



DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

Qual é o efeito do nível médio educacional e renda per capita dos domicílios sobre o cumprimento da condicionalidade de educação do Programa Bolsa Família?

Isabel Costa de Carvalho
Matrícula: 1412961

Orientador: José Márcio Camargo

Rio de Janeiro
Junho de 2018

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

Qual é o efeito do nível médio educacional e renda per capita dos domicílios sobre o cumprimento da condicionalidade de educação do Programa Bolsa Família?

Isabel Costa de Carvalho
Matrícula: 1412961

Orientador: José Márcio Camargo

Rio de Janeiro
Junho de 2018

“Declaro que o presente trabalho é de minha autoria e que não recorri para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor”

“As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor”

Dedico esta monografia aos meus pais e às minhas irmãs que me apoiaram incansavelmente durante essa longa trajetória. Ao meu irmão, Tuca, por sua paciência, conselhos e por ser o espelho de tudo que aspiro ser.

Às minhas amigas, Carol, Rebeca, Juliana, Luisa e Priscilla por estarem comigo em todos os momentos desta faculdade, fossem eles felizes, tristes, desafiadores ou vitoriosos. Sem vocês nada teria sido possível e nada teria tido a mesma graça. Desejo a vocês todo sucesso do mundo.

Por fim, agradeço imensamente ao meu orientador, José Márcio Camargo, por sua contribuição e auxílio durante a elaboração deste trabalho.

Sumário

1. Introdução.....	9
2. Revisão Bibliográfica.....	11
3. Construção e fonte de dados.....	17
4. Análise descritiva.....	23
4.1. Análise Gráfica.....	26
5. Metodologia.....	35
5.1. Estratégia de previsão dos dados – Método de Holt com 2 Parâmetros.....	40
6. Discussão acerca das variáveis.....	45
7. Tabelas de Resultados.....	49
8. Discussão acerca dos resultados.....	61
8.1. Interpretações econômicas.....	67
9. Conclusão.....	71
10. Limitações.....	72
Referências Bibliográficas.....	74

Índice de Tabelas

Tabelas

<i>Tabela 1</i> – Estatísticas Descritivas.....	21
<i>Tabela 2</i> – Resultado das previsões obtidas através do Método de Holt com 2 parâmetros.....	39
<i>Tabela 3</i> – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação.....	49
<i>Tabela 4</i> – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem DF).....	49
<i>Tabela 5</i> – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem intercepto).....	50
<i>Tabela 6</i> – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem DF e sem intercepto).....	50
<i>Tabela 7</i> – Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e adequação da formação docente no cumprimento da condicionalidade de educação (Sem o DF).....	51
<i>Tabela 8</i> – Efeito do rendimento domiciliar, escolaridade, comprometimento da renda com transporte público e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF).....	51
<i>Tabela 9</i> – Efeito do rendimento domiciliar, escolaridade, taxa de violência e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF)....	52
<i>Tabela 10</i> – Efeito do rendimento domiciliar, escolaridade e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF).....	53
<i>Tabela 11</i> - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 1 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal).....	53
<i>Tabela 12</i> - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 2 no cumprimento da condicionalidade de educação (sem o DF).....	54

Tabela 13 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 3 no cumprimento da condicionalidade de educação (sem o DF).....55

Tabela 14 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 4 no cumprimento da condicionalidade de educação (sem o DF).....55

Tabela 15 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 5 no cumprimento da condicionalidade de educação (sem o DF).....56

Tabela 16 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 1 no cumprimento da condicionalidade de educação (sem o DF).....57

Tabela 17 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 2 no cumprimento da condicionalidade de educação (sem o DF).....57

Tabela 18 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 3 no cumprimento da condicionalidade de educação (sem o DF).....58

Tabela 19 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 4 no cumprimento da condicionalidade de educação (sem o DF).....59

Tabela 20 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 5 no cumprimento da condicionalidade de educação (sem o DF).....59

Índice de Gráficos e Figuras

Figuras

<i>Figura 1</i> – Exemplar de divulgação dos dados do Sistema Presença para o mês/ano de referência de novembro/2017.....	20
<i>Figura 2</i> – Cumprimento de condicionalidade de educação do Programa Bolsa Família versus rendimento mensal domiciliar per capita.....	44
<i>Figura 3</i> – Cumprimento de condicionalidade de educação do Programa Bolsa Família versus escolaridade média (por anos de estudo).....	44

Gráficos

<i>Gráfico 1</i> – Cumprimento da condicionalidade de educação versus rendimento mensal domiciliar per capita (R\$).....	27
<i>Gráfico 2</i> – Cumprimento da condicionalidade de educação versus escolaridade média (em anos de estudo).....	28
<i>Gráfico 3</i> – Cumprimento da condicionalidade de educação versus taxa de latrocínios a cada 100 mil habitantes.....	29
<i>Gráfico 4</i> – Cumprimento da condicionalidade de educação versus taxa comprometimento da renda com a tarifa de transporte público.....	29
<i>Gráfico 5</i> – Cumprimento da condicionalidade de educação versus percentual de docentes do Grupo 1.....	30
<i>Gráfico 6</i> – Cumprimento da condicionalidade de educação versus percentual de docentes do Grupo 2.....	30
<i>Gráfico 7</i> – Cumprimento da condicionalidade de educação versus percentual de docentes do Grupo 3.....	31
<i>Gráfico 8</i> – Cumprimento da condicionalidade de educação versus percentual de docentes do Grupo 4.....	31
<i>Gráfico 9</i> – Cumprimento da condicionalidade de educação versus percentual de docentes do Grupo 5.....	32

1. Introdução

Desde as colaborações de Cangussu, Salvato e Nakabashi [Cangussu, Salvato e Nakabashi (2010)] sobre a importância da acumulação de capital humano, passando pelos efeitos da pobreza intergeracional estudados por Camargo [Camargo e Almeida (1994)], Emerson e Souza [Emerson e Souza (2003)], até às contribuições de Bestagil [Bestagil et al (2014)] e Jones (2015) acerca do papel dos incentivos na economia, é conferida demasiada relevância ao investimento em educação. A importância da acumulação de capital humano cresce ainda mais para a parcela mais carente da população que obtém retornos maiores em resposta à investimentos em educação. A consequência do comprometimento com a acumulação de capital é a interrupção da pobreza intergeracional no futuro, além do aumento do bem-estar social gerado como uma externalidade em decorrência do aumento do nível de escolaridade da população.

Os dois principais efeitos estudados são os efeitos gerados pelas diferenças no rendimento nominal mensal domiciliar per capita e pela escolaridade média em anos de estudo da população, das 27 unidades da federação, na taxa de cumprimento da condicionalidade do PBF. Para a taxa cumprimento da condicionalidade, serão usados os dados disponibilizados pelo Ministério do Desenvolvimento Social que oferecem dados mensais desde 2014, cujo método para agregação dos dados consistirá em uma média anual da taxa apresentada mensalmente para cada estado. Além das duas variáveis explicativas já mencionadas, serão testados os efeitos sobre a condicionalidade, da taxa de latrocínios dos estados (para representar a violência nos estados), da qualidade do corpo docente das escolas públicas (donde a maioria dos beneficiários do programa se dirigem e que será usada como *proxy* do nível de motivação por parte dos alunos) e do comprometimento da renda de um salário mínimo com as tarifas de transporte público.

Assim, este trabalho busca investigar as causas que explicam as variações, sejam elas positivas ou negativas, no cumprimento de condicionalidade de educação do Programa Bolsa Família por parte dos beneficiários. A metodologia utilizada será baseada em uma regressão de Mínimos Quadrados Ordinários (MQO) que busca investigar os efeitos das variáveis de interesse na variável resposta. Além do mais, na segunda parte da metodologia, serão rodadas regressões para investigar como mudanças

na composição dos grupos docentes tem participação nas variações da taxa de cumprimento da condicionalidade, tendo em conta que as diferenças entre os grupos são as diferenças de qualidade de formação dos professores. O objetivo é identificar se uma mobilidade de professores de um grupo menos qualificado para um grupo mais qualificado afeta positivamente a motivação dos alunos, sendo a variável grupo docente uma proxy para a variável “motivação dos beneficiários do programa”.

Além do mais, este trabalho aborda a questão de incentivos econômicos de forma a ilustrar como uma parte importante dos programas de transferência condicional de renda diz respeito à formulação e propagação correta destes incentivos. O trabalho visa explorar como uma via de comunicação clara entre os formuladores de política econômica e as famílias beneficiárias do programa pode ser essencial para o bom funcionamento do programa e para garantia de resultados positivos e sustentáveis. Assim, de forma geral, o trabalho pretende investigar quais os principais motivos que levam os pais a perceberem um aumento do custo de oportunidade de manterem seus filhos na escola, gerando assim, o incentivo de estimularem as crianças a irem em busca de uma alternativa que geraria um bem-estar potencialmente maior no curto prazo.

2. Revisão Bibliográfica

Uma das questões mais importantes que esse trabalho busca enfatizar é a importância da acumulação de capital humano. Através da acumulação de capital humano a população de um país pode se tornar mais produtiva e qualificada, melhorando a empregabilidade e conseqüentemente contribuindo para o crescimento econômico do país. Em especial, para o caso cujo essa monografia busca investigar, a acumulação de capital humano entre beneficiários dos programas de transferência de renda significa uma futura independência dessas mesmas pessoas do benefício monetário oferecido pelo programa, qualquer que seja ele. Dito isso, é inegável a importância do capital humano como elemento de autodestruição de tais programas de transferência de renda oferecidos pelo governo. A literatura acerca da acumulação de capital humano é extensa e em sua grande parte, defensora de sua importância como mecanismo de crescimento econômico.

Para investigar o efeito do capital humano como fator de produção, Cangussu, Salvato e Nakabashi (2010)¹ buscam separar a importância do capital humano na determinação de renda por vias diretas e indiretas. Os autores encontram que, por via direta, a acumulação de capital humano melhora a produtividade marginal dos trabalhadores, aumentando, proporcionalmente, sua remuneração e também aumentando o produto da economia, *ceteris paribus*. A incorporação do capital humano como determinante direto do produto também foi defendida por Weil, Mankiw e Romer (1992)². Por sua vez, o efeito indireto de acordo com Cangussu et al deriva da influência do capital humano na criação e na difusão de novas tecnologias. Isso é, quanto maior o nível de capital humano, maior a probabilidade de novas inovações e maior a facilidade com que essas novas tecnologias se difundem pela sociedade. Sendo tecnologia um dos determinantes de produtividade (o outro é eficiência), um aumento de tecnologia, dado um nível de eficiência, aumenta a produtividade e conseqüentemente o produto, gerando crescimento econômico. O efeito indireto do capital humano também foi enfatizado por Lucas (1988)³ e Aghion e Howitt (1992)⁴.

¹ Uma análise do capital humano sobre o nível de renda dos estados brasileiros: MRW *versus* Mincer

² A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*

³ The mechanics of economic development. *Journal of Monetary Economics*

⁴ A model of growth through creative destruction. *Econometrica*

No campo dos experimentos, Nakabashi e Salvato (2007)⁵ comprovam o efeito indireto do capital humano sobre o produto através de uma regressão que estima o efeito de diversos fatores de produção sobre o PIB *per capita*. Em uma regressão de efeitos fixos (i.e. controlando as especificidades de cada estado com dummies), Nakabashi et al encontraram uma diminuição do coeficiente do capital humano. Dito isso, os autores comprovam que, como o capital humano afeta o PIB *per capita* indiretamente, por via da tecnologia, com a introdução dos efeitos fixos, o impacto direto do capital humano sobre o nível de PIB *per capita* se reduz.

Ainda na esfera dos experimentos, Lau, Jamison, Liu e Rivkin (1993)⁶, mostram que: um ano adicional de escolaridade dos trabalhadores brasileiros tem um impacto positivo de aproximadamente 20% sobre a renda. Visto que os experimentos foram executados em nível estadual, Lau et al encontraram que anos de estudos dos trabalhadores brasileiros tem um impacto extraordinário sobre o nível de renda dos estados como um todo. Enfatizando mais uma vez, a importância do capital humano na determinação do crescimento econômico de um país.

Dito isso, Mincer (1974)⁷ mostrou através do Modelo de Mincer que: se o custo de um ano adicional de escolaridade é o custo de oportunidade de frequentar a escola (i.e. emprego como potencial fonte de renda familiar) e se o aumento proporcional de renda trazido por esse ano adicional na escola for constante, então, o log da renda seria correlacionado ao número de anos de escolaridade do indivíduo, e a inclinação dessa relação seria a taxa de retorno do investimento em educação. Graficamente, verificamos que os retornos da educação são ainda maiores para regiões com baixa quantidade de capital humano, retornos esses que se tornam decrescentes à medida que atingimos um nível elevado de capital humano. O resultado enfatiza não só a importância da acumulação de capital humano, mas a maior importância desse efeito dentre a parcela mais carente da população, justificando o valor de incentivos corretos para que tal efeito se propague.

⁵ Human capital quality in the Brazilian states. *Revista Economia*

⁶ Education and economic growth: some cross-country evidence from Brazil. *Journal of Development Economics*

⁷ *Schooling, experience, and earnings*. New York: Columbia University Press

Ademais, vale ressaltar que, o retorno da educação não se concretiza apenas para o indivíduo em questão, mas há também um retorno social que deve ser considerado. A externalidade positiva da educação é abordada por Krueger e Lindahl (2011)⁸ na forma de redução dos níveis de violência e os autores mostram que este tipo de externalidade é mais provável de ser colhida através de um investimento em educação em grupos de indivíduos de classes sociais mais baixas. Solidificando, novamente, importância do capital humano no bem-estar da sociedade e no crescimento econômico do país, ainda mais para indivíduos com baixos níveis de educação, validando a importância dos incentivos gerados pelos desenhos dos programas de transferência condicional de renda em questão.

Adentrando a literatura com relação ao Bolsa Família, iniciamos com Camargo e Almeida (1994)⁹. Os autores mostram que, se aceitamos a hipótese de que famílias de baixa renda são caracterizadas pelo baixo nível de poupança, então verificamos que essas mesmas famílias tendem a sub-investir em educação em comparação a famílias de renda mais alta. Isso ocorre porque o investimento em educação no período t , diminui o consumo da família no período t , porém só aumenta esse mesmo consumo no período $t+1$. Logo, dependendo da taxa de desconto de uma dada família, pode ser mais custoso poupar e investir em educação agora, visando retornos futuros, à medida que isso implica em uma diminuição do consumo atual (período t). Vale ressaltar que famílias com níveis mais baixos de renda tendem a ter taxas de desconto mais altas, responsáveis por gerar o fenômeno de sub-investimento em educação.

A consequência disso, verificada por Camargo et al, é o que chamamos de ciclo vicioso: crianças vindas de famílias pobres, que tem baixo nível de acumulação de capital humano, também tendem a sub-investir em capital humano, perpetuando a onda de pobreza. Dado o referido, programas de transferência condicional de renda seriam parte de um equilíbrio socialmente ótimo à medida em que a transferência monetária auferida pelos indivíduos serviria como subsídio às famílias, reduzindo o custo marginal de investir em educação.

⁸ Education for growth: why and for whom? *Journal of Economic Literature*

⁹ Human Capital Investment and Poverty

A questão da pobreza intergeracional derivada do ciclo vicioso de sub-investimento em educação também é abordada por Emerson e Souza (2002)¹⁰. Os autores também encontram o resultado de que filhos de pais com menos acumulação de capital humano, são mais propensos à investirem menos em capital humano. Contudo, os autores avançam na questão e encontram que o nível de acumulação de capital humano dos avós não necessariamente afeta o investimento em capital humano de um indivíduo, pelo menos não por vias diretas. Ressalto, novamente, a importância do desenho institucional dos programas de transferência de renda como ferramenta de quebra do ciclo vicioso intergeracional. Falaremos sobre isso mais adiante.

Uma das questões que esse trabalho procura explorar é a influência que o ambiente familiar pode ter no cumprimento ou não da condicionalidade de educação imposta pelo programa. A literatura nos diz que famílias que são detentoras de pouco capital humano ou o que chamamos de “capital escolar” são menos eficazes na hora de transmitir esse mesmo capital escolar para seus filhos.¹¹ Os autores enfatizam que para que os pais sejam engajados e comprometidos com o processo educativo de seus filhos, eles precisam ter o mínimo de capital escolar para que o retorno percebido da educação seja alto o suficiente para que os mesmos estejam empenhados em fazerem seus filhos cumprirem a condicionalidade do PBF.

É de entendimento amplo que, o descumprimento de alguma das condicionalidades impostas pelos programas (isso funcionou de forma análoga para os dois programas) implica em penalidades que vão de uma advertência às famílias até o potencial cancelamento do benefício monetário. Dentro dos limites desse escopo, Janvry, Finan e Sadoulet (2006)¹² mostraram que o cumprimento da condicionalidade de frequência escolar por meio da ameaça de perda de benefícios está fortemente associada a um impacto maior na assiduidade da mesma. Nesse caso, a literatura indica que condicionar transferências monetárias de renda pode ser uma forma eficiente de assegurar matrícula e frequência escolar das crianças.

¹⁰ Is there a child labor trap? Inter-generational persistence of child labor in Brazil

¹¹ Family context and fulfillment of school attendance condition in the Bolsa Família program in Ceará

¹² Evaluating Brazil's Bolsa Escola Program: Impact on Schooling and Municipal Roles

Com relação a duração dos programas em questão, a literatura tem visões bastante conflitantes. Alguns, como Bastagli et al (2016)¹³ acreditam que se o horizonte para esses programas fossem finitos, os indivíduos teriam mais incentivos a investirem seu tempo em atividades que possam auferir uma renda extra (i.e. empregabilidade de seus filhos) para que eles estejam preparados financeiramente para quando o benefício chegar ao fim. Isso diminuiria o investimento em capital humano, evidenciando um incentivo contraproducente por parte do desenho institucional do programa. Por outro lado, outros acreditam que, se o benefício não se estender por período longo o suficiente as famílias não terão tempo o suficiente para investirem na acumulação de capital humano de forma adequada.

Por último, mas não menos importante, ingressamos na literatura existente com relação ao papel de incentivos e desenhos de políticas públicas. Contudo, antes entrarmos em questões relacionadas ao papel dos incentivos nas políticas públicas, vale entender de que forma os custos de incentivos e retornos gerados pelos incentivos se relacionam. Paquette e Fallon (2010)¹⁴ explicam que, o custo do retorno da educação gerado pelas condicionalidades dos programas, está associado ao benefício monetário (subsídio) recebido pelas famílias beneficiadas. Logo, a conta a ser feita por parte dos formuladores de políticas públicas é, se os custos incorridos hoje são maiores do que os retornos extrapolados para o futuro.

Para fins de ilustração, Allan e Fryer (2011)¹⁵ mostram que essa conta aufere ao governo e ao país como um todo, um saldo positivo. De acordo com os autores, mesmo depois que incentivos são retirados, (i.e. mesmo caso o Programa Bolsa Família chegasse ao fim, os efeitos por ele gerados, não chegariam). O efeito conhecido como *fade out* implica que assim que incentivos são retirados (i.e. o benefício monetário), o efeito do incentivo (i.e. matrícula e frequência escolar) chegam ao fim. Allan et al provam que isso não ocorre e que mesmo que os efeitos do benefício monetário na condicionalidade da educação se reduza, o efeito de longo prazo ainda é extremamente positivo, advogando pelo bem-estar social gerado pelos programas.

¹³ Cash transfers: what does the evidence say? A rigorous review of programme impact and the role of design and implementation features

¹⁴ First nations education policy in Canada: Progress or Gridlock?

