

## **Oferta de escolas e retorno do ensino médio entre os jovens no Brasil**

Maíra Penna Franca (UFF)  
Danielle Carusi Machado (UFF)  
Carlos Henrique Corseuil (IPEA)

### **Resumo**

Este artigo tem por objetivo estimar o retorno salarial do ensino médio para jovens entre 20 e 22 anos a partir dos dados do Censo Demográfico de 2010 e dos Censos Escolares de 2003 a 2005. Assumindo a hipótese de endogeneidade da educação na equação de rendimentos foram utilizados instrumentos que captam a oferta de escolas de ensino médio no município em que o jovem morava aos 15 anos, idade de entrada no ensino médio. Os resultados apontam que a densidade de escolas possui uma relação positiva e significativa com a probabilidade de o jovem ter o ensino médio completo em 2010. As estimativas do retorno do ensino médio obtidas pelo método de variável instrumental (IV) excederam consideravelmente a estimativa pelo método de mínimos quadrados ordinários (MQO). Esse resultado é condizente com a literatura empírica que utiliza fatores institucionais relacionados ao sistema educacional como instrumento para educação. Se os retornos educacionais não foram constantes, e se a proximidade de escolas afetar mais as escolhas de indivíduos que, de outra forma, teriam uma escolaridade relativamente mais baixa, é esperado que o estimador IV superestime o retorno educacional.

**Palavras chave:** Retornos à educação, ensino médio, variáveis instrumentais

### **Abstract**

This paper aims to estimate economic returns to secondary education for young people aged 20 to 22 years old using data from the Demographic and School Census. Taking into account the hypothesis of endogeneity we use instrumental variables that capture the supply of secondary schools in the municipality where the young person lived at the secondary school starting age, 15 years old. The results indicate that the density of schools has a positive and significant effect on the probability of the young people having graduated from the secondary school in 2010. Instrumental variables estimates of the return to secondary education exceed the OLS estimates. This result is consistent with the empirical literature that explores institutional factors related to the educational system as instruments for education. If returns to education are not constant, and if the proximity of schools affects more the choices of individuals who would otherwise have a relatively lower educational level, the IV estimator is expected to overestimate the educational return.

**Keywords:** Returns to Education, secondary education, instrumental variables

**JEL:** J21; J24; J31

**Área 12 - Economia Social e Demografia Econômica**

## 1) Introdução

Nas últimas duas décadas verificou-se uma expansão considerável do sistema educacional brasileiro, o que resultou no aumento do nível de escolaridade da população de jovens e adultos. O acesso ao ensino fundamental foi praticamente universalizado, com 97% de crianças de 7 a 14 anos frequentando a escola em 2010. No entanto a cobertura do nível médio ainda se encontra aquém do desejado. Entre os jovens de 15 a 17 anos, 78% frequentavam a escola em 2010, mas somente 50,5% dos jovens de 20 a 22 anos haviam completado essa etapa de escolarização de acordo com dados do Censo Demográfico.

A despeito do abandono e evasão no ensino médio, diversos estudos [Crespo & Reis (2009), Barbosa Filho & Pessoa (2008)] apontam que o mercado de trabalho brasileiro proporciona ganhos elevados para esse nível de escolaridade. Tendo em vista que a educação é um determinante importante na elevada desigualdade de renda no país, é de extrema importância entender melhor como os rendimentos se relacionam com o nível de educação.

Este artigo tem por objetivo estimar o retorno salarial do ensino médio para os jovens no Brasil. Levando em consideração o problema da endogeneidade da educação na equação de rendimentos, o presente trabalho contribui com a literatura ao empregar um conjunto de instrumentos que captam a densidade de oferta de escolas no município quando o jovem tinha 15 anos, idade de entrada no ensino médio.

