}^2$  é a variância da qualidade do professor na escola  $s$  e  $\sigma_{\theta_s \theta_{s'}}$  é a covariância entre turmas diferentes da mesma escola para esta particularidade.

Rivkin, Hanushek e Kain (2005) enfatizam que o Relatório Coleman popularizou a discussão que aspectos familiares são mais relevantes que recursos escolares para a performance acadêmica, mas que tal questionamento é irrelevante para a formulação de políticas públicas educacionais, uma vez que o foco deveria ser verificar se os benefícios produzidos pela intervenção justificam os custos. Assim, a partir da abordagem de efeitos fixos, os autores concluíram que as características observáveis dos professores e das escolas são significativas para o desempenho acadêmico, e que são capazes de aumentar a performance de estudantes em situação socioeconômica adversa, principalmente a dos mais jovens, e também que uma sucessão de professores qualificados pode substancialmente reduzir o hiato no desempenho acadêmico entre os estudantes de classes sociais diferentes.

Corroborando com outros estudos, verificou-se que a experiência e a titulação acadêmica dos professores não foram estatisticamente significantes para a melhora na performance acadêmica nas disciplinas de Matemática e Leitura. Todavia, a quantidade de alunos apresentou-se estatisticamente significativa, porém modesto, diluindo-se com o passar dos anos.

#### 1.4.4 A estrutura curricular

Afora as características relacionadas à escola e à família, a organização do currículo escolar também é destaque na análise dos determinantes do desempenho acadêmico e nas implicações da teoria do capital humano. Em ambos, o currículo tem valor, dada a sua capacidade de agregar técnicas e habilidades que influ-

enciam tanto no curto prazo, através da melhora da performance acadêmica, quanto a longo prazo, ao tornarem os estudantes mais produtivos e melhores remunerados no mercado de trabalho.

Essas perspectivas estão centradas em duas alternativas de desenho curricular: a primeira, um núcleo comum e restrito de disciplinas que visa ser apropriada para todos; e a segunda, um conjunto diversificado e compreensivo de disciplinas que atenda e compreenda diferenças de habilidade e de interesse dos alunos (ALTONJI, 1992; LEE; SMITH, 1995; ROSE; BETTS, 2004; OAKES, 2005).

O desenho tradicional de currículo do Ensino Médio, que foi institucionalizado no início do século XX nos Estados Unidos com base na Unidade Carnegie<sup>11</sup>, enfatiza a tendência homogenizadora em relação às experiências estudantis, técnicas de aprendizagem, estruturas e regras que organizam o modo de trabalho das instituições. Essa formulação permitiria um melhor sequenciamento do processo de aprendizagem e classificação dos alunos por proficiência, promovendo a equidade no sistema ao impor que alunos da mesma sala de aula e série, seja ele pobre ou rico, menino ou menina, recebessem do mesmo modo toda a oferta de conteúdo dada pelo professor. Ao impor essa padronização, as pessoas assumem que entendem como o sistema educacional funciona e que, portanto, a convergência entre essa

---

<sup>11</sup> Sistema de medida desenvolvido por *Andrew Carnegie* (1835 - 1919) que vinculava a obtenção do crédito acadêmico ao tempo de interação direta entre aluno e professor. A unidade padrão foi definida em 120 horas de contato, i.e., uma hora de instrução por dia, cinco dias por semana por 24 semanas.

crença cultura e forma organizacional promove legitimidade e apoio popular, o que corrobora com a explicação para as similaridades entre as escolas das mais variadas localidades (TYACK; TOBIN, 1994).

Uma das primeiras análises empíricas sobre o tema foi realizada por Alexander e Pallas (1984) que verificaram uma relação positiva entre o programa curricular americano para o Ensino Médio *New Basics*<sup>12</sup> e a performance no *Scholastic Assessment Test (SAT)*<sup>13</sup>. Os autores concluíram que as disciplinas de Inglês e de Estudos Sociais não afetam o desempenho nos seus respectivos testes, mas que existe uma correlação positiva entre a conclusão dos cursos de Ciência e, principalmente, Matemática no desempenho na prova.

Para Rogers (1997), o currículo tradicional não condiz com o mundo contemporâneo e as expectativas dos estudantes, prejudicando os melhores, em razão de sua natureza fragmentada, da qualidade fixa e unidimensional do conhecimento que propagam, da passividade imposta ao aprendizado do aluno, na rigidez de tempo e espaço de aula e do distanciamento do conteúdo pedagógico em relação as demandas diárias dos alunos, características essas que demonstram a necessidade de se buscar alternativas a esse modelo.

---

<sup>12</sup> Consiste em quatro anos de Inglês, três de Matemática, Ciências e Estudos Sociais, dois de Língua Estrangeira e seis meses de Cursos de Informática.

<sup>13</sup> Exame educacional instituído em 1926 nos Estados Unidos aplicado a estudantes do Ensino Médio, que serve de parâmetro de admissão nas universidades norte-americanas, o qual é subdividido em dois: o *SAT Reasoning Test* (de raciocínio) e o *SAT Subject Test* (de matérias). O ENEM se assemelha a esse teste.

Assim, os novos padrões de organização e procedimentos com ênfase em descentralização, ampla participação e promoção de habilidades individuais caracterizam praticamente todas as estratégias de reestruturação curricular do Ensino Médio. Conforme Lee, Smith e Croninger (1997), essa perspectiva, ao ser emoldurada a partir de um modelo de escolha e do fornecimento de uma ampla gama de opções de disciplinas, faria com que a escola seguisse os preceitos democráticos ao ofertar estudos úteis a todos, ao invés de abstrações acadêmicas que interessam apenas a alguns e, assim, atingiria um dos principais objetivos das reformas educacionais: a de reduzir os problemas de repetência e evasão escolar.

Em uma perspectiva internacional, a mudança para o currículo compreensivo e diversificado se apresenta como tendência. Na Alemanha, estudantes a partir dos 10 anos de idade são direcionados a diferentes tipos de escola<sup>14</sup>, de acordo com suas capacidades e perspectivas de carreiras profissionais, enquanto que na França, Suíça e Inglaterra, essa filosofia educacional também prevalece, apesar de ainda haver debates sobre as medidas práticas referentes a transmissão de conhecimento e o agrupamento de habilidades que sugere esse desenho de currículo (TERWEL, 2005).

---

<sup>14</sup> A escola secundária alemã são de três tipos: *Hauptschule*, *Realschule*, *Gymnasium*. Para maiores detalhes, ver <<http://www.dw.com/pt-br/o-sistema-de-ensino/a-1015639-0>>.

Nos Estados Unidos, as reformas educacionais dos anos de 1980, apoiadas no currículo flexível, na avaliação de competências e de responsabilização, a diferenciação foi introduzida em todos os níveis de ensino e as disciplinas foram sequencialmente alinhadas (*tracking*), sendo acompanhadas por uma expansão do currículo, tanto horizontalmente (ofertas múltiplas para a mesma sequência de disciplinas) quanto verticalmente (níveis diferentes de exigência para a mesma disciplina) (LEE; SMITH, 1993; OAKES, 2005; TERWEL, 2005).

Em contraste com a visão majoritária dos formuladores de política sobre o currículo compreensivo, Lee e Smith (1993) demonstraram que essa estrutura cria diferenças substanciais, entre estudantes da mesma escola, no que estudam e aprendem. Para tanto, utilizaram um modelo linear hierárquico para investigar como a organização curricular do Ensino Médio afeta a performance acadêmica, controlando características regionais, escolares e familiares, para uma amostra de 3096 alunos do último ano do Ensino Médio, de 123 escolas. Os resultados evidenciaram que quanto maior a diferenciação curricular, mais prejudicados são os alunos em pior situação acadêmica, uma vez que eles acabam por optar por disciplinas com conteúdos menos exigentes e que a organização curricular, na qual a maioria dos estudantes cursem as mesmas disciplinas e que o conteúdo destas sejam acadêmicos por natureza (e.g. menos escolha e variabilidade), parece promover maiores e equânimes níveis de aprendizagem.

Com o intuito de investigar a relação entre as possibilidades de *tracking* e os resultados em testes internacionais padronizados, Hanushek et al. (2006) realizaram estimativas econométricas baseadas na metodologia de diferenças em diferenças para identificar o quão cedo é recomendável ou não a prática desse decisão curricular. Os autores fizeram uma comparação internacional entre os resultados obtidos na prova de Leitura do teste PISA<sup>15</sup> e de compreensão em Leitura do PIRLS<sup>16</sup>, assim sendo, concluíram que o início precoce da diferenciação escolar leva a um aumento sistemático da desigualdade no desempenho acadêmico entre o fim da escola primária e o termino do 8º ano, além de não encontrarem evidências significantes que essa opção de estrutura curricular influencia positivamente nos resultados dos exames acima citados.

A partir de uma base de dados que contempla mais de 180000 alunos, de 27 países, Woessmann et al. (2009) investigou os efeitos da associação entre características familiares e o momento da diferenciação curricular no desempenho nas provas de Matemática e Ciências do PISA 2003. Os resultados qualitativos mostraram que quanto mais avançada for a idade quando realizado o *tracking*, menores serão os efeitos das características

---

<sup>15</sup> *Programme for International Student Assessment*, desenvolvido pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) para avaliar as competências dos jovens de 15 anos em Leitura, Matemática e Ciências. Os testes são aplicados a cada três anos.

<sup>16</sup> *Progress in International Reading Literacy Study* é uma avaliação internacional sobre a compreensão da Leitura dos alunos do 4º ano do Ensino Fundamental, desenvolvida pela *International Association for the Evaluation of Educational Achievement (IEA)*.

sócioeconômicas sobre o desempenho dos estudantes em testes padronizados e que jovens provenientes de famílias com uma condição econômica mais favorável obtêm resultados semelhantes.

Quanto a análise empírica da relação entre a estrutura curricular e o nível de salários, uma dos primeiros exercícios que investigaram esse efeito de longo prazo foi feito por [Altonji \(1992\)](#). Para ele, a dificuldade em se realizar esse tipo de trabalho deriva, primeiro, pela pouca disponibilidade de observações que contemplam: informações sobre o currículo; experiências no mercado de trabalho uma década ou mais após a diplomação no Ensino Médio; e a grande quantidade de variáveis de controle que se deve ter para evitar problemas de viés de seleção. Segundo, que existem sérias dificuldades estatísticas em se separar o efeito real de cada bloco de disciplinas nas taxas de salários, de correlações espúrias oriundas de fatos não observáveis comuns tanto aos cursos quanto aos rendimentos. Por exemplo, a quantidade de horas dedicadas ao longo do Ensino Médio a Matemática e Ciências, por um estudante particular, é presumida ser positivamente correlacionada com sua habilidade acadêmica e suas preferências de Ensino Superior e, nãis, o número de disciplinas cursadas é provavelmente endógena a realização educacional e a equação de salários, dado que os controles disponíveis para habilidade e preferências são imperfeitos.

Tendo como objeto de informações, a base de dados de acompanhamento longitudinal de 1986, para a turma diplomada no Ensino Médio no ano acadêmico de 1971-72, de 897 escolas norte-americanas, a qual trazia observações sobre o números de

horas estudadas por semestre em várias disciplinas, características da família e da escola, resultados de provas e preferências dos pais em relação a educação. O trabalho buscou estimar os efeitos sobre os salários da variação de currículo interescolar pelo método *2SLS*, usando como variável instrumental para as disciplinas escolhidas pelos indivíduos a quantidade média de matérias cursadas por tema, enquanto controlava outras características dos estudantes e da estrutura da escola. O autor concluiu que o currículo tem um efeito extremamente fraco sobre o nível de salários e que um ano a mais de Ciências, Matemática ou Inglês levaria a um aumento de apenas 0.3%.

A visão tradicional (FENNEMA; SHERMAN, 1977; ADELMAN, 1991) sobre as preferências femininas indicam que, em média, elas decidem procurar por disciplinas mais gerais e menos técnicas, além de terem uma menor propensão a escolher carreiras técnicas do que os meninos, o que implicaria nas disparidades de rendimentos futuros entre os sexos. Sob essa perspectiva, Levine e Zimmerman (1995) propuseram identificar os efeitos das decisões de currículo tomadas por ambos os gêneros sobre seus respectivos salários e na probabilidade de alocação no mercado de trabalho, em especial, em empregos que exigem uma maior expertise, além de examinar se um ano a mais de dedicação a disciplinas de matemática e ciências conduz a maiores ganhos no mercado de trabalho. Para tal experimento, recorrem a duas bases de dados longitudinais: a primeira, a *National Longitudinal Survey of Youth (NLSY)* do ano de 1979, que contemplou informações de mais de 12000 homens e mulheres, entre as idades

de 14 e 21 anos; e, a segunda, a *High School and Beyond (HSB)*, com dados sobre mais de 10000 alunos do último ano do Ensino Médio em 1980, os quais foram novamente entrevistados em 1986. Dito isso, os modelos *OLS* propostos foram:

$$\ln W_i = \beta_{1g}X_i + \beta_{2g}F_i + \beta_{3g}C_i + e_i, \quad (1.12)$$

$$\ln W_i = \beta_{1gj}X_i + \beta_{2gj}F_i + \beta_{3gj}C_i + e_i, \quad (1.13)$$

em que os índices  $i = 1, \dots, N_{gj}$ ,  $g = 1, 2$  e  $j = 1, \dots, 3$  representam respectivamente, o indivíduo, o gênero e a titulação educacional (Ensino Médio, Ensino Superior de curta duração ou completo),  $N$  o número total de integrantes de cada subgrupo da população da pesquisa,  $\ln W_i$  o logaritmo natural dos salários,  $X_i$  o vetor de características observáveis dos indivíduos (idade, estado civil, nível de educação, etc.),  $F_i$  o vetor de características familiares (quantidade de irmãos, nível educacional dos pais, etc.) e  $C_i$  a quantidade absoluta de disciplinas cursadas, em separado, nas áreas de Matemática e Ciências.

Os autores chamam a atenção para a possibilidade de problema de viés na estimação de  $\beta_3$ , inerentes as especificações das Equações (1.12) e (1.13). Para evitar essa questão econométrica, foram utilizadas medidas extensivas para as três categorias de variáveis explicativas expressas nessas equações e adicionado um novo vetor de variáveis ( $S_i$ ) que contém os resultados de testes em diversas áreas, que pode-se ser interpretado como uma *proxy*<sup>17</sup> para a aptidão do aluno. O modelo *OLS* a ser estimado

<sup>17</sup> É uma variável que atua no lugar de uma outra não observável ou não

toma a seguinte forma:

$$\ln W_i = \beta_{1gj}X_i + \beta_{2gj}F_i + \beta_{3gj}C_i + \beta_{4gj}S_i + e_i. \quad (1.14)$$

Entretanto, a inclusão de um vetor de resultados de testes pode ocasionar problemas de endogeneidade na estimação, uma vez que é possível que estudantes que decidiram por mais cadeiras de Matemática e Ciências utilizem os conhecimentos nelas adquiridos para obterem um desempenho melhor nos exames. Formalmente, um modelo da performance nos exames seria:

$$S_i = \gamma_1 X_i + \gamma_2 F_i + \gamma_3 C_i + \gamma_4 A_i + \nu_i \quad (1.15)$$

em que  $A_i$  é um vetor de características não observáveis que influencia os salários e  $\nu_i$  um termo de erro aleatório.

Os resultados das estimações *OLS* para as diversas especificações mostraram que os efeitos de um currículo com mais disciplinas de Matemática e Ciências sobre os salários, aproximadamente 6 anos após conclusão do Ensino Médio, são dispersos para os homens, enquanto que, para as mulheres, em média, o benefício é limitado a um aumento de 2% por semestre adicional de matemática. Já para aquelas que se formaram em curso superior, apresentam um incremento salarial de mais de 5%. Quanto às cadeiras de Ciências, estas não apresentaram nenhuma influência sobre os níveis de salários tanto dos homens quanto das mulheres.

---

mensurável, e que deve apresentar forte correlação, não necessariamente linear, com a variável que se busca analisar.

Com o objetivo de identificar as consequências do currículo do Ensino Médio sobre os hiatos salariais existentes entre raças, etnias e gêneros, [Rose e Betts \(2004\)](#) estimaram os efeitos de cursos específicos de Matemática (Matemática Vocacional, Álgebra, Geometria e Cálculo) nos salários, na década seguinte a graduação neste nível de ensino, para um conjunto de 11724 alunos que eram *sophmores*<sup>18</sup> em 1981, inseridos na base *HSB*.

O modelo log-linear proposto para estimar os proventos de 1991 para o estudante  $i$  na escola  $s$  é:

$$\ln \text{earn}_{i,s} = \alpha + \beta_0 \text{Curric}_{i,s} + \beta_1 \text{Demo}_{i,s} + \beta_2 \text{Fam}_{i,s} + \beta_3 \text{Sch}_{i,s} + \beta_4 \text{HiDeg}_{i,s} + \varepsilon_{i,s}, \quad (1.16)$$

em que  $\text{Curric}_{i,s}$  quantifica o vetor de créditos recebidos em cada tipo de curso de Matemática,  $\text{Demo}_{i,s}$  se refere às informações demográficas,  $\text{Fam}_{i,s}$  e  $\text{Sch}_{i,s}$  tratam das características familiares e escolares respectivamente,  $\text{HiDeg}_{i,s}$  representa um conjunto de variáveis *dummy* que indicam a titulação acadêmica máxima atingida em 1992 e  $\varepsilon_i$  é um termo de erro do tipo *iid*, enquanto o vetor de coeficientes,  $\beta$ , pode ser interpretado como a mudança percentual nos proventos associada ao ano adicional de cursos específicos de matemática<sup>19</sup>. O procedimento econométrico utilizado foi o *OLS* e para se verificar a robustez dos parâmetros estimados e corrigir eventuais problemas de viés, os dados também foram submetidos às técnicas *IV* e de Efeitos Fixos ao modelo log-linear proposto.

<sup>18</sup> Equivalente ao 2º ano do Ensino Médio no Brasil.

<sup>19</sup> A variação percentual exata é dada por  $(e^\beta - 1) \times 100\%$  onde  $\beta$  é o coeficiente da regressão.

Inicialmente, foi analisada uma versão simplista para a Equação (1.16), que não considerou nenhuma variável de controle observável a não ser as disciplinas de Matemática, com o intuito de identificar os efeitos e significância de cada uma destas sobre as variações médias dos salários dos trabalhadores. Os coeficientes estimados foram grandes e variaram bastante em função do grau de dificuldade do curso e mostraram que o crédito/ano adicional de Cálculo prediz um aumento salarial de mais de 19%, enquanto Álgebra/Geometria de cerca de 8% e cursos vocacionais e introdutórios praticamente não têm impacto algum.

Esses resultados mudaram drasticamente ao se adicionar variáveis de controle para características escolares, familiares e demográficas. O efeito dos cursos de Álgebra e Geometria foram mitigados em até 38% e os básicos se mostraram ainda mais insignificantes, em especial, o de Matemática Vocacional que apresentou uma relação negativa estatisticamente significativa (ao nível de 5%) com o nível dos salários. O resultado negativo não implica que o ano adicional de cursos vocacionais prediz uma retração nos proventos em relação aos alunos que cursaram mais disciplinas técnicas, e sim, na comparação em relação aqueles que optaram pelo currículo médio, o qual contemplou, em média, 3 bimestres para esta categoria de curso. Assim, uma maior dedicação a cursos de baixa complexidade pode ser interpretado como um sinalizador negativo de habilidade, posto que os estudantes que optaram por esse currículo pagaram o custo de oportunidade de não frequentarem cursos mais avançados, ao não desenvolverem o nível de capital humano exigido pelo mercado de trabalho.

Para melhor esclarecer a trajetória temporal dos efeitos desse tipo de currículo, os autores incluíram no modelo um conjunto de variáveis binárias que indicavam o nível de escolaridade máxima obtida em 1992, e notaram que, através desse controle, os valores dos coeficientes do currículo se reduziram à metade. Segundo a teoria da sinalização, essa queda indica que, aproximadamente, metade do efeito geral dos cursos de matemática do Ensino Médio é verificada através do melhor preparo para ingresso no Ensino Superior e, portanto, demonstraram sua maior capacidade técnica aos futuros empregadores, enquanto que a interpretação baseada na teoria do capital humano diz que quanto maior o número de cursos de matemática, mais eficiente será o aluno, maior as suas chances de ingressar em uma universidade, de modo a aumentar a sua produtividade futura. Nesse modelo, um curso de Álgebra/Geometria aumenta os ganhos em 3,1% enquanto que o curso de Cálculo mais do que dobra esse percentual, chegando a 6,5%.

Apesar da impossibilidade de se incluir completamente aspectos de motivação e de habilidade nas análises de desempenho acadêmico e de remuneração futura, em ambas as situações, elas se correlacionam positivamente com a variável de interesse, o que pode provocar problemas de viés de estimação dos parâmetros, sobretudo em modelos de unidade desagregada, como o apresentado por [Rose e Betts \(2004\)](#), os quais adotaram duas estratégias para essa problemática, a de variáveis instrumentais e a de *OLS* com efeitos fixos para escolas. A primeira técnica, similar à adotada por [Altonji \(1992\)](#) (1992), procurou eliminar

do currículo a parte que está correlacionada com as habilidades e motivações do aluno, a partir do uso da média escolar de créditos, para cada uma das seis categorias de cursos de matemática como instrumento para a quantidade de créditos obtidos em cada uma dessas categorias por aluno. Assim, o currículo escolar médio foi utilizado como preditor do currículo atual do estudante, sendo a diferença entre eles provocada pelas variações de habilidade. Os resultados estimados por *2SLS* para a especificação *IV* identificaram um efeito negativo estatisticamente significativo sobre os salários das disciplinas de Matemática Vocacional, enquanto que os créditos em Álgebra, Geometria e Cálculo apresentaram uma magnitude similar positiva de cerca de 9%, valores estes que foram superiores aos obtidos pelas estimativas *OLS*.

O segundo teste de robustez buscou controlar os efeitos de variações de habilidade entre escolas e indicou que metade dos valores dos coeficientes curriculares são maiores que o modelo *OLS* normal, enquanto que os parâmetros de Álgebra e Geometria mantiveram-se praticamente os mesmos e o de Matemática Vocacional não é mais significativo.

O trabalho de [Rose e Betts \(2004\)](#) também examinou os efeitos das diferenças nos currículos do Ensino Médio sobre hiatos salariais entre raças, etnias e nível de renda. Os resultados mostraram que latino-americanos e negros ganham em média de 5% a 10% menos que brancos, respectivamente. Em relação às diferenças de renda entre estudantes, os da classe mais baixa predizem um provento 30% menor do que aqueles de classe média que, por sua vez recebem 10% menos que os alunos mais

abastados. O currículo enfatizado em cursos de Matemática aparentou ser responsável por cerca de um terço do hiato salarial entre estudantes da classe mais baixa e da média, mas este em nada contribuiu para a explicação das diferenças entre esse extrato social e a classe mais rica.

No que tange as ações governamentais, os resultados em relação à renda familiar mostram que mesmo havendo outros fatores que possam contribuir na melhor alocação no mercado de trabalho, estudantes de baixa renda podem consideravelmente aumentar suas perspectivas salariais através de políticas públicas que encorajem e preparem esses alunos a optarem por currículos escolares mais exigentes e técnicos.

As principais conclusões extraídas desse trabalho tratam da relevância dos cursos de Matemática nos currículos escolares do Ensino Médio. [Rose e Betts \(2004\)](#) verificaram que a quantidade e o nível de disciplinas que o estudante cursa nesta área estão fortemente correlacionadas com os seus ganhos salariais dez anos após a graduação, neste nível de ensino, e que o desenho curricular que inclua mais disciplinas de Álgebra e Geometria são preferíveis a qualquer outra proposta, quando o objeto final é o sucesso no mercado de trabalho.

Todavia, existem outras razões, mais gerais, que evidenciam a importância de se entender os efeitos da organização curricular do Ensino Médio e das políticas públicas. Primeiro, caso ele venha a ter pouca influência sobre o desempenho acadêmico, talvez uma intervenção seja necessária em um estágio de ensino anterior. Segundo, com a crescente valorização do Ensino

Superior, a redução de políticas afirmativas pode limitar o acesso de alguns grupos étnicos a ele, podendo exacerbar os problemas de desigualdade de renda em vários países. Terceiro, qual seria o efeito da especialização curricular entre as escolas de tamanhos diferentes e até que ponto essa diversificação curricular é satisfatória, dadas as discrepâncias do sistema. E, por fim, a temática da mudança é recorrente nas orientações oficiais para a educação, mas como fazer essa mudança na prática, ou seja, como trazer para a escola, como inseri-la na formação dos professores e educadores pedagógicos.

Os trabalhos apresentados acima mostraram que as questões da organização curricular do Ensino Médio são bastante complexas, uma vez que seus reflexos se manifestarão sobre as pessoas por um longo período de tempo, o que torna as políticas públicas que incidem sobre o seu formato vitais para o desenvolvimento econômico do país.

### 1.4.5 A família

Ao longo das últimas décadas, houve um incremento significativo nos investimentos destinados ao sistema educacional (OECD, 2017) mas, em contrapartida, o desempenho acadêmico não acompanhou esse ritmo de crescimento, o que evidenciou a necessidade de se analisar quais outras variáveis, que não estão ao alcance direto dos formuladores de políticas públicas, estavam afetando negativamente. Segundo Greenwald, Hedges e Laine (1996), existem inúmeras pesquisas em psicologia e educação que sustentam a hipótese de que o ambiente familiar tem um

efeito superior aos sociais de sala de aula no sucesso escolar (KELLAGHAN et al., 1993; CARTER, 2002). Essas pesquisas reforçam a interdependência entre a escola e a família, além de definirem os pais, com ênfase na mãe, como elementos essenciais no processo educacional.

As variáveis que retratam o ambiente familiar foram caracterizadas como capital social, por refletirem uma alocação consistente de tempo por parte dos pais no acompanhamento e participação das atividades educacionais dos filhos, que vão desde a leitura de livros, ajuda com o dever de casa e brincadeiras com jogos educativos até a participação em eventos escolares. Quanto as variáveis que compõem o capital, aquela que aparenta ser a mais relevante é a quantidade de tempo que a mãe se dedica ao filho, em especial, às atividades escolares.

Presumidamente, a presença do pai em casa aumenta a quantidade total de dedicação de adultos às atividades familiares, inclusive educativas, de tal modo que as crianças que são criadas apenas pela mãe teriam uma quantidade de capital social inferior em relação às que possuem os dois pais em casa.

Quanto ao tamanho da família, com a urbanização e a entrada da mulher no mercado de trabalho, a redução do número de filhos por casal pode ajudar a contrabalancear, em parte, os efeitos da menor exposição dos pais aos seus descendentes no capital social, mas que, mesmo assim, o efeito geral é negativo. Fato é que, apesar das mudanças ocorridas na estrutura familiar que levaram a perda de capital social, a performance acadêmica média não se reduziu, o que evidencia os efeitos positivos do aumento dos

investimentos em educação (GREENWALD; HEDGES; LAINE, 1996).

## 1.5 A educação no Brasil sob a perspectiva da Economia da Educação

Para discorrer sobre o desempenho acadêmico e o abandono escolar no contexto da rede pública de ensino no Brasil, é necessária a compreensão da relação entre os fatores internos e externos, alguns deles relacionados ao contexto sócioeconômico, que definem ambas as situações. Entre esses fatores que podem contribuir para o grau de interesse pelos estudos por parte dos jovens, temos: o tempo de permanência e participação em atividades realizadas na escola; a logística para se chegar à escola; sucessivas reprovações; falta de incentivo da família e da escola; necessidade de trabalhar e gerar de renda; estrutura curricular defasada, alcoolismo, gravidez precoce; violência urbana; e drogas. Além disso, a análise deve considerar também o momento no qual qualquer atitude relacionada ao continuísmo escolar é deliberada, a fim de compreender o nível de importância relativa dos vários fatores de risco ao longo do tempo (BARROS; MENDONÇA, 2008; NERI et al., 2009; NERI et al., 2015).

Com o objetivo de identificar quais são os fatores determinantes do desempenho escolar do Brasil, Menezes-Filho (2007) utilizou os dados do Sistema de Avaliação do Ensino Básico de (SAEB), de 2003, para alunos da 4<sup>a</sup> e 8<sup>a</sup> séries do Ensino Fundamental e da 3<sup>a</sup> série do Ensino Médio, em relação ao desem-

penho na disciplina de Matemática. Os exercícios econométricos, baseados na técnica de *MQO*, mostraram que os fatores mais relevantes para o desempenho acadêmico são as características do aluno e da família, tais como número de reprovações, distorção idade-série, idade de entrada no ciclo básico escolar, cor, nível de escolaridade da mãe, presença de computadores e livros em casa. Para o autor, um dos resultados mais relevantes trata da idade de entrada no sistema, os quais indicaram que aqueles alunos que cursaram a Educação Infantil apresentam um desempenho melhor nos testes nas três séries analisadas do que os que ingressaram apenas na 1ª Série.

No âmbito escolar, as variáveis que tratam do perfil etário e salarial dos professores, do processo de seleção do diretor, da estrutura física e da disponibilidade de recursos da instituição, além do número de alunos por sala, se mostraram pouco significativas para a melhora do desempenho do aluno (situação semelhante à verificada por [Hanushek \(1996\)](#) nos Estados Unidos), entretanto, o tempo que o aluno permanece na escola está positivamente correlacionado com o resultado escolar do aluno. Assim, o autor conclui que uma política educacional, que busque capacitar melhor o estudante deve priorizar o aumento da carga horária, mesmo que isso exija um maior número de alunos por turma.

Com o objetivo de verificar se as alterações no nível de salário dos professores da rede pública, introduzidas pelo Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de

Valorização do Magistério (FUNDEF)<sup>20</sup> promoveram alguma melhora na situação acadêmica do aluno, [Menezes-Filho e Pazello \(2007\)](#) utilizaram o SAEB como fonte de dados (para os anos de 1995, 1997, 1999 e 2001) e a metodologia de diferenças em diferenças<sup>21</sup> como técnica econométrica. Formalmente, a especificação que relaciona o desempenho  $y$  do aluno  $i$  do professor  $j$  da escola  $k$  no ano  $t$  com a remuneração dos professores  $w$  controlando as características dos estudantes  $x$ , professores  $p$  e escola ( $s$ ) é:

$$y_{ijkt} = \alpha + \delta w_{jt} + \gamma x_{it} + \gamma p_{jt} + \theta s_{kt} + \varepsilon_{it}. \quad (1.17)$$

Entretanto, a especificação acima pode gerar estimativas viesadas, uma vez que o termo de erro normalmente está correlacionado com as covariadas, em especial, com o salário dos professores. Assim, com o intuito de reduzir a probabilidade desse problema, e inserir as regras de definição de salários dos professores da rede pública tal como o transferência de recursos pelo FUNDEF (\$), a Equação (1.17) foi redefinida:

$$E[u_{ijkt}|w, x, t, s, \$] = \beta_t + \gamma_s + \eta_m + \beta_{tm} + \gamma_{sm} + \beta_{ts}, \quad (1.18)$$

o que significa que a diferença não observada em desempenho  $u_{ijkt}$  é definida por um efeito de tempo ( $\beta_t$ ), por um tipo de sistema de ensino ( $\gamma_s$  municipal, estadual ou privada), de municipalidade ( $\nu_m$ ), e interações entre tempo e municipalidade ( $\beta_{tm}$ ),

<sup>20</sup> Instituído pela Emenda Constitucional n.º 14, de setembro de 1996, e implantado, nacionalmente, em 1º de janeiro de 1998, com o objetivo de promover e estruturar o financiamento da educação básica pública.

<sup>21</sup> Para maiores detalhes, ver [Cameron e Trivedi \(2010\)](#).

sistema de ensino e municipalidade ( $\beta_{sm}$ ) e entre tempo e sistema ( $\beta_{ts}$ ).

Os resultados estimados identificaram uma não significância estatística para os efeitos dos salários das professores da rede pública sobre o desempenho do estudante, entretanto, sugeriram que o FUNDEF, ao aumentar o salário relativo dos professores da rede pública, atraiu novos e melhores profissionais, o que pode levar a uma melhora de performance acadêmica. Entre outros resultados relevantes, [Menezes-Filho e Pazello \(2007\)](#) não encontraram significância estatística para o grau de experiência e nível de escolaridade dos professores.

A partir de um modelo de análise de regressão hierárquico de dois níveis<sup>22</sup>, [Andrade e Soares \(2008\)](#) realizaram um exercício empírico para investigar as influências das escolas sobre o desempenho cognitivo de seus alunos, utilizando os resultados do SAEB, para alunos do 5º e 9º ano (4ª e 8ª séries) do ensino fundamental e 3º ano do Ensino Médio, nas disciplinas de Matemática e Língua Portuguesa, para os anos de 1995 a 2003. Formalmente, temos:

- Nível 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}NSE_{ij} + \beta_{2j}SEXO_{ij} + \beta_{3j}PARDO_{ij} + \beta_{4j}PRETO_{ij} + \beta_{5j}ATRASO_{ij} + \beta_{6j}DISCIPLINA_{ij} + \beta_{7j}ATITUDE_{ij} + r_{ij}. \quad (1.19)$$

<sup>22</sup> Para maiores detalhes, ver [Raudenbush e Bryk \(2002\)](#).

- Nível 2:

$$\begin{aligned}
 \beta_{0j} &= \gamma_{00} + \gamma_{01}MNSE_j + \gamma_{02}REDE_j + \gamma_{03}MATRASO_j + u_{0j} \\
 \beta_{1j} &= \gamma_{10} \\
 &\vdots \\
 \beta_{7j} &= \gamma_{70}
 \end{aligned}
 \tag{1.20}$$

tal que  $Y_{ij}$  é a nota alcançada, em uma das duas disciplinas, do  $i$ -ésimo aluno da escola  $j$ ,  $\beta_{0j}$  é a média das notas dos alunos da escola  $j$ ,  $NSE_{ij}$  e  $MNSE_j$  são, respectivamente, o nível sócio-econômico do aluno centrado na média da escola  $j$  e o nível da escola  $j$  centrada na média total. Já  $ATRASO_{ij}$ ,  $MATRASO_j$  e  $ATITUDE_{ij}$  tratam de questões de distorção idade-série e de comportamento do aluno, enquanto que  $SEXO_{it}$ ,  $PARDO_{it}$ ,  $PRETO_{it}$ ,  $DISCIPLINA_{it}$  e  $REDE_{ij}$  são variáveis *dummies*. Por fim,  $r_{ij} \sim N(0, \sigma^2)$  e  $\begin{Bmatrix} u_{0j} \\ u_{1j} \end{Bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$ , onde  $u_{0j}$  representa a medida individual do efeito da escola sobre o desempenho do aluno. Para se verificar se há diferença entre estudar em uma escola ou outra, deve-se obter o quociente entre a variância de  $u_{0j}$  e a soma das variâncias de  $U_{0j}$  e  $r_{ij}$ .

Os autores concluíram que a influência da escola sobre o desempenho do aluno ao longo do tempo permaneceu estável. Além disso nenhuma escola se sobressaiu em relação à outra e não houve grandes variações nas disciplinas ou séries avaliadas, ou seja, em geral, escolas brasileiras têm efeito muito próximos entre elas. Por fim, os autores criticam a hierarquização das escolas com base em resultados de proficiência, visto que a variância

entre as escolas é pequena e é estatisticamente incerta e sensível aos modelos econométricos construídos pelos autores. Entretanto, notaram que as escolas do Ensino Médio são mais heterogêneas em comparação as da 4ª Série do Ensino Fundamental, enquanto as escolas públicas são mais homogêneas em relação às particulares.

O exercício empírico realizado por Ferrão, Beltão e Santos (2002) também utilizou o Modelo Hierárquico de dois níveis para analisar a existência de diferenças no desempenho escolar dos alunos que possam ser atribuídas às políticas de não repetência, ou seja, verificar se há alguma evidência de que alunos com defasagem idade-série, em escolas com promoção automática, têm desempenho acadêmico diferenciado dos demais, utilizando dados de 16066 alunos da 5ª série do Ensino Fundamental, na região sudeste do Brasil, de 514 escolas. Formalmente, temos:

- Nível 1:

$$proficiencia_{ik} = \beta_{0K} + \beta_{1k}defasagem_{ik} + e_{ik}. \quad (1.21)$$

- Nível 2:

$$\begin{aligned} \beta_{0K} &= \beta_{10} + \beta_{01}prom\_aut_k + u_{0k} \\ \beta_{1K} &= \beta_{10} + \beta_{11}prom\_aut_k + u_{1k}, \end{aligned} \quad (1.22)$$

onde a variável dependente do Nível 1,  $proficiencia_{ik}$  representa o desempenho escolar do aluno  $i$  da escola  $k$  regredido contra a situação de defasagem idade-série do aluno  $ik$ . Por sua vez, a variável indicadora da política pública de aprovação automática é expressa no Nível 2 por se referir à escola  $k$ .

Os parâmetros estimados por [Ferrão, Beltão e Santos \(2002\)](#) indicaram a influência positiva e significativa do nível socioeconômico no desempenho do aluno, quanto a cor do estudante, mostraram que os negros têm rendimento inferior aos seus colegas de outras etnias e que a repetência apresentou sinal negativo e significativo sobre o desempenho acadêmico. Por fim, os resultados não mostram evidências de que alunos com defasagem escolar tenham um desempenho acadêmico pior do que aqueles alunos que frequentam escolas onde foram implementadas políticas de aprovação automática, e nem que alunos desfavorecidos socioeconomicamente tenham desempenho inferior aos seus colegas em escolas onde a aprovação automática foi adotada.

Ao lado dos baixos níveis de aprendizagem durante todo o Ciclo Básico, o abandono e a evasão escolar constituem um dos mais graves problemas do Ensino Médio. Os dados revelam que daqueles, que ingressam nessa etapa, um percentual relevante não consegue avançar e acaba desistindo: segundo [BRASIL \(2014\)](#), de cada 100 alunos dessa etapa, 8 são reprovados e 12 abandonam a escola.

[Barros e Mendonça \(2008\)](#) fizeram uso de informações do Censo Escolar, do Exame Nacional do Ensino Médio (ENEM), SAEB e da Pesquisa Nacional por Amostra de Domicílios (PNAD) com o objetivo de documentar o nível, a evolução temporal, as disparidades regionais na evasão e no abandono do Ensino Médio no Brasil e em que medida esses jovens que não estão frequentando a escola, também não estão trabalhando (nem

nem).

A pesquisa indicou que cerca de 20% dos jovens de 15 a 17 anos (idade em que estes jovens deveriam estar frequentando o Ensino Médio) não frequentam a escola hoje no Brasil e que as taxas de abandono ainda são muito elevadas, apresentando um valor médio de 15% entre os anos de 1999 e 2005, sendo que praticamente não há disparidades entre as áreas rurais e urbanas. Os autores também rejeitaram a hipótese de que o desinteresse dos jovens pela educação é uma das principais explicações para as elevadas taxas de abandono. Além disso, os resultados também mostraram que entre 1986 e 2007, a porcentagem de jovens de 15 a 17 anos que não estudam e nem trabalham aumentou, em média, 20% no país, mas que a heterogeneidade regional reduziu, sendo nas regiões Sudeste e Sul, a maior concentração desse perfil de adolescente.

Outro trabalho estatístico que buscou interpretar o desinteresse acadêmico por parte dos jovens do Ensino Médio foi desenvolvido por Pazello, Cabral e Cazolla (2008). Em sua pesquisa, os autores investigam as formas alternativas de não participação, as quais são entendidas como o continuar a frequentar escola, mas com pouca dedicação ou pouco esforço no sentido do aprendizado, por esta situação ser bastante comum entre os alunos de 15 a 17 anos. Para tanto, utilizaram a base de dados da PNAD e do SAEB (Suplemento sobre “Educação e Programas Sociais”) de 2004, e estimaram a seguinte equação por *MQO*:

$$faltas = \alpha + \beta em + \lambda X + e, \quad (1.23)$$

em que a variável dependente *faltas* representa o número de faltas bimestrais, *em* a variável binária igual a 1, caso o aluno curse o Ensino Médio, *X* representa um vetor de variáveis explicativas observáveis, como gênero, estado de residência, rede de ensino, condição de trabalho, entre outras.

Os resultados comprovam a hipótese de que os jovens que estudam e trabalham da rede pública têm maior probabilidade de se ausentarem do que aqueles que apenas estudam. Entretanto, os motivos que os levam a faltar, não estão diretamente relacionados a esses dois temas, na verdade, são oriundos de doenças e do mero desinteresse. Outra situação que chama atenção é que o fato de o jovem trabalhar implica em um número de faltas maior, mas o seu impacto na probabilidade de fazer a lição de casa não é afetado. Logo, não é possível concluir que o trabalho é um fator que leve o jovem a se dedicar menos às atividades escolares. Por fim, quando o estudante do Ensino Médio é comparado ao do Fundamental, os primeiros se dedicam menos às atividades complementares, além de faltarem mais.

Em suma, os trabalhos apresentados nesta seção mostram que os problemas relacionados ao Ensino Médio são preocupantes e que a literatura da Economia da Educação enfrenta diversas dificuldades em suas pesquisas, como especificação de funções, a falta de qualidade dos dados e problemas teóricos e metodológicos.

## 1.6 Considerações finais

Este ensaio buscou apresentar a evolução do tratamento dado a Educação, dentro do contexto econômico, desde a introdução do assunto na perspectiva dos modelos macroeconômicos de crescimento à construção de uma área na ciência dedicada inteiramente a isso, a Economia da Educação, que realça toda a sua relevância do tema em um ambiente que sempre busca o desenvolvimento e aperfeiçoamento das técnicas produtivas.