¹⁵ The Power and Pitfalls of economic incentives

Voltando à questão dos incentivos, Jones (2015) explica que um dos fundamentos da economia clássica é que incentivos podem e devem ser usados para influenciar comportamento. Para ele é exatamente isso que acontece em programas condicionais de transferência de renda: os governos oferecerem aos beneficiados uma transferência monetária, e em troca, esperam uma mudança comportamental em linha com seus objetivos. A sustentabilidade dos programas, de acordo com o autor, seria o componente responsável pela possível quebra de pobreza sistemática que já abordamos.

Bastagil et al (2016) enfatizam também a importância do papel das percepções dos beneficiados do programa com relação a validade da condicionalidade imposta pelo programa e a mensagem por ela passada. Os autores mostram a legitimidade de uma via de comunicação clara entre os formuladores de política econômica e os beneficiados, garantindo melhores resultados esperados. A literatura nos mostra que, uma condicionalidade explícita (como é o caso dos dois programas abordados por esse trabalho) diminui o custo de oportunidade relativo ao uso alternativo do tempo aumentando a adesão ao programa. Isso é resultado do efeito do aprendizado do indivíduo acrescido da mensagem passada pela condicionalidade do programa.

3. Construção e fonte de dados

Para a parte empírica deste projeto, três principais fontes de dados serão utilizadas. Primeiramente, os dados disponibilizados pelo Portal Social que diz respeito ao cumprimento da condicionalidade de educação das crianças e jovens beneficiários do Bolsa Família; a segunda fonte de dados informa a rendimento mensal nominal domiciliar per capita, cujos dados são originados pela PNAD Contínua e são disponibilizados pelo IBGE; e terceira e última fonte de dados é originada também pela PNAD e os dados disponibilizados no Portal Todos Pela Educação¹⁶. Estes dados são disponibilizados como uma serie temporal que mostra a evolução da escolaridade média da população medida em anos de estudo, para cada estado. A seguir, pretendo fazer uma descrição mais detalhada de cada uma dessas fontes de dados, desde a forma como os dados são disponibilizados, até a janela de tempo usada e como pretendo trabalhar com cada um deles.

A principal fonte de dados, e a que dá início à pesquisa, diz respeito ao cumprimento da condicionalidade de educação dos beneficiários do Bolsa Família. A construção dos dados será feita com base nos dados disponibilizados pelo Ministério do Desenvolvimento Social através da Secretaria de Avaliação e Gestão da Informação (SAGI). Os dados estão disponíveis no Portal Data Social¹⁷ e são construídos a partir dos dados computados pelo Sistema Presença. Dessa forma, escolas cujas qualquer parcela de alunos sejam beneficiários do Bolsa Família, devem se cadastrar no Sistema Presença, e fornecer mensalmente os dados de frequência escolar desses mesmos alunos. Dentro deste mesmo escopo, é importante ressaltar que, qualquer mudança de alunos entre escolas cadastradas, deve ser informada à SAGI para que o controle possa continuar sendo feito de forma correta.

Os dados do Sistema Presença não são discriminados pelos grupos de idade, logo, apenas apresentam a porcentagem de alunos que cumpriu a condicionalidade de educação, de acordo com a frequência mínima necessária. Não obstante, destaca-se que a forma de acompanhamento da frequência escolar em resposta à condicionalidade do programa é feita da seguinte maneira:

¹⁶ Portal Todos Pela Educacao (2014-2017)

¹⁷ Portal Data Social: Data CON (2014-2017)

- i. Crianças e Adolescentes acompanhados com frequência escolar igual ou superior a 85% (frequência mínima requerida para crianças e adolescentes de 6 a 15 anos)
- ii. Jovens acompanhados com frequência escolar igual ou superior a 75% (frequência mínima requerida para jovens de 16 e 17 anos)

Devido ao fato de que o Sistema Presença foi implementado em 2014, os dados disponibilizados são mais recentes. Assim, farei uso todos os dados disponibilizados mensalmente, de 2014 a 2017. Para que haja apenas um valor para cada ano de interesse, será calculada a média entre todos os dados disponibilizados para o ano em questão. Levando em conta esta janela de tempo, será necessário fazer ajustes na janela de tempo das outras fontes de dados, que serão explicados mais adiante.

Vale ressaltar que estes dados do Sistema Presença são organizados de acordo com os 27 estados do país (Região Norte (Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins), Região Nordeste (Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe), Região Centro-Oeste (Goiás, Mato Grosso, Mato Grosso do Sul e o Distrito Federal), Região Sudeste (Espírito Santo, Minas Gerais, São Paulo e Rio de Janeiro) e Região Sul (Paraná, Rio Grande do Sul e Santa Catarina)). (*Figura 1*).

A segunda fonte de dados utilizada será os dados de “Rendimento nominal mensal domiciliar per capita” da população residente de cada estado da federação. Os dados são calculados a partir de informações da PNAD Contínua¹⁸ realizada em cada ano de interesse (2014, 2015, 2016 e 2017) com base nas respostas obtidas nas quatro entrevistas realizadas trimestralmente. O rendimento nominal mensal domiciliar é calculado como a razão entre o total dos rendimentos domiciliares e o total de moradores. Vale ressaltar que são considerados os rendimentos de trabalho e de outras fontes, como por exemplo as fontes de renda de indivíduos classificados como pensionistas, empregados domésticos e parentes dos empregados domésticos.

¹⁸ IBGE: PNAD Contínua (2014-2017)

Por fim, a terceira fonte de dados será os dados de “Escolaridade média em anos de estudo (pessoas de 25 anos ou mais)” que podem ser encontrados no Portal Todos Pela Educação. O indicador é calculado com base nos dados da PNAD e as estimativas levam em consideração a idade em anos completos na data de referência da PNAD. No cálculo, consideram-se os anos de estudo da população de 25 anos ou mais, independentemente das etapas de ensino. Em outras palavras, o indicador é calculado levando em conta todos os anos em que o indivíduo esteve estudando, sem discriminar as etapas de ensino (ensino fundamental, ensino médio, ensino superior e assim por diante), e levando em consideração apenas o número total de anos. A série histórica é disponibilizada para cada um dos 27 estados brasileiros e abrange uma janela de tempo de 1992 a 2014, com exceção dos anos de 1994, 2000 e 2010. Devido ao fato de que os dados estão disponibilizados apenas até o ano de 2014, farei uso de uma estratégia de previsão para extrapolar os dados até o ano de 2017 para que pertençam à mesma janela de dados das outras fontes de dados utilizadas. Essa estratégia será explicada de forma detalhada, mais adiante, na seção de metodologia.

Levando em conta que grande parte dos dados se origina da PNAD Contínua, é válido acrescentar que essa é uma pesquisa domiciliar feita trimestralmente que, capta informações socioeconômicas e demográficas de cerca de 211 mil domicílios distribuídos em cerca de 3,5 mil municípios. Dito isso, a relevância desta fonte de dados se dá a partir da extensa quantidade de dados coletados, distribuídos de forma consideravelmente homogênea geograficamente.

Além das variáveis de interesse, renda per capita e rendimento médio nominal domiciliar, também pretendo trabalhar com algumas variáveis de controle para cada unidade da federação. Na próxima seção, será explicado como será feita a regressão usando cada uma dessas variáveis e a teoria econômica por trás das escolhas das mesmas.

A primeira fonte de dados utilizada para a construção das variáveis de controle foi o Portal DAPP (Diretoria de Análise de Políticas Públicas) da Fundação Getúlio Vargas¹⁹. O portal disponibiliza diversos indicadores de segurança pública no país, como número de policiais militares e/ou civis em um estado, número de homicídios,

¹⁹ Portal DAPP (Diretoria de Análise de Políticas Públicas): taxa de latrocínio a cada 100 mil habitantes

número de latrocínios e assim por diante. As fontes utilizadas para obtenção desses dados são: o Anuário Brasileiro de Segurança Pública e Sistema de Informação sobre Mortalidade do Sistema Único de Saúde (SIM/DATASUS). A variável de interesse nesse caso é a *taxa de latrocínios a cada 100 mil habitantes por unidade da federação*, também na janela de tempo entre os anos de 2014 e 2017.

A segunda fonte de dados utilizada foi a Plataforma Mobilidados²⁰ desenvolvida pelo Instituto de Políticas de Transporte e Desenvolvimento (ITDP), um instituto Americano que promove o transporte sustentável no mundo, e está presente no Brasil há mais de uma década. A variável utilizada é o *comprometimento da renda de um salário mínimo com a tarifa de transporte público*. O indicador é calculado com base nas tarifas de transporte público das capitais de cada unidade da federação e é apontado como um percentual do salário mínimo do referido ano. Da mesma forma que as outras variáveis, farei uso de uma série de tempo desta variável, do ano de 2014 a 2017.

A terceira e última variável de controle foi originada de dados disponibilizados pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais)²¹ e que tem como fonte os dados do Censo da Educação Básica. A variável utilizada é o *percentual de docentes por grupo do indicador de adequação da formação do docente*. A variável é categorizada em 5 grupos que variam de acordo com a adequação da formação dos docentes em relação à disciplina em que lecionam. Serão usados apenas os dados de escolas públicas, agregados para cada uma das regiões da federação. As 5 categorias mencionadas acima variam da seguinte forma:

- i. Grupo 1 - Docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona
- ii. Grupo 2 - Docentes com formação superior de bacharelado (sem complementação pedagógica) na mesma área da disciplina que leciona

²⁰ Plataforma Mobilidados - Instituto de Políticas de Transporte e Desenvolvimento: comprometimento da renda de um salário mínimo com a tarifa de transporte público

²¹ INEP utilizando dados do Censo da Educação Básica: percentual de docentes por grupo do indicador de adequação da formação do docente

- iii. Grupo 3 - Docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) em área diferente daquela que leciona
- iv. Grupo 4 - Docentes com formação superior não considerada nas categorias anteriores
- v. Grupo 5 - Docentes sem formação superior

Assim como os dados para “Escolaridade média em anos de estudo (pessoas de 25 anos ou mais)”, os dados referentes à taxa de letalidade a cada 100 mil habitantes e à adequação da formação docente, também não estão completos. As duas séries temporais vão apenas de 2014-2016, havendo novamente a necessidade de adotar uma estratégia de estimação para os dados faltantes. Essa estratégia será discutida com mais clareza na seção 5.1.

Por fim, fora do escopo da metodologia em si, serão usados dados de taxa de desemprego das unidades da federação para uma interpretação econômica dos resultados encontrados (seção 8.1). A taxa de desemprego é calculada como a proporção de pessoas capazes de exercer alguma profissão ou que estão em busca de um emprego remunerado, mas que por algum motivo não estão empregadas, sobre o a População Economicamente Ativa (PEA). Os dados foram originados da Pesquisa Mensal de Emprego (PME) disponível na página do IBGE.²²

²² Pesquisa Mensal de Emprego disponibilizada pelo IBGE: taxa de desemprego

Figura 1 – Exemplo de divulgação dos dados do Sistema Presença para o mês/ano de referência de novembro/2017:

Bolsa Família - Condicionalidade da Educação
Outubro e Novembro de 2017 (Mês referência: Novembro)

Estados	Total de Alunos Beneficiários	Alunos acompanhados*		Cumpriram condicionalidade**	
		Quant.	%	Quant.	%
Distrito Federal	104.967	95.843	91,31%	93.069	97,11%
Goiás	362.308	333.809	92,13%	319.387	95,68%
Mato Grosso	194.221	178.780	92,05%	171.742	96,06%
Mato Grosso do Sul	150.025	136.828	91,20%	128.036	93,57%
Centro Oeste	811.521	745.260	91,83%	712.234	95,57%
Acre	128.956	115.769	89,77%	113.102	97,70%
Amapá	101.140	89.743	88,73%	89.215	99,41%
Amazonas	559.583	502.309	89,76%	488.940	97,34%
Pará	1.191.073	1.113.664	93,50%	1.090.008	97,88%
Rondônia	107.434	100.642	93,68%	97.027	96,41%
Roraima	59.410	55.409	93,27%	53.942	97,35%
Tocantins	149.684	140.470	93,84%	135.938	96,77%
Norte	2.297.280	2.118.006	92,20%	2.068.172	97,65%
Paraná	374.364	352.510	94,16%	324.115	91,94%
Rio Grande do Sul	389.176	374.999	96,36%	345.684	92,18%
Santa Catarina	139.850	131.853	94,28%	126.303	95,79%
Sul	903.390	859.362	95,13%	796.102	92,64%
Alagoas	427.427	401.226	93,87%	389.864	97,17%
Bahia	1.718.055	1.565.945	91,15%	1.517.290	96,89%
Ceará	1.000.745	943.184	94,25%	908.693	96,34%
Maranhão	1.152.293	1.040.585	90,31%	1.023.551	98,36%
Paraíba	472.310	430.831	91,22%	420.550	97,61%
Pernambuco	1.081.532	1.019.049	94,22%	987.563	96,91%
Piauí	433.069	408.145	94,24%	403.878	98,95%
Rio Grande do Norte	338.819	318.449	93,99%	305.151	95,82%
Sergipe	262.589	248.409	94,60%	231.797	93,31%
Nordeste	6.886.839	6.375.823	92,58%	6.188.337	97,06%
Espírito Santo	185.423	179.075	96,58%	160.252	89,49%
Minas Gerais	1.121.223	1.036.005	92,40%	980.269	94,62%
Rio de Janeiro	843.009	766.643	90,94%	711.609	92,82%
São Paulo	1.503.501	1.444.449	96,07%	1.284.676	88,94%
Sudeste	3.653.156	3.426.172	93,79%	3.136.806	91,55%
Brasil	14.552.186	13.524.623	92,94%	12.901.651	95,39%

* Quantidade de alunos que tiveram informações de frequência cadastradas no Sistema Presença

** Quantidade de alunos que cumpriram frequência entre os que tiveram informações cadastradas no Sistema Presença

Portal Data Social (2014-2017). Acompanhamento da Taxa de Condicionalidade de Educação – Disponível em: <http://aplicacoes.mds.gov.br/sagi>

4. Análise Descritiva

Tabela 1 – Estatísticas Descritivas

	Cumprim. Condicionalidade	Rendimento mensal pc.	Escolaridade média	Violência_Taxa	Comprom. Renda Transporte
Média	96.1	1026.0	7.5	1.4	17.4
Mediana	96.9	879.0	7.6	1.4	17.0
Assimetria	-101.0	152.0	0.3	0.6	0.1
Mínimo	88.4	461.0	5.7	0.3	12.0
Máximo	99.4	2548.0	10.0	3.0	22.0
Percentil 5%	91.1	608.0	6.1	0.6	14.0
Percentil 10%	92.4	668.0	6.2	0.8	15.0
Percentil 25%	95.0	753.0	6.8	0.9	16.0
Percentil 50%	96.8	880.0	7.6	1.4	17.0
Percentil 75%	97.8	1228.0	8.2	1.8	19.0
Percentil 90%	98.8	1465.0	8.9	2.2	20.0
Percentil 95%	98.9	1719.0	9.0	2.4	21.0

(cont.)

	Docente_G1	Docente_G2	Docente_G3	Docente_G4	Docente_G5
Média	49.2	0.3	14.7	3.3	32.3
Mediana	50.9	0.3	12.9	2.3	32.5
Assimetria	-0.3	134.0	0.7	157.0	0.3
Mínimo	20.4	0.0	5.3	1.2	6.7
Máximo	77.1	1.1	30.9	11.8	66.2
Percentil 5%	24.0	0.1	5.9	1.3	8.3
Percentil 10%	25.8	0.1	6.8	1.4	9.6
Percentil 25%	39.5	0.2	9.1	1.9	19.4
Percentil 50%	50.9	0.3	12.9	2.3	32.5
Percentil 75%	58.9	0.5	19.4	4.4	43.9
Percentil 90%	64.9	0.6	25.8	5.9	55.4
Percentil 95%	68.6	1.1	28.9	7.4	64.9

Para a variável de cumprimento de condicionalidade do programa, a média da taxa de cumprimento foi de 96,1%, sendo 88,4% a taxa mínima e 99,4% a taxa máxima. Representando então uma dispersão de 10 pontos percentuais entre o maior e menor dado de taxa de cumprimento registrada. Com relação à assimetria dos dados, que diz respeito ao grau de afastamento da distribuição dos dados de uma distribuição simétrica, os dados têm distribuição assimétrica negativa (-101) o que significa que a grande

maioria dos dados se encontra na cauda direta da distribuição (i.e., mais perto da taxa de 100%). Com relação à frequência cumulativa, 50% dos dados são equivalentes ou menores à 96,8%.

Para a variável de rendimento mensal domiciliar per capita, a média nos anos analisados foi de R\$ 1.026, tendo atingido uma máxima de R\$ 2.548 e uma mínima de R\$ 461. Os dados coletados apresentam uma distribuição simétrica positiva, o que significa que a maioria dos dados foram registrados na cauda esquerda dos dados (i.e., rendimento mensal domiciliar per capita relativamente baixo). Isso é corroborado pelos dados de frequência cumulativa que nos mostra que 50% das observações são equivalentes ou menores à R\$ 880, R\$ 146 abaixo da média registrada.

Para os dados de escolaridade média (medidos em anos de estudos), a média encontrada foi de 7,5 anos, onde a medida máxima em um estado atingiu 10 anos e a mínima chegou a 5,7 anos. A medida de assimetria muito próxima de 0 (i.e. 0,3), nos diz que os dados estão concentrados no meio da distribuição, em outras palavras, os dados são bastante simétricos. Vemos isso também a partir da frequência cumulativa que mostra que 50% das observações são menores ou iguais a 7,6 anos, exatamente o valor da mediana.

Para os dados de taxa de violência nos estados (medida pelo número de latrocínios a cada 100 mil habitantes), a média registrada foi de 1,4%, tendo atingido uma máxima de 3% e uma mínima de 0,3%. Novamente, os dados têm distribuição simétrica e 50% das taxas observadas estão abaixo ou iguais 1,4%.

Quanto ao comprometimento da renda com o transporte público, a média encontrada foi de 17,4%, onde a taxa mínima foi de 12% e a máxima de 22%, representando uma diferença de 10 pontos percentuais. Mais uma vez, os dados têm distribuição simétrica e 50% das observações estão abaixo ou iguais a 17%.

Para os dados referentes à adequação da formação docente, as médias encontradas foram de 49,2%, 0,3%, 12,9%, 2,3% e 32,5%, respectivamente. Esses dados nos mostram que a grande maioria dos docentes dos estados estão concentrados nos grupos 1 e 5, que representam Grupo 1 - Docentes com formação superior de

licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona e Grupo 5 - Docentes sem formação superior. Assim, fica claro a concentração de observações nos extremos da distribuição.

4.1 Análise Gráfica

Com o objetivo de visualizar melhor a dinâmica dos dados e como eles se comportam com relação uns aos outros, foram elaborados gráficos para cada uma das variáveis de interesse utilizadas neste trabalho.

O primeiro gráfico (*Gráfico 1*) mostra como a variável y de interesse (cumprimento da condicionalidade de educação do Programa Bolsa Família) se comporta com relação ao rendimento mensal domiciliar per capita de cada estado. A relação encontrada é bastante clara: estados com maior renda per capita apresentam menores taxas de cumprimento da condicionalidade de educação e vice-versa. A organização do gráfico a seguir foi feita para que se pudesse distinguir facilmente os estados pertencentes à cada região do país. Vemos então como em sua maioria, os estados do sul e sudeste (regiões mais ricas), apresentam taxas de cumprimento menores, enquanto estados pertencentes ao norte e nordeste estão mais concentrados na parte superior à esquerda do gráfico. Vale ressaltar que os quatro pontos plotados para cada um dos estados representa os dados referentes aos quatro anos utilizados nesse estudo (2014, 2015, 2016 e 2017).