Considerando que a oferta de escolas na região aumenta o conjunto de possibilidades dos jovens de frequentarem o ensino médio, e que os indivíduos irão atribuir um peso maior às instituições que estão mais próximas, foram construídas medidas de oferta de escolas ponderadas pela distância entre municípios. Seguindo o proposto por Doyle & Skinner (2016), além da variável de oferta de escolas no próprio município de residência quando em idade escolar, agregamos uma medida de densidade da oferta escolar nos municípios próximos. A variável do número de escolas foi ponderada pela distância geodésica entre as coordenadas centrais dos municípios de forma a captar uma proxy de densidade escolar.

Os resultados do 1º estágio comprovam que a densidade de escolas possui um efeito significativo na probabilidade de conclusão do ensino médio. No 2º estágio encontramos que a conclusão do ensino médio está associada a um prêmio salarial que varia entre 15% a 20% quando incluímos no 1º estágio os valores preditos da dummy de ensino médio por um modelo probit, procedimento sugerido por Angrist & Pischke (2008) para evitar a chamada “*Forbidden Regression*”.

Na hipótese de retornos heterogêneos da educação, a estimativa produzida pelo método de variável instrumental pode ser interpretada como uma média ponderada dos retornos marginais para a educação na população, onde o peso para qualquer indivíduo é dado pelo tamanho relativo do incremento em sua escolaridade induzida pelo instrumento (Card 2001).

O artigo está dividido em seis seções, sendo a primeira esta introdução. A seção 2 aborda aspectos teóricos e apresenta uma breve revisão da literatura sobre o tema. Em seguida, apresentamos as bases de dados e fazemos uma análise descritiva das variáveis. A seção 4 descreve a estratégia empírica, enquanto que a seção 5 apresenta os resultados. A sexta e última seção apresenta as considerações finais.

## 2) Discussão Conceitual e Evidências Empíricas

A relação entre educação e salários tem sido amplamente estudada pela literatura econômica desde os trabalhos seminais de Becker (1962, 1975) e Mincer (1958, 1974). De acordo com modelo proposto por Becker (1975), os agentes irão decidir o nível de educação que irão alcançar tendo em vista os custos e benefícios de obtenção de cada ano de escolaridade. Nesse arcabouço teórico, considerando a alocação de tempo finita entre estudo e trabalho, os indivíduos incorrem em um custo de oportunidade durante a escolarização devido aos rendimentos que deixam de ser recebidos no mercado de trabalho. Por outro lado, o retorno da educação se traduz em aumentos salariais. A mensuração dos retornos educacionais baseia-se na equação de rendimentos proposta por Mincer (1974):

$$(1) \ln(y_i) = \alpha + \beta s_i + \delta X_i + \varepsilon_i$$

Onde  $\ln(y_i)$  é o logaritmo do rendimento do trabalho,  $s_i$  é a escolaridade, em geral medida por de anos de estudo,  $X_i$  é um vetor de variáveis de controle e  $\varepsilon_i$  é o distúrbio aleatório que comporta fatores não observados que influenciam os rendimentos. O coeficiente  $\beta$  é frequentemente interpretado como retorno da educação. No entanto, como apontado por Heckman et al (2006), o coeficiente minceriano só poderá ser interpretado como taxa de retorno da educação sob determinadas circunstâncias.

A identificação do efeito causal da educação sobre os rendimentos esbarra em diversas dificuldades, dentre elas o chamado “viés de variável omitida”. Segundo Card (1999) esse viés surge pela correlação entre a habilidade individual, não observada no modelo, e o custo marginal da escolaridade. Em geral, espera-se que indivíduos com maior habilidade tenham menor custo marginal de se escolarizar, e com isso, tenderão a completar um número maior de anos de estudo. Paralelo a isso, indivíduos mais habilidosos são também aqueles com maior potencial de auferir renda. Desse modo, espera-se que haja uma correlação positiva entre anos de estudo e salários mesmo na ausência de efeito causal de escolaridade sobre os rendimentos do trabalho

Para que  $\beta$  recupere o efeito causal da escolaridade é preciso que a hipótese de exogeneidade seja atendida. Um dos métodos mais utilizado para lidar com o problema é o de variáveis instrumentais (IV). Segundo Ashenfelter et al (1999) a essência do método IV é explorar uma variação natural nos dados que influencie na decisão de se escolarizar e que não seja correlacionado com o resíduo da equação de rendimentos. Por construção, a variável instrumental é correlacionada com o rendimento apenas por seu efeito na escolaridade. A maior crítica a esse método é que o instrumento pode não ser de fato totalmente exógeno na equação de salários.