Como apresentado ao longo deste ensaio, exercícios empíricos que buscam interpretar o comportamento acadêmico do estudante são frequentemente desenvolvidos de forma a medir o produto educacional através dos resultados de exames padronizados, o que se confirmou na literatura apresentada ao longo deste ensaio. Foi possível observar, portanto, que a maioria dos artigos indicou possíveis problemas na estruturação das bases de dados a serem utilizadas nestes exercícios. Assim, se faz necessária a correção de problemas metodológicos por parte dos pesquisadores, além de atenção às falhas que existem nas fontes de dados disponíveis.

Por fim, vale ressaltar que apesar das diversas técnicas estatísticas, especificações de funções, variáveis explicativas e bases de dados utilizadas, na maioria das vezes, pesquisadores chegaram à conclusões próximas das mesmas variáveis, deixando claro que variáveis como raça/cor do aluno, nível de escolaridade e participação dos pais, salário do professor, nível socioeconômico da família e necessidade de trabalhar exercem influência

tanto sobre o desempenho educacional quanto sobre a decisão de não abandonar os estudos.

## 2 Um modelo bivariado para a progressão acadêmica no Ensino Médio

### 2.1 Introdução

A forma como o Ensino Básico é tratado hoje no Brasil resulta de uma série de discussões realizadas nos últimos 30 anos, as quais se consolidaram na Constituição Federal de 1988 (CF88), na Lei de Diretrizes e Bases de 1996 (LDB) e, mais recentemente, na Medida Provisória nº 748/2016, que veio a ser conhecida como Novo Ensino Médio. Todos esses atos normativos têm como objetivos principais ampliar o acesso à escola e a construção de mecanismos que aumentem e estimulem o desejo dos alunos de permanecerem nos estudos, principalmente no Ensino Médio. Dessa forma, pretende-se dar condições para uma melhora no bem estar do cidadão e, por consequência, no desenvolvimento econômico e social do país.

Apesar de conquistas relevantes no período, como a universalização do Ensino Fundamental, a queda significativa no analfabetismo, a redução na defasagem idade-série, entre outros, ainda se observa problemas de repetência, reprovação e atraso

escolar, além do estagnação do desempenho acadêmico dos estudantes de Ensino Médio, o que gera preocupações por parte dos formuladores de políticas públicas educacionais para com este segmento da educação básica. Para ilustrar essa problemática, a Figura 1 traz as taxas de escolarização líquida<sup>1</sup>, a de distorção idade-série<sup>2</sup> e as de rendimento, com destaque para as taxas de reprovação e abandono.

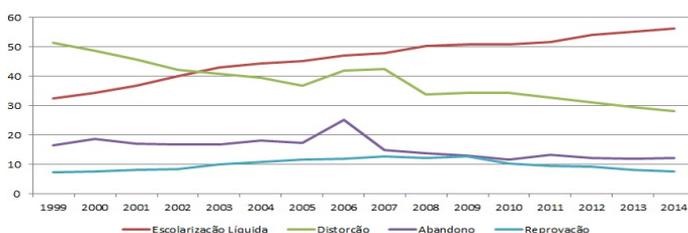


Figura 1 – Indicadores de Rendimento - Ensino Médio Brasil 1999/2014

Fonte: INEP. Formatação própria.

Entre os indicadores apresentados, a taxa de abandono apresentou, para o período analisado, um valor médio anual de 15% e uma redução de 3% ao compararmos o ano inicial com o final da amostra. A manutenção de uma elevada taxa de evasão é explicada por [Neri et al. \(2009\)](#) a partir de três motivos principais: o primeiro diz respeito a visão de curto prazo de alunos e agentes públicos, o que limita a oferta de serviços educacionais; o segundo se relaciona à falta de interesse dos alunos e de seus pais

<sup>1</sup> Indica o percentual da população em determinada faixa etária que se encontra matriculada no nível de ensino adequado à sua idade.

<sup>2</sup> Percentual de alunos, em cada ano, com idade superior a 2 anos ou mais em relação à idade recomendada para aquele ano.

quanto a educação ofertada; e por último, a limitação na operação de restrições de renda e a dificuldade de acesso ao mercado de crédito impõe barreiras à exploração dos altos retornos de longo prazo oferecidos pela educação. Além da existência de um *trade-off* entre trabalho e escola, a questão financeira é um requisito importante na decisão do jovem em continuar seus estudos, em função da demora em receber os proventos da dedicação aos estudos.

Além desses motivos, existe um outro importante questionamento que diz respeito ao momento em que tal atitude é tomada e qual a relevância dos vários fatores de risco ao longo do tempo (PLANK; DELUCA; ESTACION, 2008). Nesse contexto, os dados do IBGE<sup>3</sup> para evasão por série, indicam que existe uma grande dificuldade por parte de alguns alunos na transição do Ensino Fundamental para o Médio, e isso pode estar relacionado com a probabilidade de maior de evasão durante a primeira série da última etapa do ciclo básico.

Por sua vez, a taxa de escolarização líquida apresentou uma expansão significativa, se mantendo acima dos 50% desde 2008 e atingindo seu valor máximo em 2014, 56,3% ao, que é quase duas vezes superior ao ocorrido em 1998. A despeito desse crescimento, se a população dessa faixa etária é de cerca de dez milhões, praticamente metade dos jovens em idade adequada à escola média não estão matriculados nela, o que evidencia a incompletude do processo de democratização da escola, pois esse contingente de jovens ou ainda cursa o ensino fundamental,

---

<sup>3</sup> Instituto Brasileiro de Geografia e Estatística.

devido à reprovações ou ingresso tardio, ou simplesmente não frequenta a escola. E por fim, a taxa de distorção idade-série vem sendo reduzida ao longo das últimas duas décadas, apontando uma taxa de 47,6% inferior ao ocorrido no final do século passado.

Na Figura (2), está representado como os jovens de 15 a 17 anos de idade se ocupam. Para tanto, foram feitas 4 divisões: (i) só estudam; (ii) só trabalham, (iii) estudam e trabalham; e, por último, (iv) nem estudam e nem trabalham. É nítida a queda da porcentagem de jovens que só trabalhavam, de 26% em 1992 para somente 5% em 2015 e do aumento dos jovens que só estudavam, respectivamente 38% e 71%, o que parece indicar que as causas de evasão, ou até mesmo do desinteresse nos estudos não podem ser justificadas apenas pelos motivos tradicionais, como é a questão do trabalho.

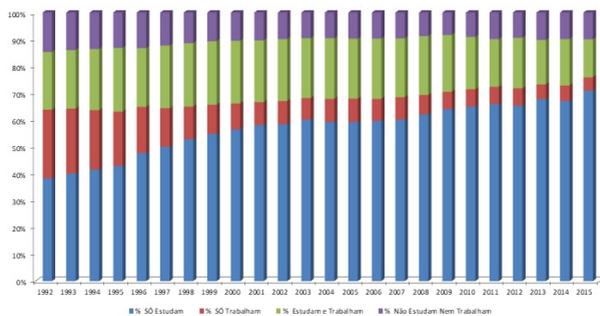


Figura 2 – Ocupação de Jovens: 15 a 17 anos

Fonte: INEP/IBGE. Elaboração própria.

Para Tinto (1975), Bean (1980) e Doll, Eslami e Walters (2013), a reprovação e, principalmente, o abandono escolar representam os eventos finais de um processo de longo prazo de descontinuidade de atendimento à educação formal, que para ser entendido se faz necessária a compreensão das características dos estudantes em relação às questões sociais, econômicas, psicológicas e também de fatores institucionais (escola, família, comunidade) que contribuem para que o jovem tome a decisão de abandonar o Ensino Médio antes que ele venha a se formar.

As consequências futuras dessa opção apresentam um alto custo, tanto para o jovem quanto para a sociedade. Os resultados negativos associados a evasão e a repetência incluem uma maior probabilidade de gravidez na adolescência, de desemprego, de ter um padrão de vida baixo e conseqüentemente, depender de serviços de assistência pública, uma maior frequência de incorrer em problemas de saúde, além de aumentar as chances de se envolver em atividades criminais (SUBBARAO; RANEY, 1995; CHIRAS; CREA, 2004; LADD, 2012).

Para se fazer uma análise mais detalhada a respeito dos motivos que levam o jovem a abandonar o Ensino Médio e da probabilidade de ocorrência deste fenômeno, é comum se mensurar a qualidade educacional a partir de um conjunto de características observáveis, como por exemplo renda, saúde, ocupação da mãe, condições físicas das escolas e distância para as mesmas, entre outras, em especial a partir de testes padronizados de desempenho escolar e abordagens estatísticas, principalmente os modelos probabilísticos paramétricos univariados e bivariados. Neste sen-

tido, a pesquisa que se pretende fazer neste ensaio, apresenta um modelo de progressão serial, com foco na transição da 1<sup>o</sup> para a 2<sup>o</sup> Série em razão deste ser o momento de maior risco para o abandono escolar ao longo dessa etapa do ciclo básico, com vistas a identificar as características individuais, familiares e escolares mais relevantes para a realização acadêmica.

Para cumprir com o objetivo proposto nesta pesquisa, as informações utilizadas provém de dois sistemas que regulam as matrículas da rede estadual de ensino de Santa Catarina, o SERIE-Escola, descontinuado em 2011, e o Sistema de Gestão Educacional de Santa Catarina (SISGESC), que é o atual sistema de gerenciamento, as quais contêm dados de mais de 400000 alunos matriculados em 650 escolas para o período de 2008 a 2012, relacionados aos aspectos individuais e familiares dos estudantes e características escolares, referentes à estrutura física e do corpo docente. A consolidação dessa base de dados traz um caráter singular e de extrema relevância para os resultados que serão estimados a partir dela nesta tese de doutorado, pois assim será possível obter um retrato fiel da educação de Ensino Médio no estado e com isso, contribuir com o poder público para a eficácia e efetividade na implementação de políticas e ações educacionais.

O ensaio está organizado em 6 seções. Além desta introdução, a Seção 2 traz uma breve descrição de modelos teóricos de evasão e progressão acadêmica, a Seção 3 trata da modelagem econométrica, a Seção 4 apresenta as variáveis independentes, a descrição da base de dados e algumas estatísticas descritivas e a Seção 5 traz o modelo econométrico proposto. Já a Seções

6 apresenta os resultados empíricos e a Seção 7 traz as conclusões finais, além da relevância do tema no desenvolvimento de políticas públicas que melhorem o desempenho acadêmico dos estudantes do Ensino Médio.

## 2.2 Modelos teóricos de evasão e persistência escolar

Em busca da melhor compreensão sobre as questões de evasão e persistência escolar, esta seção apresenta um conjunto de modelos teóricos que ilustram o processo de tomada de decisão do aluno sob uma perspectiva institucional, comportamental, dinâmica e cumulativa. Para tanto, optou-se por descrever os modelos institucionais de atrito escolar de [Tinto \(1975\)](#), [Tinto \(1982\)](#), [Tinto \(1988\)](#) e [Bean \(1980\)](#) e os econômico-econométricos de [Cameron e Heckman \(1998\)](#) e [Eckstein e Wolpin \(1999\)](#), em razão de suas aplicabilidades tanto no contexto do Ensino Superior quanto do Ensino Médio e da perspectiva individual abordada neles, que foca nos valores, nas atitudes e nas condutas dos estudantes e, ainda, por considerarem a evasão escolar como a etapa final de um processo dinâmico de descolamento acadêmico-social, em relação à escola, limitado ao contexto de vida do estudante, seja familiar, educacional ou comunitário.

Com o intuito de identificar a natureza da decisão de abandonar a escola, [Tinto \(1975\)](#) formulou um modelo teórico de integração para estudantes universitários de duas dimensões: a acadêmica (estudante) e a social (instituição de ensino). Essa

estrutura foi construída a partir da adaptação dos fundamentos da teoria do suicídio de Durkheim (1951)<sup>4</sup> ao contexto acadêmico e da análise de custo-benefício derivado da Economia da Educação.

A integração acadêmica trata do sentimento do aluno em estar inserido no contexto do curso e de suas demandas acadêmicas, a qual inclui a sua percepção sobre o seu desempenho nas notas, motivação para obtenção do diploma, desenvolvimento intelectual e pessoal e identificação com os valores do curso. Por sua vez, a integração social trata do ambiente extraclasse, isto é, da interação com outros colegas de cursos, professores e funcionários da instituição (conversas e amizades), da participação em eventos sociais na universidade (eventos culturais e esportes), além do comprometimento com a universidade. Ambos os conceitos, apesar de serem analiticamente distintos, interagem e realçam um ao outro, e caso haja uma divergência entre os valores introspectivos e morais do estudante e os atributos da instituição de ensino a probabilidade de evasão cresce significativamente.

Entretanto, Tinto (1975) também reconhece que um aluno pode abandonar os estudos por motivos pouco correlacionados com a interação aluno-universidade, os quais são percebidos pelas alterações no valor dado por ele a obtenção do diploma e no comprometimento com a universidade, a partir de uma análise

---

<sup>4</sup> A partir de observações empíricas, Durkheim propôs uma teoria acerca do suicídio, em que o ato estaria relacionado com fatores sociais como, por exemplo, fatores ligados a religião, estado civil, profissão, educação e local onde se vive. Especificamente, a probabilidade de suicídio aumenta quando dois tipos de interação não são suficientes: a moral (valores) e a coesão social.

econômica de custo-benefício. Isso ocorre porque essas variáveis refletem a integração individual aos domínios sociais e acadêmicos da instituição de ensino, as quais são também resultados das perspectivas individuais dos benefícios (satisfação pessoal, amizades, realizações acadêmicas) e custos (financeiro, tempo, insucesso acadêmico) de se frequentar uma universidade. Temos então que a percepção individual é o fator central nesse modelo de evasão.

A partir desse referencial, o autor tratou a decisão de evadir como o resultado de um processo dinâmico multidimensional da interação entre o indivíduo e a instituição de ensino, em que características específicas de cada um e o modo como elas se interagem são importantes preditores desta decisão. Entre as características individuais, as que mais se sobressaem são as relacionadas: à família (*status* sócio-econômico, raça, número de irmãos, escolaridade dos pais); ao histórico escolar; ao comprometimento e motivação com a conclusão do curso; e às características intrínsecas ao indivíduo (sexo e habilidades). Quanto as que fazem referência à instituição de ensino, temos: recursos financeiros e físicos (prédios, laboratórios, bolsas de estudos); composição demográfica de funcionários e professores; tamanho e prestígio da instituição; e, eventos sociais, etc. Por fim, o autor afirma que o estudo da decisão de permanência/evasão na educação superior é extremamente complicado, já que este não se reduz a apenas uma perspectiva, mas sim a várias, as quais englobam inúmeras características.

Vale também ressaltar os diferentes conceitos de evasão apresentados pelo autor, o qual não deve ser entendido apenas como o abandono do sistema de ensino de modo geral. Além deste conceito abrangente, tem-se o desligamento voluntário, o desligamento acadêmico e a migração, em que primeiro trata das metas e comprometimento do aluno com a instituição, o segundo com o desempenho acadêmico do estudante e o último, diz respeito aos fatores relacionados a readequação das preferências do estudante quanto à sua formação.

A busca por explicações acerca da evasão escolar no ensino superior também foi objetivo de estudo de [Bean \(1980\)](#), o qual propôs um modelo de atrito estudantil (MAE) semelhante ao de [Tinto \(1975\)](#), porém adaptado às teorias de rotatividade da mão de obra no ambiente de trabalho ([PRICE, 1977](#)) e que englobam aspectos pessoais, sociais e institucionais para a explicação desse fenômeno. A partir dessa perspectiva, a hipótese básica do modelo é a de que os motivos que levam o aluno a evasão seriam análogos a opção de um trabalhador sair de seu local de trabalho. Nesse contexto, o estudo tem um foco no particular e não no geral, o que torna o atrito estudantil específico a uma instituição e não ao sistema educacional como um todo.

Segundo [Bean \(1980\)](#), as variáveis que afetam a evasão podem ser divididas em três categorias, as quais foram definidas a partir da interação entre a economia do trabalho e a educação: a) variáveis intervenientes, que tratam da satisfação, motivação e comprometimento com a instituição de ensino; b) variáveis organizacionais da instituição; e, c) variáveis históricas do aluno,

as quais foram inseridas no modelo com vistas à refletir a influência das características pregressas do aluno na interação com o ambiente social e acadêmico da instituição, no caso a universidade. E quanto a relação entre as variáveis e evasão, tem-se por hipótese que esta é aditiva não multiplicativa, isto é, as variáveis não interagem e quantidades sucessivamente maiores (menores) de uma variável explicativa  $x$  qualquer, irá gerar probabilidades maiores de atrito escolar, *ceteris paribus*.

O MAE indica que as características do aluno devem ser levadas em consideração para que seja possível compreender o processo de interação entre o estudante e a instituição de ensino e, assim, prever a intenção de evadir ou não do aluno. Uma vez que tal atitude está balizada por conceitos sociocognitivos, como autoconfiança, motivação e autoeficácia, o nível de satisfação do aluno com a universidade (ou escola) emerge como fator essencial nesse processo decisório. Contudo, o que se observa sobre o processo de evasão escolar, é que este é um fenômeno de múltiplas dimensões, que envolve a interação de fatores individuais, institucionais, motivacionais e perspectivas diversas. Nesse sentido, a formalização do abandono escolar e o momento de sua ocorrência são apenas partes de um complexo arranjo de variáveis.

Em suma, tanto Tinto (1975) quanto Bean (1980) elencam quatro grandes categorias no processo decisório do aluno sobre evasão: a) as influências contextuais, que tratam de fatores pertencentes a uma instituição educacional que provavelmente afetará o desempenho do aluno, como o tamanho, seletividade

e recursos financeiros; b) as influências sociais, representadas pela percepção de suporte social por parte do estudante; c) o engajamento social, que inclui a integração e o sentimento de pertencer à comunidade, exemplificado por participações em atividades extraclasse e eventos; e d) o engajamento acadêmico, que inclui o compromisso com a obtenção da titulação e com os valores da escola ou universidade. Entretanto, diferem no papel exercido pelos fatores externos e não intelectuais na definição de percepções, interações e da própria permanência na escola, os quais são tidos como relevantes para o segundo, em contraposição à teoria do primeiro, que foca nos aspectos internos à instituição (universidade, no caso) e na relação entre ela e o aluno.

Em busca da compreensão desse fenômeno no Ensino Médio, [Cameron e Heckman \(1998\)](#) e [Eckstein e Wolpin \(1999\)](#) elaboraram modelos teóricos econométricos que tentam explicar a evasão e ajudar os formuladores de políticas educacionais a proporem ações afirmativas, com o intuito de prevenir o abandono escolar e promover um melhor aproveitamento no Ensino Médio.

[Eckstein e Wolpin \(1999\)](#) formulam um modelo de evasão escolar para o Ensino Médio nos Estados Unidos, onde assumem que jovens definem combinações de trabalho-escola a fim de maximizarem a utilidade de vida esperada em cada período, através de um problema de otimização dinâmica. O modelo então proposto sugere que a performance acadêmica é probabilística e que esta é função do histórico escolar, da quantidade de horas de trabalho e de características intrínsecas (habilidade e motivação) a cada estudante.

Em razão da diversidade de experiências e históricos familiares, é pouco provável que os jovens iniciem o Ensino Médio com o mesmo conjunto de preferências (por lazer e estudos), habilidades, motivações em relação à escola, ao trabalho e às expectativas quanto ao valor de um diploma. Mesmo que haja alguma alteração dinâmica nessas individualidades, a importância desses perfis iniciais é grande e persistente. Com base nessas premissas, os autores buscam responder à duas questões sobre o abandono escolar: a) o que difere os estudantes que evadem daqueles que graduam quanto às características iniciais não observadas e como elas estão relacionadas com o histórico familiar; b) por que jovens evadem, isto é, quais são as características iniciais mais importantes para a decisão de evadir, se é que elas existem. Portanto, o valor de se frequentar a escola consiste no valor de seu consumo presente, o qual é aleatório, e na utilidade esperada (incluindo potenciais ganhos salariais) percebida pelo estudante quando da diplomação.

Para responder a esses questionamentos, é necessária a compreensão da estrutura do modelo, o qual define o ambiente escolar em termos de progressão de série e exigências de créditos para a diplomação no Ensino Médio americano<sup>5</sup>, enquanto que o ambiente do mercado de trabalho é caracterizado pelas ofertas de

---

<sup>5</sup> Em cada ano do Ensino Médio, o aluno deve cursar exatamente cinco créditos (cada disciplina equivale a um crédito) e acumular um total de 20 créditos para obter a titulação, o que implica em uma duração mínima de quatro anos para a sua conclusão. O sucesso acadêmico é medido pelo índice GPA (*Grade point average*), o qual varia entre 0 e 4,0 e representa a média das notas obtidas em cada uma das disciplinas, onde a nota A=4,0, B=3,0, C=2,0, D=1,0 e F=0.

emprego, em função de disponibilidade de tempo e salário.

O ambiente escolar é definido por [Eckstein e Wolpin \(1999\)](#) em função do GPA acumulado ( $G_t$ ) e do total dos números de créditos obtidos ( $C_t$ ) até o ano  $t$ , onde a aprovação em cada série<sup>6</sup> ( $e_t$ ) corresponde a um incremento de cinco créditos (um crédito para cada disciplina cursada), ou seja,  $e_t = 8$  se  $C_t < 5$ ,  $e_t = 9$  se  $5 \leq C_t < 10$ ,  $e_t = 10$  se  $10 \leq C_t < 15$ ,  $e_t = 11$  se  $15 \leq C_t < 20$  e a graduação,  $e_t = 12$  se  $C_t \geq 20$ .

Ademais, o incremento de conhecimento adquirido ao longo dos anos do Ensino Médio é assumido como uma variável latente que determina as notas no curso, a qual é definida de modo genérico pelos conhecimentos prévios do aluno no início da nova série, suas habilidades, motivações, comprometimento e estrutura escolar, como tamanho da sala e qualidades do professor. Entretanto, em virtude da limitação da base de dados, os autores a definem com base em  $C_t$  e  $G_t$  e tratam as características da escola como fatores não observados independentes e identicamente distribuídos (*iid*).

No que tange às características do mercado de trabalho, o salário-hora ofertado ( $w_t^j$ ) a um estudante para trabalhar em período integral ( $j = f$ ) ou parcial ( $j = p$ ) depende apenas das experiências profissionais pregressas, as quais são mensuradas pelo total acumulado de horas trabalhadas e por um choque

---

<sup>6</sup>  $e_t = 9, 10, 11$  e  $12$  representam as séries do Ensino Médio, enquanto que  $e_t = 8$  a última série do Ensino Fundamental nos Estados Unidos.

idiossincrático<sup>7</sup>. Portanto, salário será:

$$w_t^j = \exp(w^j(H_t)) \exp(\varepsilon_t^j), \quad (2.1)$$

onde  $w_t$  é o salário no tempo  $t$ ,  $H_t$  as horas acumuladas de trabalho até o período  $t$  e  $(w^j(H_t))$  uma função.

Temos então que o conjunto de escolhas, para uma pessoa que seja elegível para frequentar o Ensino Médio, consiste em seis alternativas dicotômicas mutuamente exclusivas:  $d_t^k = \{d_t^{sn}, d_t^{sp}, d_t^{sf}, d_t^{nn}, d_t^{np}, d_t^{nf}\}$ , i.e., ir à escola e não trabalhar ( $d_t^{sn} = 1$ ), ir à escola e trabalhar período parcial ( $d_t^{sp} = 1$ ), ir à escola e trabalhar período integral ( $d_t^{sf} = 1$ ), não ir à escola e ao trabalho ( $d_t^{nn} = 1$ ), não ir à escola e trabalhar período parcial ( $d_t^{np} = 1$ ) e, por fim, não ir à escola e trabalhar período integral ( $d_t^{nf} = 1$ ).

Por sua vez, a utilidade contemporânea ( $U_t^k$ ) é assumida como linear e aditiva em consumo, na frequência escolar e em não trabalhar, onde o primeiro é a soma dos salários recebidos pelo estudante mais as transferências familiares. Uma vez que ir à escola envolve tanto um esforço mental quanto uma perda de horas de lazer, existe um valor psíquico atrelado à estudar,  $b_t^{sj}$ , que depende do nível de atividade laboral e que varia com a idade do aluno, sistematicamente e estocasticamente ( $b_t^{sj} = \bar{b}_t^{sj} + \varepsilon^{sj}$ ) para  $j = \{n, p, f\}$  e  $b_t^n$  o valor em moeda para o período sem trabalho ( $b_t^n = \bar{b}_t^n + \varepsilon^n$ ). Logo, dado o conjunto ( $k$ ) de escolhas alternativas e considerando  $L$  o número total de horas semanais que pode ser

<sup>7</sup> Termo de perturbação que diz que horas adicionais de trabalho afetam os salários de forma homogênea, independente se oriunda de um trabalho de período integral ou parcial.

dividido entre trabalho e não trabalho, temos a utilidade corrente em  $t$ :

$$U_t^{sn} = b_t^n L + b_t^{sn},$$

$$U_t^{sp} = b_t^n (L - h_t^p) + b_t^{sn} + b_t^{sp} + w_t^p h_t^p,$$

$$U_t^{sf} = b_t^n (L - h_t^f) + b_t^{sn} + b_t^{sf} + w_t^f h_t^f,$$

$$U_t^{nn} = b_t^n L,$$

$$U_t^{np} = b_t^n (L - h_t^p) + w_t^p h_t^p,$$

$$U_t^{nf} = b_t^n (L - h_t^f) + w_t^f h_t^f.$$

Portanto, o valor que um estudante estabelece para os anos de estudos no Ensino Médio será formalmente descrito por um problema de otimização dinâmico para um horizonte infinito:

$$V_t(S_t) = \max_{d_t} E \left[ \sum_{\tau=t}^{\infty} \beta^{\tau-t} \sum_k U_t^k d_t^k | S_t \right] \quad (2.2)$$

sendo  $V_t(S_t)$  o valor presente máximo esperado da função no período  $t$ , em um dado estado espaço  $S_t$  no tempo  $t$ , um fator de desconto  $\beta$  e  $d_t^k$ , um vetor de combinações factíveis (neste exemplo, nas alternativas entre trabalho e escola), dada as realizações correntes dos componentes estocásticos das utilidades. O estado espaço consiste em todos os aspectos da história conhecida do aluno que afetam as utilidades presentes alternativas ou as distribuições de probabilidade de futuras utilidades, como por exemplo, notas, créditos acumulados, horas de trabalho, variáveis

aleatórias não observadas, etc.

Dadas as combinações possíveis de escola e trabalho para os alunos que evadem (não vão á escola e não trabalham (nn), não vão à escola e trabalham meio período (np), não vão à escola e trabalham período integral (nf)), a função de utilidade para esse grupo será assim expressa:

$$V(S_t) = \max [V^{nn}(S_t), V^{np}(S_t), V^{nf}(S_t)] \quad (2.3)$$

Reestruturando a equação acima, o processo de otimização para alunos que abandonaram o Ensino Médio é formalizado assim:

$$V^k(S_t) = U_t^k + \beta E(V^1(S_{t+1}) | d_t^k = 1, S_t) \quad (2.4)$$

onde  $k = nn, np, nf$ . Tendo resolvido o problema de otimização, a função de verossimilhança para uma amostra de  $N$  indivíduos para um período  $t = 1, \dots, t_n$  é dada por:

$$\prod_{n=1}^N \text{Pr}(O_{1n}, O_{1n}, \dots, O_{t_n n} | S_0) \quad (2.5)$$

em que  $O_t = \{d_t^k, w_t^0, X_t\}$  é um vetor de resultados em qualquer período  $t$ , onde  $w_t^0$  significa o salário observado e  $X_t$  as variáveis do estado espaço em  $t$ .

Tinto (1975), Bean (1980) e Eckstein e Wolpin (1999) também afirmaram que é pouco provável a homogeneidade das características iniciais dos estudantes, seja nas habilidades, preferências ou no valor esperado pelo diploma de Ensino Médio e que a obtenção de medidas adequadas para estes atributos estão sujeitas a erros. Em virtude da necessidade de se contabilizar a

heterogeneidade proveniente desses fatores, os autores a trataram como não observada, além da adoção de uma metodologia não paramétrica para permitir uma quantidade finita de tipos  $M$ , cada um ponderado por uma proporção fixa da população estudantil  $\pi_m (m = 1, \dots, M)$ , análogo às proposições de [Keane e Wolpin \(1997\)](#) e [Cameron e Heckman \(1998\)](#).

Tem-se agora que qualquer tipo de heterogeneidade pode ser descrita por um vetor de parâmetros indexado a cada um dos alunos  $M$  para as respectivas variáveis não observadas como, por exemplo, a motivação e o valor esperado da graduação. Portanto, a função de verossimilhança será assim descrita:

$$\prod_{n=1}^N \sum_{m=1}^M \Pr(O_{1n}^m, O_{2n}^m, \dots, O_{t_n n}^m | tipo = m) \pi_m. \quad (2.6)$$

Logo, para calcular (2.6) uma programação dinâmica deve ser feita para cada tipo de heterogeneidade.

Os autores também tratam das formas funcionais a serem adotadas para cada uma das variáveis para a estimação do modelo. As formas funcionais adotadas não precisam ter necessariamente uma especificação *a priori*, mas podem ser definidas através de iterações específicas, baseadas no ajuste do modelo às estatísticas descritivas da base de dados. Assim, parâmetros são adicionados à medida que a estimação ocorre, com o objetivo de melhorar o ajuste do modelo aos dados. Após as especificações serem definidas, se conduz a iteração até que a verossimilhança convirja.

A partir da compreensão de que a graduação é um processo estocástico, onde as probabilidades de conclusão de cada ano escolar geram a probabilidade de titulação (BARTHOLOMEW, 1973) e das premissas do modelo logístico de transição escolar (MARE, 1980), Cameron e Heckman (1998) propõem um modelo econômico racional de transição escolar, que é assim inicialmente estruturado:

$$\Pr(D_s = 1 | \mathbf{X}_s = \mathbf{x}_s, D_{s-1} = 1) = P_{s-1,s}(\mathbf{x}_s) = \frac{\exp(\mathbf{x}_s \beta_s)}{1 + \exp(\mathbf{x}_s \beta_s)} \quad (2.7)$$

em que  $s$  representa a série escolar,  $D_s = 1$  se a pessoa completou a série  $s$  e  $D_s = 0$  caso contrário,  $\mathbf{X}_s = \mathbf{x}_s$  o conjunto de variáveis que determinam a transição da série  $s - 1$  para  $s$ .

De acordo com os autores, o modelo de transição especificado traz consigo alguns inconvenientes. Ele é fundamentalmente atemporal e, por isso, não é capaz de acomodar variáveis explicativas temporais como, por exemplo, idade e anos de estudos, além de não fazer referência aos componentes não observáveis do vetor  $\mathbf{X}_s$  que possam vir a influenciar o processo decisório do aluno (habilidade, motivação e QI). Diante disso, o modelo se expõe a problemas de viés de seleção dinâmica.

A heterogeneidade não observada, no contexto da equação (2.7), surge dos componentes omitidos de  $\mathbf{X}_s$ , os quais, em princípio, são originários de duas fontes: os que são estatisticamente série-independentes e aqueles que não são. Portanto, com o intuito de formalizar o viés de variável omitida em um modelo dinâmico, o escalar  $\Theta$  é introduzido. Assim, reordenando

os regressores tal que  $\mathbf{X}_s = (\mathbf{X}_{so}, \mathbf{X}_{su})$ , em que  $\mathbf{X}_{so}$  representa a fração observável e  $\mathbf{X}_{su}$  a omitida ou não observável de  $\mathbf{X}_s$ , então, no contexto de (2.7),  $\Theta_s = \mathbf{X}_{su}\beta_{su}$ .

No contexto do modelo inicial, o problema da heterogeneidade é bem definido com a imposição de algumas restrições:

- Hipótese 1:

$$\Pr(D_s = 1 | D_{s-1} = 1, \mathbf{X}_s = \mathbf{x}_s, \Theta_s = \theta_s) = G(\mathbf{x}_{so}\beta_{so} + \theta_s), \quad (2.8)$$

onde  $G$  é uma função de distribuição com número finito de parâmetros <sup>8</sup>. Caso seja de interesse do pesquisador, a linearidade dos parâmetros do modelo  $\mathbf{x}_{so}\beta_{so} + \theta_s$  pode ser sucedida por  $\varphi_s(\mathbf{x}_s\beta_s + \theta_s)$ , em que  $\varphi_s$  é uma função conhecida.

- Hipótese 2:

$$\Theta_s = \Theta, \quad (2.9)$$

Para o nosso contexto, os fatores não observáveis (habilidade, motivação, nível escolar dos pais, qualidade do professor) serão constantes ao longo de todo o período escolar.

---

<sup>8</sup> Condição não estritamente requerida.

Então, as hipóteses 1 e 2 definem tanto a probabilidade de completar a série inicial do ciclo escolar quanto à probabilidade de transição da série  $j$ , dado a completude da série  $j - 1$ , que são expressas, respectivamente:

$$\Pr(D_1 = 1 | \mathbf{X}_1 = \mathbf{x}_1, \Theta = \theta) = P_{0,1}(\mathbf{x}_1 + \theta), \quad (2.10)$$

$$\Pr(D_j = 1 | \mathbf{X}_j = \mathbf{x}_j, \Theta = \theta, D_{j-1} = 1) = P_{j-1,j}(\mathbf{x}_j + \theta). \quad (2.11)$$

Em decorrência da heterogeneidade oriunda de variáveis omitidas, se faz necessário caracterizar como estas variáveis afetam as estatísticas dos coeficientes estimados a partir de (2.7), tanto para a equação do ano inicial quanto para cada ano de ensino. Formalmente, para o modelo do ano inicial, assumindo o viés por  $\beta_1$  quando  $\Theta$  for omitido, que os regressores são constantes por todo o período escolar e uma regressão logística para a probabilidade de transição, temos:

$$\Pr(D_1 = 1 | \mathbf{X} = \mathbf{x}, \Theta = \theta) = \frac{\exp(\mathbf{x}\beta_1 + \theta)}{1 + \exp(\mathbf{x}\beta_1 + \theta)}. \quad (2.12)$$

Conforme [Cameron e Heckman \(1998\)](#), a grande maioria dos estudos empíricos de transição de série escolar utilizam um conjunto comum de variáveis para explicar todas as transições, mas os autores desses artigos optaram por utilizar o princípio de máxima verossimilhança para estimar um modelo *logit* não especificado que ignora  $\theta$ :

$$\Pr(D_1 = 1 | \mathbf{X} = \mathbf{x}) = \frac{\exp(\mathbf{x}\gamma_1)}{1 + \exp(\mathbf{x}\gamma_1)} \quad (2.13)$$

Além dessas situações, outros dois fatores introduzem viés nas estimativas dos parâmetros das últimas séries do Ensino Médio. Primeiro, a distribuição de  $\Theta$  se desloca para a direita com a sucessão de séries, dado que pessoas com baixa motivação, habilidade, QI (variáveis não observáveis) irão evadir e, conseqüentemente, sairão da amostra. Segundo,  $\Theta$  e  $\mathbf{X}\beta_s$  se tornam negativamente correlacionados, uma vez que indivíduos com baixo  $\mathbf{X}\beta_s$  tendem a continuar os estudos, caso tenham alto  $\Theta$ . Em outras palavras, entre os alunos que possuem características observáveis desfavoráveis (ambiente familiar, escolar e regional), aqueles com alta motivação, habilidade ou envolvimento familiar nos estudos apresentam uma maior probabilidade de finalizar os anos de estudos. Cameron e Heckman (1998), então, concluem que a seletividade educacional motivada pelas variáveis omitidas afligem os efeitos das observadas, o que leva os pesquisadores a subestimarem os seus efeitos no processo de diplomação educacional que são medidos pelos coeficientes das probabilidades logísticas de transição de série.

Até aqui, o modelo estatístico de transição apresentado por Cameron e Heckman (1998) não é consistente com um modelo racional econômico de decisão. Para haver a junção da teoria estatística com a econômica, os autores utilizam duas premissas: a primeira trata dos custos diretos da escolarização,  $c(s|\mathbf{x})$ , os quais são tidos como fracamente convexos e crescentes nos anos de estudo( $s$ ), além de serem função de subsídios escolares externos (renda e características familiares, bolsas de estudo) e assume-se  $c(0|\mathbf{x}) = 0$ ; a segunda trata da taxa de retorno inter-

temporal escolar,  $R(s)$ , a qual é assumida côncava e crescente em  $s$ , tal que  $R(0) > 0$  e que inclui na sua definição ganhos post-postos e fatores de desconto implícitos. Portanto, a escolaridade ideal,  $j$ , é definida a partir do seguinte problema de maximização:

$$\max_j \{R(j) - c(j | \mathbf{x})\}, \quad j = 0, \dots, \bar{S} \quad (2.14)$$

Caso sejam satisfeitas as premissas sobre as funções de custo e retorno, o retorno líquido,  $R(j) - c(j | \mathbf{x})$  é côncavo em  $j$ , para todo  $j$ , e positivo até o início da etapa de ensino, desde que  $R(0) > 0$  e  $c(0 | \mathbf{x}) = 0$ . Caso esses detalhes sejam ignorados, a solução ótima para os anos de escolaridade ideal será única e positiva.

Assim como no modelo expresso pela equação (2.7), a introdução da problemática de variáveis omitidas também traz implicações relevantes à confiabilidade do modelo econômico de escolaridade ideal ao afetar a relação retorno/custos. Essa hipótese é formalizada por uma variável aleatória escalar,  $\varepsilon$ , a qual é independente de  $\mathbf{X}$  e que representa uma alteração pessoa-específica na razão retorno-custo não observada. Quando a temática do viés é introduzida na função de custos relativos, esta passa a ser dependente tanto de efeitos observáveis quanto de omitidos dos alunos, da seguinte maneira:  $c(j | \mathbf{x}) = c(j) \varphi(\mathbf{x}) \varepsilon$  onde  $E(\varepsilon) = 1$ ,  $\varepsilon \geq 0$  e  $\varphi(\mathbf{x}) \geq 0$ .

O modelo estatístico de progressão escolar formalizado pelas equações (2.10), (2.11) e (2.12) não está inserido no contexto de otimização econômica do agente, e para que se possa introduzi-lo nesse contexto, se faz necessário especificar um choque série-específico,  $v_s$ , ao modelo e suprimir a hipótese de expectativas racionais (*forward-looking*) dos agentes. Dessa maneira, a regra sequencial proposta por Cameron e Heckman (1998) para que um agente continue até pelo menos a série  $j$  será:

$$R(j) - c(j) \varphi(\mathbf{x}) \varepsilon v_j \geq R(j-1) - c(j-1) \varphi(\mathbf{x}) \varepsilon v_{j-1}. \quad (2.15)$$

Todavia, existem situações em que a aplicação dessa regra de continuidade nem sempre leva a um ótimo global como, por exemplo, os estudantes podem progredir nas séries na hipótese de previsão perfeita para todos os múltiplos choques específicos de transição ( $v_j$ ) caso estes sejam conhecidos no início do período escolar, ou ainda, um baixo  $v_{j+2}$  pode tornar rentável a passagem de  $j$  para  $j+1$  no intuito cursar  $j+2$ , além da própria incerteza futura a respeito de  $v_j$ .

Para exemplificar esse último caso, considere o período regular de três anos para o Ensino Médio praticado no Brasil. Assumindo que  $V_j(\cdot)$  representa o valor de progredir até a série  $j$ , dadas as informações disponíveis em  $j$ , e que ao término da 1ª Série o aluno percebe  $v_1$ , ele irá definir se cursa a 2ª série, se e somente se:

$$E[V_2(v_2, v_1) | v_1] \geq R(1) - c(1) \varphi(\mathbf{x}) \varepsilon v_1, \quad (2.16)$$

onde:

$$V_2(v_2, v_1) = \max \left\{ \begin{array}{l} R(2) - c(2) \varphi(\mathbf{x}) \varepsilon v_2, R(3) \\ -c(3) \varphi(\mathbf{x}) \varepsilon E(v_3 | v_2, v_1) \end{array} \right\}. \quad (2.17)$$

Isso sugere uma predisposição ao comportamento míope dos estudantes secundaristas (OREOPOULOS, 2007). A regra míope de decisão sequencial expressa em (2.16) é válida para caracterizar o progresso acadêmico no caso de  $E(v_2 | v_1)$  substituir  $v_2$  no momento em que:

$$R(2) - c(2) \varphi(\mathbf{x}) \varepsilon v_2 \geq R(3) - c(3) \varphi(\mathbf{x}) \varepsilon E(v_3 | v_2, v_1), \quad (2.18)$$

caso contrário, nenhum estudante cursaria a 3ª série.

Agora, caso os valores futuros de  $v_j$  sejam incertos e se choques específicos de cada série ( $v_1$ ,  $v_2$  e  $v_3$ ) forem mutuamente independentes, os indivíduos irão se comportar em relação a  $v_j$  como um processo estocástico de martingale<sup>9</sup> tal que  $E(v_j | v_{j-1}) = v_{j-1}$ ,  $i \geq 1$  e, assim, o comportamento dos estudantes será míope. Logo, a equação (2.15) pode ser reescrita de modo a incorporar regressores como determinantes da função de custo, sendo formalizada assim:

$$\frac{R(j) - R(j-1)}{\varphi(\mathbf{x}) [c(j | \mathbf{x}) - c(j-1 | \mathbf{x})]} \geq \varepsilon v_{j-1}. \quad (2.19)$$

<sup>9</sup> A propriedade de martingale traduz-se em uma hipótese teórica do comportamento do agente, onde se quisermos estimar um valor para o período  $t$ , não haverá melhor valor que o ocorrido em  $t-1$ . Em outras palavras, o indivíduo não tem capacidade de prever o futuro.