Além da correlação negativa entre as duas variáveis, também é fácil ver que os dados referentes ao Distrito Federal são o que chamamos de “*outliers*”, podendo causar problemas na análise que este trabalho busca fazer. Os quatro dados referentes ao Distrito Federal estão localizados no canto direito superior do gráfico, indicando uma correlação positiva entre as duas variáveis. A distância desses pontos com relação aos demais também pode ser corroborada pela linha de tendência do gráfico. A linha de tendência para todos os estados possui uma inclinação significativamente menos íngreme do que a linha pontilhada que representa os mesmos dados, porém com a exclusão dos dados do Distrito Federal. Logo, é simples concluir que resultados mais robustos serão encontrados a partir de regressões que contem com a exclusão destes dados. Ainda neste tópico, vale observar que os possíveis motivos para tal *outlier* será discutido com mais clareza mais para frente no trabalho.

Seguindo para a segunda variável de interesse, podemos ver no *Gráfico 2* como a escolaridade média (medida por anos de estudo) influencia a variável y de

cumprimento da condicionalidade de educação. Novamente vemos uma correlação negativa entre as duas fontes de dados, indicando que estados com uma média de anos de estudo menor, são mais incisivos no cumprimento da condicionalidade. É fácil ver que estados pertencentes ao sul e ao sudeste estão localizados na parte inferior à direita do gráfico indicando uma média mais alta de escolaridade, porém um cumprimento menos satisfatório da condicionalidade. Assim como vemos no *Gráfico 1*, vemos aqui a mesma condição dos dados referentes ao Distrito Federal como *outliers*. Pelo isolamento dos pontos com relação aos demais estados e pela mudança na inclinação da reta de tendência, corroboramos a conclusão de que resultados mais robustos serão produzidos a partir de regressões que não possuam estes dados. Vale ressaltar que os *outliers* tem efeitos mais brandos para esses dados, em comparação com os dados do *Gráfico 1*, que pode ser visto pela mudança na inclinação da linha de tendência. Contudo, esses efeitos ainda são substanciais e não devem ser negligenciados.

O *Gráfico 3* mostra a dinâmica entre os dados de violência nos estados (medido pela taxa de latrocínio em cada 100 mil habitantes) e o cumprimento da condicionalidade de educação. Os dados apresentam uma correlação positiva entre si, muito evidenciada pela linha de tendência. Diferentemente das outras duas variáveis discutidas acima, a correlação entre os dados aqui é menos nítida. A dispersão dos dados é mais evidente pela distância dos pontos uns dos outros, e também pela ampla dispersão em volta da linha de tendência. Dessa forma, o que vemos nessa primeira análise é contra intuitivo a medida que estados com uma maior taxa de violência apresentam, na média, taxas maiores de cumprimento da condicionalidade do programa. Além disso, vemos uma maior dispersão dos dados pertencentes aos estados do Sudeste (principalmente RJ, SP e ES) que estão localizados na parte inferior do gráfico. Os pontos referentes a estes estados possuem maior variabilidade do que os demais (dentro da mesma região geográfica), variando de ~ 90% para São Paulo de até ~ 94% para o Rio de Janeiro, na taxa de cumprimento de condicionalidade. Com exceção dessa maior dispersão de pontos, que torna a correlação entre os dados mais branda, os demais pontos aparecem significativamente mais bem-comportados em volta da linha de tendência.

O *Gráfico 4* por sua vez, mostra como o cumprimento da condicionalidade de educação varia à medida que variamos a taxa comprometimento da renda de um salário

mínimo com a tarifa de transporte público de estado para estado. Vale lembrar que, o uso da taxa e não do valor absoluto, significa que a medida leva em conta a diferença de renda per capita da população de cada estado, não havendo distorções causadas por essas diferenças. É trivial ver que estados pertencentes às regiões do Sul, Sudeste e Centro-Oeste possuem uma taxa de comprometimento da renda maior, em outras palavras, uma maior proporção da renda das pessoas é consumida por tarifas de transporte público. Em contrapartida, os estados pertencentes ao Norte e ao Nordeste estão localizados entre a parte central e esquerda do gráfico. A reta de tendência com inclinação negativa nos indica que os estados com taxas de comprometimento da renda com transporte público menores, possuem taxas de cumprimento da condicionalidade consideravelmente mais satisfatórias, em linha com expectativas. Por outro lado, com exceção de algumas observações referentes ao Espírito Santo, que estão um pouco mais dispersos em volta da linha de tendência, o oposto ocorre com estados onde a população tem maior parte do seu salário comprometido com tarifas de transporte: taxas piores de cumprimento da condicionalidade. Diferentemente do comportamento da primeira variável de controle (taxa de violência nos estados), ilustrada no *Gráfico 3*, esta segunda variável de controle tem correlação negativa com a variável y e a magnitude da correlação é mais significativa, dada a maior inclinação da reta de tendência. O motivo por trás de tais observações será discutido mais para frente, na seção de discussão dos resultados.

Os gráficos 5, 6, 7, 8 e 9 mostram a dinâmica entre o percentual de docentes pertencentes à cada um dos respectivos grupos e a taxa de cumprimento de condicionalidade do programa. Para os grupos 1, 3 e 4 a correlação encontrada é negativa, para o grupo 5 é positiva e para o grupo 2 a correlação encontrada é desprezível. Nos gráficos é possível ver a distinção entre o corpo docente de estados mais pobres versus estados mais ricos. Nos primeiros 4 grupos, que representam a porcentagem de docentes que possuem algum tipo de formação superior (seja na própria área de ensino ou em outra área), os estados com maiores porcentagens são, naturalmente, em sua maioria, os estados pertencentes às regiões Sul, Sudeste e Centro-Oeste. Já os estados do Norte e Nordeste estão mais concentrados na parte esquerda do gráfico, indicando baixa porcentagem de docentes com algum tipo de formação superior. O gráfico do Grupo 5 ilustra um comportamento contrário: os estados mais ricos estão localizados na parte esquerda do gráfico e, por sua vez, os mais pobres na

parte direita. Estas observações são bastante intuitivas, à medida que esperamos que uma maior porcentagem de docentes de estados pobres não necessariamente possua ensino superior e vice-versa. Contudo, a correlação positiva entre o Grupo 5 e o cumprimento da condicionalidade, e negativa para os grupos 1, 3 e 4, é contra intuitiva, pois a hipótese usada é de que um corpo docente mais qualificado seria capaz de motivar alunos de forma melhor, aumentando a taxa de cumprimento da condicionalidade.

Gráfico 1 – Cumprimento da condicionalidade de educação versus rendimento mensal domiciliar per capita (R\$)

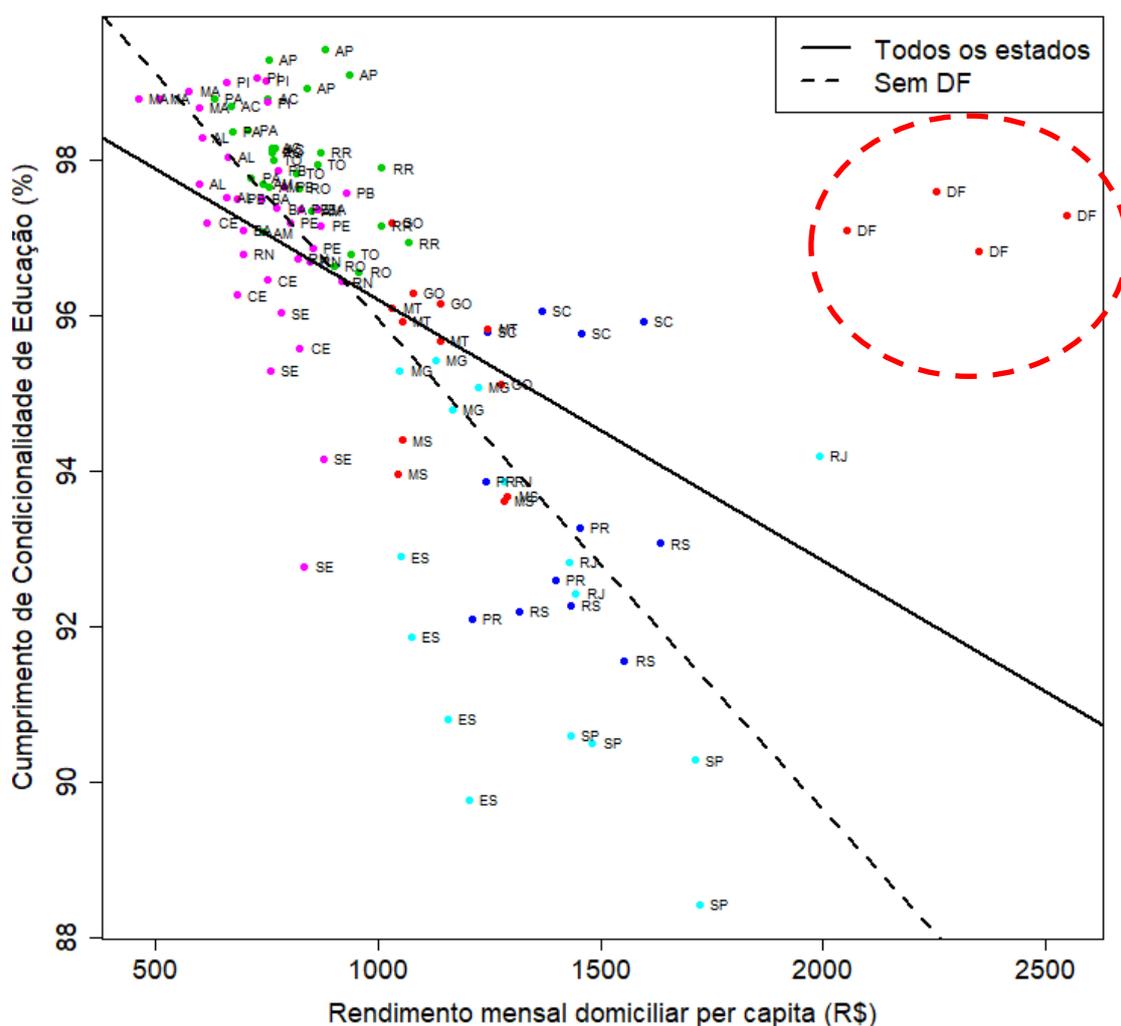


Gráfico 3 – Cumprimento da condicionalidade de educação versus taxa de latrocínios a cada 100 mil habitantes

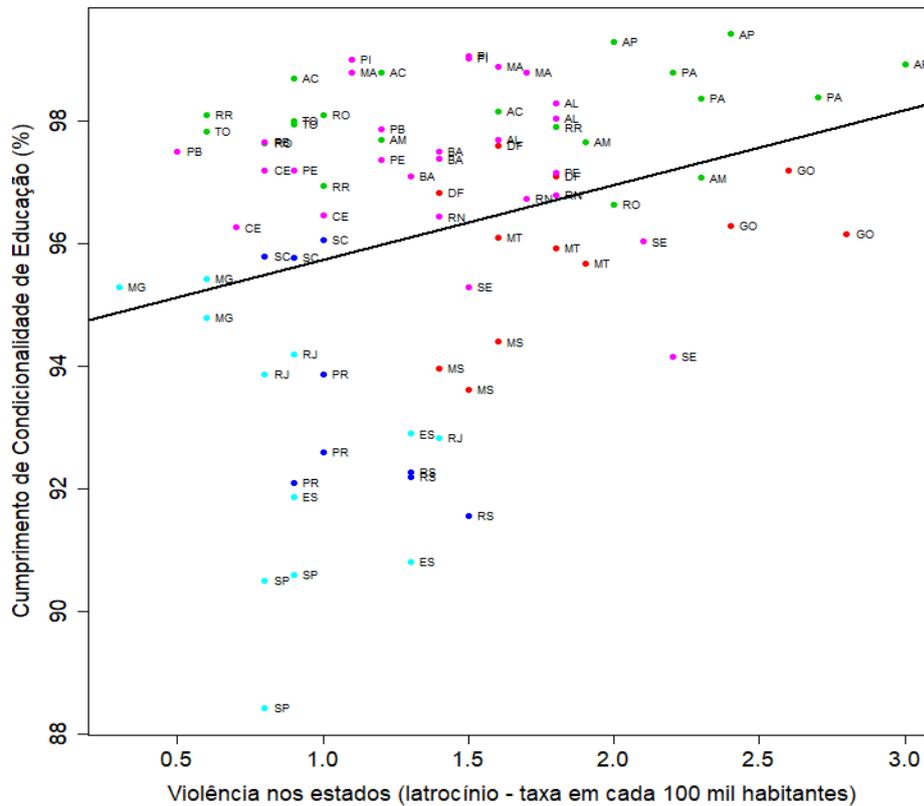


Gráfico 4 – Cumprimento da condicionalidade de educação versus taxa comprometimento da renda de um salário mínimo com a tarifa de transporte público

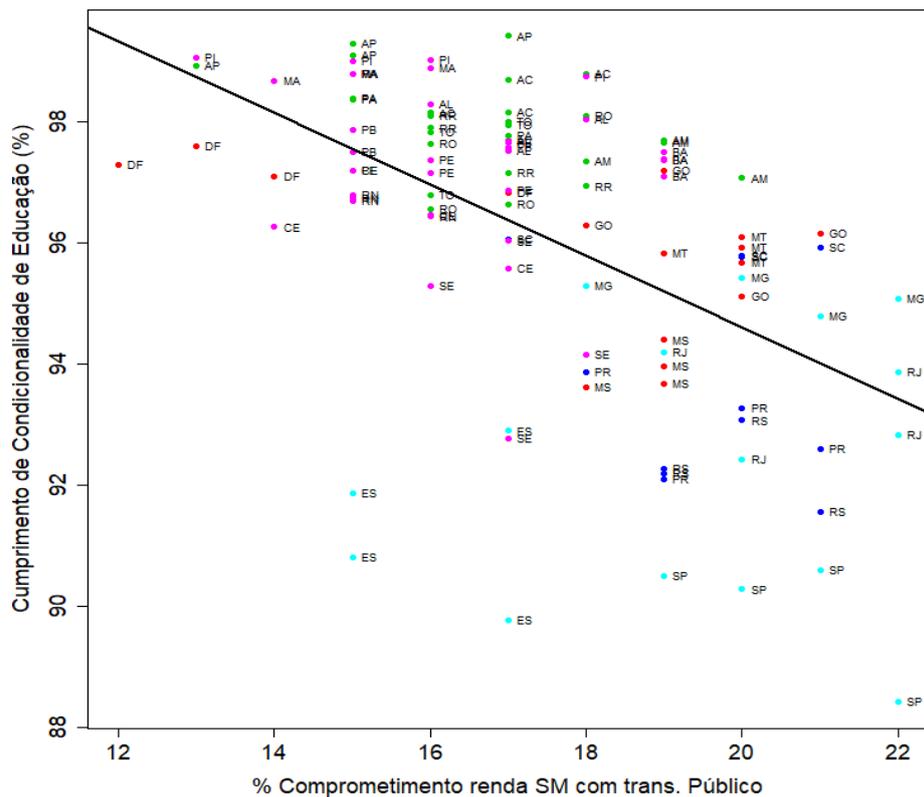


Gráfico 5 – Cumprimento da condicionalidade de educação versus percentual de docentes formação superior de licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona (Grupo 1)

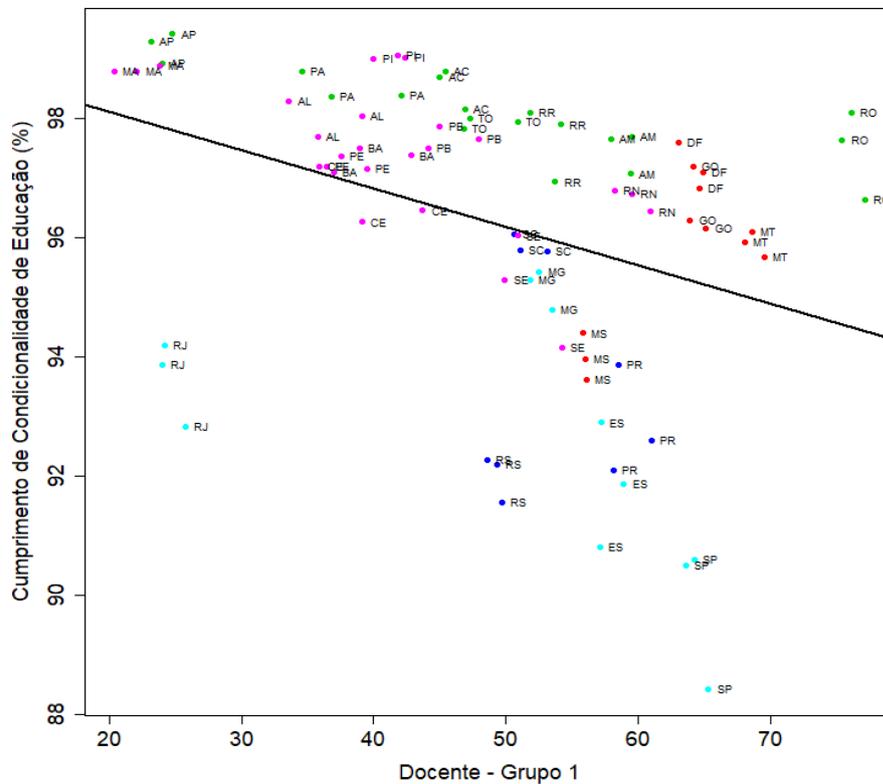


Gráfico 6 – Cumprimento da condicionalidade de educação versus percentual de docentes com formação superior de bacharelado (sem complementação pedagógica) na mesma área da disciplina que leciona (Grupo 2)

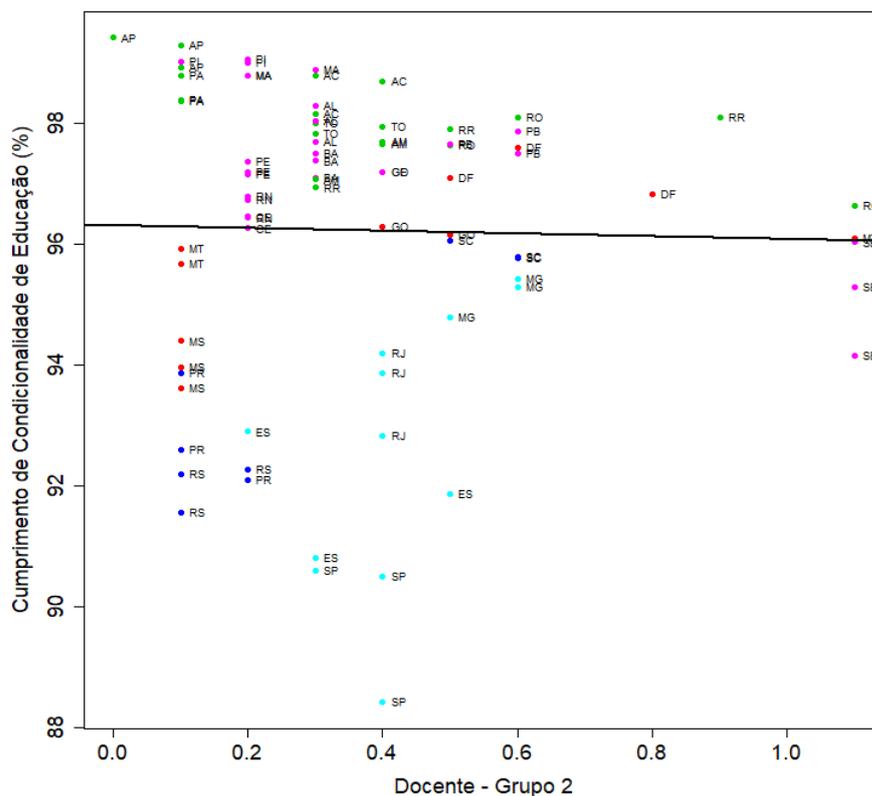


Gráfico 7 – Cumprimento da condicionalidade de educação versus percentual de docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) em área diferente daquela que leciona (Grupo 3)