Na literatura empírica tem sido recorrente a utilização de instrumentos relacionados a fatores institucionais do sistema educacional como a idade mínima de escolaridade compulsória, custos de mensalidades ou a proximidade geográfica das escolas (Card, 1999).

### 2.1) Evidências Internacionais

O estudo pioneiro de Angrist & Krueger (1991) explora uma variação exógena induzida pela combinação entre a idade de entrada na escola e a lei de escolaridade compulsória nos Estados Unidos. Os autores observaram que os indivíduos que nasciam no início do ano entravam na escola um pouco mais velhos (com 6 anos e meio) e permaneciam menos tempo até completar a idade mínima de escolaridade compulsória, de 16 anos. Utilizando o trimestre de nascimento como instrumento para escolaridade

para homens nascidos entre 1930 e 1959, Angrist & Krueger (1991) concluem que o trimestre de nascimento possui um efeito pequeno na escolaridade dos homens, e que as estimativas produzidas por MQO e por IV são semelhantes

Card (1995) utilizou como instrumento para educação a presença de instituições de nível superior no distrito onde o jovem morava aos 17 anos. A estimativa por variável instrumental (13%) ficou bem acima da estimativa convencional por mínimos quadrados ordinários (7%). O autor também considerou outra especificação, na qual ele interage a presença de instituição de ensino com medidas de background familiar. A estimativa IV dessa segunda especificação foi de 10%, ainda assim acima da estimativa de MQO.

Em uma meta-análise da literatura empírica, Card (1999) conclui que a estimativa do retorno educacional por variável instrumental tipicamente excede a estimativa por MQO correspondente, em geral por 30% ou mais. Uma possível justificativa para esse fenômeno, elaborada por Card (1999) baseia-se na hipótese de que os retornos educacionais são heterogêneos e decrescem para níveis mais altos de escolaridade. Fatores como a escolaridade obrigatória ou a proximidade das escolas têm maior probabilidade de afetar as escolhas de indivíduos que, de outra forma, teriam uma escolaridade relativamente mais baixa. Se esses indivíduos apresentarem retornos a educação acima da média, é esperado que o estimador IV baseado nesses fatores produza estimativas dos retornos a escolaridade acima das estimativas MQO.

Seguindo na linha dos instrumentos relacionados a fatores institucionais, Duflo (2001) avaliou um programa de construção de escolas implementado pelo governo da Indonésia entre 1972 e 1978. Explorando a variação exógena causada pelo aumento da oferta educacional, a autora estima um modelo de mínimos quadrados em dois estágios utilizando como instrumento para escolaridade uma interação entre dummies das coortes e o número de escolas construídas na região, que definem a exposição do indivíduo ao tratamento. A autora conclui que uma escola construída para cada mil habitantes aumenta a escolaridade dos beneficiados em 0,12 a 0,19 anos, bem como 1,5 a 2,7 por cento de aumento nos salários. Isso implica em estimativas de retornos econômicos para a educação variando de 6,8% a 10,6% por cento. Esses valores não diferiram significativamente da estimativa por MQO, contrariando as evidências encontradas por Card (1999). Fora isso, o retorno estimado para Indonésia foi similar as estimativas reportadas para países desenvolvidos. A literatura em geral reporta retornos educacionais maiores para países em desenvolvimento.

Psacharopoulos e Patrinos (2004) atualizaram o estudo de Psacharopoulos (1994) numa compilação de dados para 98 países. Os autores evidenciaram que os retornos à escolaridade são decrescentes em relação ao nível de desenvolvimento socioeconômico. As maiores taxas de retorno foram observadas na América Latina e Caribe e na África Subsaariana. No grupo de países mais desenvolvidos, pertencentes a Organização para a Cooperação e Desenvolvimento (OCDE) os retornos estimados são mais baixos. No nível médio de ensino, os retornos na América Latina foram estimados em 17% e em 11,3% nos países da OCDE.