Mas, ao assumir que  $\varphi(\mathbf{x}) = \exp(-\mathbf{x}\beta)$ ,  $\log v_{j-1}$  seja uma função logística e que  $\varepsilon$  seja identicamente definido e com valor igual a 1, se obtém o modelo probabilístico de transição escolar. Fazendo uso da expressão (2.7) por conveniência, temos:

$$\Pr(D_{s+1} = 1 | D_s = 1, \mathbf{X} = \mathbf{x}) = \frac{\exp\left[\frac{l(s) + \mathbf{x}\beta}{\sigma_s}\right]}{1 + \exp\left[\frac{l(s) + \mathbf{x}\beta}{\sigma_s}\right]}, \quad (2.20)$$

onde  $\sigma_s^2 = \text{var}(\log v_s)$  e a função  $l(s)$  representam a relação entre os retornos de martingale e custos marginais previamente parametrizados, ou seja:

$$\exp[l(s)] = \frac{R(s+1) - R(s)}{c(s+1) - c(s)}, \quad s = 1, \dots, \bar{S}. \quad (2.21)$$

Em suma, os trabalhos teóricos elencados aqui mostraram que o problema da evasão decorre de diferentes fatores tanto observáveis quanto não observáveis, como ambiente familiar, escolar, social, histórico de vida do aluno e psicológico, os quais, conjuntamente e cumulativamente, afetam tanto as taxas de abandono escolar quanto as de graduação.

## 2.3 Metodologia Econométrica

Em sintonia com os modelos teóricos de persistência escolar, esta seção descreve o modelo de regressão bivariado probabilístico, que foi utilizado para analisar um conjunto de variáveis que condicionam a probabilidade de um aluno do Ensino Médio da rede pública de Santa Catarina em progredir<sup>10</sup> de série, isto é, com o foco na transição do 1º para o 2º ano, devido a este momento ser considerado o mais crítico para a não permanência do aluno na rede pública de ensino (vide Tabela (2)). Os principais motivos que levam ao uso dessa abordagem são: a sua capacidade em modelar probabilidades conjuntas para duas variáveis, tal como progressão acadêmica e permanência na rede de ensino, além de fornecer uma estrutura de estimação que envolva questões de endogeneidade em um contexto de escolha binária (GREENE, 2003).

### 2.3.1 A modelagem bivariada *biprobit*

A estratégia empírica deste ensaio consiste na especificação e estimação de modelos probabilísticos univariados (*probit*) e bivariados (*biprobit*) para investigar questões relacionadas à progressão serial e permanência na rede pública de ensino. De acordo com Greene (2003), a abordagem *biprobit* é uma extensão natural da técnica *probit*, a qual permite mais de uma equação

---

<sup>10</sup> A base de dados fornecida pela SED-SC não nos permite ter certeza quanto à aprovação do aluno ao final do último ano do Ensino Médio. A base não possui nenhum indicador de graduação. Por essa razão, as informações referentes ao 3º ano não foram utilizadas nas estimações.

de regressão e assume uma distribuição normal padrão bivariada com média zero, variância unitária e coeficiente de correlação  $\rho$  entre os termos de erros das duas equações ( $\varepsilon_1, \varepsilon_2$ ). A seguir, segue a especificação do modelo:

$$y_1^* = \mathbf{x}'_1 \beta_1 + \varepsilon_1, \quad y_1 = 1 \text{ se } y_1^* > 0, = 0 \text{ caso contrário,} \quad (2.22a)$$

$$y_2^* = \mathbf{x}'_2 \beta_2 + \varepsilon_2, \quad y_2 = 1 \text{ se } y_2^* > 0, = 0 \text{ caso contrário;} \quad (2.22b)$$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} | \mathbf{x}_1, \mathbf{x}_2 \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]. \quad (2.23)$$

em que os vetores  $\mathbf{x}_1$  e  $\mathbf{x}_2$  contêm as variáveis explicativas e que as variáveis latentes contínuas  $y_1^*$  e  $y_2^*$  são definidas pelos resultados observados  $y_1$  e  $y_2$  e sua distribuição conjunta condicionada às covariadas.

Caso se tenha apenas a Equação (2.22a), tem-se o modelo *probit* padrão:

$$\begin{aligned} \Pr [y_{i1} = 1] &= \Pr [y_{i1}^* > 0] = \Pr [x'_{i1} \beta_1 + \varepsilon_{i1} > 0] \\ \Rightarrow \Pr [\varepsilon_{i1} > -x'_{i1} \beta_1] &= \Pr [\varepsilon_{i1} < x'_{i1} \beta_1] = \Phi (x'_{i1} \beta_1) \end{aligned} \quad (2.24)$$

onde o subscrito  $i$  indica a  $i$ -ésima unidade observada da amostra e  $\Phi(\cdot)$ , a função de distribuição Normal. Para se estimar o modelo *probit* bivariado, baseado nas equações acima, primeiro deve-se considerar quatro situações distintas de probabilidade ( $P^{y_{i1}, y_{i2}}(\theta)$ )

(CHRISTOFIDES; STENGOS; SWIDINSKY, 1997):

$$P_{11} = \Pr [y_{i1} = 1, y_{i2} = 1] = \int_{-\infty}^{x'_{i1}\beta_1} \int_{-\infty}^{x'_{i2}\beta_2} \Phi_2(z_1, z_2, \rho) dz_1 dz_2, \quad (2.25a)$$

$$P_{10} = \Pr [y_{i1} = 1, y_{i2} = 0] = \int_{-\infty}^{x'_{i1}\beta_1} \int_{x'_{i2}\beta_2}^{\infty} \Phi_2(z_1, z_2, \rho) dz_1 dz_2, \quad (2.25b)$$

$$P_{01} = \Pr [y_{i1} = 0, y_{i2} = 1] = \int_{x'_{i1}\beta_1}^{\infty} \int_{-\infty}^{x'_{i2}\beta_2} \Phi_2(z_1, z_2, \rho) dz_1 dz_2 \quad (2.25c)$$

$$P_{00} = \Pr [y_{i1} = 0, y_{i2} = 0] = \int_{x'_{i1}\beta_1}^{\infty} \int_{x'_{i2}\beta_2}^{\infty} \Phi_2(z_1, z_2, \rho) dz_1 dz_2 \quad (2.25d)$$

em que  $\Phi_2(z_1, z_2, \rho)$  denota a função de distribuição acumulada para a distribuição normal padrão bivariada com coeficiente de correlação  $\rho$ , que é definida por [Greene \(2003\)](#) assim:

$$\Phi_2(z_1, z_2, \rho) = \frac{\exp[-0,5(z_1^2 + z_2^2 - 2\rho z_1 z_2) / (1 - \rho^2)]}{2\pi(1 - \rho^2)^{1/2}}. \quad (2.26)$$

Para fins deste ensaio, utiliza-se por convenção o subscrito 1 (um) para os casos em que o aluno é aprovado (*aprov1*) e para a permanência na rede pública (*per1*) ao fim da 1ª Série e 0 caso contrário. A Tabela (1), a seguir, apresenta algumas das

notações utilizadas, bem como um sumário dos eventos relevantes e as probabilidades associadas.

Tabela 1 – Probabilidades dos eventos no Modelo *biprobit*

	<b>Permanecer na rede pública (<i>per</i>)</b>		
	Distribuição Conjunta		Marginal ( <i>per</i> )
	<i>per</i> = 1	<i>per</i> = 0	
<b>Aprovado (<i>aprov</i>)</b>			
<i>aprov</i> = 1	$\Phi_{aprov1,per1}$	$\Phi_{aprov1,per0}$	$\Phi_{per1}$
<i>aprov</i> = 0	$\Phi_{aprov0,per1}$	$\Phi_{aprov0,per0}$	$\Phi_{per0}$
Marginal ( <i>aprov</i> )	$\Phi_{aprov1}$	$\Phi_{aprov0}$	

Fonte: Elaboração própria

As intersecções das duas primeiras linhas e colunas da Tabela (1) representam as distribuições conjuntas de probabilidade, por exemplo,  $\Phi_{aprov1,per1}$  é a probabilidade de um estudante ser aprovado e permanecer na rede pública, enquanto  $\Phi_{aprov0,per1}$ , é a probabilidade de um aluno ser reprovado e continuar os estudos na sistema público estadual. Formalmente:

$$\Phi_{aprov1,per1} = \Pr (y_{aprov} = 1, y_{per} = 1 | \mathbf{x}) = \Phi (\mathbf{x}'_{aprov} \beta_{aprov}, \mathbf{x}'_{per} \beta_{per}, \rho), \quad (2.27)$$

$$\Phi_{aprov0,per1} = \Pr (y_{aprov} = 0, y_{per} = 1 | \mathbf{x}) = \Phi (-x'_{aprov} \beta_{aprov}, x'_{per} \beta_{per}, -\rho). \quad (2.28)$$

Similarmente, são definidas expressões para  $\Phi_{aprov1,per0}$  e  $\Phi_{aprov0,per0}$ . Por fim, a última linha e coluna reportam as probabilidades marginais:  $\Phi_{aprov1}$ ,  $\Phi_{aprov0}$  são as probabilidades marginais do aluno ser aprovado ou não; enquanto  $\Phi_{per1}$  e  $\Phi_{per0}$  são as probabilidades marginais do aluno permanecer na escola pública ou não.

Assim, para implementar a função log-verossimilhança (MLE) para o modelo (recursivo) *biprobit*, faz-se (GREENE, 2003):

1. Defina  $q_{i1} = 2y_{i1} - 1$  e  $q_{i2} = 2y_{i2} - 1$ , de modo que ,  
 $q_{ij} = 1$  se  $y_{ij} = 1$  e  $-1$  se  $y_{ij} = 0$  para todo  $j = 1$  e  $2$ ;
2. Assuma  $z_{ij} = \mathbf{x}'_{ij}\beta_j$  e  $w_{ij} = q_{ij}z_{ij}$  para  $j = 1, 2$ ;
3. Defina também  $\rho_{i*} = q_{i1}q_{i2}\rho$ ;
4. As probabilidades que entram na função de verossimilhança são do tipo:

$$\Pr (Y_1 = y_{i1}, Y_2 = y_{i2} | \mathbf{x}_1, \mathbf{x}_2) = \Phi_2 (w_{i1}, w_{i2}, \rho_{i*}). \quad (2.29)$$

Então, a função log-verossimilhança é assim expressa:

$$\ln L = \sum_{i=1}^n \ln \Phi_2 (w_{i1}, w_{i2}, \rho_{i*}), \quad (2.30)$$

em que as condições de primeira ordem (CPO) são:

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta_j} &= \sum_{i=1}^n \left( \frac{q_{ij}q_{ij}}{\Phi_2} \right) \mathbf{x}_{ij}, \quad j = 1, 2, \\ \frac{\partial \ln L}{\partial \rho} &= \sum_{i=1}^n \frac{q_{i1}q_{i2}\phi_2}{\Phi_2}, \end{aligned} \quad (2.31)$$

onde:

$$g_{i1} = \phi (w_{i1}) \Phi \left[ \frac{w_{i2} - \rho_{i*}w_{i1}}{\sqrt{1 - \rho_{i*}^2}} \right]. \quad (2.32)$$

$$g_{i2} = \phi (w_{i2}) \Phi \left[ \frac{w_{i1} - \rho_{i*}w_{i2}}{\sqrt{1 - \rho_{i*}^2}} \right]. \quad (2.33)$$

Na estimação de máxima verossimilhança (Equação (2.30),  $\rho$  não é obtido diretamente, e sim a tangente hiperbólica de  $\rho$ :

$$\operatorname{atanh} \rho = \frac{1}{2} \ln \left( \frac{1 + \rho}{1 - \rho} \right). \quad (2.34)$$

### 2.3.2 A correlação $\rho$

Greene (2003) sugere dois testes para se verificar a correlação entre os termos de erros na modelagem biprobit: a estatística do multiplicador Lagrange (LM) e a razão de verossimilhança (LR). O primeiro avalia a ausência de correlação, isto é, tem como hipótese nula  $H_0 : \rho = 0$ , o que sugere que modelo consiste em equações de probit independentes, que podem ser estimadas separadamente. Assim, a estatística LM para testar a hipótese nula em um modelo *probit* bivariado é:

$$LM = \frac{\left[ \sum_{i=1}^n q_{i1} q_{i2} \frac{\phi(w_{i1})\phi(w_{i2})}{\Phi(w_{i1})\Phi(w_{i2})} \right]^2}{\sum_{i=1}^n \frac{[\phi(w_{i1})\phi(w_{i2})]^2}{\Phi(w_{i1})\Phi(-w_{i1})\Phi(w_{i2})\Phi(-w_{i2})}}. \quad (2.35)$$

Quanto ao teste LR (ou Teste de Wald), caso não se rejeite a hipótese nula de  $\rho = 0$ , o modelo probit bivariado torna-se dois modelos independentes de probabilidade univariada, de modo que o valor da função de log-verossimilhança será simplesmente a soma das duas estatísticas obtidas para cada modelo. A estatística de teste é:

$$\lambda_{LR} = 2 [\ln L_{Biprob} - (\ln L_1 + \ln L_2)] \quad (2.36)$$

a qual converge para uma variável chi-quadrado com um grau de liberdade<sup>11</sup>.

### 2.3.3 Os efeitos marginais para a modelagem *birpobit*

Uma vantagem da aplicação do modelo bivariado *probit* sobre técnicas alternativas (Modelos *logit* e *SUR*) é a possibilidade de obtenção explícita da probabilidade conjunta e a maneira simplificada de se calcular os efeitos marginais sobre essa distribuição (CHRISTOFIDES; STENGOS; SWIDINSKY, 1997; HASEBE, 2013). Para Greene (2003), uma primeira estratégia seria calcular as derivadas parciais de  $Pr[y_1 = 1, y_2 = 1 | \mathbf{x}_1, \mathbf{x}_2]$ , para tanto deve-se realizar algumas transformações algébricas na Equação (2.31) e diferenciá-la em relação à  $\mathbf{x}_i$  e não  $\beta_j$ . Obtém-se:

$$\frac{\partial \Phi_2(\mathbf{x}'_1 \beta_1, \mathbf{x}'_2 \beta_2, \rho)}{\partial x_1} = \phi(\mathbf{x}'_1 \beta_1) \Phi \left( \frac{\mathbf{x}'_2 \beta_2 - \rho \mathbf{x}'_1 \beta_1}{\sqrt{1 - \rho_{i*}^2}} \right) \beta_1. \quad (2.37)$$

Note, entretanto, que a probabilidade bivariada expressa acima não é a média condicional. Assim, para se estimar o efeito

<sup>11</sup> É a quantidade de informação que os dados fornecem para se estimar os valores e as variabilidades dos parâmetros populacionais desconhecidos. É definido pela subtração do número de observações(n) pelo número de variáveis explicativas do modelo (k).

marginal médio, [Greene \(2003\)](#) define  $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$  e  $\mathbf{x}'_1\beta_1 = \mathbf{x}'\gamma_1$ . Então,  $\gamma_1$  contém todos os elementos de  $\beta_1$  diferentes de zero e possivelmente alguns zeros nos locais das variáveis de  $x$  que somente são regressores da segunda equação, ou seja, o termo em  $\gamma_1$  é zero, se  $x_i$  não estiver contido em  $\mathbf{x}_{i1}$ , e o mesmo ocorrerá para  $\gamma_2$ , caso esta situação se repita. Portanto, a probabilidade bivariada é:

$$\Pr [y_1 = 1, y_2 = 1 | x] = \Phi_2 [x'\gamma_1, x'\gamma_2, \rho]. \quad (2.38)$$

Assim, o efeito marginal médio de uma mudança em  $\mathbf{x}$  para um modelo definido pela equação acima, é dado por:

$$\frac{\partial \Phi_2}{\partial x} = g_1\gamma_1 + g_2\gamma_2 \quad (2.39)$$

onde  $g_1$  e  $g_2$  são definidos pela Equação (2.32). Nas situações que  $\rho = 0$ , tem-se o caso clássico univariado e o efeito que se pode avaliar é o impacto de uma variação de uma variável  $x_k$  na probabilidade marginal de, por exemplo,  $\Phi_{y_1}$ . Nessas condições, o efeito marginal será o mesmo do encontrado na modelagem *probit*,  $\phi(x'_{aprov}\beta_{aprov})\beta_{aprov}$ .

Assim, caso tenha-se:

$$\begin{aligned} E [y_{i1} | \mathbf{x}, y_{i2} = 1] &= \Pr [y_{i1} = 1 | \mathbf{x}, y_{i2} = 1] = \frac{\Pr[y_{i1}=1, y_{i2}|\mathbf{x}]}{\Pr[y_{i2}=1|\mathbf{x}]}, \\ &= \frac{\Phi_2(\mathbf{x}'\gamma_1, \mathbf{x}'\gamma_2, \rho)}{\Phi(\mathbf{x}'\gamma_2)}, \end{aligned} \quad (2.40)$$

os efeitos marginais médios para esta situação serão assim obtidos:

$$\frac{\partial E [y_{i1} | \mathbf{x}, y_{i2} = 1]}{\partial \mathbf{x}} = \left( \frac{1}{\Phi_2(\mathbf{x}'\gamma_2)} \right) \left[ g_1\gamma_1 + \left( g_2 - \frac{\phi(\mathbf{x}'\gamma_2)}{\Phi(\mathbf{x}'\gamma_2)} \right) \gamma_2 \right]. \quad (2.41)$$

## 2.4 A base de microdados da SEDSC

As informações utilizadas nesta pesquisa provêm da Secretaria de Educação de Santa Catarina (SEDSC) e abrangem o período de 2008 a 2012, onde há um *cross-section* de microdados representativos da época, contendo informações dos 428622 alunos matriculados no Ensino Médio na rede pública, de 649 escolas, em Santa Catarina, compostas majoritariamente por alunos do gênero feminino (53, 7%), da raça branca (34, 5%) e que moram no meio urbano (81, 2%).

Em termos de perspectiva operacional, a SEDSC dispõe de dois grandes bancos de dados: o SERIE-Escola, descontinuado em 2011; e, o Sistema de Gestão Educacional de Santa Catarina (SISGESC), que é o atual sistema de gerenciamento. O SERIE-Escola era um sistema instalado em cada escola, com banco de dados próprio (não integrado diretamente aos computadores da SED). Nesse sistema as informações da escola e dos alunos eram enviadas através de uma rotina mensal, gerando uma série de perdas de dados. Já o SISGESC é um sistema mais moderno e confiável, em que se mantém um banco de dados central acessado *online*.

Em ambos os sistemas é possível coletar e conectar diversos microdados, tais como: características da escola, contendo informações de oferta (e.g., se a escola tem laboratório de informática); características das turmas, também contendo informações de oferta (e.g., número de alunos por sala de aula e se determinada turma está na modalidade integradora); carac-

terísticas dos alunos (e.g., gênero e idade); e características de desempenho dos alunos, contendo informações de resultado (e.g., notas e faltas por disciplina). Em razão da mudança de sistema operacional ocorrida em 2011, usou-se o critério de conexão perfeita dos códigos da escola, série, turno e turma entre as bases – não havendo conexão perfeita, a informação era descartada.

A partir da conciliação das duas bases de dados, o indicador de resultado trabalhado neste ensaio é uma medida de progressão acadêmica do aluno dada pela sua manutenção na base de dados ao longo do Ensino Médio. Portanto, para aqueles alunos que permaneceram no sistema público de ensino, foram criadas duas variáveis: uma de progressão que trata do avanço do 1º para o 2º ano (*aprov*), e uma de permanência na escola pública (*per*) pois, para estes alunos, se tem certeza sobre estes dois fatos. Infelizmente, a informação sobre os alunos que cursam o 3º ano é dúbia, pois não há como inferir a condição do aluno ao final do ano (aprovação, reprovação, migração para sistema privado ou abandono). A Figura (3) e a Tabela (2) ilustram esta situação, onde o índice  $t$  identifica o ano e  $i$ , o estudante:

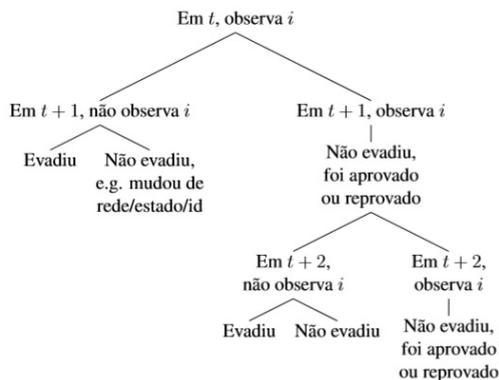


Figura 3 – Árvore de acompanhamento estudantil

Fonte: Formatação própria.

A coluna latência, na Tabela (2) abaixo, nos indica a quantidade de alunos que por algum motivo não está mais ativos na base de dados da SEDSC. Por exemplo, no ano de 2008, dos alunos que cursaram a 1<sup>o</sup> série, 3996 foram reprovados e 39518 foram aprovados<sup>12</sup>, enquanto que 15700 estão em estado de latência, isto é, ao final do ano cursado, nada podemos inferir sobre reprovação, aprovação, evasão, migração para a rede privada ou mudança de Estado, apenas que não mais constam na base de dados SEDSC. Além deste indicador de resultado, tentou-se utilizar as notas bimestrais de cada aluno. Infelizmente, muitos são os dados faltantes, principalmente em relação ao 4<sup>o</sup> bimestre do ano letivo, o que acabou por inviabilizar a utilização deste indicador. Optou-se, no final, por utilizar o indicador de progressão

<sup>12</sup> Isso pode ser afirmado com certeza, pois esses alunos permaneceram no sistema público de Ensino de Santa Catarina e a partir de uma interpretação *backward-looking*, tal suposição pode ser feita.

acadêmico como variável central na análise dos determinantes do desempenho escolar.

Tabela 2 – Desempenho acadêmico: Quantidade total de alunos. 2008 a 2012

Ano	Série	Situação			
		Reprovado	Aprovado	Latência	Total
2008	1	3996	39518	15700	59214
	2	1484	38267	7941	47692
	3	624		45972	46596
	Total	6104	77785	69613	155502
2009	1	5489	42105	12936	60530
	2	1974	39051	7133	48158
	3	565		44844	45409
	Total	8028	81156	64913	154097
2010	1	5305	38042	24178	67525
	2	2094	34045	16484	52623
	3	438		45988	46426
	Total	7837	72087	86650	166547
2011	1	2937	36451	11291	50679
	2	1304	35248	6405	42957
	3	321		36480	36801
	Total	4562	71699	54176	130437
2012	1			60634	60634
	2			54423	54423
	3			47962	47962
	Total	0	0	163019	163019

Fonte: SEDSC. Elaboração própria.

Podemos ver também na tabela, a quantidade de alunos matriculados em cada série no Ensino Médio. Entre 2009 e 2010, o número de alunos na 1ª série sofreu um aumento significativo, porém, para todo o período em análise, todas as séries seguintes apresentaram redução no número de alunos. Entre a 1ª e a 2ª série, a perda anual média foi de 17%, com ênfase em 2010, onde essa taxa foi de 22%, porém, na progressão da 2ª para a 3ª série, essa perda de alunos do sistema público reduz-se pela metade mas, em contrapartida, de 2008 e 2012, houve um aumento de 600% no número de alunos que não cursaram o último ano do Ensino Médio na rede pública.

Um outro dado que chama atenção refere-se ao ano de 2011, o qual foi o momento em que ocorreu a transição do sistema SERIE-Escola para o SISGESC. Nota-se, também, que para os anos de 2008, 2009, 2010 e 2012, a média total de alunos matriculados no Ensino Médio é de 159971, enquanto que, no ano em questão, esse número é quase 20% menor, e mais, quando comparado com o ano subsequente, há uma perda de informação para mais de 20000 alunos. Esse ruído no processo transitório evidencia que não houve uma convexão perfeita entre os dois sistemas.

Na literatura de desempenho acadêmico, nota-se que um vetor de variáveis-chave é recorrentemente analisado, o qual é passível de observação na base SEDSC e representa os impactos dos segmentos familiar, individual e escolar na decisão de evasão. O ambiente familiar de tal vetor é representado por: (i) zona de residência; (ii) pela *proxy* de renda familiar<sup>13</sup>; e, (iii) bolsa - e.g., participam ou não de algum programa de transferência direta de renda (Bolsa Família).

Por sua vez, o individual é composto por: (i) gênero; (ii) cor/raça; (iii) número de faltas cometidas no ano; (iv) transferência escolar; e, (v) repetência de série. E finalmente, o ambiente escolar: (i) número de funcionários (*proxy* para tamanho relativo de escola); (ii) idade média do professor (*proxy* para experiência profissional); (iii) existência de biblioteca; (iv) laboratório de informática e (v) quadra coberta.

Assim, com o intuito de descrever os microdados, a Tabela (3) a seguir, traz as estatísticas descritivas para vários destes indicadores de impacto.

---

<sup>13</sup> Produto Interno Bruto per capita Municipal obtido junto ao IBGE.

Tabela 3 – Estatísticas descritivas: Base de dados SEDSC

<b>Características Individuais/Familiares</b>			
<b>Variável</b>	<b>Classificação</b>	<b>Total</b>	<b>%</b>
<b>Gênero</b>	Feminino	230123	53,7
	Masculino	198196	46,2
	Não declarado	303	0,1
<b>Cor/Raça</b>	Branca	147953	34,5
	Preta	3504	0,8
	Parda	13102	3,1
	Outros	817	0,2
	Não declarado	263246	61,4
<b>Zona Residencial</b>	Rural	80192	18,8
	Urbana	345658	81,2
<b>Bolsa Família</b>	Sim	5477	1,3
	Não	423153	98,7
<b>Transferência Escolar<sup>1</sup></b>	Total	25555	5,96
	Feminino	10517	38,5
	Masculino	16798	61,5
<b>Alunos com pelo menos 1 reprovação<sup>1</sup></b>	Total	27315	6,37
	Feminino	10517	38,5
	Masculino	16798	61,5
<b>Total Matrículas</b>		<b>428622</b>	
<b>Características Escolares</b>			
<b>Biblioteca</b>	Sim	521	80,2
	Não	128	19,8
<b>Laboratório de Informática</b>	Sim	578	89,1
	Não	71	10,1
<b>Quantidade Média de Funcionários</b>		45,89 (23,94)	
<b>Quantidade Média de Alunos por sala</b>		26,05(6,52)	
<b>Idade Média Professor</b>		40,32 (2,85)	
<b>Total de Escolas</b>		<b>649</b>	

Fonte: SEDSC. Elaboração própria. Nota: Valores entre parênteses representam o desvio-padrão.  
1: Porcentagem calculada em relação ao valor total da variável.

Em termos de desempenho acadêmico, uma análise conjunta das Tabelas (2) e (3) indica um número elevado de alunos reprovados cursando o Ensino Médio na rede pública, além de uma grande quantidade de latência. Percebe-se que as taxas de reprovação estão concentradas no 1º ano sendo que, do total de casos, 66,2% ocorreram nesta série e que tal resultado se reduz drasticamente com o avançar das séries, atingindo um valor três vezes menor no 2º ano (25,4%) e atingindo ainda menos no último ano do Ensino Médio (8,4%).

A situação de risco para os alunos da 1º Série fica ainda mais evidente ao se examinar os dados para latência, pois os seus números são, em média, quase 90% superiores aos verificados para a 2ª Série. Esses indicadores mostram que, qualquer política pública que venha a impactar no Ensino Médio deve ter por objetivo reduzir o número de reprovações, pois esta parece ser um fator fundamental na decisão daqueles alunos que apresentam algum tipo de carência em continuar os estudos no sistema público de ensino.

A Figura (4) apresenta os histogramas do número de faltas para os alunos que nunca foram reprovados e para os já foram, com o objetivo de identificar um padrão de comportamento que possam decorrer da probabilidade de evadir. Logo, nota-se com clareza que os alunos reprovados faltam com mais frequência que aqueles que nunca foram reprovados, os quais são em sua maioria homens (62%), vivem no meio urbano (85,4%), suas famílias não estão inseridas em programa público de transferência de renda (98,7%) e frequentam majoritariamente os turnos matutino (45%)

e noturno (36%).

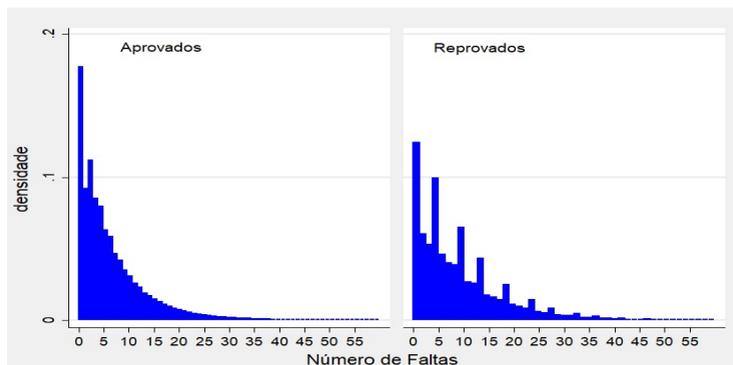


Figura 4 – Histograma do número de faltas de alunos aprovados e reprovados

Fonte dos dados: SEDSC. Elaboração própria.

Por fim, com o objetivo de se realizar um primeiro diagnóstico sobre a colinearidade entre algumas variáveis explicativas a serem utilizadas nos modelos propostos, foi construída uma matriz de correlação de Pearson<sup>14</sup>. A Figura 5 indica a presença de baixa correlação linear entre os pares de variáveis, portanto, provavelmente, as inferências baseadas nos modelos de regressão e suas aplicabilidades são críveis.

<sup>14</sup> A correlação expressa o grau de dependência linear entre duas variáveis e tem valores entre  $-1$  e  $1$ .

	Gênero	Zona Residencial	Funcionários por Escola	Biblioteca	Idade Média Professores por Escola	Alunos por turma	Distorção Idade-Série	Faltas
Gênero	1	0,006	-0,100	0,000	0,006	-0,110	0,127	0,310
Zona Residencial		1	0,194	-0,052	0,215	0,088	0,038	0,126
Funcionários por Escola			1	0,11	0,168	0,159	0,017	0,056
Biblioteca				1	-0,015	-0,018	-0,006	-0,013
Idade Média Professores por Escola					1	0,072	0,038	0,097
Alunos por turma						1	-0,52	-0,22
Distorção idade-Série							1	0,204
Faltas								1

Figura 5 – Matriz de correlação para variáveis explicativas

Fonte: Elaboração própria.

## 2.5 O modelo econométrico

A especificação dos determinantes da evasão escolar é complicada por dois motivos básicos. Primeiro, é difícil determinar e inferir a causalidade estrita dos vários fatores, onde apenas as características familiares e outros traços individuais, como sexo e raça, são fixas ou definidas com antecedência à decisão de evadir. Por exemplo, baixo rendimento acadêmico e inúmeras faltas não necessariamente levam o aluno a evadir, por outros motivos o aluno pode ter uma tendência a evadir e, por consequência, reduzir a dedicação a escola.

A segunda dificuldade trata da magnitude dos vários fatores que afetam a decisão de abandonar os estudos. Alguns dos fatores que a influenciam podem variar pelas mesmas circunstâncias, gerando problemas de endogeneidade. Por exemplo, a dedicação afeta a probabilidade de evadir, mas também é influenciada pelo ambiente familiar e outros antecedentes. A técnica

estatística de estimação que utiliza determinantes exógenos e endógenos deve ser capaz de solucionar esses entraves e, assim, expressar o real efeito líquido de todas as variáveis sobre a probabilidade de evadir.

Para examinar o desempenho acadêmico, faz-se necessário entender os fatores que o influenciam, onde as características familiares são apresentadas como um dos seus principais determinantes. Os estudos apresentados identificaram que a escolaridade dos pais, a renda familiar, o número de irmãos e a região onde vivem são variáveis importantes para a evasão, os quais sugerem que as duas primeiras são inversamente relacionadas ao abandono escolar. Em função da base de dados disponível, as características familiares inseridas nos modelos propostos retratam uma *proxy*<sup>15</sup> para a renda familiar, descrita pelo Produto Interno Bruto per Capita por cidade e um indicador binário para zona residencial, com opções para rural e urbana, com a primeira sendo a referência, dado que as taxas de abandono nas cidades são estatisticamente superiores à do meio rural (HANUSHEK, 1996; HAUSER; SIMMONS; PAGER, 2000; MENEZES-FILHO, 2007).

---

<sup>15</sup> Variável fortemente correlacionada, não necessariamente linear, com a variável de interesse.

Assim como a família, as características individuais dos estudantes também são importantes neste processo, principalmente em razão da hipótese de comportamento míope. Assim, com intuito de extrair tais qualidades, [Tinto \(1975\)](#) e [Bean \(1980\)](#) sugerem o uso de medidas de habilidade, de persistência, de histórico educacional e de gênero, enquanto [Cameron e Heckman \(1998\)](#) sugerem também o uso da localidade de residência e o uso de qualquer outra variável conhecida até o momento da decisão (notas, horas de trabalho) que possam a vir a afetar a utilidade corrente ou a futura do indivíduo.

Portanto, para que o nosso modelo seja capaz de refletir este contexto individual, tais características serão representadas por um indicador binário para gênero, sendo o feminino a categoria referência e um numérico para raça, subdividida em branca, preta, parda, amarela, indígena e não declarada, esta última sendo a característica referência. Outro binário para zona residencial, com opções para rural e urbana, a primeira sendo a referência.

Afora as características relacionadas ao indivíduo e a família, a interação dinâmica aluno-escola também é destaque em nossa análise de evasão escolar. Essa relação é medida por [Bean \(1980\)](#) através do envolvimento entre funcionários, principalmente professores e estudantes, que sugere que quanto maior o número e a diversidade de funcionários, maior será sua interação, portanto melhor o acompanhamento do desenvolvimento pedagógico desses alunos e, por consequência, menor a probabilidade de evasão. No modelo econométrico de [Eckstein e Wolpin \(1999\)](#), o ambiente escolar é especificado a partir das regras norte-

americanas para aprovação serial através do vetor de variáveis estado-espço( $S_t$ ), porém, pode ser incluído nele qualquer fator conhecido que venha a influenciar na decisão do aluno. Tratamento semelhante é dado por [Cameron e Heckman \(1998\)](#), em que o vetor  $\mathbf{X}_{so}$  representa todas aquelas variáveis observadas que influenciam a decisão do aluno ao avançar de série.

Logo, para especificar esse ambiente, foram incluídas nos modelos empíricos o número de funcionários por escola, a quantidade de alunos por turma, as características médias dos professores, como idade e perfil demográfico ([HANUSHEK; RIVKIN; TAYLOR, 1996](#); [RIVKIN; HANUSHEK; KAIN, 2005](#)), além de aspectos da estrutura física da escola, como a existência de biblioteca e laboratório de informática. A inclusão dessas variáveis nos mostra que o ambiente escolar não é definido apenas pelas relações inter alunos, mas sim pela interação de alunos, professores e funcionários administrativos da escola, mas também que as atividades extracurriculares também podem reduzir a probabilidade de reprovação, uma vez que elas oferecem prêmios sociais e educacionais que podem vir a aumentar o comprometimento do aluno com a instituição.

Em adição as características relacionadas à família, ao indivíduo e à escola, este trabalho também avaliou as implicações da dinâmica comportamental e de interesse do estudante para a predição do desempenho acadêmico. Para tanto, este trabalho incluiu algumas variáveis, como o número de alunos por turma e a quantidade de faltas, além de um indicador binário para transferência escolar, sendo a categoria de referência a permanência na escola.

Com base nessas informações, este trabalho tem como objetivo especificar um modelo de progressão serial, como foco na progressão serial da 1<sup>a</sup> para a 2<sup>a</sup> série do Ensino Médio no estado de Santa Catarina a partir de uma base de dados fornecida pela SEDSC, para o período de 2008 a 2012, com vistas à identificar as características individuais, familiares e escolares mais relevantes para a realização acadêmica. A seguir, são apresentadas essas variáveis com seus respectivos códigos de entrada no *STATA 13*::

Tabela 4 – Descrição das variáveis extraídas da SEDSC

<b>Código STATA</b>	<b>Descrição</b>
<b>aprov</b>	Variável binária, na qual possui valor 1 se o estudante é aprovado ao final da 1ª série e 0 caso contrário;
<b>d</b>	Variável binária, na qual possui valor 1 se o estudante permanece na rede pública de ensino ao fim da 1ª série e 0 caso contrário;
menino	Variável binária, na qual possui valor 1 se o estudante é menino e 0 caso contrário;
branco	Variável binária, na qual possui valor 1 se o estudante se autodeclara de cor branca e 0 caso contrário;
rural	Variável binária, na qual possui valor 1 se o estudante reside na zona rural e 0 caso contrário;
bolsa	Variável binária, na qual possui valor 1 se a família do estudante recebe Bolsa Família e 0 caso contrário;
computador	Variável binária, na qual possui valor 1 se a família do estudante tem computador em casa e 0 caso contrário;
lnpib_pc	Logaritmo natural do PIB per capita por município;
e_biblio	Variável binária, na qual possui valor 1 se a escola tem biblioteca e 0 caso contrário;
e_lab_info	Variável binária, na qual possui valor 1 se a escola tem laboratório de informática e 0 caso contrário;
i_funcionarios_razao	Razão entre número de estudantes e funcionários por escola;
prof_idade	Idade média dos professores por escola;
alunos_turma	Quantidade de alunos na turma do aluno;
faltas	Número de faltas por ano de cada estudante;
transfer	Variável binária, na qual possui valor 1 se a aluno mudou de escola ao longo do Ensino Médio e 0 caso contrário.

Fonte: Elaboração própria.

Nesse contexto, foram propostas três especificações para os modelos probabilísticos univariados (*probit*), no qual a variável binária *aprov* é a dependente e os bivariados (*biprobit*) que têm as variáveis binárias *aprov* e *d* como dependentes.. A primeira especificação traz apenas as características individuais e familiares, a segunda adiciona aspectos estruturais e pedagógicos das escolas, os quais são diretamente afetadas pela ação do Estado e, por fim, a terceira especificação, que inclui além das características individuais e escolares, particularidades da vida acadêmica do jovem. Tem-se, então:

**Especificações *probit*:**

**M1\_P:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i + \beta_6lnpib\_pc_i + \varepsilon_{ik} \end{aligned} \tag{2.42}$$

**M2\_P:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i + \beta_6lnpib\_pc_i + \beta_7e\_biblio_k + \beta_8e\_lab\_info_k \\ & + \beta_9i\_funcionarios\_razao_k + \beta_{10}prof\_idade_k + \varepsilon_{ik} \end{aligned} \tag{2.43}$$

**M3\_P:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i + \beta_6lnpib\_pc_i + \beta_7e\_biblio_k + \beta_8e\_lab\_info_k \\ & + \beta_9i\_funcionarios\_razao_k + \beta_{10}prof\_idade_k + \beta_{11}transfer_i \\ & + \beta_{12}faltas_i + \beta_{13}alunos\_turma_{ik} + \varepsilon_{ik} \end{aligned} \tag{2.44}$$

**Especificações *Biprobit*:****M1\_B:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i + \beta_6lnpib\_pc_i + \varepsilon_{ik} \end{aligned} \quad (2.45)$$

$$\begin{aligned} d_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i + \beta_6lnpib\_pc_i + \varepsilon_{ik} \end{aligned} \quad (2.46)$$

**M2\_B:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i\beta_6lnpib\_pc_i + \beta_7e\_biblio_k + \beta_8e\_lab\_info_k \\ & + \beta_9i\_funcionarios\_razaok + \beta_{10}prof\_idade_k + \varepsilon_{ik} \end{aligned} \quad (2.47)$$

$$\begin{aligned} d_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i\beta_6lnpib\_pc_i + \beta_7e\_biblio_k + \beta_8e\_lab\_info_k \\ & + \beta_9i\_funcionarios\_razaok\beta_{10}prof\_idade_k + \varepsilon_{ik} \end{aligned} \quad (2.48)$$

**M3\_B:**

$$\begin{aligned} aprov_{ik} = & \alpha + \beta_1menino_i + \beta_2branco_i + \beta_3rural_i + \beta_4bolsa_i \\ & + \beta_5i\_computador_i\beta_6lnpib\_pc_i + \beta_7e\_biblio_k + \beta_8e\_lab\_info_k \\ & + \beta_9i\_funcionarios\_razaok + \beta_{10}prof\_idade_k + \beta_{11}transfer_i \\ & + \beta_{12}faltas_i + \beta_{13}alunos\_turma_{ik} + \varepsilon_{ik} \end{aligned} \quad (2.49)$$

$$\begin{aligned}
 d_{ik} = & \alpha + \beta_1 \text{menino}_i + \beta_2 \text{branco}_i + \beta_3 \text{rural}_i + \beta_4 \text{bolsa}_i \\
 & + \beta_5 i\_computador_i + \beta_6 \ln \text{pib\_pc}_i + \beta_7 e\_biblio_k + \beta_8 e\_lab\_info_k \\
 & + \beta_9 i\_funcionarios\_razao_k + \beta_{10} \text{prof\_idade}_k + \beta_{11} \text{transfer}_i \\
 & + \beta_{12} \text{faltas}_i + \beta_{13} \text{alunos\_turma}_{ik} + \varepsilon_{ik}
 \end{aligned}
 \tag{2.50}$$

em que o subscrito  $i$  refere-se ao aluno e  $k$  à escola. O indicador de resultado, *aprov*, é uma medida de progressão acadêmica de  $i$  dado pela sua manutenção na base de dados (acompanhamento *backward-looking*) ao longo do Ensino Médio. Vale ressaltar, também, que o indicador de permanência na rede pública,  $d$ , foi proposto de maneira semelhante ao do indicador de desempenho.

## 2.6 Resultados empíricos

Neste ensaio, foram desenvolvidos dois tipos de modelos econométricos, um modelo univariado probit e um bivariado probit para se estimar a probabilidade de progressão do 1º para o 2º ano do Ensino Médio da rede pública de Santa Catarina. A Tabela (5), a seguir, traz os valores dos parâmetros estimados<sup>16</sup>, bem como seus respectivos desvios-padrão e estatísticas  $t$  para ambas as técnicas para os três modelos sugeridos.

---

<sup>16</sup> Cálculos realizados no *software STATA 13*.