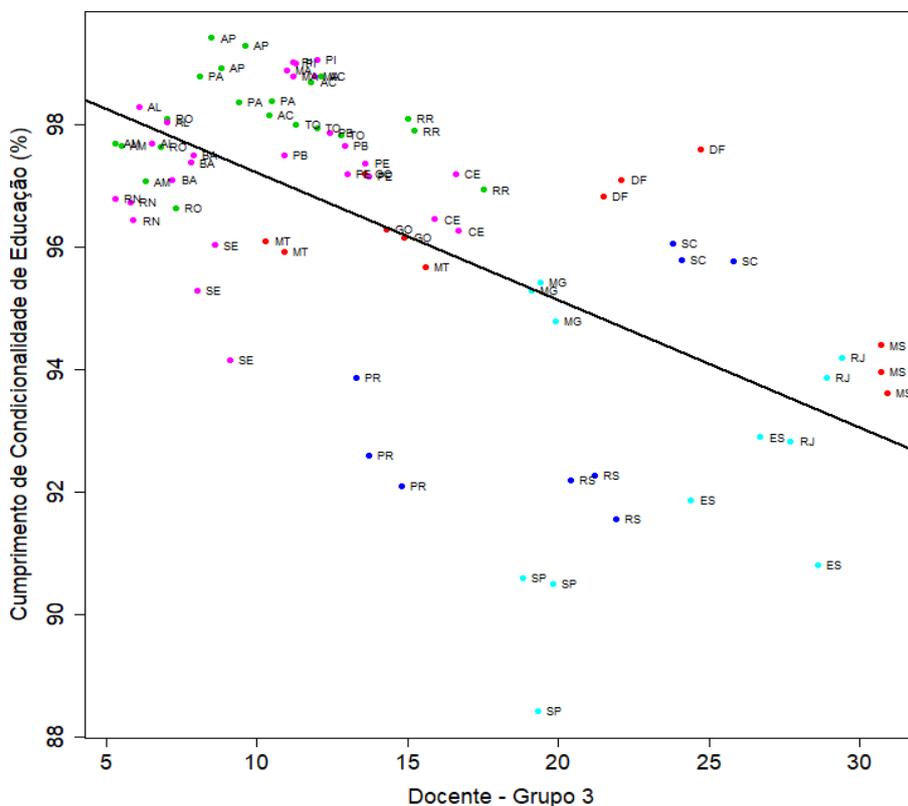


Gráfico 8 – Cumprimento da condicionalidade de educação versus percentual de docentes com formação superior não considerada nas categorias prévias (Grupo 4)

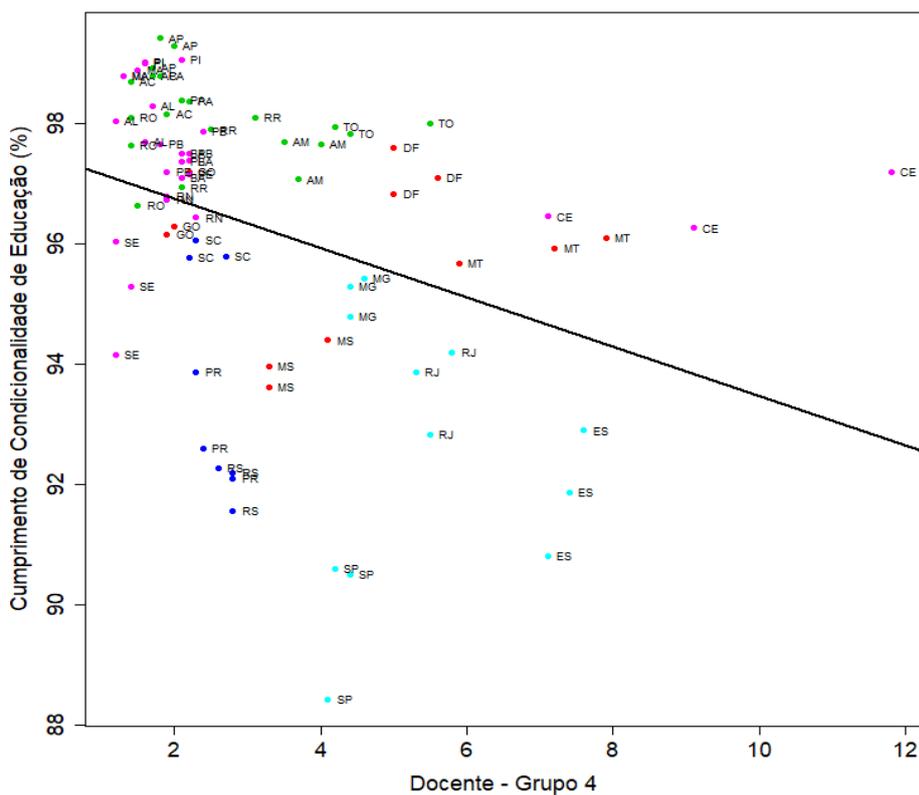
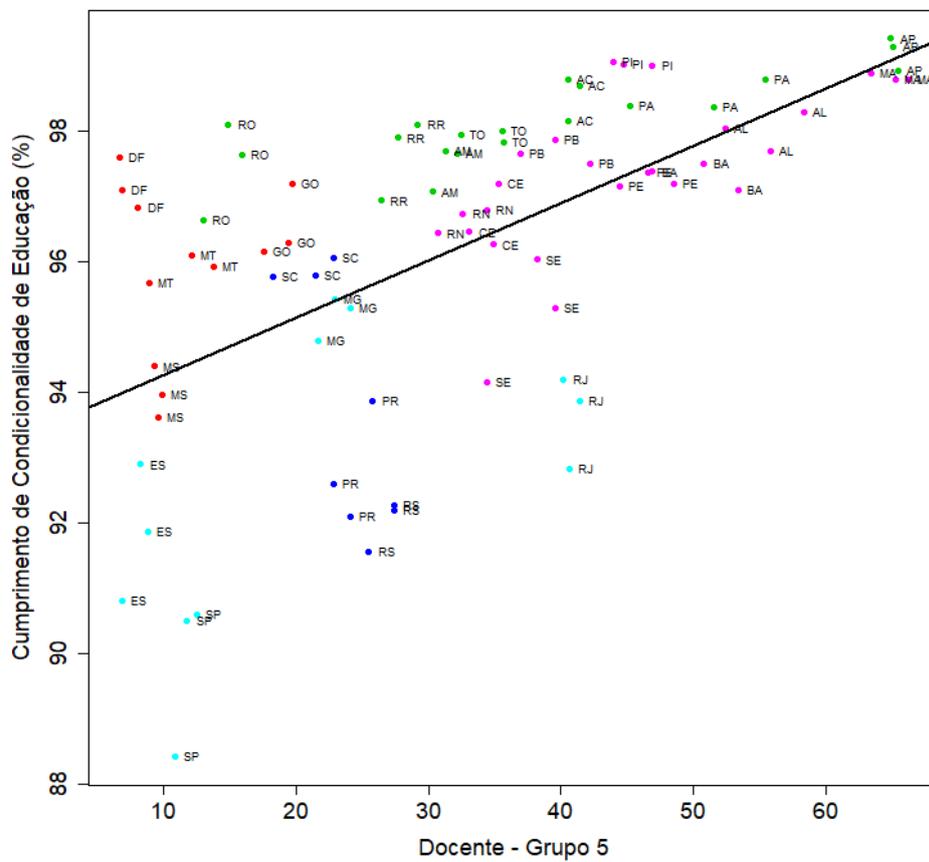


Gráfico 9 – Cumprimento da condicionalidade de educação versus percentual de docentes sem formação superior (Grupo 5)



5. Metodologia

Com o objetivo de verificar o efeito da renda domiciliar per capita e da escolaridade média em anos de estudo da população sobre o cumprimento da condicionalidade de educação do Programa Bolsa Família, pretendo rodar uma equação de Mínimos Quadrados Ordinários (MQO) em que a variável dependente será a porcentagem de beneficiários que cumpriram a condicionalidade imposta. As variáveis independentes de interesse são: i) Rendimento nominal mensal domiciliar e ii) Escolaridade média em anos de estudo (pessoas de 25 anos ou mais). Dessa forma, a primeira versão da equação de MQO é a seguinte:

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} \quad [1]$$

Após rodar a equação [1], será rodada uma segunda versão dessa equação com a inclusão de variáveis de controle para as unidades da federação. Essas variáveis são:

- i) Taxa de latrocínios a cada 100 mil habitantes por unidade da federação, denotada pela letra L;
- ii) Comprometimento da renda de um salário mínimo com a tarifa de transporte público (%), que será denotada pela letra “T”;
- iii) Percentual de docentes por grupo do indicador de adequação da formação do docente, que será denotada pela letra “M”.

Essas variáveis serão incluídas na regressão com o objetivo de separar os efeitos das variáveis de interesse de outros fatores que podem estar influenciando mudanças na variável dependente. Dito isso, com a inclusão dos controles, a regressão ficará da seguinte maneira:

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 L_i + \beta_4 T_i + \beta_5 M_i \quad [2]$$

onde M_i representa M_{i1} M_{i2} M_{i3} M_{i4} M_{i5} , para cada um dos cinco grupos de docentes.

É importante notar, contudo, uma questão crucial que chamamos de *bad control*. Em alguns casos, mais controles pode não significar um ganho de precisão na hora de

estimar os parâmetros e pode resultar em estimações enviesadas. O problema pode ser resumido da seguinte forma, usando como exemplo a variável que ilustra a violência nos estados, denotada pela letra L : a variação no cumprimento da condicionalidade de educação com relação a mudanças na renda domiciliar per capita e o nível de escolaridade, condicionado à taxa de latrocínios a cada 100 mil habitantes por unidade da federação, será o efeito causal das duas variáveis de interesse na variável dependente mais um termo que chamamos de viés de seleção. Esse termo reflete exatamente a questão central do *bad control*: as variáveis de interesse (renda domiciliar per capita e anos de escolaridade) afetam a composição e distribuição da variável taxa de latrocínios. Em outras palavras, *bad controls* são variáveis cuja composição é afetada pelas variáveis de interesse²³. Adaptando a teoria às variáveis deste trabalho é fácil ver como a taxa de latrocínio de um dado estado ou como o comprometimento da renda com transporte público podem ser afetados pelo rendimento mensal domiciliar médio deste estado e pela escolaridade média em anos de estudo. Estados com maior renda per capita e escolaridade naturalmente devem possuir taxas mais baixas de latrocínio e devem ter uma menor parte de sua renda comprometida pelos custos de transporte público.

Assim, em termos econométricos, temos:

$$E[Y_1 | L_i] - E[Y_0 | L_i] = \underbrace{E[Y_1 - Y_0 | L_{1i}]}_{\text{Efeito Causal}} + \underbrace{\{E[Y_0 | L_{1i}] - E[Y_0 | L_{0i}]\}}_{\text{Viés de Seleção}}$$

Seguindo essa linha de raciocínio, acredito que ambas as variáveis que ilustram a violência (L) e peso da tarifa no transporte público no salário (T), podem ter candidatas à serem *bad controls*. Visto que isso pode significar uma limitação da equação de MQO [2], é importante rodar três novas regressões sem a inclusão dessas variáveis, da seguinte forma:

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 T_i + \beta_4 M_i \quad [3]$$

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 L_i + \beta_4 M_i \quad [4]$$

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_5 M_i \quad [5]$$

²³ Angrist J. D., Pischke J-S., Mostly Harmless Econometrics: An Empiricist's Companion (2009) – pg 47

O objetivo é conseguir testar as duas variáveis separadamente buscando encontrar estimadores enviesados que podem implicar que estas variáveis sejam *bad controls*. Por fim, rodar uma regressão sem as duas variáveis ao mesmo tempo, buscando encontrar o mesmo efeito. Caso as estimações não sejam enviesadas, podemos concluir que as variáveis L e T não são *bad controls*.

A última parte da metodologia deste trabalho consiste em analisar não apenas o efeito que cada um dos grupos docentes tem sobre o cumprimento da condicionalidade do programa, mas também como mudanças na composição de cada um destes grupos podem ser benéficas ou maléficas para o cumprimento adequado da condicionalidade. Sabemos que cada um dos cinco grupos representam as seguintes categorias de professores componentes do corpo docente de escolas públicas brasileiras:

- i. Grupo 1 - Docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona
- ii. Grupo 2 - Docentes com formação superior de bacharelado (sem complementação pedagógica) na mesma área da disciplina que leciona
- iii. Grupo 3 - Docentes com formação superior de licenciatura (ou bacharelado com complementação pedagógica) em área diferente daquela que leciona
- iv. Grupo 4 - Docentes com formação superior não considerada nas categorias anteriores
- v. Grupo 5 - Docentes sem formação superior

É possível que as estimativas para as variáveis pertencentes à M_i , (grupos do corpo docente) produzidas pelas regressões [2], [3], [4] e [5] explicitadas acima, sejam enganosas. Isso porque, se somarmos a porcentagem de docentes em cada um dos grupos, naturalmente chegaremos à 100%, logo, à medida que todos os grupos estão sendo inseridos na equação de uma só vez, talvez não seja possível encontrar o efeito de mudanças na composição destes mesmos grupos. Para fins de ilustração, digamos que o estado de São Paulo tenha uma baixa taxa de cumprimento da condicionalidade, e digamos que essa taxa é muito atribuída pela alta porcentagem de docentes pertencentes

ao grupo 5 (docentes sem formação superior). O que aconteceria com essa taxa se houvesse uma mudança na composição dos grupos, havendo uma mobilidade maior de professores saindo do grupo 5 e entrando em algum dos outros quatro grupos? Dito isso, pretendo rodar outras duas regressões que se comportam da seguinte forma:

Primeiramente, uma regressão com todas as variáveis: i) rendimento nominal mensal domiciliar e ii) escolaridade média em anos de estudo (pessoas de 25 anos ou mais), iii) taxa de latrocínios a cada 100 mil habitantes por unidade da federação, iv) comprometimento da renda de um salário mínimo com a tarifa de transporte público (%), que será rodada cinco vezes, tendo um grupo docente zerado em cada uma das regressões. Primeiro, o grupo docente 1 será zerado, e a porcentagem referente à ele será redistribuído igualmente para os outros quatro grupos restantes. E assim por diante, funcionando de forma análoga para os grupos de docentes 2, 3, 4 e 5.

Esta mesma estratégia será repetida retirando as duas variáveis que são possíveis candidatas à serem *bad controls* (taxa de latrocínios a cada 100 mil habitantes por unidade da federação e comprometimento da renda de um salário mínimo com a tarifa de transporte público (%)).

Para fins de ilustração: se o grupo docente 1 for composto por 24% dos docentes das escolas públicas do estado em questão, ao zerarmos este grupo o valor de 24% será redistribuído de forma homogênea entre outros 4 grupos. Assim, cada um dos grupos “receberá” o valor equivalente à 6% que será somado à sua porcentagem inicial. Desta forma, o resultado será 10 modelos que testarão os efeitos no cumprimento da condicionalidade de mudanças na composição dos grupos de corpos docentes. Possibilitando assim, a avaliação do impacto de possíveis políticas de melhorias na qualidade do corpo docente das escolas públicas.

As regressões finais serão representadas da seguinte forma:

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 L_i + \beta_4 T_i + \beta_5 M_i \quad [6]$$

$$Y_{i1} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_5 M_i \quad [7]$$

onde M_i representa $M_{i1} = 0$, $M_{i2} = 0$, $M_{i3} = 0$, $M_{i4} = 0$, $M_{i5} = 0$, sendo as variáveis zeradas uma por vez e o valor conferido ao grupo zerado, redistribuído igualmente entre os demais grupos.

5.1 Estratégia de previsão dos dados – Método de Holt com 2 Parâmetros

Como estratégia de estimação para os dados faltantes, foi utilizado o Modelo de Holt com 2 Parâmetros. O método faz uso de uma estratégia de alisamento exponencial que em sua versão mais simples, é apropriada para dados que não possuem tendência e sazonalidade clara. Esse modelo se caracteriza por ponderar os valores históricos da série temporal com pesos sucessivamente menores à medida que estes dados se afastam do valor de interesse. Desta forma, os pesos comportam-se como uma função exponencial decrescente, ou seja, à observação mais recente é dado um peso maior, sendo essa redução dos pesos tão mais rápida quanto mais elevado for o valor de α , como visto a seguir:

Equação de Previsão: $Y_{t+h} = l_t + hb_t$

Equação de Nível: $l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + b_{t-1})$

Equação de Tendência: $b_t = \beta (l_t - l_{t-1}) + (1 - \beta) b_{t-1}$

Onde l_t denota uma estimativa do nível da série no momento t e b_t denota uma estimativa da tendência da série em t . Por sua vez, α é o parâmetro de alisamento de nível, e está entre $0 \leq \alpha \leq 1$ e β é o parâmetro de alisamento de tendência, também entre $0 \leq \beta \leq 1$. Fica claro que a estimativa de nível e tendência é baseada na estimativa anterior, ponderada pelos parâmetros α e β . A letra h denota o número de “passos” a frente (i.e. até que ponto do tempo queremos estimar a série temporal). Devido ao fato de que foi necessário apenas uma previsão para o ano de 2017, h seria equivalente à 1.

Tabela 2 – Resultado das previsões obtidas através do Método de Holt com 2 parâmetros

	Parâmetros - Método de Holt	Média de anos de estudo por domicílio	Violência nos estados (latrocínio - taxa em cada 100 mil habitantes)	Adequação da formação docente_G1	Adequação da formação docente_G2	Adequação da formação docente_G3	Adequação da formação docente_G4	Adequação da formação docente_G5
Distrito Federal	α	0.5	1.0	0.6	0.3	0.6	0.6	0.6
	β	0.1	0.0	1.0	0.0	1.0	1.0	1.0
	l	10.0	1.8	64.9	0.5	22.1	5.6	6.9
	b	0.0	-0.2	-1.9	0.1	2.6	-0.6	-0.2
	(MAE)	0.0	0.1	1.6	0.1	2.5	0.3	0.6
	Yt+1	10.0	1.2	64.1	0.9	22.0	4.7	8.3
Goiás	α	1.0	0.6	0.6	0.5	0.9	0.5	0.0
	β	0.0	1.0	1.0	0.1	0.2	1.0	0.0
	l	7.6	2.6	64.1	0.4	13.6	2.2	19.7
	b	0.2	-0.2	-0.2	0.0	0.7	-0.2	-0.3
	(MAE)	0.1	0.2	0.5	0.0	0.2	0.1	0.6
	Yt+1	8.2	2.8	65.3	0.5	15.5	1.8	18.5
Mato Grosso	α	1.0	0.5	0.6	0.6	0.0	0.4	0.6
	β	0.0	1.0	1.0	1.0	0.0	0.0	1.0
	l	7.7	1.6	68.6	1.1	10.3	7.9	12.1
	b	0.1	0.2	-0.6	-1.0	0.6	-0.7	1.7
	(MAE)	0.1	0.1	0.8	0.5	1.6	0.5	2.6
	Yt+1	8.2	2.0	69.6	-0.4	12.7	5.4	8.6
Mato Grosso do Sul	α	1.0	0.6	0.5	0.5	0.5	0.6	0.6
	β	0.0	1.0	1.0	0.1	0.1	1.0	1.0
	l	7.8	1.6	55.8	0.1	30.7	4.1	9.3
	b	0.1	-0.2	0.2	0.0	0.0	-0.8	0.6
	(MAE)	0.0	0.1	0.1	0.0	0.1	0.4	0.4
	Yt+1	8.1	1.4	56.2	0.1	30.8	2.9	9.8
Acre	α	1.0	1.0	0.0	0.6	0.6	0.7	0.6
	β	0.0	0.0	0.0	1.0	1.0	0.5	1.0
	l	6.9	0.9	45.0	0.4	11.8	1.4	41.4
	b	0.0	0.3	0.4	-0.1	0.3	0.3	-0.9
	(MAE)	0.0	0.1	0.5	0.1	0.7	0.1	0.5
	Yt+1	7.1	1.9	46.6	0.3	10.1	2.1	40.1

(cont.)