Tabela 5 – Estimações *probit* e *biprobit* para a transição acadêmica da 1ª série 2ª série do Ensino Médio

Variáveis	Probit		Biprobit		Probit		Biprobit	
	aprov	d	aprov	d	aprov	d	aprov	d
menino	-0,2256*** (0,0046)	-0,1436*** (0,0047)	-0,2273*** (0,0046)	-0,2274*** (0,0046)	-0,2219*** (0,0047)	-0,1446*** (0,0047)	-0,2220*** (0,0046)	-0,1387*** (0,0047)
branco	0,2331*** (0,0109)	0,1826*** (0,0109)	0,2358*** (0,0109)	0,2335*** (0,0109)	0,2279*** (0,0109)	0,1840*** (0,0109)	0,2255*** (0,0109)	0,1763*** (0,0109)
rural	0,2470*** (0,0060)	0,2462*** (0,0060)	0,2113*** (0,0061)	0,2146*** (0,0062)	0,2145*** (0,0062)	0,1819*** (0,0062)	0,2153*** (0,0062)	0,1827*** (0,0062)
bolsa	-0,0513*** (0,0197)	-0,0492*** (0,0197)	-0,0695*** (0,0197)	-0,0682*** (0,0197)	-0,0668*** (0,0198)	-0,0457*** (0,0199)	-0,0654*** (0,0197)	-0,0432*** (0,0199)
computador	0,1627*** (0,0052)	0,1637*** (0,0052)	0,1741*** (0,0053)	0,1760*** (0,0053)	0,1698*** (0,0053)	0,1696*** (0,0053)	0,1634*** (0,0053)	0,1634*** (0,0053)
lnpib_pc	-0,2993*** (0,0058)	-0,3018*** (0,0058)	-0,2590*** (0,0060)	-0,2615*** (0,0060)	-0,2690*** (0,0060)	-0,2911*** (0,0060)	-0,2713*** (0,0060)	-0,3015*** (0,0060)
e_biblio			0,0507*** (0,0058)	0,0509*** (0,0058)	0,0509*** (0,0058)	0,0486*** (0,0058)	0,0510*** (0,0058)	0,0486*** (0,0058)
e_lab_info			0,0484*** (0,0071)	0,0485*** (0,0071)	0,0507*** (0,0071)	0,0306*** (0,0071)	0,0507*** (0,0071)	0,0328*** (0,0071)
i_funcionarios_razao			-0,0116*** (0,0005)	-0,0117*** (0,0005)	-0,0116*** (0,0006)	-0,0089*** (0,0005)	-0,01159*** (0,0006)	-0,0131*** (0,0006)
prof_idade			-0,0145*** (0,0009)	-0,0145*** (0,0009)	-0,0143*** (0,0009)	-0,0141*** (0,0009)	-0,0142*** (0,0009)	-0,0139*** (0,0009)
transfer								
faltas_completas								
alunos_turma								
Constante	2,8980*** (0,0603)	2,9246*** (0,0602)	3,0960*** (0,0677)	3,1115*** (0,0676)	2,8961*** (0,0680)	3,5614*** (0,0681)	2,9178*** (0,0679)	3,3738*** (0,0684)
athrtho		5,0314 (51,5439)		5,4251 (53,9712)		5,4251 (53,9712)		5,4399 (42,9712)
Observações	298582	298582	298582	298582	298582	298582	298582	298582

Fonte: Elaboração própria.

Coef → Valor dos coeficientes estimados; se → desvio-padrão.

\*\*\*  $p < 0,01$ , \*\*  $p < 0,05$ , \*  $p < 0,1$ .

Iniciou-se a análise com a comparação dos parâmetros estimados para as técnicas *probit* e *biprobit* com a variável dependente binária *aprov.* Em ambas as metodologias, o coeficiente estimado não pode ser interpretado diretamente como a mudança marginal na probabilidade do aluno ser aprovado, apenas o sinal (indicativo da direção do efeito) e a significância estatística dos parâmetros podem ser considerados (GREENE, 2003).

Dentre as variáveis explicativas, para todos os três modelos especificados nas duas técnicas, observa-se que: todas são estatisticamente significantes a um nível de significância de 1%; que pessoas do sexo masculino têm uma menor probabilidade de serem aprovados do que as do sexo feminino; as pessoas declaradas brancas têm uma maior chance de sucesso acadêmico; jovens que residem na zona rural têm maiores chances de serem aprovados do que os que habitam regiões urbanas; quanto às variáveis que identificam a estrutura da escola, a presença de quadra e laboratório de informática influenciam positivamente o desempenho acadêmico do aluno, entretanto, professores mais velhos e uma quantidade de funcionários pequena parecem não corroborar para a aprovação do aluno; alunos que faltam mais têm uma maior probabilidade de repetir de série; e, por fim, quanto às variáveis que expressam o nível socioeconômico do aluno, *bolsa* e *lnpib\_pc* apresentaram sinais negativos, porém o sinal encontrado para a segunda variável não parece fazer muito sentido, o que é uma indicação de problemas de viés de seleção amostral, que será verificado no Ensaio 3 desta tese.

Agora observe os valores obtidos para estatística *athrho*, referentes às estimações *biprobit*<sup>17</sup> na vizinhança de +5, que se traduz em uma correlação estimada  $\hat{\rho}$  próximo a +1. Assim, realizou-se um teste LR para verificar a correlação entre os termos de erros das duas equações, em que rejeitou-se a hipótese nula de  $\rho = 0$ . Portanto, aprovação e permanência são diretamente dependentes, e as equações devem ser estimadas conjuntamente em um modelo *biprobit* ao invés de equações *probit* separadas.

As Tabelas (6) e (7), abaixo, apresentam os valores para as probabilidades marginais preditas pelas técnicas *probit* e *biprobit* para as suas respectivas variáveis dependentes. A primeira tabela traz os resultados obtidos pela metodologia univariada, mostrando que em todas as três especificações a probabilidade estimada mais elevada de aprovação foi de cerca de 80%. Todavia, a adição de aspectos que ilustram as características estruturais da escola e de participação do aluno, reduziram significativamente a probabilidade de reprovação do aluno, alcançando o terceiro modelo uma valor quase 7 vezes menor que o primeiro, que apenas contém variáveis que descrevem aspectos familiares do estudante.

Tabela 6 – Probabilidades marginais preditas: modelos *probit*

Modelo	Média	Desvio-Padrão	Mínimo	Máximo
M1_P	0,5254	0,0833	0,2145	0,7908
M2_P	0,5254	0,0876	0,1645	0,8023
M3_P	0,5203	0,0990	0,0345	0,8363

Fonte: Elaboração própria.

<sup>17</sup> Qualquer valor estimado de *athrho* superior a +1, o software STATA 13 converge o modelo para um valor próximo de  $\rho = 1$ . Para melhor entendimento, verificar Seção 3.4.2.

Como era de se esperar, as probabilidades marginais previstas para a variável *aprov*, estimadas pela técnica *biprobit*, foram muito similares às encontradas na *probit*. Quanto aos valores obtidos para a variável dependente *per*, o modelo M3\_B foi aquele que apresentou a menor probabilidade marginal.

Tabela 7 – Probabilidades marginais preditas: modelos *biprobit*

Modelo	Variáveis	Média	Desvio Padrão	Mínimo	Máximo
M1_B	$\Phi_{aprov1}$	0,5254	0,0834	0,2139	0,7915
	$\Phi_{per1}$	0,5841	0,0737	0,2824	0,8203
M2_B	$\Phi_{aprov1}$	0,5254	0,0879	0,1644	0,8033
	$\Phi_{per1}$	0,5841	0,0771	0,2498	0,8274
M3_B	$\Phi_{aprov1}$	0,5203	0,0990	0,0358	0,8367
	$\Phi_{per1}$	0,5791	0,0896	0,0696	0,8585

Fonte: Elaboração própria.

A comparação entre os resultados estimados para os modelos *probit* e *biprobit* mostram como a magnitude, os graus de significância dos coeficientes estimados e as probabilidades marginais preditas são semelhantes com a variável dependente *aprov*. Além disso, pode-se observar que os fatores que afetam a probabilidade de um indivíduo ser aprovado exercem um efeito similar sobre a probabilidade dese permanecer na rede pública de ensino. No entanto, existem alguns benefícios analíticos para o uso de métodos de modelagem de *probit* bivariados. Por exemplo, a análise de efeitos marginais nos permite considerar os impactos do vetor de variáveis explicativas em várias combinações das duas variáveis independentes. Essas combinações formam os resultados para as quatro probabilidades conjuntas,  $\Phi_{aprov0,per0}$ ,  $\Phi_{aprov0,per1}$ ,  $\Phi_{aprov1,per0}$  e  $\Phi_{aprov1,per1}$ . No caso específico deste

ensaio, apenas as combinações  $\Phi_{aprov1,per1}$  e  $\Phi_{aprov0,per1}$  são críveis (Equações (2.27) e (2.28)), devido às características da base de microdados da SEDSC.

Quanto às probabilidades conjuntas previstas, os resultados apresentados na Tabela (8) mostram que a probabilidade máxima de permanência na rede pública de ensino de um jovem reprovado é bastante reduzida, atingindo valores de cerca de 12% para as três especificações. Todavia, os valores estimados indicam que há uma maior chance de transição serial para aqueles estudantes que são aprovados. Esses resultados indicam que as políticas públicas educacionais que tenham como foco o abandono e a evasão escolar devem sugerir medidas de transição de série automáticas.

Tabela 8 – Probabilidades conjuntas previstas: modelos *biprobit*

Modelo	Cenário	Média	Desvio-Padrão	Mínimo	Máximo
M1_B	$\Phi_{aprov0,per1}$	0,0588	0,0182	0,0192	0,1152
	$\Phi_{aprov1,per1}$	0,5254	0,0834	0,2139	0,7915
M2_B	$\Phi_{aprov0,per1}$	0,0588	0,0192	0,0118	0,1186
	$\Phi_{aprov1,per1}$	0,5254	0,0879	0,1644	0,8033
M3_B	$\Phi_{aprov0,per1}$	0,0588	0,0192	0,0110	0,1233
	$\Phi_{aprov1,per1}$	0,5203	0,0990	0,0358	0,8367

Fonte: Elaboração própria.

### 2.6.1 Efeitos marginais para o modelo *biprobit*

Esta seção discorre sobre os efeitos marginais dos modelos bivariados relatados nas Tabelas (9), (10) e (11). Esses efeitos correspondem às variações na probabilidade de um por cento nas variáveis explicativas contínuas, e no caso das *dummies*, eles

referem-se a uma variação discreta de 0 para 1 (GREENE, 2003). Esse tipo de análise oferece o benefício analítico de permitir considerar os efeitos de cada variável explicativa sobre as probabilidades conjuntas observadas, no caso deste ensaio  $\Phi_{aprov0,per1}$  e  $\Phi_{aprov1,per1}$ .

Para o caso de estudantes do sexo masculino, nota-se que eles apresentam uma maior probabilidade de não terem um desempenho acadêmico satisfatório do que as estudantes do sexo feminino, isto é, na média dos três modelos, possuem uma probabilidade 9% menor de serem aprovados e continuar nos estudos. Quanto à ser branco e morar na zona rural, os resultados indicam que a chance do estudante ser aprovado e continuar os estudos aumentam em quase 10% a chance do estudante ser aprovado e permanecer na rede de ensino. Quanto ao fato da família do estudante estar inscrita no Programa Bolsa Família, tem impacto bastante reduzido no comportamento estudantil do jovem quando comparado a jovens cujas famílias não recebem o auxílio público. Já os efeitos da variável *computador*, mostrou-se de pouca relevância para o caso dos alunos reprovados que permaneceram na base de dados, entretanto, para aqueles que foram aprovados, os alunos que tiveram acesso a computador possuíram 6% maior de serem aprovados e continuar na rede pública de ensino. Por fim, os efeitos marginais médios estimados para a variável *lnpib\_pc* aparentam estar viesados, já que imagina-se que o nível socioeconômico tenha um efeito positivo no desempenho acadêmico daqueles alunos que permanecem na escola pública.

Tabela 9 – Efeitos marginais médios: M1\_B

Variáveis	Média	Pr(aprov0,per1)	Pr(aprov1,per1)
menino	0,4764	0,0338	-0,0898418
branco	0,9522	-0,0210	0,0922355
rural	0,1926	-0,0156	0,0980428
bolsa	0,0142	0,0073	-0,01958
computador	0,2759	-0,0025	0,0652118
lnpib_pc	10,1753	-0,0066	-0,1201858

Fonte: Elaboração própria.

Os efeitos marginais estimados para os modelos M2\_B e M3\_B apresentaram resultados semelhantes para as características familiares apresentadas no modelo M1\_B. A segunda especificação difere da primeira pela introdução de variáveis que descrevem a estrutura física e pedagógica da escola. Como esperado, as variáveis  $e\_biblio$  e  $e\_lab\_info$  praticamente não influenciam no desempenho dos alunos reprovados, mas promovem uma ligeira melhora, de cerca de 2% a mais de alunos aprovados em escolas que possuem esses dois equipamentos em relação aos que cursam escolas que não os têm.

Em relação ao impacto do tamanho relativo da escola, mensurado pela variável contínua  $i\_funcionarios\_razao$ , este mostrou-se de pouco efeito para a melhora e comprometimento do jovens com os estudos, assim como a idade média dos professores, a qual inclusive mostrou-se não significativa para  $\Phi_{aprov0,per1}$ . Ao se comparar os dois primeiros modelos, nota-se que o impacto das características familiares é bem mais relevante na determinação do rendimento acadêmico do que as variáveis que caracterizam o

ambiente escolar.

Tabela 10 – Efeitos marginais médios: Modelo M2\_B

Variáveis	Média	Pr(aprov0,per1)	Pr(aprov1,per1)
menino	0,4764	0,0342	-0,0906
branco	0,9522	-0,0212	0,0930
rural	0,1926	-0,0145	0,0855
bolsa	0,0142	0,0093	-0,0272
computador	0,2759	-0,0039	0,0701
lnpib_pc	10,1753	-0,0094	-0,1042
e_biblio	0,7938	-0,0013	0,0203
e_lab_info	0,8734	-0,0074	0,0193
i_funcionarios_razao	7,5292	0,0012	-0,0046
prof_idade	40,7436	0,0003*	-0,0058

Fonte: Elaboração própria.

\* Não é estatisticamente significativa.

Por fim, os resultados estimados para o Modelo M3\_B mostraram que a transferência escolar tem efeitos expressivos no rendimento escolar do estudante. Os alunos que mudaram de colégio apresentaram uma probabilidade 20% menor de serem aprovados e permanecer na rede pública de ensino do que aqueles que não tiveram que passar por essa situação ao longo do Ensino Médio.

Sobre à quantidade de faltas, elas comprometem o rendimento escolar, por consequência, aumentam a probabilidade de reprovação dos estudantes, apesar da magnitude do efeito ser bastante reduzida. As estimativas para o efeito do número de alunos por turma também apresentaram uma magnitude muito baixa, que pode ser explicada pela pouca variabilidade no tamanho das turmas entre os colégios.

Tabela 11 – Efeitos marginais médios: M3\_B

Variáveis	Média	Pr(aprov0,per1)	Pr(aprov1,per1)
menino	0,4764	0,0343	-0,0884
branco	0,9522	-0,0210	0,0898
rural	0,1926	-0,0145	0,0857
bolsa	0,0142	0,0092	-0,0260
computador	0,2759	-0,0039	0,0676
lnpib_pc	10,1753	-0,0096	-0,1080
e_biblio	0,7938	-0,0014	0,0203
e_lab_info	0,8734	-0,0074	0,0202
i_funcionarios_razao	7,5292	0,0012	-0,0063
prof_idade	40,7436	0,0002	-0,0057
transfer	0,0091	0,0091	-0,2011
faltas	0,3865	0,0014	-0,0094
alunos_turma	26,2287	-0,0001	0,0050

Fonte: Elaboração própria.

\* Não é estatisticamente significativa.

Os resultados estimados dos efeitos marginais para as três especificações evidenciaram que as características individuais e familiares são as que têm maior probabilidade de efeito sobre o comportamento médio dos estudantes do Ensino Médio de Santa Catarina. Além disso, o formador de política deve estar bastante atento ao caso específico dos alunos que mudam de colégio, pois estes apresentaram uma probabilidade de desempenho escolar bastante inferior aos alunos que cursaram todo o Ensino Médio na mesma instituição pública de ensino.

## 2.7 Considerações finais

Como apresentado no início deste ensaio, uma preocupação recorrente da literatura é compreender os principais fatores individuais, familiares e escolares que definem o desempenho acadêmico de jovens de 15 a 17 anos e, por consequência, como esta performance está relacionada com a decisão de continuidade dos estudos.

Este ensaio realizou um exercício empírico com o objetivo de identificar quais são as variáveis mais relevantes para a progressão serial do aluno da rede pública de ensino de Santa Catarina, em especial da transição da 1<sup>a</sup> para a 2<sup>a</sup> Série do Ensino Médio, usando modelos probabilísticos univariados e bivariados. Esse momento foi o foco do trabalho, pois as estatísticas descritivas mostraram que os maiores índices de reprovação e latência foram encontrados para a 1<sup>a</sup> série (Tabelas (2) e (3)), o que corrobora com as evidências da literatura nacional (FERRÃO; BELTÃO; SANTOS, 2002; BARROS; MENDONÇA, 2008; MENEZES-FILHO, 2007; NERI et al., 2015) de que a transição do Ensino Fundamental para o Médio e no primeiro ano da última etapa do ciclo básico, são os momentos em que há as maiores taxas de evasão.

A inclusão da análise de efeitos marginais de um modelo *biprobit* é um dos diferenciais deste trabalho, pois a maioria dos estudos empíricos sobre transição e permanência na rede de ensino geralmente optam por equações únicas para a predição das respectivas probabilidades. A seguir, os principais resultados:

1. A 1ª Série do Ensino Médio é a que merece mais atenção por parte dos formuladores de política, em virtude dos altos índices de reprovação e latência;
2. Os meninos apresentam uma probabilidade de rendimento escolar, em média, 10% inferior às meninas, além de terem maiores chances de reprovação;
3. Jovens que residem na zona rural têm maior probabilidade de sucesso acadêmico do que os que moram nas cidades;
4. A escola ofertar biblioteca e laboratório de informática praticamente não influencia o desempenho acadêmico dos piores alunos, isto é, daqueles que reprovaram ao final da 1ª Série;
5. Professores mais velhos têm influência negativa no desempenho acadêmico e permanência na rede pública de ensino;
6. As características familiares são as mais importantes para os indicadores de resultado escolar de um aluno da 1ª série do Ensino Médio;

7. O efeito de programas de transferência de renda são mínimos em estudantes da 1ª Série;
8. Atenção especial deve ser dedicada aos alunos de transferência, pois estes reportaram uma probabilidade 20% inferior de rendimento escolar entre aqueles alunos que foram aprovados e continuaram os estudos na rede pública;
9. A probabilidade de um estudante reprovado continuar os estudos na rede pública ao fim do 1º ano é baixa.

Ademais, concluiu-se que possíveis problemas de seleção amostral podem influenciar o comportamento esperado de algumas variáveis, como  $\ln pib\_pc$  e, por isso, deve-se adotar técnicas que corrijam tal situação.

# 3 Modelo de inferência de Heckman aplicado ao abandono escolar no Ensino Médio

## 3.1 Introdução

Na literatura da Economia da Educação, as taxas de diplomação e evasão do Ensino Médio vem sendo utilizadas como indicadores de produtividade do sistema educacional e eficácia do desenvolvimento econômico e social de um país, o que leva os agentes públicos a proporem uma variedade de programas (Novo Ensino Médio, ProEMI, Educação para a Juventude, etc. ) e medidas avaliativas (ENEM, Ideb) que diagnostiquem as dificuldades e os problemas dos estudantes, principalmente daqueles em desvantagem social, com o intuito de promover o bem estar da sociedade como um todo.

No que tange à educação brasileira, a evasão escolar está entre os temas que historicamente fazem parte de debates e reflexões no âmbito da educação pública, as quais visam buscar e encontrar respostas para questões que quantifiquem e qualifiquem o abandono, as suas causas e o que pode ser feito em termos de políticas públicas para mitigar este problema, segundo [Leon e](#)

Menezes-Filho (2002), Krawczyk et al. (2009), Neri et al. (2009), Neri et al. (2015).

A história do Ensino Médio é marcada pelo descaso do Estado brasileiro em torná-lo mais acessível as camadas mais populares da população, e em formular políticas públicas que motivem esses grupos sociais a ingressar e permanecer nesse nível de educação. Parcialmente, essa perspectiva começa a se alterar a partir da década de 1990, com a promulgação de marcos legislativos, programas específicos e reformas curriculares, como por exemplo a Lei de Diretrizes e Bases (LDB) (1996), a instituição do Exame Nacional do Ensino Médio (ENEM) (1998), a criação do Fundo de Manutenção e Desenvolvimento da Educação Básica (FUNDEB) (2006), e, mais recentemente, a reforma curricular intitulada como Novo Ensino Médio (2017), entre outros.

Para Krawczyk et al. (2009), merece destaque a mudança de tratamento dada a esse nível de ensino desde a homologação da LDB, ao expressar que “a inclusão do Ensino Médio no âmbito da educação básica e o seu caráter progressivamente obrigatório demonstram o reconhecimento da importância política e social que ele possui.” Essa mudança reproduz tanto o anseio das camadas mais pobres por mais escolarização quanto a necessidade de atendimento de maiores exigências apresentadas no contexto econômico do âmbito da concorrência no mercado de trabalho e do comércio internacional.

Apesar dos avanços ocorridos ao longo dos últimos 25 anos em relação à consolidação regulatória e normativa, universalizar e aprimorar qualitativamente o Ensino Médio se apresenta hoje como um dos principais desafios no campo das políticas educacionais. O INEP exemplifica essa realidade ao mostrar que em 2014, as taxas de escolarização líquida, abandono e reprovação eram de 56,3%, 12,2% e 7,6%, valores estes bem distantes dos encontrados nos países desenvolvidos.

Para Plank, DeLuca e Estacion (2008) e Doll, Eslami e Walters (2013), o abandono escolar representa o evento final de um processo de longo prazo de descontinuidade de atendimento à educação formal. Para entender esse fenômeno, se faz necessário a compreensão das características dos estudantes em relação às questões sociais, econômicas, psicológicas e também de fatores institucionais (escola, família, comunidade), em especial do Ensino Médio, que contribuem para que o jovem tome a decisão de abandonar os estudos antes que ele venha a se formar.

Segundo Neri et al. (2015), as principais motivações de jovens, com até 17 anos, para estarem fora da escola dizem respeito à: questões de logística para chegar na escola; necessidade de trabalho e geração de renda; falta de interesse intrínseco nas atividades escolares, motivado principalmente pelo desconhecimento dos potenciais prêmios oferecidos pela educação; entre outros, como sucessivas reprovações, gravidez, prostituição, drogas e problemas familiares.

Um dos primeiros passos de qualquer análise estatística para a educação é propor um modelo de trajetória escolar, que destaque os aspectos mais importantes deste fluxo. Assim, com base em Klein (2007) e Gonçalves (2008), foi elaborada a Figura 6, a seguir, para ilustrar o fluxo escolar para o Ensino Médio.

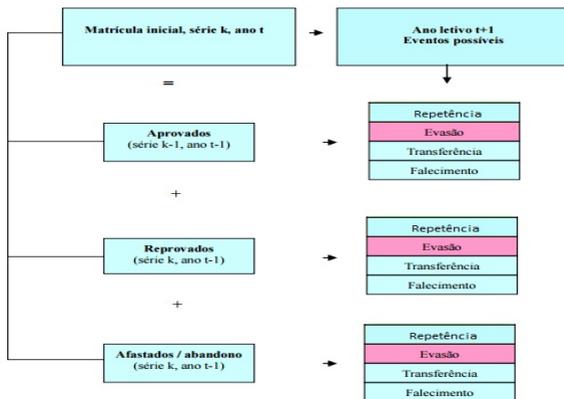


Figura 6 – Modelo de fluxo escolar

De acordo com a figura, o número de alunos com matrículas no ano letivo  $t$  representa a soma dos alunos aprovados na série imediatamente anterior ao ano  $t - 1$  e promovidos para a nova série; os alunos repetentes em  $t - 1$  e; dos alunos que abandonaram ou evadiram durante o ano letivo  $t$ . Percebe-se, também, que os alunos aprovados na série  $k$  no ano  $t$  podem incorrer nas seguintes situações: avanço de série; reprovação; evasão; transferência para escolas privadas ou supletivo; e, em casos extremos, falecimento. Quanto aos alunos que evadem, têm comportamento similar, com exceção do evento promoção.

Em outras palavras, os estudantes com matrículas ativas são os que irão incorrer no futuro em algum evento, seja aprovação, reprovação, diplomação, evasão ou transferência de rede de ensino. Entretanto, aqueles alunos que foram diplomados serão tratados como dados censurados, uma vez que não fornecem informações limitadas acerca do evento analisado, no caso em análise sobre o abandono da rede pública de ensino.

Para se fazer uma análise mais detalhada a respeito do desempenho acadêmico e dos motivos que levam o jovem a abandonar o Ensino Médio, além da probabilidade de ocorrência desse fenômeno, é comum utilizar um conjunto de características (observáveis) relacionadas à família, à escola e ao desempenho acadêmico do estudante. Após serem coletadas as informações em relação a essas variáveis, são realizados testes padronizados de desempenho escolar e abordagens estatísticas, principalmente os de modelos de seleção amostral com restrição de informação, para se verificar o nível de importância de cada um desses fatores ao longo do processo de decisão do aluno e assim, definir em qual momento tal atitude é deliberada (PLANK; DELUCA; ESTACION, 2008; DUBIN; RIVERS, 1989).

Desde o começo dos anos 1970, muito se evoluiu na literatura econométrica em torno dos modelos de tratamento-reposta ou seleção amostral, quanto à atribuição ao tratamento ou participação em algum programa ser não aleatória. O problema estatístico em tais modelos se origina da própria decisão de participação no tratamento ser endógena, o que faz com que a estimativa por mínimos quadrados ordinários (*MQO*) distorça o

seu real impacto do programa com efeitos de fatores não observados que influenciam tanto a tomada decisão quanto o resultado de interesse. Para Sartori (2003), a presença de variáveis omitidas gera um problema ainda mais grave, pois os resultados obtidos a partir de técnicas básicas não reproduzem com precisão os efeitos marginais das variáveis independentes.

Nesse ambiente de viés e dificuldade de mensuração da variável dependente, a estratégia empírica deste ensaio consiste em utilizar a técnica de correção de Heckman (1977), que expressa-se em duas equações: as equações de participação (ou seleção) e a de regressão (ou interesse). A primeira trata da probabilidade do aluno permanecer na rede pública de ensino, enquanto a segunda explica o seu desempenho acadêmico, incluindo o termo de correção derivado da distribuição do termo de erro da equação de participação. Tem-se como objetivo identificar as características individuais e escolares que mais afetam o rendimento acadêmico e a permanência do aluno na escola pública e, assim, contribuir para a análise da crise de audiência do Ensino Médio, no Brasil.

Para a realização deste exercício, a utilização do Censo Escolar se mostrou inviável, uma vez que é vedada a identificação do aluno em anos subsequentes em uma base de dados pública, limitando a capacidade de análise. No entanto, tivemos acesso á uma base de dados nunca antes observada, que foi obtida junto à Secretaria de Educação de Santa Catarina (SEDSC), e contém informações diversas sobre as características dos alunos, escolas e professores, para os anos de 2008 a 2012, vinculados à

rede pública de ensino do estado de Santa Catarina. Constatase que é possível saber, ao término de cada ano para aqueles que permaneceram no sistema público, se o aluno foi aprovado ou não, a quantidade de dias que faltou, o tamanho da turma em que estava matriculado, entre outras informações relevantes para a análise dos condicionantes, tanto do resultado acadêmico, quanto da permanência no sistema educacional.

O ensaio encontra-se organizado da seguinte maneira: além desta introdução, a Seção 2 traz uma breve explanação acerca do problema de viés de seleção e progressão acadêmica, Seção 3, faz uma revisão teórica acerca do uso de modelos de seleção amostral em experimentos sociais, a Seção 4 trata da metodologia econométrica com ênfase na técnica de máxima verossimilhança, da descrição da base de dados, além de apresentar algumas estatísticas descritivas, a Seção 5 trata da base de dados e algumas estatísticas descritivas, a Seção 6 apresenta o modelo econométrico, a Seção 7 traz os resultados empíricos, e, por fim, a Seção 8 trata das conclusões e da relevância do ensaio no auxílio do desenvolvimento de políticas públicas que melhorem o desempenho acadêmico dos estudantes do Ensino Médio.

## 3.2 A questão do viés de seleção amostral

Com o intuito de explicitar sobre a correção do viés de seleção amostral, é útil realizar uma análise intuitiva acerca do mesmo. Para isto, conceba as situações em que a estruturação de uma base de dados utilizada para uma análise econométrica é limitada pelo processos de coleta de informação, o que reduziria a possibilidade de se fazer inferências estatísticas geradas a partir da amostra para toda a população.

Existem basicamente duas formas gerais para o problema do viés de seleção. O caso clássico, que trata das situações em que não se têm informação a respeito da variável dependente para parte dos indivíduos analisados. Por exemplo, caso se deseje estimar o efeito da educação sobre o salário das mulheres, nos deparamos com o fato de que muitas delas não estão engajadas em nenhum tipo de trabalho assalariado e por consequência, não apresentam renda. Mas caso uma parte considerável desse desemprego resulte de uma taxa de retorno relativamente baixa para a educação, a estimativa de uma regressão em que a renda é a variável dependente e a educação seja uma das variáveis explicativas, pode gerar estimativas viesadas do efeito da educação sobre a renda (HECKMAN, 1977; WINSHIP; MARE, 1992).

A segunda versão do problema trata do caso em que a informação acerca da variável dependente está disponível para todos os componentes da base de dados, mas a distribuição dos indivíduos sobre as categorias da variável independente que in-

teressa ao pesquisador ocorreu de maneira seletiva. Para ilustrar esse caso, considere a situação que se busca avaliar os efeitos da migração sobre a renda, usando uma amostra aleatória de uma população, sabendo-se a renda de todos os indivíduos e se migraram ou não para outro lugar no passado. Caso se faça uma regressão em que a variável dependente seja a renda e que uma variável binária (*dummy*) indicando a ocorrência de migração no passado faça parte do conjunto de variáveis explicativas, pode-se ter uma estimativa viesada dos efeitos da migração porque a distribuição dos indivíduos nas categorias de imigrantes e não imigrantes não foi aleatória. Isso acontece porque pessoas que decidiram imigrar podem se diferenciar em muitas características (perceptíveis ou não) daqueles que não, e caso estes aspectos estejam relacionados com a variável dependente, o seu coeficiente estimado pode carregar esses efeitos e provocar o viés. O controle dessas características resolveria o problema, entretanto, geralmente isso não é possível, porque em qualquer conjunto de dados o número de fatores de controle é limitado, enquanto que o número de possíveis diferenças entre os indivíduos é infinito. Essa segunda forma de viés também é conhecido como viés de heterogeneidade.

Apesar dos exemplos serem meramente ilustrativos, fica evidente que a seletividade amostral opera tanto através de elementos não observáveis e sua correlação com os variáveis observadas (VELLA, 1998), quanto no processo de seleção e formação da base de dados, em que indivíduos podem ser divididos em dois (ou mais) grupos (empregado/desempregado, imigrantes/não imi-

grantes, alunos que saem do sistema público de ensino/alunos que não saem) e que a não aleatoriedade do processo também pode distorcer as relações de interesse. Nessa última situação, [Winship e Mare \(1992\)](#) evidenciam que há dois processos interligados, os quais podem ser descritos por duas equações, a equação de seleção e a equação de interesse, sendo que tal relação se reflete numa correlação diferente de zero entre os termos de erro de cada equação.

A possibilidade de viés de seleção amostral surge em várias aplicações econômicas e, conseqüentemente, a metodologia para controlar tal problema tornou-se comumente empregada em microeconometria.

### 3.3 Revisão teórica

A maioria dos fenômenos sociais e políticos mais interessantes são aqueles nos quais a amostra não é aleatoriamente definida ([WINSHIP; MARE, 1992](#); [SARTORI, 2003](#)) e que as informações são incompletas e restritas ([DUBIN; RIVERS, 1989](#)) como, por exemplo, a participação ou não em uma guerra, em crimes, na escolha de uma nova forma de governo após uma guerra civil, em pesquisas de emprego e salários, no volume e diversidade de investimentos de uma empresa privada, na opção por abandonar os estudos, em tratamentos médicos, entre outros. Nesse contexto, um número crescente de trabalhos faz uso de procedimentos estatísticos para evitar os efeitos do problema de viés de seleção.

Mare e Winship (1984) realizaram uma análise multivariada para duas abordagens distintas, a primeira, um modelo *probit* de equação única que prediz se um indivíduo está empregado ou não, e a segunda, um outro modelo *probit*, mas de duas equações que preveriam em conjunto a probabilidade de empregos para homens fora da escola e se ele está ou não na população economicamente ativa não militar, para investigar tendências de emprego, entre os anos de 1960 e 1980, para homens brancos e negros que estão fora da escola, da faixa etária dos 16 aos 29, utilizando como covariadas os anos matriculados na escola, o tempo fora da escola, a idade, entre outras.

Os autores concluíram que com o passar do tempo, os fatores que afetam a seletividade se modificou porque a proporção da população que estava fora da escola, principalmente de negros, reduziu, que o padrão de empregabilidade observado é viesado, em função da probabilidade de emprego e de abandono escolar serem dependentes e, por fim, que para se fazer uma interpretação substantiva sobre a perspectiva de emprego para os jovens que estão fora da escola, é necessário a análise conjunta dos níveis de emprego e matrículas escolares.

A questão do viés de seleção também foi explorada em estudos de violência, com destaque para os trabalhos de Peterson e Hagan (1984) e Zatz e Hagan (1985), nos quais o foco central da análise do crime e da punição são os determinantes das diferenças no tratamento de pessoas em contato com o sistema de justiça criminal, por exemplo, a gravidade diferencial da punição de negros e brancos. Para os autores, a amostra de pessoas que

foram condenadas é extremamente seletiva, pois, daqueles que cometeram algum tipo de crime, apenas uma parte foi presa, destes, uma proporção foi processada, destes, uma parcela foi condenada, dos quais apenas uma parte foi para a prisão, uma vez que fatores não observados podem afetar todas essas etapas do processo. Eles verificaram que nas situações em que é elevada a probabilidade de uma pessoa ser condenada, caso seja presa, podem levar a exageros e abusos da polícia no momento da prisão, além disso o grau de severidade das sentenças, dependendo do perfil da pessoa condenada, está sujeitos a viés de seleção.

No campo da Economia da Educação, [Manski e Wise \(1983\)](#) investigam os determinantes da graduação universitária, assim como a capacidade dos resultados obtidos no exame *Scholastic Aptitude Test* (SAT<sup>1</sup>) em prever as probabilidades do aluno cursar todo o ensino superior. A base de dados utilizada nesse estudo foi obtida junto à *National Center for Educational Statistics? National Longitudinal Study* (NLS), que concluiu o Ensino Médio em 1972, a qual foram aplicados modelos econométricos de variável dependente limitada, como *probit*, *logit* e *logit condicional*<sup>2</sup> sendo utilizadas as notas do SAT, a classificação acadêmica do aluno na sua classe, além do nível educacional e renda dos pais como variáveis explicativas. Concluiu-se que as notas do SAT têm uma influência maior na decisão do aluno em

---

<sup>1</sup> Exame admissional utilizado pelas universidades americanas como parte do processo de seleção de novos estudantes, composto de questões de múltipla escolha de Matemática e Inglês, administrado pelo *College Board*. É semelhante ao ENEM.

<sup>2</sup> Para maiores detalhes ver [Wooldridge \(2015\)](#).

aplicar a uma universidade do que na decisão desta em aceitá-lo, além de servirem como *proxy* para os estudantes avaliarem tanto sua própria capacidade quanto o nível de habilidade exigido em escolas alternativas. Os estudantes escolhem não se candidatarem às universidades que são muito difíceis.

Em relação à capacidade de predição dos resultados do SAT para o sucesso universitário, os autores perceberam que eles têm um efeito significativo na conclusão do ensino superior, e que estudantes que obtiveram notas baixa no exame admissional têm uma probabilidade de evadir muito maior que aqueles que obtiveram notas altas. Entretanto, o nível de qualidade e excelência da escola secundária consegue reduzir a relevância do teste nessa decisão, desde que os alunos entrem no Ensino Médio com a perspectiva de ir para a faculdade.

## 3.4 O Modelo de Heckman

Embora apresente uma trajetória decrescente ao longo dos últimos 20 anos, a evasão escolar no Ensino Médio, no Brasil, ainda é elevada, o que prejudica e retarda o desenvolvimento econômico do país, em razão da baixa taxa de acumulação de capital humano que mantém em níveis muito reduzidos a produtividade da população. Nesse contexto, a educação formal é o principal elemento de composição do capital humano, o que torna a manutenção do aluno na escola a principal meta das políticas públicas educacionais voltadas para o público dos 15 aos 17 anos.

Em termos econométricos, um dos principais desafios na estimação dos determinantes da permanência escolar é que esta é uma variável difícil de ser completamente mensurável, porque alunos saem da base, mudam de identificador dentro da base, mudam de unidade da federação, etc., o que faz com que o “ficar na escola” seja uma sucessão de resultados acadêmicos, por isso a variável dependente utilizada neste ensaio é aprovação.

Outro problema surge do fato da situação acadêmica não ser observável para quem não está matriculado na rede de ensino, como por exemplo, para os casos das crianças que apresentam maiores restrições sócioeconômicas para frequentarem a escola. Logo, ignorar o fato de que o resultado acadêmico observado é truncado pode gerar um viés de seleção na estimativa dos parâmetros da equação de aprovação.

Um número considerável de métodos estatísticos pode ser utilizado para controlar o possível viés, em que o método convencional baseia-se na estratégia proposta por Heckman (1977)<sup>3</sup>, desenvolvida usando uma estrutura econométrica para o manejo de variáveis dependentes limitadas na presença do problema de viés de seleção. Essa situação surge de duas fontes principais (BRIGGS, 2004):

1. A autoseleção dos indivíduos e/ou dos dados que são investigados;

---

<sup>3</sup> O modelo original foi projetado para abordar o problema de estimar o salário médio das mulheres usando dados coletados de uma população de mulheres em que as donas de casa foram excluídas por autoseleção.

2. Definição da amostra por parte dos pesquisadores, que se torna um processo semelhante a uma auto-seleção.

Heckman (1977) demonstra que a estimação por MQO das relações entre as covariadas gera estimadores inconsistentes. Logo, o uso de amostras compostas exclusivamente por um único tipo categórico de indivíduo, como por exemplo, alunos que estão cursando um determinado período escolar ou que estejam trabalhando pode não ser representativo de toda a população, conduzindo a um viés de seleção.

Para solucionar o problema de viés de seleção das informações, Heckman (1977) propôs um modelo de seleção amostral bivariado<sup>4</sup> em que a variável dependente  $y_{1i}$  é observada apenas para uma proporção da base de dados. Portanto, para uma base de dados  $(X, Z)$  para  $N$  indivíduos existe um subconjunto  $N_1 = N - N_0$  que se observa a variável dependente de interesse,  $y_{1i}$ , mas que não é observada para o restante  $N_0$  das observações. A seguir, o modelo é formalmente apresentado (CAMERON; TRIVEDI, 2010):

$$y_{1i} = X_i\beta + \varepsilon_{1i}, \quad (3.1)$$

$$y_{2i} = Z_i\delta + \nu_{2i}, \quad (3.2)$$

onde  $X_i = (1, x_{12}, \dots, x_{1L})$ ,  $Z_i = (1, z_{12}, \dots, z_{1K})$ ,  $\beta$  e  $\delta$  são vetores dimensionais  $L$  e  $K$  respectivamente e não necessariamente

<sup>4</sup> Também conhecido como modelo *heckit* por sua equivalência ao modelo Tobit estocástico limitado (Tobit model with stochastic threshold) desenvolvido por Nelson (1977).

iguais,  $y_{2i}$  é uma variável dependente binária e o subscrito  $i$  refere-se ao indivíduo.

Nesse contexto, [Cameron e Trivedi \(2010\)](#) representam o modelo de seleção amostral bivariado a partir de duas equações: uma de regressão padrão (Eq. 3.1) e uma de seleção ou participação (Eq. 3.2), que procura explicar como os indivíduos fazem parte da amostra. Respectivamente, ambas as equações são assim definidas:

$$y_{1i} = \begin{cases} y_{1i}^* & \text{se } y_{2i}^* > 0 \\ \cdot & \text{se } y_{2i}^* \leq 0; \end{cases} \quad (3.3)$$

$$y_{2i} = \begin{cases} 1 & \text{se } y_{2i}^* > 0 \\ 0 & \text{se } y_{2i}^* \leq 0 \end{cases} \quad (3.4)$$

Esse modelo especifica que  $y_{1i}$  será observado apenas quando  $y_{2i}^* > 0$  (isto é, quando  $y_{2i} = 1$ ), de modo que  $y_{1i}$  não precisa ter qualquer valor significativo quando  $y_{2i}^* \leq 0$  e que é função do valor da equação de seleção, a qual relaciona uma variável latente  $y_{2i}$  à algumas características observáveis  $Z_i$ . Dessa forma, para um  $y_{1i}$  positivo, a densidade observada é  $f^*(y_1^* | y_2^* > 0) \times \Pr(y_2^* > 0)$  e o modelo bivariado de seleção amostral tem a seguinte função de verossimilhança ([CAMERON; TRIVEDI, 2010](#)):

$$L = \prod_{i=1}^n \{\Pr(y_{2i}^* \leq 0)\}^{1-y_{2i}} \{f^*(y_{1i} | y_{2i}^* > 0) \times \Pr(y_{2i}^* > 0)\}^{y_{2i}}, \quad (3.5)$$

onde o primeiro termo é a contribuição discreta quando  $y_{2i}^* \leq 0$  desde que  $y_{2i} = 0$  e o segundo termo é a contribuição contínua quando  $y_{2i}^* > 0$ .

O ponto de partida do Modelo de Heckman é a equação de seleção, que descreve os motivos pelos quais um indivíduo decide por participar ou permanecer em alguma situação, no caso desta pesquisa, sobre aqueles motivos que levam o aluno a permanecer na rede pública de ensino, onde o vetor  $Z_i$  contém as variáveis explicativas relacionadas à probabilidade da pessoa (aluno, firma, imigrante) incorrer em tal evento. O termo de erro  $\nu_{2i}$  é definido como uma variável aleatória contínua latente não mensurada que tem distribuição normal padrão. Portanto, a decisão do indivíduo em participar é determinada por uma combinação linear de covariadas observadas ( $Z_i$ ) e não observadas ( $\nu_i$ ), onde o indivíduo irá participar do evento (permanecer no Ensino Médio público/) caso  $Z_i\delta + \nu_{2i} > 0$ , do contrário não.

Portanto, dadas as hipóteses de independência das covariadas em relação aos termos de erro e que estes são independentes e identicamente distribuídos (*iid*) para cada  $i$  com distribuição normal padrão, a função de participação pode ser reescrita da seguinte maneira (BRIGGS, 2004):

$$P(y_{2i} = 1 | Z_i) = P(Z_i\delta + \nu_{2i} > 0 | Z_i) \quad (3.6)$$

$$P(y_{2i} = 1 | Z_i) = P(-\nu_{2i} < Z_i\delta | Z_i) \quad (3.7)$$

$$P(y_{2i} = 1 | Z_i) = \Phi(Z_i\delta), \quad (3.8)$$

em que  $\Phi$  representa a função de distribuição cumulativa normal padrão e a Equação (3.8) irá constituir um modelo *probit*<sup>5</sup>.

<sup>5</sup> Para mais informações acerca desse grupo de modelos, ver Wooldridge

Quanto aos termos de erro,  $\varepsilon_{1i}$  e  $\nu_{2i}$ , estes seguem uma distribuição normal bivariada e são homocedásticos:

$$\begin{bmatrix} \varepsilon_1 \\ \nu_2 \end{bmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_{12} \\ \rho\sigma_{21} & 1 \end{pmatrix} \right] \quad (3.9)$$

com três parâmetros de escala  $\sigma_1$ ,  $\sigma_{12}$  e  $\sigma_{21}$  e coeficiente de correlação  $\rho$ , que é o termo que conecta as duas equações, assumidas independentes das covariadas ( $X$  e  $Z$ ) mas que apresentam dependência entre matérias. Atente que a variância do termo de erro ( $\nu_2$ ) de uma equação *probit* pode ser normalizada para igual a 1 ( $\sigma_2^2 = 1$ ) sem perda de generalidade, porque a escala da variável dependente é não observada, isto é, a variância não é identificada no modelo (WOOLDRIDGE, 2015).

Para Cameron e Trivedi (2010) deve-se considerar a média truncada da amostra selecionada apenas quando os valores de  $y_1$  estiverem sendo utilizados, de modo que:

$$\begin{aligned} E[y_1|x, y_2^* > 0] &= E[X_i\beta + \varepsilon_1 | Z_i\delta + \nu_2 > 0] \\ &= X_i\beta + \sigma E[\varepsilon_1 | \nu_2 > -Z_i\delta] \end{aligned} \quad (3.10)$$

em que  $\mathbf{x}$  representa a união de  $X$  e  $Z$ . Caso os termos de erro  $\varepsilon_1$  e  $\nu_2$  sejam independentes, o último termo simplifica para  $E[\nu_2] = 0$  e a regressão MQO de  $y_1$  em  $X$  terá um estimador consistente para  $\beta$ . Entretanto, se houver qualquer correlação entre os resíduos a média truncada não mais será  $X_i\beta$ , sendo necessário considerar a seletividade.

Assumindo as hipóteses de normalidade para os resíduos e que a distribuição conjunta de duas variáveis,  $x$  e  $y$ , seja normal bivariada, Heckman (1977) demonstra que na expectativa de  $y$  condicional a  $x$  truncada em um valor maior que  $\alpha$  qualquer, por exemplo 0, teremos:

$$E[y|x > 0] = \mu_y + \rho\sigma_y\lambda\left(\frac{0 - \mu_x}{\sigma_x}\right). \quad (3.11)$$

Logo,  $E[y_{1i}|y_{2i} > 0]$  é:

$$\begin{aligned} E[y_1|y_2 > 0] &= X\beta + E[\varepsilon_1|\nu_2 > -Z\delta] = X\beta + \rho\sigma_{12}\lambda\left(\frac{-Z\delta}{\sigma_2=1}\right) \\ &= X\beta + \rho\sigma_{12}\lambda\left(\frac{-Z\delta}{1-\Phi(-Z\delta)}\right) = X\beta + \rho\sigma_{12}\lambda\left(\frac{Z\delta}{\Phi(Z\delta)}\right), \end{aligned} \quad (3.12)$$

em que  $\lambda(\cdot)$  é uma estimativa do viés,  $E[\varepsilon_1|X, y_2]$ , definido a partir do valor esperado de uma variável aleatória truncada<sup>6</sup>, em que  $Z\delta$  é o ponto onde a distribuição foi truncada. Essa estimativa é conhecida como a Razão Inversa de Mills, que pode ser expressa pela razão das funções de densidade normal padrão  $\phi(Z\delta)$  e de densidade cumulativa normal padrão  $\Phi(Z\delta)$ . Entre as características que mais se destacam desta variável temos: a de que o denominador da razão representa a probabilidade da população com características geradoras de viés (neste caso, os alunos que evadiram do sistema público de ensino) sejam selecionados na amostra e;  $\lambda_t$  é função monotônica decrescente da função de probabilidade de seleção amostral  $\Phi(t)$  (HECKMAN, 1977). Na prática,  $\lambda(Z\delta)$  não é conhecido, mas dadas as hipótese de que  $\varepsilon_{1i}$  e  $\nu_{2i}$  têm distribuições normais padrão,  $\hat{\lambda}$  pode ser estimado como

<sup>6</sup> Para mais detalhes, ver Greene (2003).

uma função dos parâmetros estimados  $\hat{\delta}$  na equação de seleção (3.2).

Agora, assumindo que a relação entre  $y_{1i}$  e  $y_{2i}$  é devida a  $X$  e que todo o viés de seleção é originado por  $\hat{\lambda}$ , então ao regredir  $y_1$  em relação a uma constante,  $y_2$ , as covariadas e  $\hat{\lambda}$  irão quase sempre controlar o viés na estimativa, de tal modo que  $\hat{\beta}$  irá convergir para  $\beta$  assintoticamente. Então,  $\hat{\beta}$  será tendencioso, mas consistente (HECKMAN, 1977; CAMERON; TRIVEDI, 2010).

A interpretação do modelo apresentado na Equação (3.12) diz que a reta de regressão para  $y$  em  $X$  será viesada para cima quando  $\rho$  é positivo e para baixo quando o contrário, desde que a razão inversa de Mills ( $\lambda$ ) seja sempre positiva. O grau do viés depende da magnitude da correlação, da variância relativa do erro ( $\sigma_1$ ) e da rigidez do truncamento ( $\lambda$ , e será maior quanto menor o valor de corte de  $Z\delta$ ). Note, ainda, que quando  $\rho = 0$  não haverá viés e, por consequência, o estimador será não tendencioso, mas caso a Equação (3.1) seja estimada por MQO, o segundo termo será omitido e as estimativas dos coeficientes serão inconsistentes (BRIGGS, 2004; SARTORI, 2003).

### 3.4.1 A estimação de dois estágios de Heckman

Heckman (1977) propõe uma metodologia para se estimar a função de verossimilhança expressa na Equação (3.5) a partir de um método de dois estágios, em que o mecanismo de

seleção amostral tem um resultado dicotômico e a equação de interesse tem uma variável dependente contínua. Intuitivamente, a ideia básica é inserir a variável omitida ilustrada na Equação (3.12) na função de regressão (3.1). Conforme os autores, essa técnica é conhecida como estimador *heckit*, a qual é baseada na observação de que o valor condicional esperado de  $y_{1i}$ , dado  $y_{2i} = 1$  é:

$$E(y_{1i} | y_{2i} = 1) = X'_i \beta + \rho \sigma_1 \lambda_i(Z_i \delta), \quad (3.13)$$

que especifica o modelo de regressão:

$$y_i = X'_i \beta + \rho \sigma_1 \lambda_i(Z_i \delta) + v_i. \quad (3.14)$$

O procedimento de dois estágios é o seguinte:

1. Faz-se a regressão *probit* para a Equação (3.2) e os valores previstos são retidos como estimativas de  $Z_i \hat{\delta}$ ;
2. A partir dos valores de  $Z_i \hat{\delta}$ , estima-se para cada observação  $i$  da amostra a razão inversa Mills:

$$\frac{\phi(Z_i \hat{\delta})}{\Phi(Z_i \hat{\delta})}; \quad (3.15)$$

3. Estima-se  $\hat{\beta}$  por MQO de  $y_{1i}$ ,  $X'_i \beta$  e  $\hat{\lambda}_i$  produzindo estimativas consistentes de  $\beta$ . O modelo linear será:

$$y_{1i} = X'_i \beta + \rho \sigma_{12} \lambda(Z'_i \hat{\delta}) + \varepsilon_{1i}, \quad (3.16)$$

Percebe-se que a regressão acima não fornece diretamente a estimativa de  $\hat{\sigma}_1^2$ . Para tanto, utiliza-se a fórmula da variância truncada definida a seguir:

$$\hat{\sigma}_1^2 = \frac{1}{N} \sum_i \left[ \hat{\varepsilon}_1^2 + \hat{\sigma}_{12}^2 \hat{\lambda}_i \left( X'_{i\beta} + \hat{\lambda}_i \right) \right], \quad (3.17)$$

onde  $\hat{\varepsilon}_i$  é o resíduo MQO da Equação (3.16) e  $\hat{\lambda}_i = \lambda \left( Z'_i \hat{\delta} \right)$ , sendo possível notar que  $V(\varepsilon_1)$  não é constante, variando tanto com  $i$  quanto com  $Z$ . Diante da necessidade de se obter um estimador consistente para a matriz de variância-covariância assintótica, Heckman (1977) propôs a realização de um teste-t em relação a  $\lambda$  em que a hipótese nula é a não existência de viés de seleção, enquanto que Lee (1982) sugere o uso do método de White<sup>7</sup> por não necessitar de nenhuma premissa acerca da variância não constante dos resíduos.

Por sua vez, a correlação ( $\rho$ ) entre os termos de erro das Equações (3.1) e (3.2) será estimada por:

$$\hat{\rho} = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1}. \quad (3.18)$$

Uma maneira para se verificar a correlação entre os resíduos e por consequência, da necessidade de se utilizar a correção de seletividade amostral, ou seja, se  $\sigma_{12} = 0$  ou  $\rho = 0$ , é o teste estatístico de Wald<sup>8</sup> baseado em  $\hat{\sigma}_{12}$ , que é o valor do coeficiente estimado da razão inversa de Mills (CAMERON; TRIVEDI, 2010).

<sup>7</sup> Para maiores detalhes ver Greene (2003)

<sup>8</sup> Para mais detalhes ver (KODDE; PALM, 1986).

No que diz respeito a especificação a função de seleção, Heckman (1977) não fornece nenhuma indicação sobre quais variáveis incluir nesta equação, apenas assume que  $\{Z_i : i = 1, \dots, n\}$  é independente de  $\{\nu_{2i} : i = 1, \dots, n\}$ . Por sua vez, Greene (2003) e Wooldridge (2015) sugerem que a matriz de covariadas dessa função deva conter uma ou mais variáveis relacionadas à probabilidade de participar do evento, as quais devem ser excluídas da predição da variável dependente de interesse, enquanto que Briggs (2004) propõem a separação total das covariadas da equação de seleção das covariadas da equação de interesse. Portanto, a decisão sobre quais variáveis incluir em  $Z_i$  parece, em grande parte, ser uma questão de garantir que o modelo esteja bem identificado e de reduzir a colinearidade entre os regressores, assim como o desvio padrão e com isso obter estimativas mais robustas.

Outro fator relevante ao modelo diz respeito à estimação da razão inversa de Mills. De acordo com Bushway, Johnson e Slocum (2007), pelo fato do termo de correção do viés de seletividade amostral<sup>9</sup>  $\lambda$  ser estimado por um modelo *probit* não linear, ele não terá correlação perfeita com  $X$ , mesmo na ausência da restrição de exclusão. Esse atributo do estimador de dois estágios de Heckman permite identificar o modelo sem restrições de exclusão, porque caso a equação de seleção fosse estimada por um método linear, o modelo não seria identificado e, por consequência, não poderia ser estimado. Portanto, a identificação de um modelo bivariado de seletividade amostral com erros nor-

---

<sup>9</sup> Os elementos selecionados da população para compor a amostra não são selecionados de maneira aleatória.

mais é alcançada a partir de pressupostos de formas funcionais (CAMERON; TRIVEDI, 2010).

Conclui-se que Heckman (1977) caracterizou a questão da seleção amostral como um caso especial do problema de variável omitida, em que  $\lambda$  representaria esse elemento, caso a técnica de MCO fosse utilizada na subamostra para a qual  $y_i^* > 0$ . Logo, enquanto  $\nu_2$  tiver uma distribuição normal e  $\varepsilon_1$  for independente de  $\lambda$ , o estimador *heckit* será consistente, mas não eficiente, pois  $\varepsilon_1$  será heterocedástico. Apesar de tal problema, o estimador *heckit* é recomendável para trabalhos empíricos, além de fornecer bons valores iniciais para a estimação de máxima verossimilhança (PUHANI, 2000).

### 3.4.2 O método de máxima verossimilhança

A estimação de um modelo para a situação em que a variável dependente é quantitativa ou dicotômica, pode ser realizada de duas maneiras, por dois estágios, mostrado acima, ou por máxima verossimilhança (MLE).

Assim, cada grupo de observações terá uma função de log-verossimilhança média diferente. Por exemplo, para o modelo de seleção amostral, existem dois tipos de observações (VELLA, 1998; HALL, 2002):

1. Cenário na qual  $y_1$  é observada e se sabe que  $y_2 > 0$ , em que a função de ML será a probabilidade conjunta destes eventos. Com o auxílio da Regra de Bayes, a probabilidade para a  $i$ -ésima observação pode ser assim escrita:

$$\begin{aligned}
 \Pr(y_{1i}, y_{2i} > 0 | X, Z) &= f(y_{1i}) \Pr(y_{2i} > 0 | y_{1i}, X, Z) \\
 &= f(\varepsilon_{1i}) \Pr(\nu_{2i} > -Z_i \delta | \varepsilon_{1i}, X, Z) \\
 &= f(\varepsilon_{1i}) \Pr(\nu_{2i} > -Z_i \delta | \varepsilon_{1i}, X, Z) \\
 &= \frac{1}{\sigma_1} \phi\left(\frac{y_{1i} - X_i \beta}{\sigma_1}\right) \int_{-Z_i \delta}^{\infty} f(\nu_{2i} | \varepsilon_{1i}) d\nu_{2i} \\
 &= \frac{1}{\sigma_1} \phi\left(\frac{y_{1i} - X_i \beta}{\sigma_1}\right) \int_{-Z_i \delta}^{\infty} \phi\left(\frac{\nu_{2i} - \frac{\rho}{\sigma_1}(y_{1i} - X_i \beta)}{\sqrt{1 - \rho^2}}\right) d\nu_{2i} \\
 &= \frac{1}{\sigma_1} \phi\left(\frac{y_{1i} - X_i \beta}{\sigma_1}\right) \left[1 - \Phi\left(\frac{-Z_i \delta - \frac{\rho}{\sigma_1}(y_{1i} - X_i \beta)}{\sqrt{1 - \rho^2}}\right)\right]
 \end{aligned} \tag{3.19}$$

$$\begin{aligned}
 &\Pr(y_{1i}, y_{2i} > 0 | X, Z) \\
 &= \frac{1}{\sigma_1} \phi\left(\frac{y_{1i} - X_i \beta}{\sigma_1}\right) \Phi\left(\frac{Z_i \delta + \frac{\rho}{\sigma_1}(y_{1i} - X_i \beta)}{\sqrt{1 - \rho^2}}\right).
 \end{aligned} \tag{3.20}$$

Intuitivamente, a probabilidade de um evento para o qual se tem a informação é a função de densidade no ponto  $y_1$  multiplicado pela probabilidade de distribuição condicional de  $y_2$  dado que o valor de  $y_1$  tenha sido observado.

2. Situação na qual  $y_{1i}$  é não observado e se sabe que  $y_{2i} \leq 0$  e nesse caso, a função ML é simplesmente a probabilidade marginal de  $y_2 \leq 0$ . Como não há informação sobre  $y_1$ , a probabilidade é escrita da seguinte forma:

$$\begin{aligned}
 \Pr(y_{2i} \leq 0) &= \Pr(\nu_{2i} \leq -Z_i \delta) \\
 &= \Phi(-Z_i \delta) = 1 - \Phi(Z_i \delta).
 \end{aligned} \tag{3.21}$$

Então, a função log-máxima verossimilhança média para todo o conjunto de observações e que se aplica a restrição ( $-1 < \rho < 1$ ), será:

$$\log L(\beta, \delta, \rho, \sigma) = \sum_{i=1}^{N_0} \log(1 - \Phi(Z_i \delta)) + \sum_{i=N_0+1}^N \left[ -\log \sigma_1 + \log \phi\left(\frac{y_{1i} - X_i \beta}{\sigma_1}\right) + \log \Phi\left(\frac{Z_i \delta + \frac{\rho}{\sigma_1}(y_{1i} - X_i \beta)}{\sqrt{1 - \rho^2}}\right) \right], \quad (3.22)$$

em que  $N_0$  é o subconjunto de observações em que não se observa  $y_1$  e  $N_1$  o que se observa ( $N_0 + N_1 = N$ ). As estimativas dos parâmetros para o modelo de seleção amostral podem ser obtidas pela maximização dessa função em relação aos seus argumentos, os quais serão consistentes e assintoticamente eficientes, sob o pressuposto de normalidade e homocedasticidade dos distúrbios não censurados (GRILICHES; HALL; HAUSMAN, 1978).

Mas, caso  $y_{2i}$  seja não observado, o valor da função (3.22) será (VELLA, 1998):

$$\log L = \log(1 - \Phi(Z_i \delta)) = \log \Phi(-Z_i \delta). \quad (3.23)$$

Nessa técnica, a  $\sigma$  e  $\rho$  não são estimados diretamente, e sim  $\ln \sigma$  e a tangente hiperbólica inversa (*atanh*) de  $\rho$ :

$$\operatorname{atanh} \rho = \frac{1}{2} \ln \left( \frac{1 + \rho}{1 - \rho} \right), \quad (3.24)$$

o que faz com que essa verossimilhança não seja necessariamente côncava global em  $\rho$ , e por consequência, o método do gradiente de maximização pode não encontrar um máximo global para uma amostra finita (CHIBURIS; LOKSHIN et al., 2007).

Conclui-se que ambas as metodologias fornecem boas aproximações acerca dos valores dos parâmetros, apesar de que a estimação por ML obtém os resultados mais robustos por produzir estimadores eficientes. Entretanto, o processo por dois estágios é uma boa alternativa por gerar estimadores consistentes, além de ser menos computador intensiva que a primeira. Este ensaio realizou as duas abordagens.

## 3.5 A base de microdados

Para a realização dos estudos econométricos, propostos neste ensaio, foi utilizada a base de dados fornecida pela SEDSC, a mesma aplicada ao experimento do ensaio anterior, porém submetida a outros tipos de tratamentos que permitissem a aplicação da metodologia de Heckman (1977) para a análise dos efeitos da permanência na rede pública sobre a probabilidade de aprovação acadêmica. Apenas para relembrar, os dados obtidos junto ao órgão retratam características da população de estudantes catarinenses do Ensino Médio da rede pública, assim como das estruturas físicas e pedagógicas das escolas do estado, para os anos de 2008 a 2012, e representam a junção de dois grandes blocos de dados: o SERIE-Escola, descontinuado em 2011; e o Sistema de Gestão Educacional de Santa Catarina (SISGESC), que é o atual sistema de gerenciamento.

Devido à necessidade de superar algumas limitações impostas à técnica econométrica pela base de dados, algumas modificações foram feitas nela. Entre elas, destaca-se a exclusão das observações para a 3ª Série do Ensino Médio, uma vez que o resultado acadêmico obtido ao final da série é dúbios, isto é, nada se pode afirmar a respeito de aprovação, graduação, abandono ou outras possíveis situações. Ao final desse procedimento, a base contemplou um universo de 359029 alunos matriculados no Ensino Médio da rede pública, em 649 escolas, em Santa Catarina, composto majoritariamente por alunos do gênero feminino (53,56%), da raça branca (95,61%), dos quais apenas 6,99% destes sofreram ao menos uma reprovação, em sua maioria meninos. Esta última estatística evidencia que o número de reprovados ainda é elevada e a discrepância de comportamento entre os gêneros, em que o subgrupo dos meninos, é o mais vulnerável ao abandono escolar precoce, motivados pelo maior desinteresse pelos estudos, maior probabilidade de inserção no mercado de trabalho e outros fatores externos à escola.

Quanto às características físicas das escolas, a grande maioria apresenta boas condições estruturais, apresentando biblioteca, laboratório de informática e quadra coberta. A Tabela (12), a seguir, traz as estatísticas descritivas para várias características desses estudantes e do ambiente escolar.

Tabela 12 – Estatísticas descritivas: Base de dados SEDSC

<b>Características Individuais/Familiares</b>			
Variável	Classificação	Total	%
Gênero	Feminino	192286	53,56
	Masculino	166743	46,44
Cor/Raça	Branca	343273	95,61
	Outra	15756	4,39
Zona Residencial	Rural	69192	19,27
	Urbana	289837	80,73
Bolsa Família	Sim	4748	1,32
	Não	354281	98,68
Alunos com pelo menos 1 reprovação e que permaneceram na rede pública	Total	25114	6,99
	Feminino	9723	38,72
	Masculino	15391	61,28
Total Matrículas		359029	
<b>Características Escolares</b>			
Biblioteca	Sim	521	80,28
	Não	128	19,72
Laboratório de Informática	Sim	578	89,06
	Não	71	10,94
Quadra Coberta	Sim	556	85,67
	Não	93	14,33
Ensino Fundamental	Sim	594	91,53
	Não	55	8,47
Ensino Fundamental e Pré-Escolar	Sim	513	80,04
	Não	136	20,96
Quantidade Média Funcionários		45,87 (23,95)	
Idade Média Professor		40,32 (2,85)	
Total de Escolas		649	

Fonte: SEDSC. Elaboração própria. Nota: Valores entre parênteses representam o desvio-padrão. 1: Porcentagem calculada em relação ao valor total da variável.

Com o objetivo de se realizar um primeiro diagnóstico sobre a colinearidade entre as variáveis explicativas a serem utilizadas na função de interesse, foi construída uma matriz de correlação de Pearson<sup>10</sup>. A Tabela (3.5) indica a presença de baixa correlação linear entre os pares de variáveis e, assim sendo, provavelmente, as inferências baseadas nos modelos de regressão e suas aplicabilidades são críveis.

<sup>10</sup> A correlação expressa o grau de dependência linear entre duas variáveis e tem valores entre  $-1$  e  $1$ .

Tabela 13 – Matriz de correlação: Equação de interesse

	menino	branco	rural	bolsa	lnpib_pc	e_biblio	i_funcionarios_razao	prof_idade	alunos_turma
menino	1,0000	-0,0001	-0,0084	-0,0129	0,0024	0,0014	-0,0066	0,0044	-0,0152
branco		1,0000	0,0015	-0,0783	0,0039	0,0015	0,0170	-0,0120	0,0168
rural			1,0000	0,0481	-0,1747	0,0492	-0,1209	-0,2116	-0,0694
bolsa				1,0000	-0,0491	0,0060	-0,0405	-0,0280	-0,0144
lnpib_pc					1,0000	-0,0513	0,1870	0,1986	0,1147
e_biblio						1,0000	-0,0350	-0,0137	-0,0166
i_funcionarios_razao							1,0000	0,1208	0,2537
prof_idade								1,0000	0,0571
alunos_turma									1,0000

Fonte: Formatação própria.

### 3.6 O modelo econométrico

A educação é um componente central no desenvolvimento humano sustentável, influenciando diretamente na integração social, na construção da cidadania e na habilidade dos países em se posicionar perante o contexto global de produtividade. Desse modo, um dos requerimentos principais para que uma sociedade se adapte às mudanças de paradigmas tecnológicos é que o maior número possível de crianças e adolescentes tenham uma vivência educacional adequada, de modo que a simples manutenção destes na escola reduz a probabilidade de incorrerem em comportamentos problemáticos.

Ao longo desta tese, foi apresentado que as implicações do abandono escolar não se resumem apenas ao contexto do jovem, mas abrangem também toda uma conjuntura familiar, cultural e econômica, e que esta situação raramente ocorre como uma situação inesperada. A decisão de abandono escolar é tido como um processo dinâmico, em que o risco de tal evento ocorrer é motivada por dificuldades crescentes em termos de desempenho acadêmico e adaptação, formando uma combinação complexa

de fatores individuais, familiares, pedagógicos e institucionais (TINTO, 1975; RUMBERGER; LIM, 2008; NERI et al., 2015).

Então, com base nas evidências apresentadas e na base de dados disponível, somos levados a procurar os motivos que levam à aprovação acadêmica e à permanência na rede pública de ensino conjuntamente. A seguir, serão apresentadas estas variáveis com seus respectivos códigos de entrada no *STATA 13*, as quais são:

Tabela 14 – Descrição das variáveis extraídas da SEDSC

Código STATA	Descrição
<b>aprov</b>	Variável binária, na qual possui valor 1 se o estudante é aprovado e 0 caso contrário;
<b>d</b>	Variável binária, na qual possui valor 1 se estudante permanece na rede pública de ensino e 0 caso contrário;
menino	Variável binária, na qual possui valor 1 se o estudante é menino e 0 caso contrário;
branco	Variável binária, na qual possui valor 1 se o estudante se autodeclara de cor branca e 0 caso contrário;
rural	Variável binária, na qual possui valor 1 se o estudante reside na zona rural e 0 caso contrário;
bolsa	Variável binária, na qual possui valor 1 se a família do estudante recebe Bolsa Família e 0 caso contrário;
lnpib_pc	Logaritmo natural do PIB per capita por município;
e_preescolar	Variável binária, na qual possui valor 1 se a escola oferta Pré-escolar e 0 caso contrário;
e_fundamental	Variável binária, na qual possui valor 1 se a escola oferta Ensino Fundamental e 0 caso contrário;
e_biblio	Variável binária, na qual possui valor 1 se a escola tem biblioteca e 0 caso contrário;
e_lab_info	Variável binária, na qual possui valor 1 se a escola tem laboratório de informática e 0 caso contrário;
e_quadra	Variável binária, na qual possui valor 1 se a escola tem quadra coberta e 0 caso contrário;
i_funcionarios_razao	Razão entre número de estudantes e funcionários por escola;
prof_idade	Idade média dos professores por escola;
prof_idade_dp	Desvio-padrão da idade média dos professores por escola;
alunos_turma	Quantidade de estudantes por sala de aula;
faltas_completas	Número de faltas por ano de cada estudante.

Fonte: Elaboração própria.

Portanto, para que o nosso modelo expresse as características individuais do aluno, foram utilizados indicadores binários para menino, branco e rural, os quais indicam, respectivamente, o gênero, a cor e o local onde vive. Por sua vez, o ambiente sócioeconômico no qual ele está inserido foi representado pelas variáveis bolsa e *lnpib\_pc*. Em razão das características econômicas do Brasil, estudos mostram que quanto maior a renda da região em que vive o indivíduo, maior a probabilidade da família dar suporte aos estudos, todavia, maior a probabilidade do jovem não cursar a rede pública, além de crescerem as chances de abandono escolar devido à oferta de empregos por parte daqueles alunos em risco sócioeconômico (NERI et al., 2009; UNIBANCO, 2016).

Quanto aos fatores relacionados ao interesse do aluno e ao ambiente de sala de aula, foi proposto o uso das variáveis *faltas\_completas* e *alunos\_turma*, as quais podem mudar de ano para ano. Em relação à primeira variável, a literatura indica que esta é um dos primeiros sinais de evasão e desinteresse escolar (FERNANDES, 2013), enquanto que para a segunda, a análise empírica promovida mostrou que o tamanho médio da sala não é estatisticamente significativo para a melhora do desempenho acadêmico do estudante (COATES, 2003). Entretanto salas e escolas menores mostraram ter um impacto importante na redução da evasão escolar (LEE; BURKAM, 2003).

Afora as características relacionadas ao indivíduo e a família, a interação dinâmica aluno-escola também é destaque em nossa análise. Essa relação é medida pelas variáveis: (i) *i\_funcionarios\_razao*, que retrata o relacionamento entre os funcionários de uma escola e seus alunos, além de permitir uma comparação relativa entre os diferentes tamanhos de instituição de ensino, em que quanto menor for o valor desse indicador melhores serão as condições pedagógicas ofertadas pela escola; (ii) *prof\_idade* e *prof\_idade\_dp*, onde ambas retratam o perfil etário do professor, nesse sentido a literatura sugere que professores mais novos estão associados com taxas menores de abandono (SUMMERS; WOLFE, 1977; RIVKIN; HANUSHEK; KAIN, 2005); (iii) *e\_biblioteca*, *e\_lab\_info* e *e\_quadra* apresentam a estrutura física das escolas, nessa perspectiva, a literatura diz que impacto marginal de uma mudança de investimento, independente do contexto sócio-econômico, qualidades acadêmicas e até mesmo gasto escolar, não garantem a melhora do desempenho acadêmico do estudante (HANUSHEK, 1996); e, (iv) *e\_preescolar* e *e\_fundamental*, identificam se a Escola de Ensino Médio oferta também as etapas iniciais do ciclo básico de educação, fato comum em pequenas cidades do Estado.

Nesse contexto, a equação de regressão linear (Equação 3.1) para o Modelo de Heckman de MLE em que a variável binária *aprov* é a dependente, foi assim proposta:

$$\begin{aligned} aprov_i = & \alpha + \beta_1menino + \beta_2branco + \beta_3rural + \beta_4bolsa \\ & + \beta_5 \ln pibpc + \beta_6e\_biblio\beta_7i\_funcionarios\_razao \\ & + \beta_8prof\_idade + \beta_9alunos\_turma + \varepsilon_i \end{aligned} \quad (3.25)$$

em que o subscrito *i* refere-se à estudante e o indicador de resultado<sup>11</sup>, *aprov*, é uma medida de progressão acadêmica do aluno dado pela sua manutenção na base de dados ao longo do Ensino Médio. Convém lembrar que o indicador de permanência na rede pública, *d*, também foi formulada a partir do acompanhamento *bakward-looking* do aluno na mesma base de dados.

### 3.6.1 Especificando a função de seleção

Para se estimar a probabilidade de aprovação utilizando a metodologia de máxima verossimilhança de Heckman, primeiro deve-se especificar uma função de seleção, a partir de um conjunto de covariadas  $Z_i$ , para predizer se um estudante *i* está matriculado ou não na rede pública de ensino.

Com o objetivo de identificar qual seria a melhor especificação, foram consideradas quatro funções diferentes, FS1 (semelhante ao sugerido por Heckman (1977)), FS2, FS3 e FS4, as quais foram submetidas à metodologia *probit*<sup>12</sup> para avalia-

<sup>11</sup> Processo de construção análogo ao processo descrito no segunda ensaio desta tese.

<sup>12</sup> Para maiores detalhes, ver Wooldridge (2015).

ção (BRIGGS, 2004). Os preditores de cada especificação são apresentados na Figura (15) a seguir:

Tabela 15 – Especificações para a função de seleção

Modelo	Especificação
FS1	$Z_i = X_i$
FS2	$Z_i = X_i, e\_preescolar, e\_fundamental$
FS3	$Z_i = e\_preescolar, e\_fundamental, prof\_idade\_dp, e\_lab\_info, e\_quadra$
FS4	$Z_i = e\_preescolar, e\_fundamental, prof\_idade\_dp, e\_lab\_info, e\_quadra, faltas\_completas$

Qual dessas especificações é a correta para a função de seleção? Na primeira sugestão, FS1, todas as variáveis explicativas especificadas como possíveis determinantes da equação de seleção estão incluídas como preditores da equação de interesse. Este caso retrata o uso mecânico da metodologia de Heckman quando a análise dos dados não nos permite inferir nenhuma regra sobre os motivos que levaram os estudantes a decidir por permanecer ou não na rede de ensino (VELLA, 1998).

Por sua vez a FS2 adiciona duas novas covariadas binárias à função de seleção proposta anteriormente, *e\_preescolar* e *e\_fundamental*, as quais retratam a oferta de todo o ciclo básico da educação na mesma escola. Aqui, se supõe que se for o caso, o aluno mais predisposto a sair da rede pública terá maior probabilidade de concluir o Ensino Médio, uma vez que serão reduzidos os efeitos de uma eventual adaptação e logística à nova escola. Essa função não representa nenhuma teoria *per se*, apenas uma simples melhoria em comparação à primeira sugestão.

A especificação FS3 é um caso extremo a FS1, pois nenhuma covariada definida em  $\mathbf{X}_i$  está em  $\mathbf{Z}_i$ , isto é, sob a terceira sugestão de função de seleção, existem agora cinco variáveis relacionadas às características acadêmicas e estruturais (e\_preescolar, e\_fundamental, prof\_idade\_dp, e\_lab\_info, e\_quadra) que se pensa preverem a permanência na rede pública, mas não o resultado acadêmico. Por último, a FS4 é mais um exemplo mecânico para a metodologia de Heckman, em que a única diferença para a FS3 é a inclusão da variável faltas\_completas. A adição dessa covariada foi motivada pela relação teórica, empírica e antagônica existente entre ela e a permanência na escola.

A Tabela (16) apresenta os parâmetros estimados, além de algumas estatísticas descritivas a partir de uma modelagem *probit* para cada uma das quatro especificações sugeridas:

Tabela 16 – Funções de Seleção - Coeficientes estimados usando Modelo *Probit*

	FS1		FS2		FS3		FS4	
log-verossimilhança	-360073,37		-360024,93		-365733,27		-365454,15	
Pseudo R2 (McFadden's)	0,0166		0,0167		0,00113		0,00189	
% Covariadas significantes	89% (8/9)		82%(9/11)		100% (5/5)		100% (6/6)	
Variáveis	coef	se	coef	se	coef	se	coef	se
menino	-0,1093***	0,0035	-0,1096***	0,0035				
branco	0,1951***	0,0084	0,1964***	0,0084				
rural	0,0848***	0,0045	0,0855***	0,0045				
bolsa	0,0133	0,0152	0,0141	0,0152				
lnpib_pc	-0,3278***	0,0045	-0,3296***	0,0045				
e_biblio	0,0601***	0,0043	0,0604***	0,0043				
i_funcionarios_razao	-0,0104***	0,0004	-0,0118***	0,0004				
prof_idade	-0,0161***	0,0007	-0,0165***	0,0007				
alunos_turma	0,0108***	0,0003	0,0110***	0,0003				
e_preescolar			-0,0057	0,0048	-0,0494***	0,0048	-0,0491***	0,0048
e_fundamental			-0,0509***	0,0062	0,0839***	0,0057	0,0837***	0,0057
prof_idade_dp					-0,0202***	0,0013	-0,0202***	0,0013
e_lab_info					0,0770***	0,0052	0,0772***	0,0052
e_quadra					0,0222***	0,0047	0,0224***	0,0047
faltas_completas							-0,0217***	0,0009
constante	3,8482***	0,0501	3,9383***	0,0512	0,3626***	0,0150	0,3699***	0,0151

Fonte: Elaboração própria.

Coef → Valor dos coeficientes estimados; se → desvio-padrão.

\*\*\*  $p < 0, 01$ ; \*\*  $p < 0, 05$ ; \*  $p < 0, 1$ .

Nota-se que a definição da melhor especificação para ser utilizada na metodologia de Heckman por critérios estatísticos não é tão trivial quanto em um modelo de regressão linear, onde o ajuste do modelo é frequentemente avaliado com base em  $R^2$ , pois não há tal medida para o modelo *probit*.

Uma comparação pode ser feita a partir da análise dos valores do pseudo  $R^2$ , em que seu valor é calculado a partir da fórmula de McFadden's ( $pseudo - R^2 = 1 - \frac{L}{L_0}$ ), em que  $L$  é o valor do log-verossimilhança estimado para uma dada especificação da função de seleção, enquanto que  $L_0$  é o valor log-verossimilhança para um modelo base, composto apenas da constante.<sup>13</sup>, os quais mostram que as especificações FS1 e FS2 são as que mais melhoram o grau de ajustamento relativo ao modelo,  $L_0$ , apesar do grau de melhoria ser pequeno.

Briggs (2004) sugere um outro critério para a escolha da melhor especificação, preferir aquela com maior proporção de covariadas estatisticamente significativas. Segundo o autor, isto é relevante, porque o próximo passo do modelo Heckman é calcular a relação inversa de *Mills* como uma função dos coeficientes estimados, sejam eles significativos ou não.

A Tabela (16) mostra que as especificações FS3 e FS4 são as mais indicadas, dado que todos os coeficientes são estatisticamente significativos, o que prova que o critério de seleção adotado nas escolhas das variáveis explicativas está correto.

---

<sup>13</sup> Metodologia de cálculo utilizada no *software* Stata 13. Para maiores informações ver: <<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/>>.

Quanto às alternativas FS1 e FS2, ambas apresentaram, respectivamente 89% e 81% dos coeficientes estimados estatisticamente significativos porém em ambas, a variável bolsa apresentou-se estatisticamente não significativa na decisão do aluno em permanecer na rede pública de ensino.

Também pode-se examinar as probabilidades previstas em permanecer na rede pública de ensino para cada uma das funções de seleção. Para FS1 e FS2, a probabilidade estimada mais elevada é cerca de 18% maior que as alcançadas pelas funções FS3 e FS4, atingindo ambas o valor de 0,83. Abaixo, a Tabela (17) apresenta os resultados estimados para as probabilidades das quatro funções de seleção:

Tabela 17 – Funções de Seleção - Probabilidades previstas

Função de Seleção	Probabilidades Previstas	
	Mínimo	Máximo
FS1	0,237	0,834
FS2	0,238	0,836
FS3	0,517	0,666
FS4	0,104	0,670

Fonte: Elaboração própria.

A intenção em se realizar tais comparações é que, na maioria das aplicações do modelo Heckman, pouco se é debatido quanto às especificações da função de seleção, visto que raramente são propostas alternativas a ela. Conclui-se que a decisão de quais preditores incluir ou excluir da função de seleção é não trivial e pode ter ramificações substanciais nos parâmetros estimados gerados pelo Modelo Heckman.

## 3.7 Resultados empíricos

As Tabelas (18) a (21), abaixo, apresentam os valores dos parâmetros estimados, bem como seus respectivos desvios-padrão e estatísticas *t*, para as técnicas *MQO* e *Heckman MLE*<sup>14</sup>, dados os quatro modelos sugeridos, especificamente, para a 1ª Série, período matutino (com exceção dos resultados estimados por *MLE* nesta seção, os demais estão expostos no Anexo I).