	Parâmetros - Método de Holt	Média de anos de estudo por domicílio	Violência nos estados (latrocínio - taxa em cada 100 mil habitantes)	Adequação da formação docente_G1	Adequação da formação docente_G2	Adequação da formação docente_G3	Adequação da formação docente_G4	Adequação da formação docente_G5
Amapá	α	1.0	0.6	1.0	0.5	0.5	0.6	0.6
	β	0.0	1.0	0.0	0.1	1.0	1.0	1.0
	l	8.2	2.0	23.2	0.1	9.6	2.0	65.1
	b	0.4	1.0	0.8	0.0	-0.8	-0.3	0.3
	(MAE)	0.1	0.7	0.3	0.0	0.3	0.2	0.3
	Yt+1	9.4	2.7	25.6	0.0	8.1	1.7	64.9
Amazonas	α	1.0	0.6	0.6	0.5	0.0	0.6	0.6
	β	0.0	0.7	1.0	0.1	0.0	1.0	1.0
	l	8.7	1.2	59.5	0.4	5.3	3.5	31.3
	b	-0.4	0.7	-1.6	0.0	0.2	0.5	0.9
	(MAE)	0.1	0.2	1.4	0.0	0.3	0.4	1.1
	Yt+1	7.7	2.7	59.0	0.3	6.1	3.9	30.3
Pará	α	1.0	0.0	0.0	0.5	0.8	0.6	0.7
	β	0.0	0.0	0.0	0.1	0.2	1.0	0.0
	l	6.8	2.2	34.6	0.1	8.1	1.8	55.4
	b	0.2	0.1	2.2	0.0	1.3	0.4	-3.9
	(MAE)	0.1	0.1	1.8	0.0	0.4	0.2	2.4
	Yt+1	7.3	2.6	43.3	0.1	11.6	2.3	42.0
Rondônia	α	1.0	0.6	0.6	0.6	0.6	0.5	0.6
	β	0.0	1.0	1.0	1.0	1.0	0.1	1.0
	l	7.3	1.0	76.1	0.6	7.0	1.4	14.9
	b	-0.1	-0.2	-0.7	-0.1	-0.2	0.0	1.0
	(MAE)	0.1	0.5	1.0	0.3	0.3	0.0	1.5
	Yt+1	6.9	2.2	77.2	1.2	7.3	1.5	12.8
Roraima	α	1.0	0.6	0.6	0.5	0.0	0.7	0.9
	β	0.0	1.0	1.0	1.0	0.0	0.5	0.2
	l	8.3	0.6	51.8	0.9	15.0	3.1	29.2
	b	0.1	1.2	2.3	-0.4	0.2	-0.6	-1.5
	(MAE)	0.1	0.9	1.4	0.1	0.8	0.2	0.5
	Yt+1	8.8	1.4	54.7	0.1	15.8	1.7	25.1
Tocantins	α	1.0	0.6	0.6	0.5	0.6	0.5	0.6
	β	0.0	1.0	1.0	0.1	1.0	1.0	1.0
	l	7.1	0.9	47.3	0.3	11.3	5.5	35.6
	b	0.1	-0.3	-0.5	0.0	1.5	-1.1	0.1
	(MAE)	0.0	0.3	1.6	0.0	1.1	0.5	1.1
	Yt+1	7.4	0.8	51.7	0.4	12.5	3.7	31.8
Paraná	α	0.5	0.6	0.0	0.6	0.6	0.6	0.6
	β	0.1	1.0	0.0	1.0	1.0	1.0	1.0
	l	8.1	0.9	58.1	0.2	14.8	2.8	24.1
	b	0.0	0.1	0.4	-0.1	-1.5	-0.5	1.7
	(MAE)	0.0	0.1	0.8	0.1	0.9	0.3	1.9
	Yt+1	8.1	1.0	59.7	0.1	13.1	2.2	22.9
Rio Grande do Sul	α	1.0	0.5	0.6	0.6	0.9	0.6	0.5
	β	0.0	0.1	1.0	1.0	0.1	1.0	0.1
	l	8.0	1.3	49.3	0.1	20.4	2.8	27.4
	b	0.3	0.0	-0.7	0.1	0.8	-0.2	0.0
	(MAE)	0.1	0.1	0.7	0.1	0.3	0.2	0.6
	Yt+1	8.8	1.4	49.6	0.1	22.6	2.8	26.4
Santa Catarina	α	0.5	0.6	0.6	0.6	0.6	0.5	0.6
	β	0.1	1.0	1.0	1.0	1.0	1.0	1.0
	l	8.2	0.8	51.1	0.6	24.1	2.7	21.5
	b	0.0	0.2	-0.5	-0.1	-0.3	-0.4	1.3
	(MAE)	0.0	0.1	1.1	0.1	0.8	0.2	2.2
	Yt+1	8.2	1.0	53.5	0.6	26.1	2.0	17.9

(cont.)

	Parâmetros - Método de Holt	Média de anos de estudo por domicílio	Violência nos estados (latrocínio - taxa em cada 100 mil habitantes)	Adequação da formação docente_G1	Adequação da formação docente_G2	Adequação da formação docente_G3	Adequação da formação docente_G4	Adequação da formação docente_G5
Alagoas	α	1.0	0.6	1.0	0.5	1.0	0.0	1.0
	β	0.0	1.0	0.0	0.1	0.0	0.0	0.0
	l	5.7	1.8	33.6	0.3	6.1	1.7	58.3
	b	0.2	-0.2	2.2	0.0	0.4	-0.1	-2.5
	(MAE)	0.1	0.2	1.1	0.0	0.2	0.1	1.1
	Yt+1	6.3	1.8	41.3	0.3	7.4	1.3	49.9
Bahia	α	1.0	0.6	0.3	0.5	0.6	0.5	1.0
	β	0.0	1.0	0.0	0.1	1.0	0.1	0.0
	l	6.6	1.3	37.0	0.3	7.2	2.1	53.4
	b	0.2	0.1	1.9	0.0	0.7	0.0	-2.6
	(MAE)	0.1	0.1	1.4	0.0	0.4	0.0	1.3
	Yt+1	7.2	1.4	44.0	0.3	8.1	2.2	44.3
Ceará	α	1.0	0.6	1.0	0.6	0.6	0.7	0.0
	β	0.0	1.0	0.0	1.0	1.0	0.3	0.0
	l	6.4	0.8	35.9	0.4	16.6	11.8	35.3
	b	0.1	-0.1	3.2	-0.2	0.1	-2.7	-0.4
	(MAE)	0.0	0.2	1.5	0.1	0.3	0.9	0.6
	Yt+1	6.7	1.0	46.9	0.1	15.8	5.1	33.7
Maranhão	α	1.0	0.6	1.0	0.5	0.5	0.5	0.5
	β	0.0	1.0	0.0	0.1	1.0	0.1	0.0
	l	6.1	1.1	20.4	0.2	11.9	1.3	66.2
	b	0.0	0.6	1.7	0.0	-0.7	0.0	-1.0
	(MAE)	0.0	0.4	0.6	0.0	0.3	0.1	0.7
	Yt+1	6.0	1.9	25.5	0.3	10.7	1.4	62.7
Paraíba	α	1.0	0.6	0.0	0.5	0.5	0.6	1.0
	β	0.0	1.0	0.0	0.1	1.0	1.0	0.0
	l	6.5	0.5	44.1	0.6	10.9	2.2	42.2
	b	0.0	0.7	0.9	0.0	1.5	0.2	-2.6
	(MAE)	0.0	0.5	1.0	0.0	0.6	0.3	0.9
	Yt+1	6.7	1.0	47.7	0.5	13.6	1.8	34.3
Pernambuco	α	1.0	0.3	0.4	0.5	0.5	0.5	1.0
	β	0.0	0.0	0.0	0.1	1.0	1.0	0.0
	l	6.9	0.9	36.4	0.2	13.0	1.9	48.5
	b	0.0	0.3	1.1	0.0	0.6	0.2	-1.9
	(MAE)	0.0	0.2	0.8	0.0	0.3	0.1	0.7
	Yt+1	7.1	2.0	40.3	0.2	14.0	2.3	42.5
Piauí	α	1.0	0.6	0.5	0.5	0.6	0.6	0.6
	β	0.0	1.0	1.0	0.1	1.0	1.0	1.0
	l	5.8	1.1	40.0	0.2	11.3	1.6	46.9
	b	0.3	0.4	1.8	0.0	0.7	0.5	-3.0
	(MAE)	0.1	0.2	0.7	0.0	0.6	0.4	1.9
	Yt+1	6.6	1.7	43.3	0.1	11.4	1.7	43.5
Rio Grande do Norte	α	1.0	0.0	1.0	0.5	0.5	0.5	1.0
	β	0.0	0.0	0.0	0.1	1.0	0.1	0.0
	l	6.6	1.8	58.2	0.2	5.3	1.9	34.4
	b	0.3	-0.1	1.3	0.0	0.5	0.0	-1.8
	(MAE)	0.1	0.1	0.5	0.0	0.2	0.1	0.6
	Yt+1	7.5	1.4	62.2	0.2	6.1	2.1	28.9
Sergipe	α	1.0	0.5	0.0	0.5	0.8	0.6	0.0
	β	0.0	1.0	0.0	0.1	0.2	1.0	0.0
	l	6.2	1.5	49.9	1.1	8.0	1.4	39.6
	b	0.4	0.6	1.0	0.0	0.6	-0.2	-1.4
	(MAE)	0.1	0.3	1.1	0.0	0.2	0.1	1.3
	Yt+1	7.3	2.5	53.9	1.1	9.6	1.1	34.0

(cont.)

	Parâmetros - Método de Holt	Média de anos de estudo por domicílio	Violência nos estados (latrocínio - taxa em cada 100 mil habitantes)	Adequação da formação docente_G1	Adequação da formação docente_G2	Adequação da formação docente_G3	Adequação da formação docente_G4	Adequação da formação docente_G5
Espírito Santo	α	1.0	0.6	0.6	0.6	0.6	1.0	0.6
	β	0.0	1.0	1.0	1.0	1.0	0.0	1.0
	I	7.8	1.3	57.2	0.2	26.7	7.6	8.3
	b	0.1	-0.4	1.7	0.3	-2.3	-0.2	0.5
	(MAE)	0.0	0.4	1.5	0.2	2.7	0.1	0.9
	Yt+1	8.1	1.2	57.5	0.4	28.5	6.9	6.7
Minas Gerais	α	1.0	0.6	1.0	0.5	0.6	0.6	1.0
	β	0.0	1.0	0.0	0.1	0.0	1.0	0.0
	I	7.5	0.3	51.8	0.6	19.1	4.4	24.1
	b	0.3	0.3	0.7	0.0	0.3	0.2	-1.2
	(MAE)	0.1	0.2	0.3	0.0	0.2	0.2	0.4
	Yt+1	8.3	0.7	54.2	0.5	20.1	4.4	20.5
Rio de Janeiro	α	1.0	0.6	0.6	0.5	0.1	0.6	0.6
	β	0.0	1.0	1.0	0.1	0.0	1.0	1.0
	I	8.9	0.9	24.2	0.4	29.4	5.8	40.2
	b	0.0	-0.1	-0.2	0.0	-0.5	-0.5	1.2
	(MAE)	0.0	0.3	0.7	0.0	0.4	0.3	0.9
	Yt+1	9.1	1.5	26.1	0.4	27.4	5.3	41.0
São Paulo	α	1.0	0.6	0.6	0.6	0.6	0.6	1.0
	β	0.0	1.0	1.0	1.0	1.0	1.0	0.0
	I	8.9	0.9	64.2	0.3	18.8	4.2	12.5
	b	0.0	-0.1	-0.6	0.1	1.0	0.2	-0.7
	(MAE)	0.0	0.1	0.9	0.1	0.7	0.2	0.3
	Yt+1	9.1	0.8	65.4	0.4	19.7	4.1	10.2

A fim de verificar a adequação da estratégia utilizada, foi calculado o MAE (Mean Absolute Error) de cada uma das estimações. O MAE é uma medida de erro que estima o erro da previsão obtida como a diferença entre as previsões e a observação atual. Devido ao fato de que a medida usa o erro absoluto, a direção dos erros não é levada em consideração, logo, deve-se levar em conta que erros positivos e negativos serão cancelados uns com os outros. Desta forma, através do MAE não é possível observar com precisão a magnitude de um erro, por exemplo, um erro muito grande que seria causado por um evento de pouca frequência. Contudo, para os fins necessários e pela falta de eventos de impacto, a medida é devidamente adequada.

Como é possível ver pelas tabelas de resultados acima, os valores correspondentes ao MAE são consideravelmente baixos, o que nos leva a concluir que o modelo de previsão obtido pelas Estimativas de Holt foi preciso para obtenção dos dados faltantes.

6. Discussão acerca das variáveis

Dada a metodologia exposta na seção acima, é válido realizar uma discussão acerca das escolhas das variáveis e porque elas foram construídas da forma explicada acima. Vale ainda, explicar porque algumas variáveis foram usadas em detrimento de outras e como elas podem gerar valor ao estudo.

A escolha das duas variáveis de interesse (*rendimento nominal mensal domiciliar* e *escolaridade média em anos de estudo (pessoas de 25 anos ou mais)*) foi feita com base em estudos preliminares em relação ao cumprimento da condicionalidade de educação do programa. Com base nos dados disponibilizados pelo Sistema Presença (*Figura 1*) e pela PNAD Contínua, como descritos na seção III, ficou evidente uma relação inversa entre a taxa de cumprimento da condicionalidade e os indicadores para as duas variáveis para cada unidade da federação. Em outras palavras, uma visualização inicial dos dados mostrou que estados com indicadores de escolaridade média por anos de estudo e rendimento mensal domiciliar maior, em geral, reportavam taxas menores de cumprimento da condicionalidade de educação do programa. É interessante, portanto, observar que esse resultado vai contra a intuição comum, que nos diz que estados com famílias de maior renda e maior nível de escolaridade seriam mais propícios a satisfazerem as condicionalidades da forma exigível. Esse resultado preliminar pode ser visto nas *figuras 2 e 3*, abaixo, e explica a escolha da regressão inicial do estudo [1], apresentada na seção anterior.

Figura 2 – Cumprimento de condicionalidade de educação do Programa Bolsa Família versus rendimento mensal domiciliar per capita

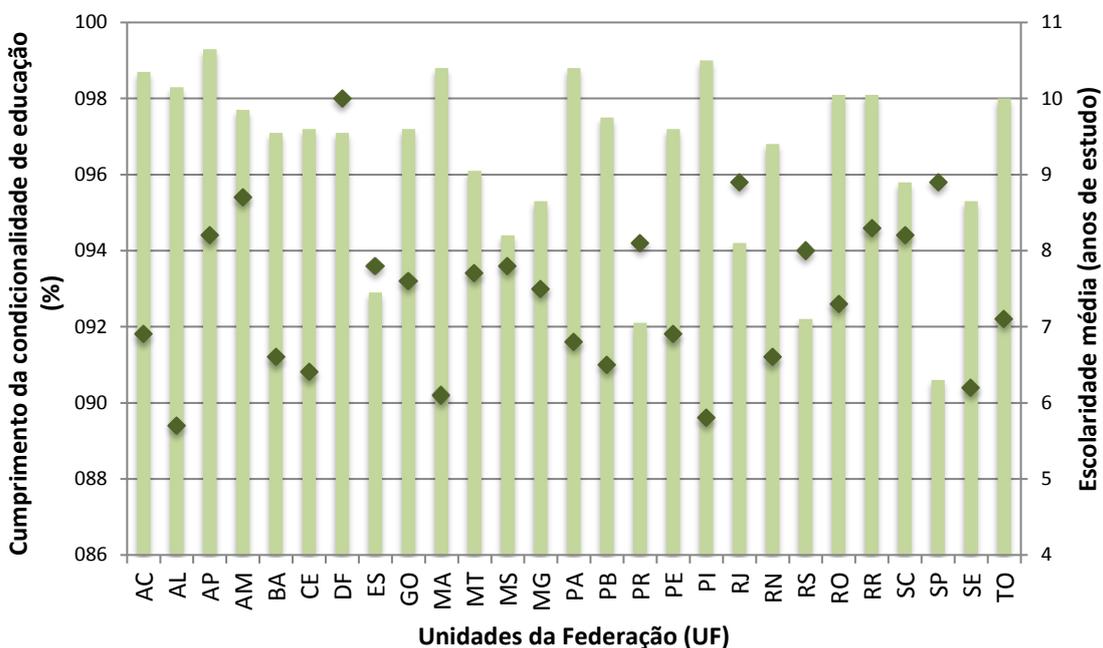
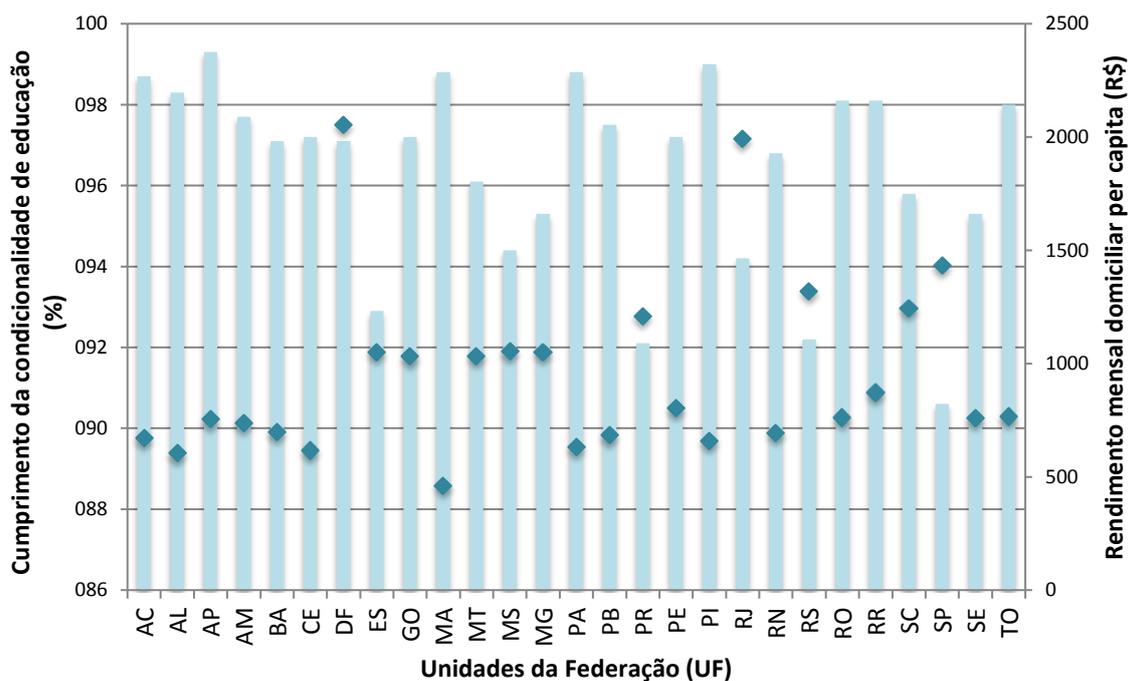


Figura 3 – Cumprimento de condicionalidade de educação do Programa Bolsa Família versus escolaridade média (por anos de estudo)



Com o objetivo de apresentar uma estatística descritiva, para cada um dos gráficos acima foi calculado o Coeficiente de Correlação dos dados de interesse. O Coeficiente de Correlação, ρ , é um valor entre $0 \leq \rho \leq 1$ que pode ser interpretado como o grau de correlação entre duas variáveis. A correlação encontrada entre o “Cumprimento da condicionalidade de educação” versus “Rendimento mensal domiciliar per capita” e “Escolaridade média (anos de estudo)” foi de -0,587 e -0,441, respectivamente. Com base nesses resultados preliminares concluímos que ambos os indicadores têm aparente correlação negativa com o cumprimento da condicionalidade, e que essa correlação é mais forte para os dados de “Rendimento mensal domiciliar per capita”.

Mais adiante, nos deparamos com as escolhas das variáveis de controle para as unidades da federação. A ideia era encontrar outras variáveis que podem influenciar no alto ou baixo índice de cumprimento da condicionalidade de educação por parte das crianças beneficiárias do programa. A partir de relatórios trimestrais divulgados pelo Ministério do Desenvolvimento Social (MDS) sobre a evolução e funcionamento do programa, ficou claro que algumas das principais razões pela qual beneficiários deixam de cumprir a condicionalidade de educação são: preços das tarifas de transporte, violência nas cidades e falta de motivação por parte dos alunos.

Para representar o papel da violência neste cenário, foi escolhida a variável de *taxa de latrocínios a cada 100 mil habitantes, por unidade da federação*. Dentre outras opções de indicadores de violência como número de homicídios, furtos e roubos ou lesão corporal, foi escolhida a variável que indica o número de latrocínios pela natureza do crime. Por ser um crime de natureza não-passional, à qual todos os cidadãos estão sujeitos e que representam uma ameaça suficiente para prevenirem cidadãos de áreas com maior propensão para esse tipo de crime, a realizarem atividades cotidianas. A escolha de utilizar a variável em taxa ao invés do número absoluto de casos se deu com o objetivo de neutralizar o viés que seria causado pelo tamanho da população em cada estado. Naturalmente, estados com maior população teriam um número maior de casos de latrocínio, não necessariamente indicando um nível maior de violência no estado, mas sim consequência de uma população maior. Estados que apresentam maiores taxas de latrocínio são considerados de mais alto risco e é esperado um efeito negativo dessa variável sobre a taxa de cumprimento da condicionalidade de educação.

Pela pouca disponibilidade de dados de mobilidade e transporte público nos estados brasileiros, foi adotado nesse trabalho uma perspectiva alternativa quanto ao papel do transporte público na decisão sobre a frequência escolar dos alunos. A escolha da variável foi a de *comprometimento da renda de um salário mínimo com a tarifa de transporte público (%)*. A ideia é que, mesmo que esses dados não se apliquem diretamente às famílias beneficiárias do programa, ele mostra o peso das tarifas de transporte no orçamento dos cidadãos brasileiros. Extrapolando os resultados, podemos encontrar até que ponto o alto custo de transporte nos estados brasileiros pode estar prejudicando o cumprimento da condicionalidade de educação do Bolsa Família. De acordo com relatórios do MDS, algumas famílias beneficiárias do programa relatam a falta de dinheiro para o transporte público como motivo para a baixa frequência dos alunos nas escolas, logo, é importante incluir essa variável de controle para cada unidade da federação.

Por fim, o último motivo citado pelas famílias pode ser aproximado por diversas variáveis. Neste trabalho, será usada a variável do *percentual de docentes por grupo do indicador de adequação da formação do docente*. A visão adotada é de que professores com ensino superior, especialmente na área da disciplina em que lecionam, tem a formação adequada e assim são mais bem preparados, podendo motivar melhor seus alunos. Dessa forma, estados com uma alta porcentagem de docentes pertencentes ao grupo 5 (docentes sem formação superior) por exemplo, deve apresentar uma taxa de cumprimento mais baixa do que outros estados com um maior número docentes pertencentes aos 4 primeiros grupos por exemplo.

7. Tabelas com os resultados das regressões

Tabela 3 – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade, data =
dados_novos)

Residuals:
    Min       1Q   Median       3Q      Max
-5.749 -1.013  0.383  1.094  6.269

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  99.630548   1.989468   50.08 < 0.0000000000000002 ***
Rendimento  -0.003327   0.000901   -3.69   0.00036 ***
Escolaridade -0.012284   0.351663   -0.03   0.97220
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.09 on 105 degrees of freedom
Multiple R-squared:  0.291, Adjusted R-squared:  0.277
F-statistic: 21.5 on 2 and 105 DF, p-value: 0.0000000148

                                ME RMSE MAE      MPE MAPE  MASE
Training set 0.000000000000000131 2.06 1.5 -0.0472 1.58 0.769
[1] "AIC: 470.347939993325"
[1] "BIC: 481.076464901822"
```

Tabela 4 – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem DF)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade, data =
dados_Sem_DF)

Residuals:
    Min       1Q   Median       3Q      Max
-4.899 -0.665  0.242  0.768  4.856

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept) 100.361626   1.499218   66.94 < 0.0000000000000002 ***
Rendimento  -0.007125   0.000792   -9.00   0.0000000000000015 ***
Escolaridade  0.357504   0.267934    1.33   0.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.56 on 101 degrees of freedom
Multiple R-squared:  0.615, Adjusted R-squared:  0.607
F-statistic: 80.5 on 2 and 101 DF, p-value: <0.0000000000000002

                                ME RMSE MAE      MPE MAPE  MASE
Training set 0.000000000000000273 1.54 1.12 -0.0266 1.18 0.562
[1] "AIC: 392.848324908443"
[1] "BIC: 403.425888505009"
```

Tabela 5 – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem intercepto)

```
lm(formula = Condicionalidade ~ -1 + Rendimento + Escolaridade,
   data = dados_novos)

Residuals:
    Min       1Q   Median       3Q      Max
-29.42  -3.60   2.10   7.88  22.33

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
Rendimento  -0.03224    0.00344   -9.38  0.0000000000000014 ***
Escolaridade 16.88211    0.49297   34.25 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.4 on 106 degrees of freedom
Multiple R-squared:  0.989, Adjusted R-squared:  0.988
F-statistic: 4.6e+03 on 2 and 106 DF, p-value: <0.0000000000000002

              ME RMSE  MAE   MPE MAPE  MASE
Training set  1.02 10.3  8.03 0.969 8.31 4.11
[1] "AIC: 815.487738877004"
[1] "BIC: 823.534132558376"
```

Tabela 6 – Efeito do rendimento domiciliar e escolaridade no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal e sem intercepto)

```
lm(formula = Condicionalidade ~ -1 + Rendimento + Escolaridade,
   data = dados_Sem_DF)

Residuals:
    Min       1Q   Median       3Q      Max
-30.07  -3.41   2.24   7.63  22.14

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
Rendimento  -0.03491    0.00452   -7.72  0.0000000000000081 ***
Escolaridade 17.21721    0.61279   28.10 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.5 on 102 degrees of freedom
Multiple R-squared:  0.988, Adjusted R-squared:  0.988
F-statistic: 4.33e+03 on 2 and 102 DF, p-value: <0.0000000000000002

              ME RMSE  MAE   MPE MAPE  MASE
Training set  1.05 10.4  8.02 1.01 8.29 4.02
[1] "AIC: 787.591716201808"
[1] "BIC: 795.524888899232"
```

Tabela 7 – Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e adequação da formação docente no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
Violencia_Taxa +
  Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
  Docente_G5, data = dados_Sem_DF)

Residuals:
    Min       1Q   Median       3Q      Max
-4.350  -0.749   0.100   0.958   3.644

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  123.51997   23.13776    5.34 0.00000065 ***
Rendimento   -0.00525    0.00105   -4.99 0.00000272 ***
Escolaridade  0.37296    0.26366    1.41  0.161
Violencia_Taxa -0.04540    0.26767   -0.17  0.866
Transporte   -0.02338    0.09231   -0.25  0.801
Docente_G1   -0.24683    0.23306   -1.06  0.292
Docente_G2   -0.47811    0.56375   -0.85  0.399
Docente_G3   -0.27626    0.22935   -1.20  0.231
Docente_G4   -0.42632    0.24533   -1.74  0.086 .
Docente_G5   -0.21333    0.23067   -0.92  0.357
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.44 on 94 degrees of freedom
Multiple R-squared:  0.696, Adjusted R-squared:  0.667
F-statistic: 23.9 on 9 and 94 DF, p-value: <0.00000000000000002

              ME RMSE MAE      MPE MAPE MASE
Training set 0.000000000000000273 1.37 1.05 -0.0211  1.1 0.525
[1] "AIC: 382.098732820707"
[1] "BIC: 411.187032711263"
```

Tabela 8 – Efeito do rendimento domiciliar, escolaridade, comprometimento da renda com transporte público e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Transporte
+
  Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5,
  data = dados_Sem_DF)

Residuals:
    Min       1Q   Median       3Q      Max
-4.391  -0.762   0.094   0.952   3.634

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  123.66409   23.00365    5.38 0.00000054 ***
Rendimento   -0.00522    0.00104   -5.05 0.00000217 ***
```

```

Escolaridade    0.36109    0.25291    1.43    0.157
Transporte     -0.02281    0.09178   -0.25    0.804
Docente_G1     -0.24870    0.23161   -1.07    0.286
Docente_G2     -0.47059    0.55912   -0.84    0.402
Docente_G3     -0.27705    0.22812   -1.21    0.228
Docente_G4     -0.42476    0.24390   -1.74    0.085 .
Docente_G5     -0.21516    0.22923   -0.94    0.350
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.43 on 95 degrees of freedom
Multiple R-squared:  0.696, Adjusted R-squared:  0.671
F-statistic: 27.2 on 8 and 95 DF,  p-value: <0.00000000000000002

```

```

                                ME RMSE  MAE      MPE MAPE  MASE
Training set -0.000000000000000137 1.37 1.05 -0.0211  1.1 0.524
[1] "AIC: 380.130553765898"
[1] "BIC: 406.574462757311"

```

Tabela 9 – Efeito do rendimento domiciliar, escolaridade, taxa de violência e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF)

```

lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
  Violencia_Taxa +
    Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5,
  data = dados_Sem_DF)

```

```

Residuals:
  Min      1Q  Median      3Q      Max
-4.343 -0.782  0.088  0.970  3.704

```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  123.332259  23.011690    5.36 0.00000058 ***
Rendimento   -0.005348   0.000973   -5.50 0.00000032 ***
Escolaridade  0.365587   0.260754    1.40  0.164
Violencia_Taxa -0.042947   0.266177   -0.16  0.872
Docente_G1   -0.247842   0.231878   -1.07  0.288
Docente_G2   -0.488257   0.559546   -0.87  0.385
Docente_G3   -0.276281   0.228213   -1.21  0.229
Docente_G4   -0.426808   0.244114   -1.75  0.084 .
Docente_G5   -0.213963   0.229515   -0.93  0.354
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.43 on 95 degrees of freedom
Multiple R-squared:  0.696, Adjusted R-squared:  0.671
F-statistic: 27.2 on 8 and 95 DF,  p-value: <0.00000000000000002

```

```

                                ME RMSE  MAE      MPE MAPE  MASE
Training set -0.000000000000000137 1.37 1.05 -0.0211  1.11 0.526
[1] "AIC: 380.169662730279"
[1] "BIC: 406.613571721693"

```

Tabela 10 – Efeito do rendimento domiciliar, escolaridade e adequação da formação docente no cumprimento da condicionalidade de educação (sem o DF)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
  Docente_G1 +
    Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
  dados_Sem_DF)

Residuals:
    Min       1Q   Median       3Q      Max
-4.382 -0.789  0.083  0.965  3.692

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  123.473089  22.878185   5.40 0.00000049 ***
Rendimento   -0.005322   0.000954  -5.58 0.00000022 ***
Escolaridade  0.354521    0.250292   1.42  0.160
Docente_G1   -0.249589    0.230448  -1.08  0.281
Docente_G2   -0.480898    0.554848  -0.87  0.388
Docente_G3   -0.277027    0.227006  -1.22  0.225
Docente_G4   -0.425314    0.242697  -1.75  0.083 .
Docente_G5   -0.215681    0.228102  -0.95  0.347
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.42 on 96 degrees of freedom
Multiple R-squared:  0.696, Adjusted R-squared:  0.674
F-statistic: 31.4 on 7 and 96 DF, p-value: <0.00000000000000002

              ME RMSE MAE      MPE MAPE MASE
Training set 0.0000000000000000273 1.37 1.05 -0.0211  1.1 0.526
[1] "AIC: 378.198158499906"
[1] "BIC: 401.997676592179"
```

Tabela 11 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 1 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
  Violencia_Taxa +
    Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
    Docente_G5, data = dados_Sem_DF_G1)

Residuals:
    Min       1Q   Median       3Q      Max
-4.6258 -0.7674  0.1125  0.9718  3.5670

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  127.214090  22.418261   5.675 1.50e-07 ***
Rendimento   -0.005244   0.001048  -5.003 2.58e-06 ***
Escolaridade  0.389250    0.261802   1.487  0.1404
Violencia_Taxa -0.020691    0.264405  -0.078  0.9378
Transporte    -2.634142    9.194099  -0.287  0.7751
```

```

Docente_G2      -0.134713    0.243341   -0.554    0.5812
Docente_G3      -0.313985    0.221809   -1.416    0.1602
Docente_G4      -0.453103    0.241417   -1.877    0.0636 .
Docente_G5      -0.251583    0.222974   -1.128    0.2620
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.433 on 95 degrees of freedom
Multiple R-squared:  0.6948, Adjusted R-squared:  0.669
F-statistic: 27.03 on 8 and 95 DF, p-value: < 2.2e-16

```

```

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -6.83185e-16 1.369887 1.039999 -0.02119984 1.096328
0.5216998
[1] "AIC: 380.602703481679"
[1] "BIC: 407.046612473093"

```

Tabela 12 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 2 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```

lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
Violencia_Taxa +
  Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
  Docente_G5, data = dados_Sem_DF_G2)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-4.4908 -0.7705  0.1019  0.9264  3.5902

```

```

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  125.234865  22.433464   5.583 2.24e-07 ***
Rendimento   -0.005254   0.001046  -5.022 2.39e-06 ***
Escolaridade  0.384035    0.260257   1.476  0.1434
Violencia_Taxa -0.037919    0.265445  -0.143  0.8867
Transporte   -2.538740    9.167209  -0.277  0.7824
Docente_G1   -0.265943    0.224605  -1.184  0.2393
Docente_G3   -0.293618    0.222107  -1.322  0.1894
Docente_G4   -0.443124    0.238799  -1.856  0.0666 .
Docente_G5   -0.231170    0.223113  -1.036  0.3028
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.431 on 95 degrees of freedom
Multiple R-squared:  0.6959, Adjusted R-squared:  0.6703
F-statistic: 27.17 on 8 and 95 DF, p-value: < 2.2e-16

```

```

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -5.466485e-16 1.367354 1.042479 -0.02112813 1.098839
0.5229438
[1] "AIC: 380.217747338917"
[1] "BIC: 406.661656330331"

```

Tabela 13 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 3 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
Violencia_Taxa +
  Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
  Docente_G5, data = dados_Sem_DF_G3)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5190 -0.7597  0.1223  0.9348  3.5988

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  125.463969  22.606790   5.550 2.58e-07 ***
Rendimento   -0.005186   0.001036  -5.003 2.58e-06 ***
Escolaridade  0.381703    0.261771   1.458  0.1481
Violencia_Taxa -0.035527   0.265574  -0.134  0.8939
Transporte   -2.662468    9.161499  -0.291  0.7720
Docente_G1   -0.268832    0.226560  -1.187  0.2384
Docente_G2   -0.260598    0.261085  -0.998  0.3208
Docente_G4   -0.435712    0.243341  -1.791  0.0766 .
Docente_G5   -0.233335    0.225090  -1.037  0.3025
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.431 on 95 degrees of freedom
Multiple R-squared:  0.6956, Adjusted R-squared:  0.67
F-statistic: 27.14 on 8 and 95 DF, p-value: < 2.2e-16

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -6.83185e-16 1.367952 1.042975 -0.02114603 1.099425
0.5231927
[1] "AIC: 380.308697523727"
[1] "BIC: 406.752606515141"
```

Tabela 14 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 4 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
Violencia_Taxa +
  Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
  Docente_G5, data = dados_Sem_DF_G4)

Residuals:
    Min       1Q   Median       3Q      Max
-4.0474 -0.7026  0.0487  0.9381  3.7497

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
```

```

(Intercept)    120.098884  22.635618   5.306 7.32e-07 ***
Rendimento     -0.005121   0.001035  -4.948 3.23e-06 ***
Escolaridade   0.331565    0.257278   1.289  0.2006
Violencia_Taxa -0.039511    0.266949  -0.148  0.8826
Transporte     -1.922185    9.193417  -0.209  0.8348
Docente_G1     -0.209534    0.227220  -0.922  0.3588
Docente_G2     -0.803218    0.361703  -2.221  0.0288 *
Docente_G3     -0.247602    0.225651  -1.097  0.2753
Docente_G5     -0.177338    0.225146  -0.788  0.4329
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.434 on 95 degrees of freedom
Multiple R-squared:  0.6944, Adjusted R-squared:  0.6687
F-statistic: 26.98 on 8 and 95 DF, p-value: < 2.2e-16

```

```

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 1.366303e-16 1.370688 1.056746 -0.02124294 1.113605
0.5301004
[1] "AIC: 380.724248613504"
[1] "BIC: 407.168157604918"

```

Tabela 15 - Efeito do rendimento domiciliar, escolaridade, taxa de violência, comprometimento da renda com transporte público e retirada do grupo docente 5 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```

lm(formula = Condicionalidade ~ Rendimento + Escolaridade +
Violencia_Taxa +
  Transporte + Docente_G1 + Docente_G2 + Docente_G3 + Docente_G4 +
  Docente_G5, data = dados_Sem_DF_G5)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-4.7622 -0.7443  0.1155  0.9637  3.5444

```

```

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  128.692927  22.535945   5.711 1.28e-07 ***
Rendimento   -0.005271   0.001051  -5.015 2.46e-06 ***
Escolaridade  0.398293    0.262380   1.518  0.1323
Violencia_Taxa -0.008452    0.265022  -0.032  0.9746
Transporte    -2.904964    9.212015  -0.315  0.7532
Docente_G1    -0.304378    0.225654  -1.349  0.1806
Docente_G2     0.028085    0.236081   0.119  0.9056
Docente_G3    -0.328364    0.223184  -1.471  0.1445
Docente_G4    -0.464552    0.242239  -1.918  0.0581 .
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.437 on 95 degrees of freedom
Multiple R-squared:  0.6931, Adjusted R-squared:  0.6672
F-statistic: 26.81 on 8 and 95 DF, p-value: < 2.2e-16

```

```

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -5.466485e-16 1.373661 1.034907 -0.02130786 1.091114
0.5191454

```