Tabela 18 – Modelo 1: Estimações *MQO* e *Heckman MLE* para a 1ª série, turno da manhã

Variáveis	2008.1.1			2009.1.1			2010.1.1			2011.1.1		
	OLS aprov	Heckman aprov	MLE d									
menino	-0,0727*** (0,0037)	-0,0440*** (0,0036)	0,1543*** (0,0125)	-0,0786*** (0,0040)	-0,0481*** (0,0039)	0,1502*** (0,0121)	-0,0694*** (0,0042)	-0,0336*** (0,0039)	0,0994*** (0,0116)	-0,0375*** (0,0034)	-0,0235*** (0,0033)	0,0941*** (0,0133)
branco	0,0610*** (0,0097)	0,0393*** (0,0091)	-0,1377*** (0,0319)	0,0658*** (0,0101)	0,0477*** (0,0096)	-0,1490*** (0,0301)	0,0499*** (0,0099)	0,0248*** (0,0092)	-0,0734*** (0,0270)	0,0543*** (0,0085)	0,0261*** (0,0081)	-0,1047*** (0,0325)
rural	0,0283*** (0,0051)	0,0074 (0,0050)	-0,0260 (0,0174)	0,0343*** (0,0056)	0,0236*** (0,0055)	-0,0736*** (0,0171)	0,0251*** (0,0058)	-0,0009 (0,0056)	0,0026 (0,0167)	0,0085 (0,0041)	-0,0203*** (0,0040)	0,0816*** (0,0161)
bolsa	0,0144 (0,0132)	0,0101 (0,0128)	-0,0356 (0,0448)	0,0098 (0,0145)	0,0245* (0,0139)	-0,0765* (0,0435)	-0,1643*** (0,0251)	0,0350* (0,0202)	-0,1034* (0,0598)	0,0145 (0,0225)	0,1140*** (0,0188)	-0,4573*** (0,0755)
lnpib_pc	0,0145*** (0,0045)	0,0174*** (0,0043)	-0,0610*** (0,0150)	-0,0187*** (0,0051)	-0,0159*** (0,0050)	0,0497*** (0,0155)	-0,0146** (0,0060)	0,0125** (0,0055)	-0,0370** (0,0161)	-0,0312*** (0,0049)	-0,0244*** (0,0047)	0,0979*** (0,0190)
e_biblio	0,0018 (0,0045)	-0,0049 (0,0043)	0,0174 (0,0151)	0,0019 (0,0049)	-0,0019 (0,0047)	0,0061 (0,0148)	0,0269*** (0,0052)	-0,0097** (0,0048)	0,0287** (0,0143)	0,0017 (0,0043)	-0,0009 (0,0041)	0,0034 (0,0165)
i_funcionarios_razao	-0,0007 (0,0005)	0,0003 (0,0004)	-0,0011 (0,0015)	-0,0015*** (0,0005)	-0,0011** (0,0005)	0,0035** (0,0016)	-0,0023*** (0,0005)	-0,0023*** (0,0005)	0,0068*** (0,0014)	-0,0028*** (0,0004)	-0,0014*** (0,0004)	0,0057*** (0,0016)
prof_fidade	-0,0023*** (0,0008)	0,0012* (0,0007)	-0,0042* (0,0025)	-0,0010 (0,0008)	0,0020*** (0,0008)	-0,0061*** (0,0024)	-0,0015* (0,0008)	0,0043*** (0,0008)	-0,0128*** (0,0023)	-0,0005 (0,0007)	0,0001 (0,0006)	0,0003 (0,0025)
alunos_turma	0,0039*** (0,0003)	0,0016*** (0,0003)	-0,0056*** (0,0011)	0,0025*** (0,0004)	0,0009** (0,0004)	-0,0029** (0,0012)	0,0005 (0,0004)	0,0022*** (0,0004)	-0,0065*** (0,0011)	0,0008** (0,0003)	0,0004 (0,0003)	0,0015 (0,0012)
Constant	0,7234*** (0,0543)	0,6947*** (0,0520)	1,0711*** (0,1823)	1,0214*** (0,0602)	0,9847*** (0,0584)	0,0480 (0,1826)	1,0533*** (0,0664)	0,6522*** (0,0613)	1,0282*** (0,1811)	1,2553*** (0,0533)	1,2176*** (0,0513)	-0,8725*** (0,2059)
athrho		-10,6858*** (0,3440)			-10,4663*** (0,7947)			-10,4598*** (0,3430)			-10,6046*** (0,5033)	
insigma		-1,2550 (0,0000)			-1,1398*** (0,0045)			-1,0837*** (0,0039)			-1,3888*** (0,0054)	
Observações <sup>1</sup>	21862	27957	27957	24588	29521	29521	23033	33804	33804	20290	24482	24482

1: Número de observações utilizadas para a estimação do modelo. Quantidade se mantém para os 4 modelos.  
\*\*\* *p* < 0,01; \*\* *p* < 0,05; \* *p* < 0,1. Fonte: Elaboração própria.

Os números para o Modelo 1 indicam que as duas técnicas produzem resultados semelhantes em termos de significância de covariadas. em que as variáveis *menino*, *branco* e *lnpib\_pc* foram estatisticamente significativas em ambos os casos, para todo o período estudado. Note que essa última variável possui

<sup>14</sup> A técnica de dois estágios também foi aplicada, porém as características do estimador de MLE são superiores, por isso optou-se por não apresentar apenas os resultados referentes à esta técnica.

um comportamento complexo, pois ao mesmo tempo que indica que a região onde o indivíduo mora possui melhor infraestrutura e economia, o que permitiria as famílias manter os jovens na escola pública ou migrá-los para a escola privada, cria também incentivos financeiros para os jovens em condição de risco sócio-econômico buscarem o mercado de trabalho, enquanto que ser *menino* afeta negativamente tanto o resultado acadêmico quanto a permanência na escola (resultado mais coerente com a realidade do que o estimado no Ensaio 2).

Em relação às variáveis que expõem a estrutura da escola, ter biblioteca foi significativo apenas em um ano, enquanto que o tamanho relativo da escola (*i\_funcionarios\_razao*) indica que escolas com uma quantidade maior de alunos são prejudiciais ao resultado acadêmico, entretanto, têm efeitos positivos na manutenção do mesmo na rede de ensino. Por sua vez, o tamanho da sala mostrou-se prejudicial para o desenvolvimento educacional do jovem.

Além disso, também observe que os valores dos coeficientes e desvios-padrão estimados por *MQO* são, na maioria dos casos, superiores aos obtidos por *Heckman MLE*, o que indica que os resultados atingidos na primeira técnica podem levar a erros na formulação de políticas públicas de educação, ao aumentar o efeito do impacto das covariadas sob a variável dependente.

Tabela 19 – Modelo 2: Estimções MQO e Heckman MLE para a 1ª série, turno da manhã

Variáveis	2008.1.1		2009.1.1		2010.1.1		2011.1.1	
	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d
menino	-0,0726*** (0,0037)	-0,0439*** (0,0125)	-0,0786*** (0,0040)	-0,0481*** (0,0039)	-0,0694*** (0,0042)	-0,0335*** (0,0116)	-0,0376*** (0,0034)	-0,0235*** (0,0033)
branco	0,0617*** (0,0097)	0,0392*** (0,0319)	0,0658*** (0,0101)	0,0477*** (0,0096)	0,0499*** (0,0099)	0,0247*** (0,0271)	0,0347*** (0,0085)	0,0261*** (0,0081)
rural	0,0280*** (0,0051)	0,0074 (0,0050)	0,0339*** (0,0056)	0,0236*** (0,0055)	0,0252*** (0,0058)	-0,0010 (0,0056)	0,0058 (0,0041)	-0,0204*** (0,0040)
bolsa	0,1140 (0,0132)	0,0101 (0,0128)	0,0095 (0,0145)	0,0245* (0,0139)	-0,1643*** (0,0251)	-0,1044* (0,0202)	0,0149 (0,0225)	0,1141*** (0,0188)
lnpib_pc	0,0119*** (0,0045)	0,0174*** (0,0043)	-0,0192*** (0,0052)	-0,0159*** (0,0050)	-0,0147*** (0,0060)	0,0126*** (0,0054)	-0,0312*** (0,0049)	-0,0244*** (0,0047)
e_biblio	0,0026 (0,0045)	-0,0050 (0,0043)	0,0174 (0,0049)	-0,0019 (0,0047)	0,0268*** (0,0052)	-0,0098*** (0,0048)	0,0013 (0,0043)	-0,0009 (0,0041)
i_funcionarios_razao	-0,0008* (0,0005)	0,0003 (0,0004)	-0,0013** (0,0005)	-0,0011** (0,0005)	-0,0024*** (0,0006)	-0,0023*** (0,0005)	-0,0030*** (0,0005)	-0,0014*** (0,0004)
prof_idade	-0,0025*** (0,0008)	0,0012* (0,0007)	-0,0042* (0,0025)	-0,0061*** (0,0008)	-0,0014** (0,0008)	-0,0129*** (0,0023)	-0,0005 (0,0007)	0,0001 (0,0006)
alunos_turma	0,0040*** (0,0003)	0,0016*** (0,0003)	0,0025*** (0,0004)	0,0009** (0,0004)	0,0005 (0,0004)	0,0022*** (0,0004)	0,0008*** (0,0003)	0,0004 (0,0003)
e_preescolar	0,0232*** (0,0000)	0,0000 (0,0000)	0,0067 (0,0055)	0,0000 (0,0000)	-0,0030 (0,0050)	-0,0000 (0,0001)	0,0005 (0,0050)	0,0000 (0,0000)
e_fundamental	-0,0209*** (0,0067)	0,0000 (0,0000)	0,0019 (0,0070)	-0,0000 (0,0001)	0,0002 (0,0079)	0,0000 (0,0001)	-0,0072 (0,0067)	-0,0000 (0,0001)
Constant	0,7542*** (0,0553)	0,6947*** (0,0520)	1,0207*** (0,0612)	0,9841*** (0,0585)	1,0568*** (0,0670)	1,0284*** (0,0612)	1,2626*** (0,0537)	-0,8732*** (0,2059)
athrho		-10,8604*** (0,2656)		-9,9662*** (0,3825)		-10,3644*** (0,4002)		-11,4779*** (0,6733)
Insigma		-1,2551 (0,0000)		-1,1380*** (0,0046)		-1,0857*** (0,0048)		-1,3892*** (0,0062)

Fonte: Elaboração própria.

\*\*\* $p < 0,01$ , \*\* $p < 0,05$ , \* $p < 0,1$ .

O Modelo 2, acima, difere-se do anterior em relação as características da escola, ao incluir as variáveis *e\_preescolar* e *e\_fundamental* nas estimações. Apesar de serem estatisticamente positivas para a regressão *OLS* no ano de 2009, a oferta de todo o ciclo básico mostrou-se irrelevante na metodologia de *Heckman MLE*, ou seja, o aluno ter tido a possibilidade de frequentar todos os níveis de ensino não impacta na continuidade dos estudos auferidos ao longo da 1ª Série. Essas evidências corroboram com a literatura (HECKMAN, 1977; CAMERON; TRIVEDI, 2010) de que a presença de viés compromete a validade interna<sup>15</sup> e externa<sup>16</sup> dos estimadores de *MQO*, inviabilizando a utilidade prática do modelo.

A Tabela (20) apresenta os resultados para o modelo que utiliza a função de seleção FS3. Note que assim como ocorrera nos dois primeiros modelos, os parâmetros estimados são quase sempre muito próximos e apresentam o mesmo sinal, exceto para o ano de 2010. Além disso, tal como aparece na literatura, os coeficientes das variáveis menino (NERI et al., 2015; UNIBANCO, 2016) e *i\_funcionarios\_razao* são sempre negativos (HANUSHEK, 1996), enquanto o coeficiente estimado para bolsa expressou-se, via de regra, estatisticamente não significativo.

---

<sup>15</sup> Define, até que ponto, os resultados de um estudo são corretos para a amostra de indivíduos estudados e é aplicada a condições daquele grupo em particular, e não, necessariamente, a outros grupos.

<sup>16</sup> Referente a aplicabilidade, ou de generalização, dos resultados de um estudo em particular, para outros contextos.

Tabela 20 – Modelo 3: Estimações MQO e Heckman MLE para a 1ª série, turno da manhã

Variáveis	2008.1.1		2009.1.1		2010.1.1		2011.1.1	
	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d	OLS aprov	Heckman MLE d
menino	-0,0728*** (0,0037)	-0,0726*** (0,0037)	-0,0788*** (0,0040)	-0,0786*** (0,0040)	-0,0694*** (0,0042)	-0,0000 (0,0000)	-0,0375*** (0,0034)	-0,0375*** (0,0034)
branco	0,0632*** (0,0097)	0,0610*** (0,0097)	0,0676*** (0,0101)	0,0658*** (0,0101)	0,0508*** (0,0099)	0,0000 (0,0001)	0,0352*** (0,0085)	0,0343*** (0,0085)
rural	0,0278*** (0,0051)	0,0283*** (0,0051)	0,0336*** (0,0056)	0,0343*** (0,0056)	0,0245*** (0,0058)	0,0000 (0,0000)	0,0057 (0,0041)	0,0055 (0,0041)
bolsa	0,0124 (0,0132)	0,0143 (0,0132)	0,0099 (0,0145)	0,0098 (0,0145)	-0,1653*** (0,0251)	-0,0000 (0,0002)	0,0123 (0,0225)	0,0145 (0,0225)
lnpib_pc	0,0113*** (0,0045)	0,0144*** (0,0045)	-0,0189*** (0,0052)	-0,0187*** (0,0051)	-0,0140*** (0,0060)	-0,0000 (0,0000)	-0,0307*** (0,0049)	-0,0312*** (0,0049)
e_biblio	-0,0009 (0,0046)	0,0018 (0,0045)	0,0010 (0,0050)	0,0019 (0,0049)	0,0244*** (0,0053)	0,0000 (0,0000)	-0,0016 (0,0043)	0,0017 (0,0043)
i_funcionarios_razao	-0,0008* (0,0005)	-0,0007 (0,0005)	-0,0013** (0,0005)	-0,0015*** (0,0005)	-0,0023*** (0,0006)	-0,0000 (0,0000)	-0,0029*** (0,0005)	-0,0028*** (0,0004)
prof_idade	-0,0037*** (0,0009)	-0,0024*** (0,0008)	-0,0021** (0,0009)	-0,0010 (0,0008)	-0,0022** (0,0005)	0,0000 (0,0000)	-0,0013* (0,0007)	-0,0005 (0,0007)
alunos_turma	0,0040*** (0,0003)	0,0039*** (0,0003)	0,0025*** (0,0004)	0,0025*** (0,0004)	0,0005 (0,0004)	0,0000 (0,0000)	0,0006** (0,0003)	0,0008** (0,0003)
e_preescolar	0,0223*** (0,0052)	-0,0121 (0,0235)	0,0067 (0,0055)	0,0584** (0,0238)	-0,0028 (0,0059)	-0,0000 (0,0001)	-0,0000 (0,0050)	0,0991*** (0,0270)
e_fundamental	-0,0218*** (0,0067)	0,1167*** (0,0277)	0,0009 (0,0070)	-0,0530* (0,0289)	-0,0011 (0,0079)	0,0000 (0,0002)	-0,0077 (0,0067)	-0,0221 (0,0342)
prof_idade_dp	0,0045*** (0,0016)	-0,0367*** (0,0063)	0,0046*** (0,0017)	-0,0463*** (0,0065)	0,0041** (0,0018)	-0,0000 (0,0000)	0,0046*** (0,0015)	-0,0334*** (0,0075)
e_lab_info	0,0133** (0,0058)	0,0549** (0,0254)	0,0098 (0,0061)	0,0307 (0,0261)	0,0220*** (0,0067)	0,0000 (0,0001)	0,0323** (0,0055)	0,0222 (0,0303)
e_quadra	0,0135*** (0,0051)	-0,0757*** (0,0225)	-0,0124** (0,0054)	-0,0688*** (0,0233)	-0,0020 (0,0055)	0,0000 (0,0001)	0,0142*** (0,0045)	-0,0406 (0,0247)
Constant	0,7508*** (0,0557)	0,7222*** (0,0543)	1,0091*** (0,0616)	1,0205*** (0,0602)	1,0226*** (0,0683)	1,0102*** (0,0021)	1,2170*** (0,0547)	1,2265*** (0,0533)
athrto		0,0445 (0,0354)		0,0213 (0,0499)		-10,1774*** (0,6596)		0,0313 (0,0395)
Insigma		-1,2974*** (0,0048)		-1,1760*** (0,0045)		-1,0826*** (0,0048)		-1,4137*** (0,0050)

Fonte: Elaboração própria.  
 \*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

A Tabela (21) chama a atenção pelo valor obtido para os coeficientes e desvios-padrão estimados pela metodologia de Heckman para os anos de 2009 a 2011, onde todos eles são iguais a zero. Estes resultados evidenciam problemas de validade externa do modelo, dado que as estimativas da amostra restrita são viesadas para a inferência da população, além de realçar as diferenças entre as técnicas de *MQO* e máxima-verossimilhança de Heckamn.

Tabela 21 – Modelo 4: Estimações MQO e Heckman MLE para a 1ª série, turno da manhã

Variáveis	2008.1.1		2009.1.1		2010.1.1		2011.1.1	
	OLS	Heckman MLE	OLS	Heckman MLE	OLS	Heckman MLE	OLS	Heckman MLE
variáveis	aprov	d	aprov	d	aprov	d	aprov	d
menino	-0,0729*** (0,0037)	-0,0727*** (0,0037)	-0,0787*** (0,0040)	-0,0000 (0,0000)	-0,0694*** (0,0042)	-0,0000 (0,0000)	-0,0375*** (0,0034)	-0,0000 (0,0000)
branco	0,0632*** (0,0097)	0,0610*** (0,0097)	0,0674*** (0,0101)	0,0000 (0,0000)	0,0508*** (0,0099)	0,0000 (0,0000)	0,0350*** (0,0085)	0,0000 (0,0000)
rural	0,0276*** (0,0051)	0,0283*** (0,0051)	0,0333*** (0,0056)	0,0000 (0,0000)	0,0240*** (0,0058)	0,0000 (0,0000)	0,0036 (0,0054)	0,0000 (0,0000)
bolsa	0,0119 (0,0132)	0,0144 (0,0132)	0,0102 (0,0145)	0,0000 (0,0001)	-0,1640*** (0,0251)	-0,0000 (0,0001)	0,0121 (0,0225)	-0,0000 (0,0001)
lnpib_pc	0,0114** (0,0045)	0,0144*** (0,0045)	-0,0186*** (0,0052)	-0,0000 (0,0000)	-0,0140** (0,0060)	-0,0000 (0,0000)	-0,0304*** (0,0049)	-0,0000 (0,0000)
e_biblio	-0,0011 (0,0046)	0,0018 (0,0045)	0,0012 (0,0050)	0,0000 (0,0000)	0,0244*** (0,0053)	0,0000 (0,0000)	-0,0017 (0,0043)	0,0000 (0,0000)
i_funcionarios_razao	-0,0009* (0,0005)	-0,0007 (0,0005)	-0,0013** (0,0005)	-0,0000 (0,0000)	-0,0023*** (0,0006)	-0,0000 (0,0000)	-0,0029*** (0,0005)	-0,0000 (0,0000)
prof_idade	-0,0036*** (0,0009)	-0,0023*** (0,0008)	-0,0021** (0,0009)	-0,0000 (0,0000)	-0,0022** (0,0009)	0,0000 (0,0000)	-0,0013* (0,0007)	-0,0000 (0,0000)
alunos_turma	0,0040*** (0,0003)	0,0039*** (0,0003)	0,0024*** (0,0004)	0,0000 (0,0000)	0,0005 (0,0004)	0,0000 (0,0000)	0,0006** (0,0003)	0,0000 (0,0000)
e_preescolar	0,0221*** (0,0052)	-0,0106 (0,0236)	0,0069 (0,0055)	0,0000 (0,0001)	-0,0027 (0,0059)	-0,0000 (0,0050)	-0,0000 (0,0050)	0,0000 (0,0001)
e_fundamental	-0,0210*** (0,0067)	0,1153*** (0,0278)	0,0010 (0,0070)	-0,0000 (0,0001)	-0,0014 (0,0079)	0,0000 (0,0001)	-0,0079 (0,0067)	-0,0000 (0,0001)
prof_idade_dp	0,0045*** (0,0016)	-0,0364*** (0,0063)	0,0046*** (0,0017)	-0,0000 (0,0000)	0,0043** (0,0018)	-0,0000 (0,0000)	0,0046*** (0,0015)	-0,0000 (0,0000)
e_lab_info	0,0137** (0,0058)	0,0577** (0,0254)	0,0098 (0,0061)	0,0000 (0,0001)	0,0223*** (0,0067)	0,0000 (0,0001)	0,0325*** (0,0055)	-0,0000 (0,0001)
e_quadra	0,0135*** (0,0051)	-0,0746*** (0,0226)	-0,0124*** (0,0054)	-0,0000 (0,0001)	-0,0021 (0,0055)	0,0000 (0,0000)	0,0143*** (0,0045)	-0,0000 (0,0001)
faltas_completas	-0,0065*** (0,0015)	-0,0270*** (0,0056)	-0,0043*** (0,0012)	-0,0000 (0,0000)	-0,0065*** (0,0014)	-0,0000 (0,0000)	-0,0036*** (0,0011)	-0,0000 (0,0000)
Constante	0,7502*** (0,0557)	0,7228*** (0,0543)	1,0410*** (0,0750)	0,9490*** (0,0023)	1,0238*** (0,0683)	1,0085*** (0,0002)	1,2161*** (0,0547)	0,9802*** (0,0017)
atrhuo		0,0171 (0,0554)		-10,3446*** (0,3698)		-10,0134*** (0,1977)		-10,7796*** (0,6974)
Insigma		-1,2980*** (0,0048)		-1,1329*** (0,0053)		-1,0786*** (0,0046)		-1,3860*** (0,0050)

Fone: Elaboração própria.  
\*\*\*p < 0,01; \*\* p < 0,05; \* p < 0,1.

Agora observe os parâmetros associados à correlação entre os termos de erro  $\varepsilon_{1i}$  e  $\nu_{21}$  e ao desvio-padrão do resíduo da equação de regressão, respectivamente, *athrho* e *lnsigma*, obtidos nas estimações de *Heckman MLE*. Como descrito na Seção 4.2,  $\rho$  e  $\sigma$  não são diretamente estimados na função de verossimilhança, sendo a primeira obtida a partir de uma transformação numérica descrita na Equação (3.24), e a segunda é calculado a partir da transformação de  $\ln\sigma$ , e que dependendo da função de seleção especificada, o modelo exhibe resultados diferentes sobre a natureza do viés de seleção na variável  $d$ .

Os Modelos 1 e 2 expressam um valor estimado de *athrho*<sup>17</sup> na vizinhança de  $-10$ , que se traduz em uma correlação estimada  $\hat{\rho}$  próximo a  $-1$  e estatisticamente não significativo. Assim sendo, para estes dois casos, não rejeitou-se a hipótese nula de  $\rho = 0$ , logo, o uso de um modelo de seleção amostral não é necessário. Por fim, um valor negativo de  $\rho$  próximo ao seu limite inferior, implica que qualquer componente do termo de erro da equação de seleção (ou viés) que aumente a probabilidade de um jovem propenso ao abandono da rede pública nela permanecer, reduz a probabilidade deste mesmo aluno em atingir um resultado acadêmico satisfatório. Ou seja, o perfil de quem reprova é bastante correlacionado com o perfil de quem sai da base.

---

<sup>17</sup> Saída do STATA 13 que representa  $\text{atan}\rho$ . Qualquer valor estimado para este termo inferior a  $-1$ , o *software* converge o modelo para um valor de  $\rho = -1$ .

Em contrapartida, para os anos de 2008, 2009 e 2011 do Modelo 3 e 2008 do Modelo 4, os valores de  $\text{athrho}$  indicam um valor de  $\rho$  positivo e estatisticamente significativo, o que sugere que o efeito da permanência na escola sem a correção amostral de Heckman seria viesado para cima. Para estes casos específicos, o perfil do aluno aprovado ao fim do ano letivo é pouco correlacionado ao daquele que decide abandonar a escola pública.

### 3.7.1 Comparação entre turnos

Com o objetivo de melhor identificar as variáveis que definem a permanência na escola pública e traçar um perfil mais realista do sistema educacional brasileiro, este ensaio realizou uma análise comparativa entre os alunos dos turnos da manhã e da noite (indicadores 1 e 3, respectivamente), pois sabe-se que é no período noturno que os alunos mais frágeis em termos sócioeconômicos e propensos ao trabalho estão matriculados. Para tanto, as Tabelas de (22) a (25) apresentam os valores dos coeficientes estimados, bem como seus respectivos desvios-padrão e estatísticas  $t$ , pela técnica de máxima verossimilhança de Heckman, para os alunos de ambos os turnos, da 1<sup>a</sup> e 2<sup>a</sup> séries, nos anos de 2008 e 2010.

A Tabela (22) apresenta os resultados para o Modelo 1. Note primeiro, que os parâmetros  $\hat{\beta}$  e  $\hat{\gamma}$  estimados para a *dummy* menino apresentaram os resultados esperados, estatisticamente significante e com o sinal negativo em ambos os turnos, e evidenciam que há uma maior probabilidade de reprovação e abandono

escolar por parte dos alunos matriculados no período noturno, além de que o efeito dessa característica se reduz com o avançar das séries. Quanto a ser branco e morar na zona rural, via de regra, estes não se mostraram estatisticamente significantes.

A estimativa do parâmetro para a *dummy* bolsa, apesar da não significância estatística para o ano de 2008, tem um valor, em média, 1,5 vezes maior para os alunos do noturno do que os do período da manhã, para a equação de aprovação. Quanto ao parâmetro estimado para PIB per capita municipal, este mostrou ser mais relevante para os alunos do turno da manhã.

Quanto aos parâmetros estimados para as características escolares, as que mais se sobressaem são as variáveis *prof\_idade* e *alunos\_turma*. A primeira, mostrou-se significativa para os alunos da 1ª série nos dois anos analisados e, como esperado, sinal positivo para aprovação e negativo para permanência, ou seja, escolas com professores mais experientes melhoram o desempenho acadêmico do aluno, enquanto que os professores mais jovens têm um maior efeito no combate à saída da escola pública. Para os alunos do noturno, o efeito dessa variável para a permanência tem efeito semelhante, mas com um valor maior. A segunda variável, apresentou-se significativa em praticamente todos os modelos, exceto para a 2ª Série, do período noturno, de 2008 (2008.2.3), e sinal positivo e negativo, respectivamente, para as equações de aprovação e permanência dos alunos do turno da manhã. Note também que, para o período noturno, o sinal e a significância se alteram ao longo do tempo, reflexo de mudanças de perfil de turma ou aluno, o que torna a sua interpretação nada trivial.

Tabela 22 – Modelo 1: Alunos da manhã × Alunos do noturno

Variáveis	2008.1.1		2008.1.3		2008.2.1		2008.2.3		2010.1.1		2010.1.3		2010.2.1		2010.2.3	
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d
menino	-0,0440*** (0,0036)	0,1543*** (0,0125)	-0,0748*** (0,0061)	-0,2626*** (0,0193)	0,1067*** (0,0025)	0,1067*** (0,0145)	-0,0199*** (0,0032)	0,0914*** (0,0147)	0,0994*** (0,0116)	0,0994*** (0,0116)	-0,0344*** (0,0069)	0,0911*** (0,0161)	-0,0238*** (0,0032)	0,1014*** (0,0136)	-0,0193*** (0,0041)	0,0731*** (0,0154)
branco	0,0393*** (0,0091)	-0,1377*** (0,0319)	0,0642*** (0,0152)	0,1177*** (0,0499)	-0,0095 (0,0089)	0,0543 (0,0506)	-0,0050 (0,0111)	0,0231 (0,0507)	0,0248*** (0,0092)	-0,0734*** (0,0270)	0,0036 (0,0133)	-0,0096 (0,0356)	0,0125 (0,0081)	0,0845 (0,0345)	-0,0012 (0,0106)	0,0045 (0,0399)
rural	0,0074 (0,0050)	-0,0260 (0,0174)	0,0331*** (0,0078)	0,3392*** (0,0277)	0,0021 (0,0035)	-0,0117 (0,0197)	-0,0039 (0,0044)	0,0181 (0,0201)	-0,0009 (0,0350)	0,0026 (0,1034)	-0,0150* (0,0880)	0,0402* (0,0213)	-0,0068 (0,0044)	0,0292 (0,0189)	-0,0086 (0,0055)	0,0326 (0,0209)
bolsa	0,0101 (0,0128)	-0,0356 (0,0448)	0,0324 (0,0198)	0,1271* (0,0764)	0,0070 (0,0099)	-0,0396 (0,0633)	0,0140 (0,0181)	-0,0643 (0,0832)	0,0350* (0,0202)	-0,1034* (0,0598)	0,1038*** (0,0261)	-0,2774*** (0,0696)	0,0387*** (0,0116)	-0,1651*** (0,0494)	0,0341*** (0,0151)	-0,1289*** (0,0569)
lnpbp_pe	0,0174*** (0,0043)	-0,0610*** (0,0150)	0,0246*** (0,0067)	-0,0391* (0,0226)	0,0099*** (0,0030)	-0,0565*** (0,0169)	0,0157*** (0,0038)	-0,0721*** (0,0175)	0,0125*** (0,0055)	-0,0370*** (0,0161)	-0,0005 (0,0081)	0,0013 (0,0217)	0,0074* (0,0044)	-0,0315* (0,0186)	-0,0004 (0,0055)	0,0017 (0,0207)
e_biblio	-0,0049 (0,0003)	0,0174 (0,0151)	0,0055 (0,0068)	0,0285 (0,0232)	-0,0038 (0,0031)	0,0219 (0,0174)	-0,0046 (0,0039)	0,0210 (0,0177)	-0,0097** (0,0048)	0,0287** (0,0143)	-0,0279*** (0,0073)	0,0747*** (0,0196)	-0,0020 (0,0040)	0,0085 (0,0169)	-0,0199*** (0,0050)	0,0754*** (0,0188)
i_funcionarios_razao	0,0003 (0,0004)	-0,0011 (0,0005)	-0,0039*** (0,0008)	-0,0221*** (0,0025)	-0,0004 (0,0003)	0,0025 (0,0017)	-0,0004 (0,0004)	0,0019 (0,0019)	-0,0023*** (0,0005)	0,0068*** (0,0014)	0,0411*** (0,0008)	-0,0110*** (0,0021)	-0,0018*** (0,0004)	0,0076*** (0,0016)	0,0028*** (0,0005)	-0,0107*** (0,0019)
prof_idade	0,0012* (0,0007)	-0,0042* (0,0025)	-0,0004 (0,0011)	-0,0330*** (0,0036)	-0,0016*** (0,0005)	0,0093*** (0,0029)	-0,0013*** (0,0006)	-0,0058** (0,0029)	0,0043*** (0,0008)	-0,0128*** (0,0023)	0,0056*** (0,0012)	-0,0150*** (0,0031)	0,0033*** (0,0006)	-0,0140*** (0,0027)	0,0041*** (0,0008)	-0,0155*** (0,0030)
alunos_turno	0,0016*** (0,0003)	-0,0056*** (0,0011)	0,0023*** (0,0005)	0,0160*** (0,0015)	0,0007*** (0,0002)	-0,0040*** (0,0015)	0,0000 (0,0003)	-0,0001 (0,0003)	0,0022*** (0,0004)	-0,0065*** (0,0011)	-0,0026*** (0,0005)	0,0071*** (0,0012)	0,0007** (0,0003)	-0,0031** (0,0013)	-0,0010*** (0,0003)	0,0037*** (0,0012)
Constant	0,6947*** (0,0520)	1,0711*** (0,1823)	0,6378*** (0,0750)	1,8410*** (0,2604)	0,9624*** (0,0360)	0,2148 (0,2050)	0,8107*** (0,0445)	0,8683*** (0,2039)	0,6522*** (0,0613)	1,0282*** (0,1811)	0,9165*** (0,0864)	0,2241 (0,2307)	0,8036*** (0,0491)	0,8379*** (0,2095)	0,9054*** (0,0600)	0,3577 (0,2269)
atfho	-10,6858*** (0,3440)	-0,0032 (0,0575)	-11,6973*** (0,9449)	-0,5585*** (0,3372)	-1,1697*** (0,3772)	-1,1697*** (0,9449)	-1,1697*** (0,9449)	-1,1697*** (0,9449)	-10,4598*** (0,3430)	-0,1811*** (0,0830)	-8,9910*** (0,8325)	-11,1136*** (0,8378)	-11,1136*** (0,8378)	-11,1136*** (0,8378)	-10,5406*** (1,0475)	-10,5406*** (1,0475)
Insignia	-1,2550 (0,0000)	-1,2550 (0,0065)	-1,5232*** (0,0061)	-1,5232*** (0,0061)	-1,7407 (0,0000)	-1,7407 (0,0000)	-1,5232*** (0,0061)	-1,5232*** (0,0061)	-1,0837*** (0,0059)	-0,9825*** (0,0078)	-0,9825*** (0,0078)	-1,4508*** (0,0053)	-1,4508*** (0,0053)	-1,3298*** (0,0064)	-1,3298*** (0,0064)	-1,3298*** (0,0064)
Observações <sup>1</sup>	27957	27957	18455	18455	20677	20677	19977	19977	33804	33804	18866	18866	25111	25111	19619	19619

1: Número de observações utilizado para a estimação do modelo. Quantidade se mantém para os 4 modelos.

Fonte: Elaboração própria.

\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

A Tabela (23), a seguir, mostra os valores estimados para o Modelo 2. A diferença em relação ao modelo anterior trata da possibilidade da escola ofertar todo o ciclo básico, por consequência, do aluno cursar todo ele no mesmo local. A análise dos resultados para essas duas variáveis, mostra que elas praticamente não interferem na decisão do aluno continuar os seus estudos, salvo períodos atípicos, como em 2008 (2008.2.3).

Tabela 23 – Modelo 2: Alunos da manhã × Alunos do noturno

Variáveis	2008.1.1		2008.2.1		2008.2.3		2010.1.1		2010.1.3		2010.2.1		2010.2.3			
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d		
meio	-0,0439*** (0,00357)	0,154*** (0,0125)	-0,0752*** (0,00597)	-0,263*** (0,0193)	-0,0187*** (0,00254)	0,107*** (0,0145)	-0,0199*** (0,00321)	0,0913*** (0,0148)	-0,0335*** (0,00393)	0,0993*** (0,0116)	-0,0380*** (0,00604)	0,101*** (0,00618)	-0,0238*** (0,00318)	0,101*** (0,00618)	-0,0403*** (0,00782)	0,0926*** (0,0187)
branco	0,0392*** (0,00969)	-0,138*** (0,0319)	0,0644*** (0,0152)	0,117** (0,0499)	-0,00948 (0,00888)	0,0541 (0,0506)	-0,00516 (0,0110)	0,0237 (0,0508)	0,0247*** (0,00913)	-0,0731*** (0,0271)	-0,0273 (0,0355)	0,0127 (0,0345)	0,0075 (0,00810)	-0,0540 (0,0345)	0,0275 (0,0172)	0,156*** (0,0468)
rural	0,00738 (0,00496)	-0,0259 (0,0174)	0,0336*** (0,00765)	0,342*** (0,0278)	0,00239 (0,00346)	-0,0119 (0,0197)	-0,00403 (0,00438)	0,0185 (0,0201)	-0,009971 (0,00592)	0,00288 (0,0167)	-0,0119 (0,00800)	0,0316 (0,0213)	-0,00684 (0,00443)	0,0292 (0,0189)	0,0252 (0,0190)	0,287*** (0,0271)
bolsa	0,001 (0,0128)	-0,0355 (0,0448)	0,0326 (0,0198)	0,125 (0,0764)	0,00695 (0,00448)	-0,0396 (0,0111)	0,0140 (0,0181)	-0,0042 (0,0852)	0,0533* (0,0202)	0,104* (0,0598)	0,0991*** (0,0261)	-0,264*** (0,0694)	0,0387*** (0,0116)	-0,165*** (0,0494)	0,0383** (0,0167)	-0,0064 (0,0091)
hpbh_pc	0,010427 (0,00497)	-0,00427 (0,0150)	0,0246*** (0,00665)	0,0327 (0,0158)	0,00926 (0,00383)	-0,0366 (0,0118)	0,00926 (0,0132)	-0,0775 (0,0211)	0,0154 (0,00543)	-0,0721 (0,0217)	0,00816 (0,0244)	-0,00146 (0,00862)	0,00437 (0,00320)	0,00186 (0,00865)	0,00692 (0,00424)	0,0221 (0,0144)
e_biblio	0,00431 (0,00431)	0,0151 (0,00681)	0,00654 (0,00324)	0,0325 (0,0234)	0,00383 (0,00396)	0,0218 (0,0174)	0,00383 (0,00866)	0,0175 (0,0177)	-0,00979** (0,00484)	0,0290** (0,0143)	-0,0234*** (0,00735)	0,00855 (0,01095)	0,00437 (0,00372)	0,00865 (0,01169)	-0,0024 (0,0109)	0,114*** (0,0225)
i_funcionarios_ratio	0,00321 (0,00431)	-0,0151 (0,00681)	-0,00391*** (0,00258)	-0,0234*** (0,00258)	-0,000440 (0,00295)	0,00251 (0,00421)	-0,00409 (0,00421)	0,0188 (0,0194)	-0,00231*** (0,00471)	0,00685*** (0,0139)	-0,00991*** (0,00774)	-0,00991*** (0,00205)	-0,00179*** (0,00372)	0,00762*** (0,00159)	-0,00275 (0,00261)	-0,0278*** (0,00234)
prof_idade	0,00120* (0,00723)	-0,00422* (0,0254)	-0,00146 (0,00109)	-0,0311*** (0,00369)	-0,00164*** (0,000503)	0,00936*** (0,00287)	0,00127** (0,00634)	-0,00582** (0,00291)	0,00434*** (0,00779)	-0,0129*** (0,00231)	0,00539*** (0,00116)	-0,0144*** (0,00308)	0,00328*** (0,000623)	-0,0140*** (0,00271)	-0,000426 (0,00271)	-0,0342*** (0,00568)
alunos_turma	0,00158*** (0,000326)	-0,00554*** (0,0014)	0,00234*** (0,000459)	0,0159*** (0,00154)	0,000710*** (0,000223)	-0,00404*** (0,00127)	2,51e-05 (0,000258)	-0,000115 (0,00119)	0,00218*** (0,000383)	-0,00047*** (0,00113)	-0,00242*** (0,000467)	0,00645*** (0,00124)	0,000738** (0,000294)	-0,00315** (0,00125)	-6,97e-05 (0,000627)	0,00649*** (0,00447)
e_preescolar	4,45e-07 (2,75e-05)	0,45e-07 (0,0269)	0,000459 (0,000459)	-0,0145 (0,0269)	0,000223 (0,000223)	1,61e-07 (6,95e-05)	-2,99e-07 (6,95e-05)	-0,00119 (0,00119)	0,000383 (7,17e-05)	-2,64e-06 (4,98e-05)	-3,77e-06 (0,000222)	-1,51e-06 (9,80e-05)	0,000294 (0,000294)	-1,51e-06 (9,80e-05)	-0,0750** (0,0301)	0,00649*** (0,00447)
e_fundamental	0,695*** (0,0570)	1,071*** (0,182)	0,639*** (0,0750)	1,951*** (0,269)	0,962*** (0,0360)	0,215 (0,205)	0,811*** (0,0444)	0,868*** (0,204)	0,653*** (0,0612)	1,028*** (0,181)	0,894*** (0,0866)	0,839*** (0,230)	0,803*** (0,0492)	0,839*** (0,209)	1,064*** (0,0795)	1,394*** (0,280)
alfa_rho	-1,886*** (0,266)	0,0134 (0,0511)	-10,63*** (0,713)	-0,034 (0,1643)	-1,740*** (1,643)	0,205 (0,00565)	-11,779*** (1,643)	-10,36*** (1,643)	-10,36*** (0,400)	-10,399*** (0,400)	-10,399*** (0,400)	-10,75*** (0,804)	-10,75*** (0,804)	-10,75*** (0,804)	-0,0212 (0,518)	-0,0212 (0,518)
beta_rho	-1,255 (0)	-1,93*** (0,00648)	-1,740*** (0,00538)	-1,525*** (0,00565)	-1,740*** (0,00538)	-1,525*** (0,00565)	-1,740*** (0,00538)	-1,525*** (0,00565)	-1,086*** (0,00479)	-0,979*** (0,00756)	-0,979*** (0,00756)	-1,450*** (0,00534)	-1,450*** (0,00534)	-1,450*** (0,00534)	-1,384*** (0,00855)	-1,384*** (0,00855)

Fonte: Elaboração própria.

\*\*\*p < 0,01; \*\*p < 0,05; \*p < 0,1.

A inclusão das *dummies* *e\_lab\_info*, *e\_quadra* e da variável *prof\_idade\_dp* pouco acrescentaram à análise. Os números apresentados na Tabela (24) indicam uma grande oscilação de grandeza e significância para os parâmetros estimados das três variáveis. Entretanto, percebe-se que embora a *dummy* *prof\_idade\_dp* não tenha se mostrado estatisticamente significativa em todos os momentos, seu sinal manteve-se sempre negativo, isto é, as escolas onde o perfil etário é mais discrepante em relação à média, os alunos do período da manhã estão mais propensos à não permanecer na escola. Além disto, percebe-se que a estrutura física escolar mais relevantes para os alunos é a presença de um laboratório de informática.