```
[1] "AIC: 381.174927832704"
[1] "BIC: 407.618836824118"
```

Tabela 16 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 1 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Docente_G1 +
  Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
  dados_Sem_DF_G1)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-4.6357 -0.7818  0.1278  0.9794  3.5587
```

Coefficients: (1 not defined because of singularities)

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  127.084628  22.190589   5.727 1.15e-07 ***
Rendimento   -0.005344   0.000951  -5.619 1.84e-07 ***
Escolaridade  0.376467    0.247510   1.521  0.1315
Docente_G2   -0.140881    0.239183  -0.589  0.5572
Docente_G3   -0.314565    0.219572  -1.433  0.1552
Docente_G4   -0.453206    0.238537  -1.900  0.0604 .
Docente_G5   -0.253252    0.220686  -1.148  0.2540
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.419 on 97 degrees of freedom
Multiple R-squared: 0.6945, Adjusted R-squared: 0.6756
F-statistic: 36.75 on 6 and 97 DF, p-value: < 2.2e-16

```
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -1.366136e-16 1.37051 1.043548 -0.02121099 1.099767
0.5234801
```

```
[1] "AIC: 376.697285124374"
[1] "BIC: 397.852412317505"
```

Tabela 17 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 2 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Docente_G1 +
  Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
  dados_Sem_DF_G2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-4.5179 -0.7899  0.0895  0.9562  3.6449
```

Coefficients: (1 not defined because of singularities)

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.252e+02  2.220e+01   5.638 1.69e-07 ***
Rendimento   -5.340e-03  9.484e-04  -5.630 1.75e-07 ***
```

```

Escolaridade  3.672e-01  2.463e-01  1.491  0.1392
Docente_G1   -2.689e-01  2.222e-01  -1.210  0.2292
Docente_G3   -2.946e-01  2.199e-01  -1.340  0.1834
Docente_G4   -4.428e-01  2.360e-01  -1.876  0.0636 .
Docente_G5   -2.336e-01  2.208e-01  -1.058  0.2925
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.417 on 97 degrees of freedom
Multiple R-squared:  0.6956, Adjusted R-squared:  0.6767
F-statistic: 36.94 on 6 and 97 DF, p-value: < 2.2e-16

      ME      RMSE      MAE      MPE      MAPE      MASE
Training set -2.73342e-16 1.36803 1.045384 -0.0211403 1.101597
0.5244009
[1] "AIC: 376.320534915688"
[1] "BIC: 397.475662108819"

```

Tabela 18- Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 3 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```

lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Docente_G1 +
  Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
  dados_Sem_DF_G3)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5453 -0.7753  0.1288  0.9679  3.5944

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.254e+02  2.238e+01  5.605 1.96e-07 ***
Rendimento   -5.276e-03  9.449e-04 -5.584 2.14e-07 ***
Escolaridade  3.652e-01  2.481e-01  1.472  0.1444
Docente_G1   -2.719e-01  2.241e-01  -1.214  0.2279
Docente_G2   -2.603e-01  2.585e-01  -1.007  0.3165
Docente_G4   -4.354e-01  2.407e-01  -1.809  0.0735 .
Docente_G5   -2.359e-01  2.227e-01  -1.059  0.2920
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.417 on 97 degrees of freedom
Multiple R-squared:  0.6953, Adjusted R-squared:  0.6764
F-statistic: 36.89 on 6 and 97 DF, p-value: < 2.2e-16

      ME      RMSE      MAE      MPE      MAPE      MASE
Training set -1.366136e-16 1.368666 1.046128 -0.02115909 1.102434
0.5247745
[1] "AIC: 376.417208345956"
[1] "BIC: 397.572335539087"

```

Tabela 19 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 4 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Docente_G1
+
  Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
dados_Sem_DF_G4)

Residuals:
    Min       1Q   Median       3Q      Max
-4.0822 -0.7553  0.0377  0.9437  3.7363

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.201e+02  2.238e+01   5.368 5.44e-07 ***
Rendimento   -5.182e-03  9.333e-04  -5.552 2.46e-07 ***
Escolaridade  3.167e-01  2.445e-01   1.295  0.1983
Docente_G1   -2.127e-01  2.245e-01  -0.947  0.3458
Docente_G2   -7.981e-01  3.533e-01  -2.259  0.0261 *
Docente_G3   -2.489e-01  2.233e-01  -1.115  0.2678
Docente_G5   -1.801e-01  2.225e-01  -0.810  0.4202
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.42 on 97 degrees of freedom
Multiple R-squared:  0.6942, Adjusted R-squared:  0.6753
F-statistic: 36.7 on 6 and 97 DF, p-value: < 2.2e-16

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set  0 1.371146 1.058487 -0.02124996 1.115208 0.530974
[1] "AIC: 376.793810007269"
[1] "BIC: 397.9489372004"
```

Tabela 20 - Efeito do rendimento domiciliar, escolaridade, e retirada do grupo docente 5 no cumprimento da condicionalidade de educação (sem os dados referentes ao Distrito Federal)

```
lm(formula = Condicionalidade ~ Rendimento + Escolaridade + Docente_G1
+
  Docente_G2 + Docente_G3 + Docente_G4 + Docente_G5, data =
dados_Sem_DF_G5)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7647 -0.7720  0.1399  0.9784  3.5452

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.286e+02  2.231e+01   5.763 9.79e-08 ***
Rendimento   -5.391e-03  9.518e-04  -5.663 1.51e-07 ***
Escolaridade  3.883e-01  2.480e-01   1.566  0.1207
Docente_G1   -3.068e-01  2.233e-01  -1.374  0.1726
Docente_G2   2.501e-02  2.321e-01   0.108  0.9144
Docente_G3   -3.294e-01  2.209e-01  -1.491  0.1393
```

Docente_G4 -4.657e-01 2.394e-01 -1.945 0.0546 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.423 on 97 degrees of freedom

Multiple R-squared: 0.6927, Adjusted R-squared: 0.6737

F-statistic: 36.45 on 6 and 97 DF, p-value: < 2.2e-16

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	2.73342e-16	1.374382	1.03933	-0.02132164	1.095448	0.5213641

[1] "AIC: 377.284073099135"

[1] "BIC: 398.439200292266"

8. Discussão acerca dos resultados

A primeira regressão rodada [1] foi a versão mais simples do modelo, incluindo apenas as duas variáveis explicativas de maior interesse: i) Rendimento nominal mensal domiciliar e ii) Escolaridade média em anos de estudo (pessoas de 25 anos ou mais). Além do mais, foram incluídos os dados do Distrito Federal para fins de comparação com um modelo que será rodado mais para frente sem a inclusão dos mesmos.

De acordo com a *Tabela 3* podemos ver que a variável rendimento mensal domiciliar obteve resultados bastante significativos, sendo significante a um nível menor do que 0.01% e apresentando um p-valor extremamente baixo. Para este valor de significância, o p-valor é maior do que a estatística t encontrada, o que significa que rejeitamos a hipótese nula de que o coeficiente β_1 é zero que apontaria para uma falta de correlação entre as variáveis – indicando a validade interna do resultado. Para essa variável, o coeficiente encontrado foi negativo (-0.003327), indicando que o impacto do rendimento mensal domiciliar per capita na taxa de cumprimento da condicionalidade é negativo. Em outras palavras, um aumento na renda per capita de um estado, potencialmente diminui a taxa de cumprimento apresentada pelos beneficiários do programa neste mesmo estado. De forma contrária, o coeficiente encontrado para a segunda variável (escolaridade média em anos de estudo), foi não significativo. Nesse caso, a hipótese nula de que o coeficiente é zero, não será rejeitada, nos levando a concluir que mudanças na escolaridade da população dos estados, pouco afeta a taxa de cumprimento da condicionalidade apresentada pelos mesmos. Vale ressaltar ainda, a significância do intercepto, também a um nível menor do que 0.01%, indicando que a constante deve ser mantida na regressão.

Para verificar validade do ajuste dos dados no modelo, utilizamos a variável R^2 Ajustada. O R^2 , ou coeficiente de determinação, é uma medida de ajuste de um modelo estatístico em relação aos valores observados. O R^2 varia entre 0 e 1 e indica o quanto as variáveis explicativas conseguem explicar os valores observados para a variável dependente, y. Quanto maior o R^2 , mais explicativo é o modelo e melhor ele se ajusta à amostra. O R^2 sempre aumenta à medida que mais variáveis explicativas são adicionadas ao modelo, nem sempre significando que há melhor ajuste das variáveis, portanto, para uma medida menos enviesada, usaremos o R^2 ajustado, que ajusta o

coeficiente ao número de preditores do modelo. O R^2 ajustado do *Modelo 3* foi de 0.277, indicando que as variáveis escolhidas explicam 27.7% das variações na variável y . O R^2 ajustado baixo significa que as variáveis não explicam muito das variações na resposta, indicando a necessidade de ajustes.

Ainda no *Modelo 3*, podemos fazer uso dos critérios de seleção para comparar os modelos e assim decidir qual se ajusta melhor aos dados. Os critérios de seleção buscam o modelo mais parcimonioso, ou seja, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados, mas que explique bem o comportamento da variável resposta. Para esse estudo, faremos uso do Critério de Informação de Akaike (AIC) e o Critério Bayesiano de Schwarz (BIC). Similares, os dois critérios penalizam modelos sobre parametrizados, contudo, o critério BIC penaliza de forma mais rigorosa. De modo geral, quanto menor o valor do critério de seleção, melhor o modelo. No *Modelo 3* o AIC encontrado foi de 470.35 e o BIC de 481.08.

Seguindo adiante, a *Tabela 4* mostra os resultados do modelo ainda com as mesmas variáveis do *Modelo 3*, porém sem a inclusão dos dados do Distrito Federal. A partir da análise gráfica feita na seção 4.1 deste estudo, a inclusão dos dados referentes ao DF estaria enviesando os resultados encontrados na medida em que as observações do estado são consideradas *outliers* pela distância desses pontos com relação aos demais e também pela mudança significativa nas linhas de tendências dos gráficos. Mais para frente nesta seção será analisado os possíveis motivos para tal comportamento. Do *Modelo 4* concluímos que não há mudanças na interpretação das duas variáveis dado que a variável rendimento permanece significativa e a variável escolaridade não significativa. Contudo, pode-se observar que a exclusão das observações em questão levou à um ajuste melhor do modelo. O R^2 ajustado subiu de 0.277 para 0.607, indicando que agora as observações explicam 60.7% das variações na variável de resposta. Além do mais, os critérios de seleção AIC e BIC apresentados são menores, 329.84 e 403.43, respectivamente, indicando à um modelo mais parcimonioso.

Para fins de comparação, os *Modelos 5 e 6* foram rodados utilizando as mesmas variáveis dos modelos anteriores, porém desta vez sem a inclusão do intercepto, para ambos os modelos, com e sem os dados do DF. Os resultados do *Modelo 5* nos mostram que agora ambas as variáveis (rendimento e escolaridade) são significativas a um nível

menor do que 0.01%. Para o rendimento, o coeficiente encontrado foi novamente negativo, e para a variável de escolaridade, o coeficiente encontrado foi positivo. A interpretação seria que estados com população mais escolarizada em média, deveria apresentar taxas mais satisfatórias do cumprimento da condicionalidade de educação. Este resultado encontrado segue em linha com as expectativas apresentadas anteriormente neste trabalho. O motivo pela qual ambas as variáveis se tornam bastante significativas é de fato a não inclusão do intercepto, dando maior peso as variáveis que agora se tornam mais explicativas de mudanças na variável resposta. Isso também pode ser visto pelo aumento substantivo do coeficiente R^2 ajustado, que agora se aproxima de 1 (0.988). À primeira vista, este coeficiente pode indicar uma explicação quase perfeita dos dados, mas como sabemos que isso é pouco provável, atribuímos esse valor à exclusão indevida do intercepto. Os critérios AIC e BIC também corroboram esta conclusão: ambos sofreram aumentos de quase duas vezes comparados aos valores encontrados para os modelos anteriores, chegando a 815.49 e 823.53, respectivamente. Por fim, podemos também observar um salto das medidas de erro (RMSE, MAE, MAPE e MASE em particular), que são medidas “*negatively oriented*” onde quanto menor o valor melhor o modelo, indicando um ajuste desfavorável do modelo às variáveis.

Algo similar ocorre para o *Modelo 6*. Ambas as variáveis explicativas se tornam significativas de forma enganosa e o coeficiente R^2 ajustado também aumenta de forma não crível. Para este modelo, devido à não inclusão das observações do DF, as medidas de erro citadas são observadas em patamares um pouco menores e os critérios de seleção seguem o mesmo padrão. Ao ponderar os resultados encontrados, podemos concluir, baseado neste modelo e no *Modelo 5* acima, que deve ser feita a inclusão do intercepto para os demais modelos.

A *Tabela 7* ilustra os resultados encontrados para a regressão [2] que inclui todas as variáveis explicativas para um modelo completo. Novamente, o coeficiente β_1 da variável rendimento se mostra significativa (à 0.01%) e com um p-valor extremamente baixo. A variável escolaridade apresenta um coeficiente positivo, porém não significativo. As variáveis taxa de latrocínio, comprometimento da renda com tarifa de transporte público e os cinco grupos de docentes apresentam coeficientes negativos e também não significativos, com exceção do grupo de docentes 4. Na *Tabela 7*, podemos

observar que o coeficiente para o grupo Docente_4 é significativo ao nível de 10%. A interpretação deste resultado é que um aumento na porcentagem de “docentes com formação superior não considerada nas categorias anteriores” gera um efeito negativo na taxa de cumprimento da condicionalidade do programa. Devido à semelhança deste grupo com os outros três primeiros grupos, e dado que nenhum dos três apresentaram coeficientes negativos, podemos atribuir esse resultado à um possível erro de medida nas observações. Iremos verificar essa hipótese mais para frente. O R^2 ajustado encontrado foi de 0.667, sendo levemente maior do que o R^2 ajustado encontrado para o *Modelo 4*, até então o melhor ajuste ponderado pelas implicações. Assim, podemos concluir que a inclusão das variáveis de controle foi positiva, gerando um modelo com melhor ajuste das observações. Vale ressaltar a maior importância devida ao uso do R^2 ajustado nesse caso, dada a inclusão de um número grande de variáveis, que poderiam não necessariamente significar um modelo mais bem ajustado. As medidas de erro e critérios de seleção encontrados também são menores do que os do *Modelo 4*, indicando um modelo com a interpretação muito similar das variáveis, porém com melhor ajuste dos dados, mais parcimonioso e com menos erros.

O *Modelo 8* ilustra a regressão [3], um modelo similar ao anterior, porém sem a inclusão da variável taxa de latrocínios a cada 100 mil habitantes, candidata à *bad control*. Novamente a interpretação das variáveis não varia muito, sendo a variável rendimento mensal domiciliar per capita ainda a única significativa (ao nível 0.01%) se assumirmos a interpretação anterior para a significância da variável Grupo Docente 4. A não inclusão da variável taxa de latrocínio nos levou à um modelo com um R^2 ajustado ainda maior (0.671). As medidas de erro mudam pouco de um modelo para o outro, implicando que a inclusão desta variável não contribui para erros de medida nas observações. Por último, há uma diminuição ainda maior nos critérios de seleção AIC e BIC, que apresentam os valores 380.13 e 406.57, respectivamente.

A *Tabela 9* apresenta os resultados da regressão [4] que não inclui a variável comprometimento da renda de um salário mínimo com a tarifa de transporte público (%), a segunda variável candidata à *bad control*. Aqui, como no modelo anterior, a interpretação das variáveis se mantém a mesma, onde a variável rendimento mensal domiciliar per capita é a única estatisticamente significativa. O R^2 ajustado se mantém

invariável em 0.671 e as medidas de erro também. O critério de AIC aumenta levemente e de forma contrária, o critério de BIC diminui de forma tênue.

Seguindo em frente, no *Modelo 10* é rodado o mesmo modelo completo da *Tabela 7*, porém por fim, sem as duas variáveis candidatas à *bad controls* (regressão [5]). Como esperado, a interpretação das variáveis segue a mesma, contudo podemos observar algumas melhoras no ajuste do modelo. O R^2 ajustado encontrado foi, até o momento, o melhor de todos (0.674), indicando o modelo com melhor ajuste das observações. Apesar das medidas de erro terem se mantido constantes, o modelo apresenta critérios de seleção mais baixos comparado aos dos modelos anteriores, o AIC encontrado foi de 378.20 e o BIC de 401.99. O que podemos concluir com base nestes resultados é que a inclusão das possíveis variáveis *bad controls* não altera em nada a interpretação dos resultados que seguem não enviesados, porém, um modelo sem essas variáveis implica em um ajuste melhor das observações. Dito isso, a partir dos critérios de seleção, este seria o modelo com melhor ajuste e, portanto, o melhor modelo até o momento.

Para a segunda parte da metodologia, foi adotada uma estratégia em que, para o modelo completo principal (*Modelo 7*) serão rodadas cinco regressões, tendo o valor de cada grupo docente zerado por vez, e tendo este mesmo valor redistribuído igualmente entre os demais grupos. Como explicado, o objetivo dessa estratégia é observar efeitos de mudanças na composição destes grupos na variável de resposta y . As *tabelas 11-15* mostram os resultados encontrados. Na prática, os efeitos encontrados não se diferenciam dos resultados que já havíamos encontrado anteriormente: a variável rendimento mensal domiciliar per capita continua sendo negativa e significativa (ao nível de 0.015), o intercepto naturalmente se mantém significativo, e o coeficiente do Grupo Docente 4 também se mostra significativo ao nível de 10% de significância. Abordando a mesma hipótese já abordada, este último efeito seria atribuído à uma possível medida de erro. Comparando estes modelos ao *Modelo 10* (modelo preferido), é fácil ver que há uma diminuição, mesmo que ínfima, do coeficiente R^2 ajustado e um aumento leve dos critérios de seleção AIC e BIC. Além do mais, as principais medidas de erro, RMSE, MAE, MAPE e MASE, apresentaram aumentos significativos. Dito isso, o que podemos concluir é que, em um modelo com a inclusão de todas as variáveis

de controle, mudanças na composição dos grupos não afetam a taxa de cumprimento da condicionalidade.

Para fins de comparação, os últimos cinco modelos (*tabelas 16-20*) representam a regressão [7]. Esses modelos utilizam a mesma estratégia apresentada acima, porém sem a inclusão das demais variáveis de controle, assim voltamos à um modelo similar ao modelo mais simples (Modelo [4]): apenas com as variáveis: i) Rendimento nominal mensal domiciliar e ii) Escolaridade média em anos de estudo (pessoas de 25 anos ou mais) e iii) composição dos grupos docentes utilizando a estratégia descrita.

Para início de análise, podemos observar que a interpretação dos coeficientes permanece a mesma, com exceção da *Tabela 19* que ilustra os resultados após a retirada do grupo docente 4. A interpretação deste resultado é que, uma mobilidade de professores do grupo docente 4 (docentes com formação superior não considerada nas categorias anteriores) para o grupo docente 2 (docentes com formação superior de bacharelado na mesma área da disciplina que leciona), diminui a taxa de cumprimento da condicionalidade de educação. Isso pode ser visto pelo coeficiente da variável *Docente_G2* que é negativo e significativo ao nível de 5%. Dito isso, podemos chegar à uma possível conclusão: que a especificidade e qualidade da formação do corpo docente não é absolutamente determinante do comparecimento ou não do aluno à escola, em outras palavras, a característica da formação do docente não diz respeito à sua capacidade de motivar os alunos. Além disso, no campo estatístico, observamos que o coeficiente R^2 ajustado e as medidas de erro se mantem praticamente constantes. Com relação aos critérios de seleção, podemos observar uma pequena melhora a medida que o AIC e BIC diminuem levemente.