Tabela 24 – Modelo 3: Alunos da manhã × Alunos do noturno

Variáveis	2008.1.1		2008.1.3		2008.2.1		2008.2.3		2010.1.1		2010.1.3		2010.2.1		2010.2.3			
	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d		
variavel	-0,0726*** (0,00373)		-2,48e-06 (3,75e-05)		-0,0293*** (0,00263)		-0,0387*** (0,00339)		-2,38e-06 (3,32e-05)		-3,82e-06 (1,87e-05)		-3,57e-07 (8,93e-06)		-0,0408*** (0,00442)			
branco	0,0610*** (0,00967)		3,20e-06 (8,35e-05)		0,00652 (0,00946)		0,0227* (0,0121)		1,53e-06 (7,22e-05)		1,80e-06 (4,02e-05)		3,07e-07 (2,09e-05)		0,0283*** (0,0119)			
rural	0,0283*** (0,00507)		1,81e-06 (5,21e-05)		0,0133*** (0,00352)		0,0196*** (0,00450)		1,04e-06 (4,85e-05)		2,34e-06 (2,47e-05)		2,03e-07 (1,32e-05)		0,0265*** (0,00577)			
bolsa	0,0143 (0,0132)		2,06e-06 (0,00151)		0,00609 (0,0114)		0,0211 (0,0188)		-4,75e-06 (0,00156)		-5,70e-06 (8,29e-05)		2,61e-07 (3,96e-05)		0,0382*** (0,0162)			
lnpbib_pe	0,0144*** (0,00451)		1,18e-06 (4,45e-05)		0,0115*** (0,00309)		0,0156*** (0,00405)		0,00842 (0,00405)		4,77e-07 (2,57e-05)		4,12e-08 (2,27e-07)		-0,0102* (0,00611)			
e_biblio	0,00183 (0,00087)		7,11e-07 (4,49e-05)		0,000842 (0,00181)		8,42e-07 (4,09e-05)		0,000560 (0,00409)		-1,05e-07 (2,31e-05)		2,27e-07 (1,10e-05)		-0,00556 (0,00549)			
i_funcionarios_razao	-0,00237*** (0,000459)		-4,77e-08 (5,30e-06)		-0,00121*** (0,000311)		-0,00208*** (0,000456)		0,000456 (4,35e-06)		-1,99e-07 (2,57e-06)		-2,47e-08 (1,09e-06)		-0,00290*** (0,000595)			
prof_idade	0,00393*** (0,000757)		1,22e-07 (7,99e-06)		0,000572 (0,00133***)		-0,000704 (0,000672)		0,000704 (7,41e-06)		2,92e-07 (4,02e-06)		2,35e-09 (2,02e-06)		-0,000597 (0,000868)			
alunos_turma	0,00341 (0,000341)		1,22e-07 (2,93e-06)		0,00133*** (0,000232)		0,000746*** (0,000271)		0,000746*** (0,000271)		-2,10e-08 (1,44e-06)		-2,10e-08 (1,44e-06)		-3,35e-05 (0,000347)			
e_presecolar	0,0445 (0,0354)		-0,0121 (0,0255)		0,0151 (0,0314)		-0,0175 (0,0314)		-0,0940*** (0,0283)		-2,78e-06 (0,000143)		-8,21e-06 (7,35e-05)		-0,115*** (0,0261)			
e_fundamental	0,117*** (0,0277)		5,79e-06 (0,000206)		0,0625* (0,0348)		0,120*** (0,0350)		0,120*** (0,0350)		3,34e-06 (0,000180)		1,74e-05 (9,60e-05)		0,532*** (6,85e-05)			
prof_idade_dp	-0,0367*** (0,00630)		-7,64e-07 (4,40e-05)		-0,0529*** (0,00823)		-0,0453*** (0,00730)		-0,0453*** (0,00730)		-9,42e-07 (4,27e-05)		-1,98e-06 (1,92e-05)		-0,0465*** (0,00674)			
e_lab_info	0,0549** (0,0254)		4,13e-06 (0,00162)		0,0964*** (0,0336)		0,110*** (0,0326)		0,0964*** (0,0326)		2,09e-06 (0,000150)		7,25e-06 (7,43e-05)		1,94e-07 (5,71e-05)		0,125*** (0,0280)	
e_quadra	0,0722*** (0,0543)		1,030*** (0,00272)		1,030*** (0,00803)		0,0307 (0,0166)		0,0307 (0,0166)		0,000133 (0,000133)		3,42e-06 (7,12e-05)		1,01e-06 (5,46e-05)		-0,0590** (0,0271)	
Constant	0,722*** (0,0543)		1,030*** (0,00272)		1,030*** (0,00803)		0,809*** (0,0470)		0,809*** (0,0470)		1,010*** (0,00211)		1,059*** (0,00599)		1,020*** (0,00158)		1,067*** (0,00661)	
atrito	0,0445 (0,0354)		-0,0121 (0,0255)		0,0151 (0,0314)		-0,0175 (0,0314)		-0,0940*** (0,0283)		-2,78e-06 (0,000143)		-8,21e-06 (7,35e-05)		-0,115*** (0,0261)			
Insigima	-1,297*** (0,00484)		-1,113*** (0,00658)		-1,755*** (0,00529)		-1,449*** (0,00481)		-1,449*** (0,00481)		-0,975*** (0,00601)		-1,449*** (0,00601)		-1,384*** (0,00626)			

Fonte: Elaboração própria.  
\*\*\*p < 0,01; \*\*p < 0,05; \*p < 0,1.

Por fim, os resultados para o Modelo 4 expostos na Tabela (25) mostram que, independente do turma, a variável *faltas\_completas* mesmo não sendo estatisticamente significativa para todas as subamostras, sempre traz um sinal negativo.

Tabela 25 – Modelo 4: Alunos da manhã × Alunos do noturno

Variáveis	2008.1.1		2008.1.3		2008.2.1		2008.2.3		2010.1.1		2010.1.3		2010.2.1		2010.2.3	
	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d
menino	-0,0727*** (0,00373)		-2,10e-06 (1,23e-05)		-0,0293*** (0,00263)		-3,53e-06 (1,16e-05)		-4,04e-06 (1,16e-05)		-5,33e-06 (1,61e-05)		-6,31e-07 (7,31e-06)		-0,0405*** (0,00442)	
branco	0,0610*** (0,00067)		1,57e-06 (2,97e-05)		0,00651 (0,000946)		2,13e-06 (2,54e-05)		2,13e-06 (2,54e-05)		2,13e-06 (2,54e-05)		1,88e-06 (6,10e-05)		0,0283** (0,0119)	
rural	0,0283*** (0,00506)		9,15e-07 (1,65e-05)		0,0133*** (0,00352)		1,57e-06 (5,22e-06)		1,57e-06 (1,71e-05)		1,57e-06 (1,71e-05)		2,08e-06 (4,10e-05)		0,0262*** (0,00577)	
bolsa	0,0144 (0,0132)		1,04e-06 (4,75e-05)		0,00609 (0,0115***)		1,33e-07 (2,65e-05)		6,29e-06 (5,62e-05)		6,29e-06 (5,62e-05)		-3,90e-06 (4,48e-07)		0,0382** (0,0162)	
lnphb_pc	0,0144*** (0,00449)		7,49e-07 (1,46e-05)		0,0115*** (0,00309)		7,26e-08 (6,34e-06)		5,35e-07 (1,70e-05)		5,35e-07 (1,70e-05)		4,48e-07 (4,15e-05)		-0,0102* (0,00611)	
e_biblio	0,00185 (0,00451)		6,57e-08 (1,52e-05)		0,000841 (0,00318)		-6,34e-06 (4,74e-06)		1,21e-06 (1,44e-05)		1,21e-06 (1,44e-05)		4,46e-07 (3,60e-05)		-0,00355 (0,00548)	
i_funcionarios_razao	-0,000678 (0,00459)		-8,66e-08 (1,72e-06)		-0,00121*** (0,00311)		-8,39e-09 (5,08e-07)		-8,39e-08 (1,53e-06)		-8,39e-08 (1,53e-06)		-8,68e-08 (4,11e-06)		-0,00291*** (0,006594)	
prof_idade	-0,00234*** (0,00758)		2,80e-08 (2,64e-06)		-0,00280*** (0,00422)		9,33e-10 (8,48e-07)		7,14e-08 (2,60e-06)		7,14e-08 (2,60e-06)		2,48e-07 (6,45e-06)		-0,000601 (0,00868)	
alunos_turno	0,00393*** (0,00341)		5,83e-08 (9,58e-07)		0,00133*** (0,000232)		3,18e-09 (3,04e-07)		1,85e-08 (1,17e-06)		1,85e-08 (1,17e-06)		8,61e-09 (2,29e-06)		-2,74e-08 (7,13e-07)	
e_presscolar	-0,0106 (0,0236)		-8,09e-07 (5,20e-05)		-0,0179 (0,0314)		-0,0179 (2,52e-05)		-4,04e-06 (5,00e-05)		-4,04e-06 (5,00e-05)		-5,33e-06 (0,000117)		-1,59e-06 (4,61e-05)	
e_fundamental	0,115*** (0,0278)		2,33e-06 (7,05e-05)		0,0614* (0,0348)		1,20e-07 (3,32e-05)		4,72e-06 (6,29e-05)		4,72e-06 (6,29e-05)		1,61e-05 (0,000148)		8,35e-07 (5,60e-05)	
prof_idade_dp	-0,0364*** (0,00631)		-6,89e-07 (1,44e-05)		-0,0521*** (0,00824)		-6,91e-08 (6,98e-06)		-1,35e-06 (1,50e-05)		-1,35e-06 (1,50e-05)		-1,40e-06 (3,09e-05)		-5,40e-07 (1,35e-05)	
e_lab_info	0,0577** (0,0254)		1,22e-06 (5,58e-05)		0,0956*** (0,0326)		1,93e-07 (2,64e-05)		2,94e-06 (5,24e-05)		2,94e-06 (5,24e-05)		7,59e-06 (1,82e-06)		3,37e-07 (4,68e-05)	
e_quadra	-0,0746*** (0,0226)		8,80e-09 (5,60e-05)		0,0162 (0,0294)		9,31e-08 (2,64e-05)		2,36e-06 (4,66e-05)		2,36e-06 (4,66e-05)		-1,82e-06 (0,000116)		1,79e-06 (4,45e-05)	
fallus_complexas	-0,0270*** (0,00557)		-4,92e-07 (1,02e-05)		-0,0177*** (0,00650)		-3,39e-08 (4,97e-06)		-6,41e-07 (1,11e-05)		-6,41e-07 (1,11e-05)		-1,82e-06 (2,44e-05)		-1,56e-07 (1,03e-05)	
Constante	0,723*** (0,0543)		1,028*** (0,000243)		1,539*** (0,0574)		0,999*** (0,00185)		0,00270 (0,000843)		0,00270 (0,00197)		1,059*** (0,00304)		-0,0825*** (0,00906)	
athlho	0,0171 (0,0554)		-10,54*** (0,359)		0,0126 (0,0504)		-12,63*** (1,368)		-10,21*** (0,198)		-10,21*** (0,198)		-10,21*** (0,548)		-1,11e-06 (0,544)	
Insgima	-1,298*** (0,00480)		-1,109*** (0,00615)		-1,756*** (0,00529)		-1,521*** (0,00646)		-1,079*** (0,00460)		-1,079*** (0,00460)		-0,975*** (0,00676)		-1,447*** (0,00704)	

Fonte: Elaboração própria.  
 \*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

## 3.8 Considerações finais

Como apresentado no início deste ensaio, uma preocupação recorrente da literatura que estuda o sistema educacional no Brasil, em especial o Ensino Médio, são os motivos que levam o estudante a abandonar a rede pública de ensino. Logo, criar formas de enfrentamento a esta decisão a partir de uma perspectiva de sistema família-escola-aluno e do aluno, que defina os fatores mais importantes nesse contexto é vital para a redução dos índices de evasão escolar. Mas qual é o peso desses fatores? Qual a diferença de perfil entre o aluno do turno da manhã e do noturno? O perfil do aluno que reprova é semelhante ao perfil daquele que abandona a escola? Essas perguntas foram elucidadas nesse trabalho.

Para tanto, recorreu-se à técnica proposta por Heckman (1977) que indica uma maneira de tratar o problema do viés de seleção, em um exercício empírico aplicado a uma base de dados de Santa Catarina. E até chegar as respostas finais, outros resultados foram considerados:

1. Quando se compara a regressão *MCO* com a de Heckman MLE, a primeira superestima os coeficientes e erros-padrão estimados para as covariadas propostas;
2. Ser menino, branco e morador de cidade com alto PIB *per capita* apresenta maior probabilidade de abandonar a rede pública de ensino;

3. A família estar incluída em programa de transferência de renda contribui para a melhora de desempenho do aluno do período noturno. Porém, para os alunos do turno da manhã, não é significativo;
4. Os professores mais jovens associados à uma menor oscilação etária aparentam ter maior eficácia no combate ao abandono escolar, principalmente nos alunos do período noturno;
5. A presença de laboratório de informática foi a mais importante entre as *dummies* que caracterizam a estrutura da escola;
6. O jovem ter a possibilidade de cursar etapas diferentes do ciclo básico na mesma instituição não demonstra ter relevância nem para o desempenho e nem para a permanência na rede pública;
7. A quantidade de alunos em sala de aula tem influência restrita no desempenho acadêmico. Já na decisão de permanência, mostrou-se pouco relevante sobretudo, para os alunos do período noturno;
8. O nível de impacto das características individuais e familiares se reduzem com o avançar das séries.

Ademais, concluiu-se que a probabilidade de um jovem propenso ao abandono da rede pública nela permanecer, reduz a probabilidade desse mesmo aluno em atingir um resultado acadêmico satisfatório. Ou seja, o perfil de um estudante que apresenta um franco rendimento escolar é bastante correlacionado com os daqueles que abandonam a rede pública.



# Conclusão

Ao cabo desta pesquisa, existem três considerações finais a se tirar: a primeira, em relação ao período de maior risco para o abandono escolar; a segunda, da correlação entre reprovação e abandono escolar da rede pública; e a terceira, sobre a problemática de atrito na base de dados.

Em relação à primeira, verificou-se que é na 1ª Série o momento de maior probabilidade de abandono escolar na rede pública, devido, sobretudo, às fragilidades carregadas do Ensino Fundamental, às necessidades de trabalho e aos aspectos familiares. Todavia, o alcance das características estruturais e pedagógicas da escola mostraram-se bastante reduzidas tanto no processo de decisão de permanência, quanto no desempenho acadêmico.

Quanto à correlação entre reprovação e abandono escolar, os exercícios empíricos mostraram, em termos gerais, que o perfil do aluno que reprova é semelhante ao perfil do aluno que abandona a escola, embora existam situações em que os alunos do turno da manhã se diferenciam dos alunos do turno da noite. Diante dessas evidências, ficou claro que um desempenho acadêmico desfavorável, normalmente, resulta em abandono escolar.

Por último, observou-se que há muito atrito nas bases de dados da SEDSC. Nesse sentido, o modelo de Heckman (1977) se mostrou mais acurado do que as técnicas probabilísticas bivariadas para a análise dos fatores que condicionam simultaneamente a aprovação e a permanência na rede pública de ensino – por consequência, como resultado dessa pesquisa, recomenda-se que o modelo de correção de atrito seja usado em pesquisas futuras dessa natureza.

Em suma, esta pesquisa mostrou que, de fato, há uma forte correlação entre reprovação e abandono escolar, e que é preciso desenhar uma política que assegure a permanência dos jovens na escola, para que o país melhore não apenas em aspectos econômicos, substanciado por externalidades ocasionadas por maior estoque de capital humano, mas também em aspectos sociais, com a redução da violência, da gravidez precoce e dos índices de doenças crônicas, sobretudo oriundas do uso de drogas, álcool e de má higiene.

# Referências

ACEMOGLU, D.; ANGRIST, J. How large are human-capital externalities? evidence from compulsory schooling laws. *NBER macroeconomics annual*, MIT Press, v. 15, p. 9–59, 2000.

Citado na página 30.

ADELMAN, C. Women at thirtysomething: Paradoxes of attainment. ERIC, 1991. Citado na página 70.

AKERLOF, G. A.; KRANTON, R. E. Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, v. 40, n. 4, p. 1167–1201, 2002. Citado na página 43.

ALEXANDER, K. L.; PALLAS, A. M. Curriculum reform and school performance: An evaluation of the "new basics". *American Journal of Education*, The University of Chicago Press, v. 92, n. 4, p. 391–420, 1984. Citado na página 65.

ALLEN, D.; FRASER, B. J. Parent and student perceptions of classroom learning environment and its association with student outcomes. *Learning Environments Research*, Springer, v. 10, n. 1, p. 67–82, 2007. Citado na página 31.

ALTONJI, J. G. *The effects of high school curriculum on education and labor market outcomes*. [S.l.], 1992. Citado 3 vezes nas páginas 64, 69 e 75.

ANDRADE, R. J. de; SOARES, J. F. O efeito da escola básica brasileira. *Estudos em avaliação educacional*, v. 19, n. 41, p. 379–406, 2008. Citado na página 83.

- ARROW, K. J. et al. Higher education as a filter. *Journal of public economics*, Amsterdam, v. 2, n. 3, p. 193–216, 1973. Citado 2 vezes nas páginas 30 e 37.
- BAKER, G. Human capital: A theoretical and empirical analysis. *New York: National Bureau*, 1964. Citado 3 vezes nas páginas 28, 30 e 33.
- BARBOSA-FILHO, F. d. H.; PESSÔA, S. d. A. Educação e crescimento: O que a evidência empírica e teórica mostra? *Revista EconomiA*, v. 11, n. 2, p. 265–303, 2010. Citado 3 vezes nas páginas 34, 36 e 38.
- BARROS, R. P. d.; MENDONÇA, R. Abandono e evasão no ensino médio no brasil: magnitude e tendências. *A crise de audiência no Ensino Médio*. São Paulo: Instituto Unibanco, p. 7–42, 2008. Citado 4 vezes nas páginas 31, 80, 86 e 152.
- BARTHOLOMEW, D. J. *Stochastic models for social processes*. [S.l.]: Wiley New York, 1973. Citado na página 109.
- BEAN, J. P. Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in higher education*, Springer, v. 12, n. 2, p. 155–187, 1980. Citado 6 vezes nas páginas 95, 97, 100, 101, 107 e 136.
- BETTS, J. R. *The role of homework in improving school quality*. [S.l.]: Citeseer, 1996. Citado na página 50.
- BLAUG, M. Where are we now in the economics of education? *Economics of education review*, Elsevier, v. 4, n. 1, p. 17–28, 1985. Citado 2 vezes nas páginas 30 e 43.
- BOARDMAN, A. E.; DAVIS, O. A.; SANDAY, P. R. A simultaneous equations model of the educational process. *Journal of Public Economics*, Elsevier, v. 7, n. 1, p. 23–49, 1977. Citado na página 48.

BRASIL, M. *Relatório Educação para Todos no Brasil 2000-2015: versão preliminar*. 2014. Citado na página [86](#).

BRIGGS, D. C. Causal inference and the heckman model. *Journal of Educational and Behavioral Statistics*, Sage Publications Sage CA: Los Angeles, CA, v. 29, n. 4, p. 397–420, 2004. Citado 6 vezes nas páginas [168](#), [171](#), [174](#), [177](#), [190](#) e [192](#).

BUSHWAY, S.; JOHNSON, B. D.; SLOCUM, L. A. Is the magic still there? the use of the heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, Springer, v. 23, n. 2, p. 151–178, 2007. Citado na página [177](#).

CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics using stata*. [S.l.]: Stata press College Station, TX, 2010. v. 2. Citado 8 vezes nas páginas [82](#), [169](#), [170](#), [172](#), [174](#), [176](#), [178](#) e [197](#).

CAMERON, S. V.; HECKMAN, J. J. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political economy*, The University of Chicago Press, v. 106, n. 2, p. 262–333, 1998. Citado 9 vezes nas páginas [97](#), [102](#), [108](#), [109](#), [111](#), [112](#), [114](#), [136](#) e [137](#).

CARTER, S. The impact of parent/family involvement of student outcomes: An annotated bibliography of research from the past decade. ERIC, 2002. Citado na página [79](#).

CHIBURIS, R.; LOKSHIN, M. et al. Maximum likelihood and two-step estimation of an ordered-probit selection model. *Stata Journal*, StataCorp LP, v. 7, n. 2, p. 167–182, 2007. Citado na página [180](#).

CHIRAS, D.; CREA, D. The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *The American Economic Review*, American Economic Association, v. 94, n. 1, p. 155–189, 2004. Citado na página [95](#).

CHRISTOFIDES, L. N.; STENGOS, T.; SWIDINSKY, R. On the calculation of marginal effects in the bivariate probit model. *Economics Letters*, Elsevier, v. 54, n. 3, p. 203–208, 1997. Citado 2 vezes nas páginas 119 e 123.

COATES, D. Education production functions using instructional time as an input. *Education Economics*, Taylor & Francis, v. 11, n. 3, p. 273–292, 2003. Citado 3 vezes nas páginas 50, 51 e 187.

COHN, E. et al. Input-output analysis in public education. ERIC, 1975. Citado na página 48.

COLEMAN, J. The concept of equality of educational opportunity. *Harvard Educational Review*, Harvard Education Publishing Group, v. 38, n. 1, p. 7–22, 1968. Citado 3 vezes nas páginas 43, 46 e 48.

COLLIER, J. K. *An educational production function analysis of selected Texas school districts*. Tese (Doutorado), 1994. Citado 2 vezes nas páginas 45 e 49.

CORREA, H. The microeconomic theory of education. *International Journal of Educational Research*, Elsevier, v. 23, n. 5, p. 405–472, 1995. Citado 2 vezes nas páginas 31 e 43.

DOLL, J. J.; ESLAMI, Z.; WALTERS, L. Understanding why students drop out of high school, according to their own reports: Are they pushed or pulled, or do they fall out? a comparative analysis of seven nationally representative studies. *Sage Open*, SAGE Publications Sage CA: Los Angeles, CA, v. 3, n. 4, p. 2158244013503834, 2013. Citado 2 vezes nas páginas 95 e 157.

DUBIN, J. A.; RIVERS, D. Selection bias in linear regression, logit and probit models. *Sociological Methods & Research*, Sage Publications, v. 18, n. 2-3, p. 360–390, 1989. Citado 2 vezes nas páginas 159 e 164.

DURKHEIM, E. *Suicide: A study in sociology. Translated by John A. Spaulding and George Simpson. Edited with an introduction by George Simpson.* (Glencoe, Illinois: The Free Press, 1951), 1951. Citado na página [98](#).

ECKSTEIN, Z.; WOLPIN, K. I. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, Wiley Online Library, v. 67, n. 6, p. 1295–1339, 1999. Citado 5 vezes nas páginas [97](#), [102](#), [104](#), [107](#) e [136](#).

FENNEMA, E.; SHERMAN, J. Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American educational research journal*, Sage Publications Sage CA: Los Angeles, CA, v. 14, n. 1, p. 51–71, 1977. Citado na página [70](#).

FERNANDES, R. *Ensino Médio: Como aumentar a atratividade e evitar a evasão?* [S.l.], 2013. Citado na página [187](#).

FERRÃO, M. E.; BELTÃO, K. I.; SANTOS, D. P. dos. Políticas de não-repetência e a qualidade da educação: evidências obtidas a partir da modelagem dos dados da 4ª série do saeb-99. *Estudos em avaliação educacional*, n. 26, p. 47–74, 2002. Citado 3 vezes nas páginas [85](#), [86](#) e [152](#).

GONÇALVES, M. E. *Análise de sobrevivência e modelos hierárquicos logísticos longitudinais: uma aplicação à análise da trajetória escolar (4ª a 8ª série-ensino fundamental)*. Tese (Doutorado) — Tese (Doutorado)–Universidade Federal de Minas Gerais, Belo Horizonte, 2008. Cap. 1-2, 7.[Links], 2008. Citado na página [158](#).

GREENE, W. H. *Econometric analysis*. [S.l.]: Pearson Education India, 2003. Citado 11 vezes nas páginas [117](#), [119](#), [121](#), [122](#), [123](#), [124](#), [144](#), [148](#), [173](#), [176](#) e [177](#).

GREENWALD, R.; HEDGES, L. V.; LAINE, R. D. The effect of school resources on student achievement. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 66, n. 3, p. 361–396, 1996. Citado 3 vezes nas páginas 31, 78 e 80.

GRILICHES, Z.; HALL, B. H.; HAUSMAN, J. A. Missing data and self-selection in large panels. In: JSTOR. *Annales de l'INSEE*. [S.l.], 1978. p. 137–176. Citado na página 180.

HALL, B. *Notes on sample selection models*. [S.l.], 2002. Citado na página 178.

HANUSHEK, E. Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, JSTOR, v. 61, n. 2, p. 280–288, 1971. Citado 3 vezes nas páginas 55, 58 e 60.

HANUSHEK, E. A. The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, JSTOR, v. 24, n. 3, p. 1141–1177, 1986. Citado 4 vezes nas páginas 31, 43, 47 e 49.

HANUSHEK, E. A. School resources and student performance. *Does money matter? The effect of school resources on student achievement and adult success*, p. 43–73, 1996. Citado 5 vezes nas páginas 52, 81, 135, 188 e 197.

HANUSHEK, E. A. et al. Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, Wiley Online Library, v. 116, n. 510, 2006. Citado na página 68.

HANUSHEK, E. A.; RIVKIN, S. G.; TAYLOR, L. L. *Aggregation and the estimated effects of school resources*. [S.l.], 1996. Citado 2 vezes nas páginas 54 e 137.

HANUSHEK, E. A.; WÖSSMANN, L. *The role of school improvement in economic development*. [S.l.], 2007. Citado na página 40.

HARMON, C.; OOSTERBEEK, H.; WALKER, I. The returns to education: Microeconomics. *Journal of economic surveys*, Wiley Online Library, v. 17, n. 2, p. 115–156, 2003. Citado na página 34.

HASEBE, T. Marginal effects of a bivariate binary choice model. *Economics Letters*, Elsevier, v. 121, n. 2, p. 298–301, 2013. Citado na página 123.

HAUSER, R. M.; SIMMONS, S. J.; PAGER, D. I. High school dropout, race-ethnicity, and social background from the 1970s to the s. ERIC, 2000. Citado na página 135.

HECKMAN, J. J. *Sample selection bias as a specification error (with an application to the estimation of labor supply functions)*. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 1977. Citado 15 vezes nas páginas 24, 160, 162, 168, 169, 173, 174, 176, 177, 178, 181, 189, 197, 211 e 216.

JENCKS, C. et al. Inequality: A reassessment of the effect of family and schooling in america. ERIC, 1972. Citado 2 vezes nas páginas 47 e 48.

KEANE, M. P.; WOLPIN, K. I. The career decisions of young men. *Journal of political Economy*, The University of Chicago Press, v. 105, n. 3, p. 473–522, 1997. Citado na página 108.

KELLAGHAN, T. et al. *The home environment and school learning: Promoting parental involvement in the education of children*. [S.l.]: Jossey-Bass, 1993. Citado na página 79.

KLEES, S. J. *The economics of education: Is that all there is?* [S.l.]: University of Chicago Press, 1991. Citado na página 43.

KLEIN, R. Produção e utilização de indicadores educacionais: metodologia de cálculo de indicadores do fluxo escolar da educação básica. *Revista Brasileira de Estudos Pedagógicos*, v. 84, n. 206-07-08, 2007. Citado na página 158.

KODDE, D. A.; PALM, F. C. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1243–1248, 1986. Citado na página 176.

KRAWCZYK, N. et al. O ensino médio no brasil. Ação educativa, 2009. Citado 2 vezes nas páginas 21 e 156.

LADD, H. F. Education and poverty: Confronting the evidence. *Journal of Policy Analysis and Management*, Wiley Online Library, v. 31, n. 2, p. 203–227, 2012. Citado na página 95.

LEE, L.-F. Some approaches to the correction of selectivity bias. *The Review of Economic Studies*, Wiley-Blackwell, v. 49, n. 3, p. 355–372, 1982. Citado na página 176.

LEE, V. E.; BURKAM, D. T. Dropping out of high school: The role of school organization and structure. *American Educational Research Journal*, Sage Publications, v. 40, n. 2, p. 353–393, 2003. Citado na página 187.

LEE, V. E.; SMITH, J. B. Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education*, JSTOR, p. 164–187, 1993. Citado na página 67.

LEE, V. E.; SMITH, J. B. Effects of high school restructuring and size on early gains in achievement and engagement. *Sociology of education*, JSTOR, p. 241–270, 1995. Citado na página 64.

LEE, V. E.; SMITH, J. B.; CRONINGER, R. G. How high school organization influences the equitable distribution of learning in

mathematics and science. *Sociology of education*, JSTOR, p. 128–150, 1997. Citado na página 66.

LEON, F. L. L. d.; MENEZES-FILHO, N. A. Reprovação, avanço e evasão escolar no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2002. Citado na página 156.

LEVIN, H. M. Mapping the economics of education: An introductory essay. *Educational Researcher*, Sage Publications Sage CA: Thousand Oaks, CA, v. 18, n. 4, p. 13–17, 1989. Citado na página 43.

LEVINE, P. B.; ZIMMERMAN, D. J. The benefit of additional high-school math and science classes for young men and women. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 13, n. 2, p. 137–149, 1995. Citado na página 70.

LUCAS, R. E. On the mechanics of economic development. *Journal of monetary economics*, Elsevier, v. 22, n. 1, p. 3–42, 1988. Citado 2 vezes nas páginas 30 e 39.

MANSKI, C. F.; WISE, D. A. *College choice in America*. [S.l.]: Harvard University Press, 1983. Citado na página 166.

MARE, R. D. Social background and school continuation decisions. *Journal of the american statistical association*, Taylor & Francis, v. 75, n. 370, p. 295–305, 1980. Citado na página 109.

MARE, R. D.; WINSHIP, C. The paradox of lessening racial inequality and joblessness among black youth: Enrollment, enlistment, and employment, 1964-1981. *American Sociological Review*, JSTOR, p. 39–55, 1984. Citado na página 165.

MENEZES-FILHO, N.; PAZELLO, E. Do teachers' wages matter for proficiency? evidence from a funding reform in brazil. *Economics of Education Review*, Elsevier, v. 26, n. 6, p. 660–672, 2007. Citado 2 vezes nas páginas 82 e 83.

MENEZES-FILHO, N. A. *Os determinantes do desempenho escolar do Brasil*. [S.l.]: IFB, 2007. Citado 3 vezes nas páginas 80, 135 e 152.

MINCER, J. Investment in human capital and personal income distribution. *Journal of political economy*, The University of Chicago Press, v. 66, n. 4, p. 281–302, 1958. Citado na página 36.

MINCER, J. The distribution of labor incomes: a survey with special reference to the human capital approach. *Journal of economic literature*, JSTOR, v. 8, n. 1, p. 1–26, 1970. Citado na página 44.

MINCER, J. Schooling, experience, and earnings. human behavior & social institutions no. 2. ERIC, 1974. Citado 3 vezes nas páginas 28, 30 e 36.

NELSON, F. D. Censored regression models with unobserved, stochastic censoring thresholds. *Journal of econometrics*, Elsevier, v. 6, n. 3, p. 309–327, 1977. Citado na página 169.

NELSON, R. R.; PHELPS, E. S. Investment in humans, technological diffusion, and economic growth. *The American economic review*, JSTOR, v. 56, n. 1/2, p. 69–75, 1966. Citado na página 30.

NERI, M. et al. Motivos da evasão escolar. 2009. Citado 4 vezes nas páginas 80, 92, 156 e 187.

NERI, M. et al. Motivos da evasão escolar. 2015. Citado 7 vezes nas páginas 31, 80, 152, 156, 157, 185 e 197.

OAKES, J. *Keeping track*. [S.l.]: Yale University Press, 2005. Citado 2 vezes nas páginas 64 e 67.

OECD. Education spending (indicator). OECD, 2017. Citado na página 78.

OREOPOULOS, P. Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of public Economics*, Elsevier, v. 91, n. 11, p. 2213–2229, 2007. Citado na página 115.

OZTURK, I. The role of education in economic development: a theoretical perspective. 2001. Citado 2 vezes nas páginas 32 e 33.

PAZELLO, E.; CABRAL, R.; CAZOLLA, M. Conceitos: mapeando as formas alternativas de não participação. *A crise de audiência no Ensino Médio. São Paulo: Instituto Unibanco*, p. 43–52, 2008. Citado na página 87.

PETERSON, R. D.; HAGAN, J. Changing conceptions of race: Towards an account of anomalous findings of sentencing research. *American Sociological Review*, JSTOR, p. 56–70, 1984. Citado na página 165.

PLANK, S. B.; DELUCA, S.; ESTACION, A. High school dropout and the role of career and technical education: A survival analysis of surviving high school. *Sociology of Education*, SAGE Publications Sage CA: Los Angeles, CA, v. 81, n. 4, p. 345–370, 2008. Citado 3 vezes nas páginas 93, 157 e 159.

PRICE, J. L. *The study of turnover*. [S.l.]: Iowa State Press, 1977. Citado na página 100.

PSACHAROPOULOS, G. Returns to investment in education: A global update. *World development*, Elsevier, v. 22, n. 9, p. 1325–1343, 1994. Citado na página 43.

PSACHAROPOULOS, G.; PATRINOS, H. A. Returns to investment in education: a further update. *Education economics*, Taylor & Francis, v. 12, n. 2, p. 111–134, 2004. Citado 4 vezes nas páginas 27, 30, 43 e 44.

PUHANI, P. The heckman correction for sample selection and its critique. *Journal of economic surveys*, Wiley Online Library, v. 14, n. 1, p. 53–68, 2000. Citado na página 178.

RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: Applications and data analysis methods*. [S.l.]: Sage, 2002. v. 1. Citado na página 83.

RIVKIN, S. G.; HANUSHEK, E. A.; KAIN, J. F. Teachers, schools, and academic achievement. *Econometrica*, Wiley Online Library, v. 73, n. 2, p. 417–458, 2005. Citado 7 vezes nas páginas 31, 55, 59, 60, 63, 137 e 188.

ROGERS, B. Informing the shape of the curriculum: New views of knowledge and its representation in schooling. *Journal of Curriculum Studies*, Taylor & Francis, v. 29, n. 6, p. 683–710, 1997. Citado na página 65.

ROMER, P. M. Increasing returns and long-run growth. *Journal of political economy*, The University of Chicago Press, v. 94, n. 5, p. 1002–1037, 1986. Citado na página 30.

ROSE, H.; BETTS, J. R. The effect of high school courses on earnings. *Review of Economics and Statistics*, MIT Press, v. 86, n. 2, p. 497–513, 2004. Citado 5 vezes nas páginas 64, 73, 75, 76 e 77.

RUMBERGER, R. W.; LIM, S. A. *Why students drop out of school: A review of 25 years of research*. [S.l.], 2008. Citado na página 185.

SARTORI, A. E. An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis*, Cambridge University Press, v. 11, n. 2, p. 111–138, 2003. Citado 3 vezes nas páginas 160, 164 e 174.

- SCHULTZ, T. W. Investment in human capital. *The American economic review*, JSTOR, v. 51, n. 1, p. 1–17, 1961. Citado 3 vezes nas páginas 28, 30 e 33.
- SPENCE, M. Job market signaling. In: *Uncertainty in Economics*. [S.l.]: Elsevier, 1978. p. 281–306. Citado 3 vezes nas páginas 30, 37 e 38.
- STIGLITZ, J. E. The theory of "screening," education, and the distribution of income. *The American economic review*, JSTOR, v. 65, n. 3, p. 283–300, 1975. Citado 2 vezes nas páginas 30 e 37.
- SUBBARAO, K.; RANEY, L. Social gains from female education: a cross-national study. *Economic development and cultural change*, University of Chicago Press, v. 44, n. 1, p. 105–128, 1995. Citado na página 95.
- SUMMERS, A. A.; WOLFE, B. L. Do schools make a difference? *The American Economic Review*, JSTOR, v. 67, n. 4, p. 639–652, 1977. Citado 5 vezes nas páginas 55, 57, 58, 60 e 188.
- TANZI, V.; SCHUKNECHT, L. *Public spending in the 20th century: A global perspective*. [S.l.]: Cambridge University Press, 2000. Citado 2 vezes nas páginas 32 e 33.
- TERWEL, J. Curriculum differentiation: Multiple perspectives and developments in education. *Journal of Curriculum Studies*, Taylor & Francis, v. 37, n. 6, p. 653–670, 2005. Citado 2 vezes nas páginas 66 e 67.
- THOMAS, J. A. *Efficiency in education: A study of the relationship between selected inputs and mean test scores in a sample of senior high schools*. [S.l.]: School of Education, Stanford University, 1962. Citado na página 45.

TINTO, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 45, n. 1, p. 89–125, 1975. Citado 8 vezes nas páginas 95, 97, 98, 100, 101, 107, 136 e 185.

TINTO, V. Limits of theory and practice in student attrition. *The journal of higher education*, Taylor & Francis, v. 53, n. 6, p. 687–700, 1982. Citado na página 97.

TINTO, V. Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education*, Taylor & Francis, v. 59, n. 4, p. 438–455, 1988. Citado na página 97.

TYACK, D.; TOBIN, W. The “grammar” of schooling: Why has it been so hard to change? *American educational research journal*, Sage Publications, v. 31, n. 3, p. 453–479, 1994. Citado na página 65.

UNIBANCO, I. *Quem são os jovens fora da escola*. [S.l.], 2016. Citado 2 vezes nas páginas 187 e 197.

VANDENBUSSCHE, J.; AGHION, P.; MEGHIR, C. Growth, distance to frontier and composition of human capital. *Journal of economic growth*, Springer, v. 11, n. 2, p. 97–127, 2006. Citado na página 41.

VELLA, F. Estimating models with sample selection bias: a survey. *Journal of Human Resources*, JSTOR, p. 127–169, 1998. Citado 4 vezes nas páginas 163, 178, 180 e 190.

WALBERG, H. J.; FRASER, B. J.; WELCH, W. W. A test of a model of educational productivity among senior high school students. *The Journal of Educational Research*, Taylor & Francis, v. 79, n. 3, p. 133–139, 1986. Citado na página 50.

WALTENBERG, F. D. Teorias econômicas de oferta de educação: evolução histórica, estado atual e perspectivas. *Educação e Pesquisa*, SciELO Brasil, v. 32, n. 1, 2006. Citado na página 37.

WINSHIP, C.; MARE, R. D. Models for sample selection bias. *Annual review of sociology*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 18, n. 1, p. 327–350, 1992. Citado 2 vezes nas páginas 162 e 164.

WOESSMANN, L. et al. School accountability, autonomy and choice around the world. *Books*, Edward Elgar Publishing, 2009. Citado na página 68.

WOOLDRIDGE, J. M. *Introductory econometrics: A modern approach*. [S.l.]: Nelson Education, 2015. Citado 4 vezes nas páginas 166, 172, 177 e 189.

ZATZ, M. S.; HAGAN, J. Crime, time, and punishment: An exploration of selection bias in sentencing research. *Journal of Quantitative Criminology*, Springer, v. 1, n. 1, p. 103–126, 1985. Citado na página 165.



# ANEXO A – Informações complementares: Ensaio 3

Tabela 26 – Modelo 1: Estimativas Heckman MLE: 2008 e 2009

Variáveis	2008.1.2		2008.2.2		2009.1.2		2009.1.3		2009.2.1		2009.2.2		2009.2.3	
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d
menino	-0,0342*** (0,0057)	0,1134*** (0,0187)	-0,0238*** (0,0042)	0,1539*** (0,0250)	-0,0458*** (0,0060)	0,1387*** (0,0184)	-0,0780*** (0,0062)	-0,2042*** (0,0201)	-0,0283*** (0,0029)	0,1333*** (0,0139)	-0,0167*** (0,0050)	0,0868 (0,0859)	-0,0299*** (0,0038)	-0,1299*** (0,0210)
branco	0,0176 (0,0143)	-0,0582 (0,0475)	0,0035 (0,0142)	-0,0211 (0,0848)	0,0291** (0,0147)	-0,0880*** (0,0444)	0,0575*** (0,0157)	0,2453*** (0,0466)	-0,2622*** (0,0077)	-0,1237*** (0,0362)	0,0060 (0,0135)	-0,0280 (0,1875)	0,0368*** (0,0105)	0,1074* (0,0564)
rural	0,0049 (0,0069)	-0,0164 (0,0229)	0,0047 (0,0048)	-0,0280 (0,0284)	0,0300*** (0,0074)	-0,0907*** (0,0222)	0,0284*** (0,0081)	0,3495*** (0,0283)	0,0033 (0,0040)	-0,0155 (0,0190)	0,0194*** (0,0056)	-0,0932 (0,1173)	0,0140*** (0,0052)	0,3683*** (0,0322)
bolsa	-0,0523*** (0,0193)	0,1733*** (0,0638)	0,0100 (0,0152)	-0,0599 (0,0908)	0,0118 (0,0186)	-0,0356 (0,0561)	0,0042 (0,0206)	-0,0029 (0,0707)	0,0054 (0,0107)	-0,0255 (0,0507)	-0,0302* (0,0158)	0,1584 (0,3171)	0,0227 (0,0139)	-0,0446 (0,0819)
Implic_pc	0,0312*** (0,0067)	-0,1034*** (0,0222)	0,0017 (0,0050)	-0,0102 (0,0296)	-0,0154*** (0,0079)	0,0467** (0,0238)	0,0131* (0,0077)	0,0215 (0,0256)	-0,0119*** (0,0038)	0,0560*** (0,0179)	0,0076 (0,0068)	-0,0371 (0,1094)	-0,0003 (0,0048)	-0,0600*** (0,0256)
e_biblio	-0,0014 (0,0074)	0,0045 (0,0245)	-0,0070 (0,0059)	0,0419 (0,0355)	0,0016 (0,0083)	-0,0048 (0,0251)	0,0069 (0,0072)	-0,0478** (0,0242)	-0,0005 (0,0036)	0,0025 (0,0168)	0,0034 (0,0070)	-0,0147 (0,1128)	-0,0014 (0,0045)	0,0690*** (0,0251)
i_funcionarios_ratio	-0,0012** (0,0006)	0,0041** (0,0020)	0,0006 (0,0005)	-0,0028 (0,0038)	0,0009 (0,0007)	0,0027 (0,0020)	-0,0049*** (0,0009)	-0,0287*** (0,0026)	-0,0024*** (0,0004)	0,0112*** (0,0017)	-0,0019*** (0,0006)	0,0095 (0,0095)	-0,0021*** (0,0005)	-0,0212*** (0,0027)
prof_idade	0,0017 (0,0011)	-0,0037 (0,0035)	-0,0003 (0,0008)	0,0017 (0,0046)	0,0008 (0,0012)	-0,0025 (0,0035)	-0,0012 (0,0012)	-0,0202*** (0,0038)	-0,0019*** (0,0006)	0,0091*** (0,0027)	-0,0009 (0,0009)	0,0116 (0,0164)	-0,0001 (0,0007)	-0,0230*** (0,0041)
alunos_turno	-0,0011* (0,0005)	0,0056** (0,0017)	-0,0003 (0,0003)	0,0018 (0,0020)	0,0008 (0,0005)	-0,0023 (0,0016)	0,0009* (0,0005)	0,0115** (0,0015)	-0,0009*** (0,0003)	0,0015*** (0,0012)	-0,00223** (0,0004)	0,0012*** (0,0072)	0,0003 (0,0003)	0,0076*** (0,0017)
alhho	-10,6128*** (0,6371)	-10,2475*** (8,8102)	-9,6606*** (5,5114)											
Insigma	-1,1975*** (0,0083)	-1,7865*** (0,0100)	-1,1069*** (0,0078)											
Constante	0,6380*** (0,0770)	1,1991*** (2,2550)	0,9399*** (0,0569)	0,0045 (0,3393)	1,0447*** (0,0864)	-0,1347 (0,2614)	0,7719*** (0,0846)	0,8806*** (0,2859)	1,1566*** (0,0443)	-0,7390*** (0,2089)	0,9745*** (0,0743)	1,1262 (1,2493)	0,9352*** (0,0551)	2,1953*** (0,3000)

Fonte: Elaboração própria.

\*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

Tabela 27 – Modelo 1: Estimaciones Heckman MLE: 2010 e 2011

Variáveis	2010.1.2		2010.2.2		2011.1.2		2011.1.3		2011.2.1		2011.2.2		2011.2.3	
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d
merinho	-0,0627***	-0,1271***	-0,0181***	0,0833***	-0,0243***	0,0929***	-0,0125**	0,0388**	-0,0157***	0,0915***	-0,0177***	0,1190***	-0,0153***	0,0692***
	(0,0071)	(0,0215)	(0,0053)	(0,0243)	(0,0051)	(0,0197)	(0,0057)	(0,0176)	(0,0025)	(0,0146)	(0,0039)	(0,0261)	(0,0036)	(0,0162)
branco	0,0707***	0,1444***	0,0310**	-0,1429**	0,0158	-0,0606	0,0338***	-0,1049***	0,0183***	-0,1068***	0,0056	-0,0376	0,0022	-0,0100
	(0,0160)	(0,0467)	(0,0128)	(0,0590)	(0,0127)	(0,0486)	(0,0130)	(0,0402)	(0,0059)	(0,0347)	(0,0052)	(0,0619)	(0,0083)	(0,0375)
rural	0,0565***	0,3425***	-0,0107*	0,0494*	-0,0172***	0,0657***	-0,0542***	0,1683***	0,0062*	-0,0360*	0,0094**	-0,0630**	0,0050	-0,0227
	(0,0089)	(0,0289)	(0,0061)	(0,0280)	(0,0056)	(0,0216)	(0,0068)	(0,0212)	(0,0034)	(0,0196)	(0,0044)	(0,0296)	(0,0048)	(0,0217)
bolsa	-0,0780*	-0,6022***	0,0618***	-0,2851***	0,0605***	-0,2316**	0,2127***	-0,6601***	0,0786***	-0,4581***	-0,0200	0,1341	0,1174***	-0,5312***
	(0,0408)	(0,1013)	(0,0154)	(0,0708)	(0,0267)	(0,1023)	(0,0277)	(0,0859)	(0,0225)	(0,1310)	(0,0239)	(0,1602)	(0,0238)	(0,1078)
lnpib_pc	-0,0284***	-0,1805***	0,0187***	-0,0861***	-0,0098	0,0375	-0,0199**	0,0616**	-0,0053	0,0309	0,0107**	-0,0719**	0,0031	-0,0142
	(0,0098)	(0,0289)	(0,0071)	(0,0328)	(0,0070)	(0,0270)	(0,0079)	(0,0247)	(0,0035)	(0,0206)	(0,0053)	(0,0359)	(0,0049)	(0,0221)
e_biblio	0,0206**	0,2273***	-0,0076	0,0351	-0,0010	0,0039	-0,0008	0,0026	0,0050	-0,0289	0,0121**	-0,0811**	0,0044	-0,0197
	(0,0098)	(0,0275)	(0,0071)	(0,0326)	(0,0068)	(0,0262)	(0,0071)	(0,0221)	(0,0051)	(0,0182)	(0,0053)	(0,0358)	(0,0045)	(0,0202)
i_funcionarios_razao	-0,0042***	-0,0122***	-0,0001	0,0005	-0,0020***	0,0076***	0,0028***	-0,0086***	-0,0019***	0,0109***	-0,0021***	0,0140***	-0,0006	0,0028
	(0,0008)	(0,0021)	(0,0006)	(0,0027)	(0,0006)	(0,0022)	(0,0008)	(0,0023)	(0,0003)	(0,0018)	(0,0005)	(0,0031)	(0,0005)	(0,0023)
prof_idade	0,0009	-0,0448***	-0,0007	0,0031	-0,0011*	0,0063*	-0,0004	0,0011	-0,0001	0,0007	0,0012*	-0,0079*	0,0011	-0,0052
	(0,0014)	(0,0042)	(0,0010)	(0,0046)	(0,0010)	(0,0037)	(0,0011)	(0,0035)	(0,0005)	(0,0029)	(0,0007)	(0,0048)	(0,0007)	(0,0032)
alunos_turma	0,0011*	-0,0090***	0,0008*	-0,0035*	0,0018***	-0,0070***	-0,0003	0,0008	0,0003	-0,0020	-0,0000	0,0001	-0,0011***	0,0051***
	(0,0006)	(0,0019)	(0,0004)	(0,0020)	(0,0005)	(0,0019)	(0,0005)	(0,0015)	(0,0002)	(0,0014)	(0,0003)	(0,0023)	(0,0003)	(0,0014)
athrto	0,0149	-0,09422***	-0,0000	-0,0000	-10,6823***	-8,7340***	-8,7340***	-10,3394***	-6,8581***	-10,3394***	-6,8581***	-11,7170***	-11,7170***	-11,7170***
	(0,0599)	(0,6184)	(0,6184)	(0,6184)	(0,8681)	(0,1375)	(0,1375)	(0,6044)	(0,2356)	(0,6044)	(0,2356)	(1,1398)	(1,1398)	(1,1398)
lnsigma	-1,1043***	-1,5288	-0,0000	-0,0000	-1,3430***	-1,1330	-1,1330	-1,7633***	-1,9030***	-1,7633***	-1,9030***	-1,5092***	-1,5092***	-1,5092***
	(0,0073)	(0,0000)	(0,0000)	(0,0000)	(0,0076)	(0,0000)	(0,0000)	(0,0055)	(0,0055)	(0,0055)	(0,0055)	(0,0064)	(0,0064)	(0,0064)
Constant	1,0518***	4,0507***	0,8250***	0,8075**	1,1233***	-0,4722	1,2111***	-0,6546***	1,0379***	-0,2207	0,8406***	1,0770***	0,9490***	0,2310
	(0,1072)	(0,3198)	(0,0763)	(0,3519)	(0,0765)	(0,2929)	(0,0822)	(0,2553)	(0,0396)	(0,2308)	(0,3842)	(0,0573)	(0,0525)	(0,2377)

Fonte: Elaboração própria.

\*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

Tabela 28 – Modelo 2: Estimacões Heckman MLE: 2008 e 2009

Variáveis	2008.1.2		2008.2.2		2009.1.2		2009.1.3		2009.2.1		2009.2.2		2009.2.3	
	aprox	d	aprox	d	aprox	d	aprox	d	aprox	d	aprox	d	aprox	d
menino	-0,0316*** (0,00577)	0,107*** (0,0188)	-0,0258*** (0,00419)	0,154*** (0,0230)	-0,0450*** (0,00602)	0,119*** (0,0181)	-0,0287*** (0,00569)	0,0739*** (0,0160)	-0,0283*** (0,00294)	0,133*** (0,0139)	-0,0219*** (0,00466)	0,115*** (0,0246)	-0,0290*** (0,00375)	-0,1130*** (0,0210)
branco	0,0116 (0,0146)	-0,0373 (0,0475)	0,00358 (0,0142)	-0,0214 (0,0848)	0,0292** (0,0144)	-0,0884** (0,0457)	-0,000868 (0,0140)	0,0395 (0,0395)	0,0262*** (0,00767)	-0,124*** (0,0362)	-0,00665 (0,0124)	0,0351 (0,0655)	0,0568*** (0,0104)	0,107* (0,0564)
rural	0,00397 (0,00705)	-0,0128 (0,0229)	0,00475 (0,00476)	-0,0283 (0,0284)	0,0301*** (0,00738)	-0,0911*** (0,0223)	-0,0259*** (0,00749)	0,0730*** (0,0211)	0,000332 (0,00403)	-0,0156 (0,0190)	0,00623 (0,00653)	0,0329 (0,0280)	0,0142*** (0,00513)	0,368*** (0,0322)
bolsa	-0,0501** (0,0196)	0,162** (0,0638)	0,0100 (0,0152)	-0,0599 (0,0908)	0,0116 (0,0185)	-0,0352 (0,0561)	-0,000588 (0,0198)	-0,00166 (0,0558)	0,00539 (0,0107)	-0,0255 (0,0507)	-0,0386** (0,0150)	0,204** (0,0792)	0,0226 (0,0138)	-0,0447 (0,0819)
lnpph_pc	0,0313*** (0,00683)	-0,102*** (0,0222)	0,000171 (0,00497)	-0,1012 (0,0296)	-0,0154*** (0,00785)	0,0467** (0,0238)	0,00707 (0,00731)	-0,01199 (0,0206)	-0,01199*** (0,00379)	0,0559*** (0,0179)	0,00994 (0,00654)	-0,0525 (0,0335)	-0,000308 (0,00477)	-0,0589** (0,0267)
e_bhbio	-0,00297 (0,00754)	0,00972 (0,0245)	-0,00703 (0,00659)	0,0419 (0,0355)	0,00163 (0,00791)	-0,00492 (0,0239)	0,0126* (0,00885)	-0,0354* (0,0195)	-0,000522 (0,00336)	0,00246 (0,0168)	-0,00242 (0,00651)	0,0128 (0,0344)	-0,00132 (0,00455)	0,0576*** (0,0252)
i_functionarios_mzao	-0,00112* (0,000622)	0,00361* (0,00202)	0,000629 (0,000476)	-0,00375 (0,00284)	-0,000906 (0,000658)	0,00274 (0,00199)	0,00146* (0,000775)	-0,00410* (0,00218)	-0,00236*** (0,000362)	0,0112*** (0,00171)	-0,00217*** (0,000533)	0,0115*** (0,00281)	-0,00209*** (0,000508)	-0,0209*** (0,00284)
prof_idade	0,00222** (0,00107)	-0,00725** (0,00349)	-0,000286 (0,000770)	0,00171 (0,00459)	0,000808 (0,00116)	-0,00244 (0,00351)	-0,00681** (0,00308)	-0,00193*** (0,00308)	-0,00193*** (0,000582)	0,00911*** (0,00274)	-0,00172** (0,000881)	0,00908* (0,00465)	-0,000121 (0,000121)	-0,0227*** (0,00414)
alunos_lumna	-0,00126** (0,000513)	0,00411** (0,00166)	-0,000308 (0,000336)	0,00183 (0,00200)	0,000774 (0,000519)	-0,00234 (0,00157)	-0,00128*** (0,000436)	0,00360*** (0,00123)	0,000887*** (0,000239)	-0,00418*** (0,00122)	-0,00257 (0,000379)	0,000486 (0,00200)	0,000304 (0,000299)	0,00758*** (0,00172)
e_presecolar	9,53e-06 (0,00123)	5,98e-06 (0,00132)	9,53e-06 (0,000271)	5,52e-06 (0,000293)	9,63e-09 (0,000109)	5,38e-06 (0,000123)	-1,30e-06 (5,63e-05)	-1,30e-06 (1,00e-06)	1,91e-07 (5,82e-05)	4,86e-05 (1,09e-07)	-2,38e-07 (0,000130)	1,51e-06 (0,000146)	-0,00130 (0,00293)	0,0187 (0,0385)
e_fundamental	-7,461*** (0,372)	-9,967*** (1,049)	-9,927*** (0,402)	-10,209*** (1,354)	-11,933*** (1,354)	-10,209*** (1,354)	-11,933*** (1,354)	-10,209*** (1,354)	-11,933*** (1,354)	-10,209*** (1,354)	-11,933*** (1,354)	-10,209*** (1,354)	-11,933*** (1,354)	-10,209*** (1,354)
alhho	-1,177*** (0,00751)	-1,786*** (0,00917)	-1,107*** (0,0103)	-1,664*** (0,00509)	-1,551*** (0,00509)	-1,664*** (0,00521)	-1,551*** (0,00521)	-1,664*** (0,00521)	-1,551*** (0,00521)	-1,664*** (0,00521)	-1,551*** (0,00521)	-1,664*** (0,00521)	-1,551*** (0,00521)	-1,664*** (0,00521)
Insigma	0,019*** (0,0785)	1,242*** (0,255)	0,0999*** (0,0569)	0,00496 (0,339)	1,044*** (0,0862)	-0,133 (0,261)	0,864*** (0,0811)	0,383* (0,228)	1,579*** (0,0443)	-0,739*** (0,209)	0,976*** (0,0697)	0,127 (0,368)	0,936*** (0,0530)	2,167*** (0,309)
Constante														

Fonte: Elaboração própria.

\*\*\*p &lt; 0,01; \*\*p &lt; 0,05; \*p &lt; 0,1.

Tabela 29 – Modelo 2: Estimaciones Heckman MLE: 2010 e 2011

Variáveis	2010.1.2		2010.2.2		2011.1.2		2011.1.3		2011.2.1		2011.2.2		2011.2.3	
	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d
menino	-0,0623*** (0,00710)	-0,127*** (0,0215)	-0,0182*** (0,00535)	0,0826*** (0,0247)	-0,0242*** (0,00514)	0,0929*** (0,0197)	-0,0104* (0,00574)	0,0316* (0,0176)	-0,0157*** (0,00250)	0,0915*** (0,0146)	-0,0177*** (0,00388)	0,119*** (0,0261)	-0,0153*** (0,00357)	0,0694*** (0,0162)
branco	0,0703*** (0,0161)	0,146*** (0,0467)	0,0317*** (0,0130)	-0,144*** (0,0602)	0,0158 (0,0127)	-0,0607 (0,0486)	0,0325** (0,0131)	-0,0998** (0,0402)	0,0183*** (0,00594)	-0,107*** (0,0347)	0,00551 (0,00919)	-0,0371 (0,0618)	0,00211 (0,00825)	-0,00958 (0,0375)
rural	0,0557*** (0,00911)	0,343*** (0,0289)	-0,00981 (0,00621)	0,0445 (0,0286)	-0,0172*** (0,00564)	0,0659*** (0,0216)	-0,0465*** (0,00707)	0,143*** (0,0215)	0,00616* (0,00336)	-0,0359* (0,0196)	0,00947** (0,00440)	-0,0637** (0,0296)	0,00494 (0,00478)	-0,0224 (0,0217)
bolsa	-0,0762* (0,0410)	-0,603*** (0,101)	0,0614*** (0,0156)	-0,279*** (0,0725)	0,0605** (0,0267)	-0,232** (0,102)	0,230*** (0,0282)	-0,705*** (0,0865)	0,0786*** (0,0224)	-0,459*** (0,131)	-0,00961 (0,0239)	0,0646 (0,160)	0,127*** (0,0239)	-0,576*** (0,109)
lnpib_pc	-0,0280*** (0,00987)	-0,180*** (0,0290)	0,0186*** (0,00722)	-0,0848** (0,0332)	-0,00977 (0,00704)	0,0374 (0,0270)	-0,0122 (0,00803)	0,0370 (0,0353)	-0,00529 (0,0206)	0,0309 (0,0206)	0,0105** (0,00533)	-0,0708** (0,0359)	0,00318 (0,0487)	-0,0145 (0,0221)
e_biblio	0,0200** (0,00990)	0,228*** (0,0276)	-0,00765 (0,00720)	0,0321 (0,0331)	-0,00103 (0,00683)	0,00395 (0,0262)	-0,00229 (0,00720)	0,00704 (0,00221)	0,00495 (0,00312)	-0,0289 (0,0182)	0,0121** (0,00531)	-0,0813** (0,0358)	0,00432 (0,00443)	-0,0197 (0,0202)
i_funcionarios_razao	-0,00419*** (0,000760)	-0,0115*** (0,00238)	-0,000133 (0,000587)	0,000598 (0,00271)	-0,00199*** (0,000581)	0,00764*** (0,00223)	0,00266*** (0,000772)	-0,00816*** (0,00236)	-0,00186*** (0,000303)	0,0109*** (0,00177)	-0,00209*** (0,000455)	0,0141*** (0,00306)	-0,000596 (0,000498)	0,00271 (0,00227)
prof_idade	0,00100 (0,00144)	-0,0447*** (0,00427)	-0,000707 (0,00102)	0,00321 (0,00470)	-0,00164* (0,000968)	0,00628* (0,00371)	0,00184 (0,00116)	-0,00578 (0,00359)	-0,00122 (0,000489)	0,000709 (0,00285)	0,00117* (0,000707)	-0,00787* (0,00476)	0,00116 (0,000713)	-0,00528 (0,00324)
alunos_turno	0,00116* (0,000603)	-0,00930*** (0,00189)	0,000768* (0,000437)	-0,00349* (0,00201)	0,00181*** (0,000485)	-0,000695*** (0,00186)	0,000491 (0,000489)	0,00151 (0,00150)	0,000337 (0,000236)	-0,00196 (0,00138)	0,000334 (0,000341)	3,03e-05 (0,00229)	-0,00113*** (0,000304)	0,00515*** (0,00138)
e_preescolar	0,00796 (0,0299)	-7,44e-06 (0,00729)	3,89e-08 (9,24e-05)	0,00729 (0,00842)	0,000485 (0,000113)	0,000100 (0,000113)	0,000100 (0,000123)	-0,000116 (0,00120)	0,000100 (0,00120)	1,42e-06 (0,000124)	-0,00038 (0,000124)	-1,40e-06 (0,000273)	-0,000304 (0,000333)	0,00796 (0,000390)
e_fundamental	8,53e-06 (0,0699)	-8,29 (5,381)	-0,0182*** (0,00535)	0,0826*** (0,0247)	-0,0242*** (0,00514)	0,0929*** (0,0197)	-0,0104* (0,00574)	0,0316* (0,0176)	-0,0157*** (0,00250)	0,0915*** (0,0146)	-0,0177*** (0,00388)	0,119*** (0,0261)	-0,0153*** (0,00357)	0,0694*** (0,0162)
athro	0,0699 (0,00729)	4,014*** (0,0289)	0,822*** (0,0775)	0,810** (0,358)	1,123*** (0,0764)	-0,472 (0,293)	1,045*** (0,0845)	-0,125 (0,259)	1,038*** (0,0396)	-0,221 (0,231)	0,843*** (0,0571)	1,057*** (0,384)	0,950*** (0,0523)	0,230 (0,238)
Insigma	-1,105*** (0,00729)	-1,514*** (0,0104)	-1,344*** (0,0104)	-1,514*** (0,0104)	-1,344*** (0,00767)	-1,122*** (0,00778)	-1,764*** (0,00565)	-1,764*** (0,00565)	-1,906*** (0,00957)	-1,906*** (0,00957)	-1,906*** (0,00957)	-1,515*** (0,00615)	-1,515*** (0,00615)	-1,515*** (0,00615)
Constante	1,045*** (0,109)	4,014*** (0,326)	0,822*** (0,0775)	0,810** (0,358)	1,123*** (0,0764)	-0,472 (0,293)	1,045*** (0,0845)	-0,125 (0,259)	1,038*** (0,0396)	-0,221 (0,231)	0,843*** (0,0571)	1,057*** (0,384)	0,950*** (0,0523)	0,230 (0,238)

Fonte: Elaboração própria.

\*\*\*p &lt; 0,01; \*\*p &lt; 0,05; \*p &lt; 0,1.

Tabela 30 – Modelo 3: Estimções Heckman MLE: 2008 e 2009

Variáveis	2008.1.2		2008.2.2		2009.1.2		2009.11.3		2009.2.1		2009.2.2		2009.2.3	
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d
menino	-0,0218*** (0,00638)		-2,43e-06 (2,03e-05)		-0,0752*** (0,00616)		-2,23e-06 (1,99e-05)		-4,44e-07 (6,59e-06)		-0,0268*** (0,00380)		-0,0298*** (0,00370)	
branco	0,00255 (0,0167)		9,62e-07 (5,67e-05)		0,0542*** (0,0151)		1,38e-06 (4,47e-05)		2,96e-07 (1,19e-05)		0,00397 (0,0130)		0,0567*** (0,0104)	
rural	0,00260 (0,00755)		1,53e-06 (2,54e-05)		0,0597*** (0,00742)		7,91e-07 (2,39e-05)		4,19e-07 (3,73e-06)		0,0160*** (0,00540)		0,0140*** (0,00489)	
bolsa	-0,00694 (0,0215)		1,01e-06 (9,07e-05)		0,00912 (0,0190)		3,73e-08 (6,79e-05)		1,36e-07 (3,09e-05)		-0,0392*** (0,0153)		0,0227 (0,0139)	
lnppb_pc	0,0249*** (0,00756)		1,37e-07 (2,40e-05)		-0,0149* (0,00808)		4,09e-07 (2,48e-05)		-2,00e-07 (7,82e-06)		0,00627 (0,00657)		0,00139 (0,00477)	
e_bolhio	0,00408 (0,00834)		6,71e-07 (3,07e-05)		0,00751 (0,00817)		1,74e-07 (2,37e-05)		-2,51e-09 (7,40e-06)		-0,000177 (0,00673)		-0,00134 (0,00453)	
l_funcionarios_razaa	-0,00102 (0,000688)		4,31e-08 (2,36e-06)		-0,00304*** (0,000684)		-1,05e-07 (2,74e-06)		-2,63e-08 (6,34e-07)		-0,00306*** (0,000552)		-0,00209*** (0,000497)	
prof_idade	0,000559 (0,00118)		-1,09e-07 (4,53e-06)		-0,00269*** (0,00119)		2,12e-08 (4,16e-06)		-5,96e-08 (1,44e-06)		-0,00282*** (0,000911)		-0,000127 (0,000714)	
alunos_turma	0,000351 (0,000560)		8,90e-09 (1,68e-06)		0,00315*** (0,000535)		1,37e-08 (1,48e-06)		1,72e-08 (5,26e-07)		1,72e-08 (0,000390)		0,000502 (0,000297)	
e_prescolar	-0,0165 (0,0619)		-2,35e-06 (0,000166)		-0,00984 (0,0356)		-1,26e-06 (7,66e-05)		5,40e-07 (3,83e-05)		-0,0809 (0,0531)		-0,0322 (0,0292)	
e_fundamental	0,0207 (0,0635)		3,18e-06 (0,000181)		0,243*** (0,0381)		2,64e-06 (0,000103)		-9,59e-07 (4,57e-05)		0,208*** (0,0579)		0,207*** (0,0352)	
prof_idade_dip	0,000944 (0,0173)		4,20e-07 (5,07e-05)		-0,0364*** (0,00950)		-4,94e-07 (2,15e-05)		2,35e-07 (1,22e-05)		-0,0390*** (0,0144)		-0,0219*** (0,00753)	
e_lab_info	0,0169 (0,0618)		-1,78e-07 (0,000183)		0,0186 (0,0382)		8,24e-07 (8,32e-05)		-8,92e-08 (4,27e-05)		0,0735 (0,00625)		0,0504 (0,0309)	
e_quadra	-0,0216 (0,0384)		1,47e-06 (0,000149)		0,0554 (0,0346)		-1,36e-06 (8,13e-05)		-7,50e-08 (3,78e-05)		0,0724 (0,0505)		-0,0883*** (0,0300)	
atirho	-13,98 (0)		-10,45*** (1,026)		0,0371 (0,0578)		-10,64*** (0,580)		-12,41*** (1,771)		0,00953 (0,0916)		0,0299 (0,0440)	
hsigma	-1,180*** (0,00797)		-1,78e-03 (0,00927)		-1,153*** (0,000687)		-1,034*** (0,000884)		-1,545*** (0,00508)		-1,678*** (0,000890)		-1,483*** (0,00578)	
Constante	0,708*** (0,0860)	0,171 (0,207)	0,992*** (0,00216)	0,0471*** (0,0126)	1,041*** (0,0884)	0,115 (0,115)	1,002*** (0,00205)	-0,00631 (0,00856)	0,977*** (0,00148)	0,107*** (0,00088)	1,015*** (0,0719)	1,428*** (0,117)	0,934*** (0,0550)	0,878*** (0,0858)

Fonte: Elaboração própria.

\*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

Tabela 31 – Modelo 3: Estimações Heckman MLE: 2010 e 2011

Variáveis	2010.1.2		2010.2.2		2011.1.2		2011.1.3		2011.2.1		2011.2.2		2011.2.3	
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d
metano	-2,60e-06 (4,52e-05)	-0,0275*** (0,00565)	-0,0533*** (0,00534)	-2,16e-06 (4,67e-05)	-0,0196*** (0,00258)	-0,0167*** (0,00258)	-0,0238*** (0,00375)	-0,0275*** (0,00565)	-0,0196*** (0,00258)	-0,0167*** (0,00258)	-0,0238*** (0,00375)	-0,0275*** (0,00565)	-0,0196*** (0,00258)	-0,0167*** (0,00258)
branco	2,64e-06 (9,38e-05)	0,0653*** (0,0144)	0,0176 (0,0132)	1,72e-06 (9,82e-05)	0,0245*** (0,00619)	0,0267 (0,00941)	0,0151* (0,00884)	0,0653*** (0,0144)	0,0245*** (0,00619)	0,0267 (0,00941)	0,0151* (0,00884)	0,0653*** (0,0144)	0,0245*** (0,00619)	0,0267 (0,00941)
rural	3,00e-06 (5,96e-05)	0,0243*** (0,00633)	0,0145** (0,00573)	7,48e-07 (5,47e-05)	0,0118*** (0,00343)	0,0141*** (0,00450)	0,0185*** (0,00493)	0,0243*** (0,00633)	0,0145** (0,00573)	0,0141*** (0,00450)	0,0185*** (0,00493)	0,0243*** (0,00633)	0,0145** (0,00573)	0,0141*** (0,00450)
bolsa	-2,76e-06 (0,000234)	0,0370** (0,0172)	-0,0649*** (0,0328)	-1,97e-06 (0,000279)	-0,0738** (0,0310)	-0,0738** (0,0290)	-0,0783** (0,0332)	0,0370** (0,0172)	-0,0649*** (0,0328)	-0,0738** (0,0310)	-0,0738** (0,0290)	-0,0783** (0,0332)	0,0370** (0,0172)	-0,0649*** (0,0328)
lnpib_pc	-5,83e-07 (6,39e-05)	0,00347 (0,00796)	-0,0309*** (0,00732)	-1,80e-06 (6,48e-05)	-0,00696* (0,00364)	-0,00696* (0,00347)	-0,009923 (0,00510)	-5,83e-07 (6,39e-05)	0,00347 (0,00796)	-0,0309*** (0,00732)	-0,00696* (0,00364)	-0,009923 (0,00510)	-0,00696* (0,00347)	-0,009923 (0,00510)
e_biblio	7,60e-07 (6,11e-05)	0,0304*** (0,00792)	0,00627 (0,00710)	2,04e-07 (5,82e-05)	0,00627 (0,00710)	0,00474 (0,00547)	0,00197 (0,00466)	7,60e-07 (6,11e-05)	0,0304*** (0,00792)	0,00627 (0,00710)	0,00474 (0,00547)	0,00197 (0,00466)	0,00627 (0,00710)	0,00474 (0,00547)
i_funcionarios_razao	-1,69e-07 (5,13e-06)	-0,00337*** (0,000650)	-0,00475*** (0,000623)	-1,05e-07 (6,71e-06)	-0,00228*** (0,000319)	-0,00228*** (0,000319)	-0,00363*** (0,000475)	-1,69e-07 (5,13e-06)	-0,00337*** (0,000650)	-0,00475*** (0,000623)	-0,00228*** (0,000319)	-0,00363*** (0,000475)	-0,00337*** (0,000650)	-0,00475*** (0,000623)
prof_idade	3,22e-07 (1,04e-05)	-0,00311*** (0,00108)	-0,00129 (0,00100)	-2,35e-07 (1,01e-05)	-0,00131*** (0,000504)	-0,00131*** (0,000504)	-0,00152** (0,000605)	3,22e-07 (1,04e-05)	-0,00311*** (0,00108)	-0,00129 (0,00100)	-0,00131*** (0,000504)	-0,00152** (0,000605)	-0,00311*** (0,00108)	-0,00129 (0,00100)
alunos_turma	6,64e-08 (4,04e-06)	-0,000173 (0,000468)	0,00221*** (0,000504)	-5,91e-08 (3,93e-06)	0,00207 (0,000244)	0,000307 (0,000244)	-0,00140*** (0,000350)	6,64e-08 (4,04e-06)	-0,000173 (0,000468)	0,00221*** (0,000504)	0,00207 (0,000244)	-0,00140*** (0,000350)	-0,00140*** (0,000350)	-0,00140*** (0,000350)
e_prescolar	-3,98e-07 (0,000171)	-0,0891*** (0,0407)	0,0175 (0,0367)	-3,30e-06 (0,000207)	0,0711*** (0,0324)	0,0711*** (0,0324)	0,0580* (0,0309)	-3,98e-07 (0,000171)	-0,0891*** (0,0407)	0,0175 (0,0367)	0,0711*** (0,0324)	0,0580* (0,0309)	0,0711*** (0,0324)	0,0580* (0,0309)
e_fundamental	1,60e-07 (0,000214)	0,207*** (0,0435)	0,111*** (0,0409)	7,31e-06 (0,000287)	0,0909** (0,000287)	0,0909** (0,000287)	0,117*** (0,0402)	1,60e-07 (0,000214)	0,207*** (0,0435)	0,111*** (0,0409)	0,0909** (0,000287)	0,117*** (0,0402)	0,0909** (0,000287)	0,117*** (0,0402)
prof_idade_dp	-2,61e-06 (5,29e-05)	-0,109*** (0,0108)	-0,0580*** (0,0102)	-0,109*** (0,0102)	1,47e-07 (6,07e-05)	1,47e-07 (6,07e-05)	-0,0379*** (0,00827)	-2,61e-06 (5,29e-05)	-0,109*** (0,0108)	-0,0580*** (0,0102)	-0,0379*** (0,00827)	-0,0379*** (0,00827)	-0,0379*** (0,00827)	-0,0379*** (0,00827)
e_lab_info	3,27e-06 (0,000188)	-0,00530 (0,0468)	-0,0253 (0,0419)	-2,69e-06 (0,000222)	-0,0464 (0,0375)	-0,0464 (0,0375)	-0,0363 (0,0358)	3,27e-06 (0,000188)	-0,00530 (0,0468)	-0,0253 (0,0419)	-0,0464 (0,0375)	-0,0363 (0,0358)	-0,0363 (0,0358)	-0,0363 (0,0358)
e_quadra	1,81e-07 (0,000165)	-0,0197 (0,0386)	0,0109 (0,0568)	4,37e-07 (0,000217)	-0,157*** (0,0287)	-0,157*** (0,0287)	-0,0757** (0,0317)	1,81e-07 (0,000165)	-0,0197 (0,0386)	0,0109 (0,0568)	-0,157*** (0,0287)	-0,0757** (0,0317)	-0,0757** (0,0317)	-0,0757** (0,0317)
athro	-10,08*** (0,759)	0,0450 (0,0539)	0,0283 (0,0590)	-9,937*** (0,766)	0,0262 (0,0369)	0,0262 (0,0369)	0,0232 (0,0504)	-10,08*** (0,759)	0,0450 (0,0539)	0,0283 (0,0590)	0,0232 (0,0504)	0,0232 (0,0504)	0,0232 (0,0504)	0,0232 (0,0504)
Insgma	-1,066*** (0,00741)	-1,57*** (0,00952)	-1,376*** (0,00739)	-1,139*** (0,00722)	-1,776*** (0,00532)	-1,776*** (0,00532)	-1,538*** (0,0620)	-1,066*** (0,00741)	-1,57*** (0,00952)	-1,376*** (0,00739)	-1,139*** (0,00722)	-1,538*** (0,0620)	-1,538*** (0,0620)	-1,538*** (0,0620)
Constante	1,028*** (0,00339)	-0,0750*** (0,00895)	1,277*** (0,0796)	1,023*** (0,00297)	1,082*** (0,0407)	1,082*** (0,0407)	1,069*** (0,0545)	1,028*** (0,00339)	-0,0750*** (0,00895)	1,277*** (0,0796)	1,023*** (0,00297)	1,082*** (0,0407)	1,069*** (0,0545)	1,117*** (0,0973)

Fonte: Elaboração própria.

\*\*\*p &lt; 0,01; \*\*p &lt; 0,05; \*p &lt; 0,1.

Tabela 32 – Modelo 4: Estimáveis Heckman MLE: 2008 e 2009

Variáveis	2008.1.2		2008.2.2		2009.1.2		2009.1.3		2009.2.1		2009.2.2		2009.2.3		
	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	aprov	d	
menino	-2,93e-06 (2,80e-05)		-1,75e-06 (1,29e-05)		-0,07572*** (0,00616)		-3,91e-05 (0,00041)		-1,11e-06 (6,86e-06)		-1,34e-06 (1,76e-05)		-0,0298*** (0,00370)		
branco	2,29e-06 (6,11e-05)		6,90e-07 (3,54e-05)		0,0541*** (0,0151)		2,51e-05 (0,00177)		7,83e-07 (1,28e-05)		1,47e-07 (4,20e-05)		0,0367*** (0,0104)		
rural	2,32e-06 (3,78e-05)		1,09e-06 (1,62e-05)		0,0597*** (0,00741)		1,39e-05 (0,00109)		9,81e-07 (1,38e-05)		1,27e-06 (2,62e-05)		0,0140*** (0,00489)		
bolso	-3,02e-06 (8,00e-05)		7,04e-07 (5,79e-05)		0,00911 (0,0190)		3,14e-07 (0,00284)		3,28e-07 (3,17e-05)		-1,70e-06 (4,30e-05)		0,0227 (0,0139)		
lnpb_pc	1,72e-06 (3,44e-05)		9,81e-08 (1,52e-05)		-0,0148* (0,00809)		7,39e-06 (0,00103)		-5,05e-07 (8,31e-06)		1,16e-07 (2,27e-05)		-0,000282 (0,00477)		
e_hbhfio	1,59e-07 (3,71e-05)		-4,86e-07 (1,93e-05)		0,00750 (0,00816)		3,12e-06 (0,000982)		-2,74e-09 (7,85e-06)		-1,11e-07 (2,32e-05)		-0,00136 (0,00453)		
i_funcionarios_razao	-1,02e-07 (3,14e-06)		3,15e-08 (1,50e-06)		-0,00300*** (0,000691)		-1,89e-06 (0,000112)		-6,89e-08 (6,81e-07)		-1,41e-07 (1,89e-06)		-0,00209*** (0,000498)		
prof_idade	-1,56e-08 (6,42e-06)		-7,92e-08 (2,87e-06)		-0,00264** (0,00120)		3,66e-07 (0,000172)		-1,50e-07 (1,52e-06)		-2,14e-07 (3,98e-06)		-0,000114 (0,000715)		
alunos_turma	2,86e-08 (2,54e-06)		6,12e-09 (1,06e-06)		0,00314*** (0,000535)		2,36e-07 (6,16e-05)		4,42e-08 (3,59e-07)		3,41e-08 (1,46e-06)		0,000301 (0,000297)		
e_presscolar		-2,08e-06 (0,000138)		-1,67e-06 (0,000105)		-0,00913 (0,0358)		-2,28e-05 (0,00226)		1,33e-06 (4,07e-05)		1,28e-06 (0,000125)		-0,0309 (0,0292)	
e_fundamental		3,11e-06 (0,000151)		2,26e-06 (0,000114)		0,242*** (0,0382)		4,91e-05 (0,00428)		-2,47e-06 (4,88e-05)		-2,45e-06 (0,000142)		0,206*** (0,0352)	
prof_idade_dfp		-5,10e-07 (4,00e-05)		2,96e-07 (3,21e-05)		-0,0358*** (0,000909)		-8,38e-06 (0,000320)		6,01e-07 (1,30e-05)		9,03e-07 (3,95e-05)		-0,0215*** (0,000754)	
e_lab_info		3,94e-06 (0,000138)		4,34e-07 (0,000115)		0,0194 (0,0384)		1,46e-05 (0,00351)		-2,20e-07 (4,54e-05)		-3,15e-07 (0,000147)		0,0488 (0,0310)	
e_quadra		-1,36e-06 (0,000123)		9,42e-05 (2,39e-05)		0,0547 (0,0351)		-2,42e-05 (0,00340)		-1,50e-07 (4,02e-05)		1,21e-06 (0,000121)		-0,0851*** (0,0301)	
faltas_completas		-6,18e-07 (3,18e-05)		-2,96e-07 (2,35e-05)		-0,0178*** (0,00657)		5,48e-06 (0,000509)		-1,34e-07 (9,64e-06)		-4,23e-07 (2,67e-05)		-0,0196*** (0,00500)	
alinho		-10,84*** (0,961)		-10,84*** (0,961)		0,00577 (0,109)		-7,86e*** (1,483)		-11,49*** (0,757)		-10,99*** (1,356)		0,0134 (0,0681)	
Insgima		-1,192*** (0,00724)		-1,192*** (0,00724)		-1,514*** (0,00681)		-1,048*** (0,00709)		-1,545*** (0,00599)		-1,658*** (0,00916)		-1,483*** (0,00577)	
Constante		0,985*** (0,00287)		0,992*** (0,00258)		1,065*** (0,115)		1,009*** (0,0121)		0,977*** (0,00178)		0,984*** (0,00240)		0,935*** (0,0531)	
		0,0483*** (0,00931)		0,0487*** (0,0150)		1,009*** (0,115)		-0,0209 (0,0140)		0,107*** (0,00798)		0,0822*** (0,0124)		0,883*** (0,0859)	

Fonte: Elaboração própria.

\*\*\* $p < 0,01$ ; \*\* $p < 0,05$ ; \* $p < 0,1$ .

Tabela 33 – Modelo 4: Estimções Heckman MLE: 2010 e 2011

Variáveis	2010.1.2		2010.2.2		2011.1.2		2011.1.3		2011.2.1		2011.2.2		2011.2.3		
	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	aprov.	d	
menino	-2,80e-06 (3,48e-05)		-0,02759*** (0,005965)		-1,42e-06 (2,01e-05)		-2,16e-06 (2,46e-05)		-0,0196*** (0,00258)		-0,0167*** (0,00399)		-0,0238*** (0,00375)		
branco	3,70e-06 (7,21e-05)		0,0653*** (0,0144)		5,78e-07 (4,66e-05)		2,41e-06 (5,02e-05)		0,0245*** (0,00619)		0,00267 (0,00941)		0,0151* (0,00883)		
rural	3,86e-06 (4,52e-05)		0,0243*** (0,00633)		7,64e-07 (2,33e-05)		1,08e-06 (2,92e-05)		0,0118*** (0,00343)		0,0141*** (0,00450)		0,0185*** (0,00493)		
bolsa	-3,51e-06 (0,000184)		0,0370*** (0,0172)		-2,48e-06 (9,13e-05)		-1,96e-06 (0,000148)		-0,0738*** (0,0310)		-0,0969*** (0,0290)		-0,0783*** (0,0332)		
lnpb_pc	-4,75e-07 (4,91e-05)		0,00350 (0,00796)		-1,28e-06 (2,85e-05)		-2,07e-06 (3,43e-05)		-0,00695* (0,00364)		0,00506 (0,00547)		-0,000933 (0,00510)		
e_biblio	1,41e-06 (4,64e-05)		0,0304*** (0,00792)		-9,74e-08 (2,79e-05)		6,77e-07 (3,02e-05)		0,00474 (0,00322)		0,0128*** (0,00547)		0,00197 (0,00466)		
i_funcionarios_razao	-1,59e-07 (3,99e-06)		-0,00336*** (0,000650)		-1,85e-07 (2,29e-06)		-8,84e-08 (3,56e-06)		-0,00228*** (0,000319)		-0,00269*** (0,000475)		-0,00364*** (0,000552)		
prof_idade	4,68e-07 (7,95e-06)		-0,00310*** (0,00108)		-4,61e-08 (4,57e-06)		-2,65e-07 (5,32e-06)		-0,00131*** (0,000504)		0,000609 (0,000733)		-0,00152*** (0,000748)		
alunos_turma	9,76e-08 (3,08e-06)		-0,000173 (0,000468)		8,82e-08 (1,84e-06)		-4,92e-08 (2,06e-06)		0,000307 (0,000244)		0,000137 (0,000350)		-0,00140*** (0,000319)		
e_presscolar		3,42e-07 (0,000130)		-0,0886** (0,0407)		4,77e-07 (0,000105)		-2,80e-06 (0,000108)		0,0701*** (0,0324)		-0,0319 (0,0556)		0,0583* (0,0309)	
e_fundamental		3,54e-06 (0,000162)		0,208*** (0,0435)		-3,54e-06 (0,000132)		9,83e-06 (0,000147)		-0,0183 (0,0397)		0,0559 (0,0625)		0,118*** (0,0402)	
prof_idade_dp		-3,63e-06 (4,06e-05)		-0,109*** (0,0108)		-4,28e-07 (3,43e-05)		3,47e-07 (3,18e-05)		-0,0368*** (0,00888)		-0,0641*** (0,0147)		-0,0378*** (0,00827)	
e_lab_info		6,53e-06 (0,000141)		-0,00600 (0,0468)		-4,74e-07 (0,000118)		-1,45e-06 (0,000114)		0,00447 (0,0375)		0,0196 (0,0656)		-0,0374 (0,0358)	
e_quadra		6,45e-07 (0,000128)		-0,0195 (0,0386)		5,63e-07 (0,000105)		9,72e-07 (0,000115)		-0,157*** (0,0287)		-0,160*** (0,0514)		-0,0760** (0,0318)	
faltas_completas		-1,17e-06 (3,10e-05)		-0,0132 (0,0105)		-6,17e-07 (2,32e-05)		-5,99e-07 (2,00e-05)		-0,0176*** (0,00597)		-0,0105 (0,0125)		-0,0122** (0,00514)	
athro	9,817*** (0,449)		0,0408 (0,0538)		-10,71*** (0,877)		-9,967*** (0,417)		0,0197 (0,0421)		0,0226 (0,0659)		0,0282 (0,0458)		
Insigma	-1,008*** (0,00761)		-1,573*** (0,00951)		-3,141*** (0,00756)		-1,037*** (0,00878)		-1,176*** (0,00532)		-1,917*** (0,0938)		-1,538*** (0,0620)		
Constante	1,028*** (0,000339)		0,983*** (0,0843)		0,981*** (0,00265)		0,0726*** (0,00993)		1,082*** (0,00344)		1,608*** (0,108)		1,893*** (0,185)		
			1,634*** (0,136)		0,981*** (0,00265)		0,0726*** (0,00993)		1,082*** (0,00344)		1,608*** (0,108)		1,893*** (0,185)		

Fonte: Elaboração própria.

\*\*\*p &lt; 0,01; \*\*p &lt; 0,05; \* p &lt; 0,1.