8.1 Interpretações econômicas

Vistos os resultados produzidos pelas regressões, devemos agora buscar interpretações econômicas para os efeitos encontrados. A primeira questão que deve ser discutida diz respeito à exclusão dos dados do Distrito Federal. Ficou bastante claro através dos *Gráficos 1 e 2* e dos resultados produzidos pelas regressões rodadas sem estes dados, que os dados referentes ao Distrito Federal, de fato são *outliers*. A não inclusão destes dados torna o modelo mais preciso (i.e. com menores erros), melhor ajustado (maior R^2 ajustado) e produz melhores critérios de seleção. O que vemos nos *Gráficos 1 e 2* é que além do Distrito Federal apresentar uma alta taxa de cumprimento da condicionalidade, ele também apresenta alto rendimento mensal domiciliar per capita e um grau elevado de escolaridade da população (medido em anos de estudo). Assim, os dados contra intuitivos e distantes dos demais estados apontam para uma característica única do Distrito Federal – característica cujos efeitos são prejudiciais para o nosso modelo.

É trivial concluir que o principal possível motivo para essa discrepância dos dados do Distrito Federal esteja associado ao funcionalismo público ali presente. De acordo com dados do Tesouro Nacional o Distrito Federal gasta o dobro do que é gasto com servidores públicos em média no Brasil (dados de 2016). Enquanto a média do país é de R\$ 1,7 mil por habitante, o Distrito Federal gasta em torno de R\$ 3,4 mil²⁴. Em 2013, segundo a Relação Anual de Informações Sociais (Rais) do Ministério do Trabalho, o DF foi a unidade da federação que mais aumentou seu corpo de funcionários públicos. Neste ano, houve um aumento de 16,73% no número de funcionários públicos, equivalente à 71 mil novos servidores²⁵. Esse fato explica bastante do alto valor de renda mensal domiciliar per capita da UF: normalmente, servidores públicos tem cargos vitalícios, estabilidade de salário e aposentadoria integral. Além disso, funcionários públicos, são em grande parte, concursados e possuem alto grau de escolaridade elevando assim o nível de escolaridade da população desta unidade da federação.

²⁴ Correio Braziliense (2018)

²⁵ O Estado de São Paulo: Economia e Negócios (2014)

Com relação aos resultados em si, a primeira conclusão obtida é de que: estados cuja população tem maior rendimento nominal mensal domiciliar em média, ou seja, estados mais ricos, cumprem de forma menos satisfatória a condicionalidade de educação do programa. Os resultados para os coeficientes de rendimento para a grande maioria dos modelos rodados são negativos e bastante significativos, indicando que um aumento na renda do estado, diminui a sua taxa de cumprimento da condicionalidade. Indicando, conseqüentemente, que estados mais pobres, cumprem a condicionalidade de forma mais adequada.

Um primeiro motivo que pode ser atribuído à esse feito é a taxa de desemprego nos estados. É intuitivo pensar que estados com taxas de desemprego mais baixas, possuem melhores oportunidades de emprego, e assim há incentivos para que os pais dos alunos beneficiários do programa tirem seus filhos da escola para que possam trabalhar. Em estados com boas oportunidades de emprego é possível que a renda obtida através do trabalho seja maior do que o benefício financeiro oferecido pelo programa. Isso corrobora o resultado encontrado na maioria dos modelos deste trabalho, que estados mais ricos cumprem pior a condicionalidade do programa. O contrário ocorre para estados com altas taxas de desemprego: em um cenário de desemprego, poucas oportunidades de emprego estão disponíveis, e as poucas que estão, são normalmente ocupadas por pessoas com qualificação mais elevada. Como consequência, o benefício financeiro do programa se torna não só essencial, como insubstituível para as famílias beneficiárias.

Com o objetivo de testar esta relação entre taxa de desemprego e renda per capita domiciliar dos estados, a correlação destes dados (também para os anos de 2014 até 2019) foi calculada. O coeficiente de correlação obtido foi de -0.251 indicando uma correlação negativa dos dados, em linha com expectativas. O *Gráfico 10* mostra a relação entre os dados, onde cada ponto representa uma unidade da federação para um ano específico. Dado que o coeficiente de correlação, ρ , é um valor entre $0 \leq \rho \leq 1$, podemos concluir que a correlação entre os dados não é tão alta quando poderia ser. Logo, as diferenças em taxas de desemprego não explicam em sua totalidade o efeito encontrado, porém, serve para explicar ao menos uma parte deste resultado.

A segunda interpretação que podemos fazer deste resultado diz respeito ao custo de vida nas unidades da federação. Estados cujas cidades possuem um custo de vida mais alto exigem que uma maior proporção da renda da população seja gasta com despesas corriqueiras. Seguindo esta linha de raciocínio, um custo de vida mais alto geraria uma necessidade maior de uma renda mais elevada para sustentar os custos da vida na cidade. Dessa forma, em linha com o que foi dito acima, haveria incentivos para que as famílias não cumprissem corretamente a condicionalidade de educação e a frequência escolar seria rapidamente substituída por um emprego que gerasse uma renda mais elevada do que o benefício financeiro do programa. O Índice de Custo de Vida, calculado com base em modelos gerados por dados obtidos através de “*crowdsourcing*” mostra que, nos últimos 18 meses, das dez cidades brasileiras com o custo de vida mais altos, oito são pertencentes às regiões sul e sudeste, e as outras duas são pertencentes ao Distrito Federal. Na ponta oposta, das dez cidades brasileiras com o custo de vida mais baixo, seis são pertencentes as regiões norte e nordeste do país. Este índice engloba custos de vida devidos nas seguintes seis categorias: alimentação, vestuário, moradia, transporte, saúde e entretenimento, inseridos pelos próprios usuários. O modelo estatístico então coleta esses dados, eliminando entradas equivocadas e gerando o índice de interesse. Dito isso, fica bastante claro os motivos pelos quais a taxa de cumprimento da condicionalidade é consideravelmente inferior nas regiões sul e sudeste em comparação ao resto do país, especialmente em comparação às regiões norte e nordeste.

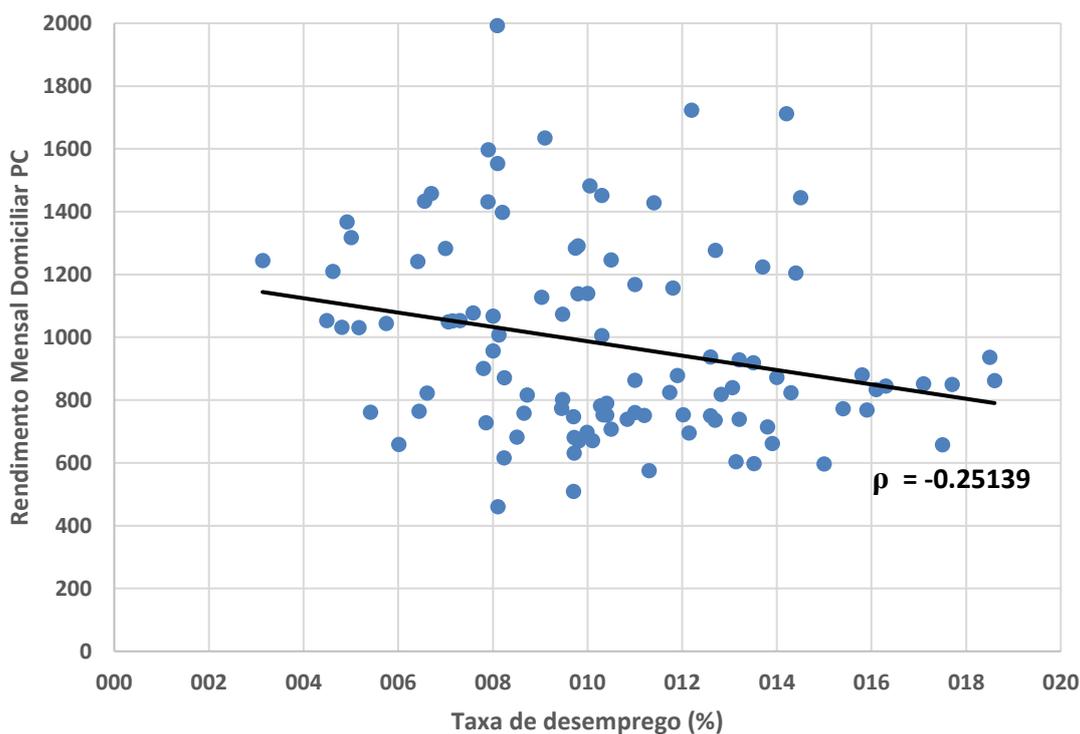
No mais, através dos resultados obtidos pudemos concluir que o nível de escolaridade da população não tem efeitos significativos sobre a taxa de cumprimento da condicionalidade. Logo, ao contrário do que esperávamos, a escolaridade dos pais não é determinante para uma frequência assídua das crianças à escola e isso nos leva a concluir que os resultados deste estudo contrapõem a literatura que mostra que famílias detentoras de pouco capital humano ou “capital escolar”, são menos eficazes na hora de transmitir esse mesmo capital escolar para seus filhos.

Seguindo a diante, as variáveis de controle incluídas nos primeiros modelos também não se mostraram serem significativas, são elas, qualidade do corpo docente (usada com uma variável *proxy* para motivação dos alunos), taxa de latrocínio a cada 100 mil habitantes e comprometimento da renda de um salário mínimo com a tarifa de transporte público. Contudo, após a implementação da segunda parte da estratégia, que

buscava investigar os efeitos de uma mudança na composição dos grupos de corpo docente, observamos que uma mobilidade de professores do grupo docente 4 para o grupo docente 2, diminui a taxa de cumprimento da condicionalidade de educação. Dado que o grupo docente 2 é mais qualificado do que o grupo docente 4, concluímos que característica da formação do docente não diz respeito à sua aptidão em motivar os alunos.

Assim, de forma geral, concluímos que a única variável que tem um efeito significativo sobre a taxa de cumprimento da condicionalidade do programa é a renda mensal domiciliar per capita da população de cada estado.

Gráfico 10 – Taxa de desemprego dos estados versus rendimento mensal domiciliar per capita



9. Conclusão

A partir deste trabalho foi possível verificar os motivos alegados para a falta de cumprimento da condicionalidade de educação do PBF e assim estudar quais são as principais variáveis que afetam o cumprimento desta mesma condicionalidade por parte de seus beneficiários. É importante entender quais são as causas por trás da variação das taxas de cumprimento apresentadas pelos diversos estados brasileiros, podendo assim, elaborar soluções que podem vir a aperfeiçoar o funcionamento, o alcance e a precisão do programa.

O ponto mais importante a ser destacado é o efeito significativo que a renda mensal domiciliar per capita da população dos estados tem sobre a taxa de cumprimento da condicionalidade. O efeito encontrado atesta para a relevância não só da renda per capita dos estados, mas também do custo de vida e em grande parte da taxa de desemprego, no cumprimento pouco satisfatório da condicionalidade. Assim, como estes são fatores exógenos, é importante que haja estímulo contínuo por parte das entidades governamentais e instituições do país com o objetivo de impulsionar o investimento em educação, mesmo em cidades cujo alto custo de vida e as diversas oportunidades de emprego possam abrir espaço para o desinvestimento em educação.

Este trabalho visa enfatizar a importância da acumulação de capital humano como mecanismo de crescimento econômico. Um aumento no nível de capital humano da população, geraria um aumento em produtividade, eficiência, difusão tecnológica e como consequência disso, um aumento de renda e um crescimento econômico sustentável no longo prazo. O Programa Bolsa Família tem como objetivo erradicar a pobreza intergeracional que advém do ciclo vicioso de sub-investimento em educação e por isso, a importância de entender os fatores determinantes ou não do cumprimento da condicionalidade de educação imposta pelo programa.

Dito isto, programas de transferência condicional de renda, como o Bolsa Família, seriam parte de um equilíbrio socialmente ótimo à medida em que a transferência monetária auferida pelos beneficiários serviria como subsídio às famílias, reduzindo o custo marginal de investir em educação e abrindo espaço para um comprometimento de longo prazo com a educação do país.

10. Limitações

O primeiro tipo de limitação que deve ser levado em conta quando se faz uso de dados de fontes originadas pelo governo é a questão da veracidade dos dados. Para os dados de cumprimento da condicionalidade de educação, que são disponibilizados pelo Ministério do Desenvolvimento Social, devemos levar em consideração que esses são dados coletados por entidades governamentais, logo, não é possível atestar sobre a exatidão destes dados. Nesta mesma esfera, a coleta dos dados por parte das escolas cadastradas pode também não ser feita com a regularidade prometida e nem com o rigor exigido pelo programa.

Com relação a cobertura dos dados, uma das limitações diz respeito ao número de anos na série de tempo utilizada. Como mencionado na seção 3, os dados usados começam em 2014 e terminam em 2017, totalizando quatro anos. É importante ressaltar que um número maior de anos poderia oferecer resultados mais robustos a medida que mais observações estariam disponíveis aumentando a capacidade dos modelos de captarem efeitos mais significativos. Ainda nesse aspecto, apesar da disponibilidade de dados dos anos 2014 até 2017, o trabalho não é capaz de identificar os efeitos das variáveis ao longo do tempo e são apenas medidos os efeitos das variáveis em um ponto fixo do tempo. Além do mais, os dados utilizados são separados apenas por unidade da federação (i.e. 27 estados). Novamente, caso os dados fossem discriminados por municípios, por exemplo, os efeitos encontrados seriam potencializados, podendo ser mais consistentes.

Ainda no contexto da cobertura dos dados, vale notar que o uso de dados agregados, ao invés de dados individuais, traz ao trabalho pontos positivos, mas também causa perdas. Quando usamos dados agregados por estados, perdemos a heterogeneidade destes dados, a medida que os efeitos de cada indivíduo, ou núcleo familiar, são agregados para formar um só dado que representa o estado em questão. Dessa forma, não é difícil ver que a heterogeneidade existente entre as famílias é perdida, causando uma generalização dos resultados encontrados. Por outro lado, dados agregados mitigam o efeito de erros de medida nos dados utilizados. Normalmente, dados podem apresentar pequenos erros de medida que são suavizados à medida que os dados são agregados e estes erros, positivos ou negativos, são anulados na média.

Seguindo esta linha de raciocínio os dados utilizados para todas as variáveis com exceção da variável de resposta, y , taxa de cumprimento da condicionalidade de educação, não são específicos para as famílias beneficiárias do programa. Assim, as características captadas por estes dados dizem respeito a toda população dos estados, causando uma possível perda de precisão nos efeitos encontrados a medida em que eles não dizem respeito exclusivamente ao grupo de interesse deste trabalho.

Por último, em um alcance mais amplo, vale observar que este trabalho não engloba a questão da qualidade de ensino das escolas frequentadas por alunos do PBF e busca apenas investigar os motivos por trás do cumprimento ou não da condicionalidade de educação imposta pelo programa. Além disso, este trabalho também não aborda a questão de fraudes existentes no funcionamento do programa. Fraudes estas que podem estar ligadas ao controle de faltas cedido pelas escolas cadastradas, aos avisos e advertências enviadas pelo governo para as famílias beneficiárias com o desígnio de incentivar a frequência dos alunos à escola, assegurando a suspensão do benefício financeiro caso a frequência não seja retomada e assim por diante. É reconhecida a possibilidade de que tais espécies de fraude ocorram, porém, não é do escopo deste trabalho investigar motivos ou possíveis soluções.

Referências Bibliográficas

Cangassu, R. , Salvato, M., Nakabashi, L. (2010). Uma análise do capital humano sobre o nível de renda dos estados brasileiros: MRW *versus* Mincer

Mankiw G., Romer D., Weil D., (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*

Lucas, R. (1988). The mechanics of economic development. *Journal of Monetary Economics*

Aghion, P., Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*

Nakabashi, L., Salvato, M. (2007). Human capital quality in the Brazilian states. *Revista Economia*

Lau, L., Jamison, D., Liu, S., Rivkin, S. (1993). Education and economic growth: some cross-country evidence from Brazil. *Journal of Development Economics*

Mincer, J. (1974). Schooling, experience, and earnings. *New York: Columbia University Press*

Krueger, A., Lindahl, M. (2001). Education for growth: why and for whom? *Journal of Economic Literature*

Camargo, J., Almeida, H. (1994). Human Capital Investment and Poverty. *Departamento de Economia PUC-Rio*

Emerson, P., Souza, A. (2003). Is there a child labor trap? Inter-generational persistence of child labor in Brazil. *The University of Chicago Press*

Abreu, D., Alves de Aquino, J. (2017), Contexto familiar e o cumprimento da condicionalidade de frequência escolar no PBF no Ceará, *Educar em Revista, Curitiba*

Bestagil, F., Hagen-Zanker, J., Harman, L., Sturge, G., Barca, V., Schmidt T., Pellerano, L. (2016). Cash transfers: what does the evidence say? A rigorous review of programme impact and the role of design and implementation features

Paquette, J., Fallon, G. (2010). First nations education policy in Canada: Progress or Gridlock? *University of Toronto Press*

Allan, B., Fryer, R. (2011). The Power and Pitfalls of economic incentives.

Portal Todos pela Educação (2014-2016). Escolaridade medida em anos de estudo – Disponível em: <http://www.todospelaeducacao.org.br/indicadores-da-educacao> - Último acesso em: 21/06/2018

Portal Data Social (2014-2017). Acompanhamento da Taxa de Condicionalidade de Educação – Disponível em: <http://aplicacoes.mds.gov.br/sagi> - Último acesso em: 21/06/2018.

IBGE: Pnad Contínua (2014-2016). Rendimento nominal mensal domiciliar – Disponível em: https://ww2.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pnad_continua - Último acesso em: 21/06/2018.

Portal DAPP (Diretoria de Análise de Políticas Públicas) da Fundação Getúlio Vargas (2014-2016). Acompanhamento da taxa de latrocínio nos estados brasileiros – Disponível em: <http://dapp.fgv.br/seguranca-e-cidadania/datacrime/#crimes> – Último acesso em: 23/06/2018

Plataforma *Mobilizados* - Instituto de Políticas de Transporte e Desenvolvimento: comprometimento da renda de um salário mínimo com a tarifa de transporte público (2014-2017) – Disponível em <http://mobilizados.org.br/database> - Último acesso em: 23/06/2018

INEP utilizando dados do Censo da Educação Básica: percentual de docentes por grupo do indicador de adequação da formação do docente (2014-2016) - Disponível em <http://portal.inep.gov.br/web/guest/educacao-basica> - Último acesso em: 23/06/2018

Pesquisa Mensal de Emprego disponibilizada pelo IBGE: taxa de desemprego (2014-2017) - Disponível em - <https://www.ibge.gov.br/estatisticas-novoportal/sociais/trabalho/9180-pesquisa-mensal-de-emprego.html?=&t=series-historicas> - Último acesso em: 23/06/2018

Angrist J. D., Pischke J-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*

Holt, C.C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, (Vol. 52), *Carnegie Institute of Technology*

Passarelli H., Distrito Federal ganhou 71 mil funcionários públicos em 2013. *O Estado de São Paulo* (2014) – Disponível em <https://economia.estadao.com.br/blogs/mural-dos-concursos/distrito-federal-ganhou-71-mil-novos-funcionarios-publicos-em-2013/> - Último acesso em: 23/06/2018

Pacheco L., Despesa com funcionários públicos por habitante em Brasília é o dobro da média do país. *Correio Braziliense* (2018) - Disponível em <http://blogs.correiobraziliense.com.br/papodeconcurseiro/despesa-com-funcionarios-publicos-por-habitante-em-brasilia-e-o-dobro-da-media-do-pais/> - Último acesso em: 23/06/2018