Por fim, Doyle & Skynner (2016), baseando-se no trabalho de Card (1995), utilizam a variação geográfica para identificar o retorno educacional de nível superior nos Estados Unidos para diferentes coortes de jovens que tinham entre 23 a 30 anos entre 2007 e 2010. Os autores implementaram um grande conjunto de instrumentos, considerando o local onde o jovem morava aos 17 anos. Os instrumentos são: presença de uma faculdade de quatro ou dois anos no distrito, logaritmo do inverso da distância para as faculdades do Estado, matrículas ponderadas pela distância, mensalidades ponderadas pela distância e o inverso da distância para todas as faculdades do país. Na

estimação preferida, um ano adicional de escolaridade superior resulta em um aumento de 9,7% nos rendimentos anuais.

## **2.2) Evidências Nacionais**

No Brasil os primeiros estudos que investigaram o retorno educacional foram realizados na década de 1970 por Castro (1973) e Langoni (1974). Nos dois estudos se comprova que os retornos educacionais eram elevadíssimos devido a baixa escolaridade da população e ao processo de desenvolvimento econômico que demandava mão de obra qualificada.

Barbosa Filho e Pessôa (2008) atualizam o estudo de Langoni (1974) e estimam a taxa interno de retorno (TIR<sup>1</sup>) por anos de estudo e por ciclo de escolaridade no Brasil usando dados da Pesquisa de Orçamento Familiar (POF) e da Pesquisa Nacional por Amostra de Domicílios (PNAD) entre 1980 e 2004. Os autores corroboram a existência do chamado efeito-diploma, que aparece na forma de um retorno maior para o ano de educação em que se completa um ciclo. A TIR do ensino médio foi estimada em 20,2% em 1981, tendo decrescido para 13,9% em 2004.

Para o mesmo período de análise, Crespo e Reis (2009) também constata uma queda no retorno salarial para o nível médio. Os autores utilizam uma metodologia mais flexível para equação de rendimento, que além de um termo linear para os anos de escolaridade inclui dummies que permitem captar mudanças de inclinação e saltos para graus completos do ciclo educacional. Os resultados indicam que o prêmio salarial associado ao ensino médio era de 32% em 1982, tendo decrescido para 27% em 2004.

Uma linha de estudos brasileiros se preocupou em lidar com o problema de seletividade amostral na equação de rendimentos, decorrente do fato que a participação no mercado de trabalho não ocorre de maneira aleatória. Nesse sentido, os estudos de Kassouf (1994, 1998), Sachsida et al (2004) e Resende & Willie (2006) aplicam o modelo de seleção amostral desenvolvido por Heckman (1979) para estimar os retornos educacionais.

Outra linha de estudos busca captar a influência de fatores familiares no retorno à educação. Lam & Schoeni (1993), para uma amostra de homens casados da PNAD de 1982, incluem a escolaridade dos pais e dos sogros na equação de rendimentos. As estimativas de retorno à educação diminuem em cerca de um terço quando as medidas de escolaridade da família são incluídas na regressão.

Já Ueda & Hoffman (2002) consideram características familiares, escolaridade dos pais e a ocupação do pai quando o indivíduo estava em idade escolar, como instrumentos na equação de rendimentos. Utilizando dados da PNAD de 1996, os autores estimam retornos por anos de estudo de 12% com o método MQO e de 9,8% com o método de variável instrumental.

Num estudo mais recente, Sulliano & Siqueira (2012) estimam os retornos educacionais para as regiões Nordeste e Sudeste, usando dados da PNAD de 2001 a 2006. Os autores utilizam como instrumento o número de pessoas na família. Na estimativa por IV, o retorno salarial de um ano adicional de escolaridade foi de 16% no Nordeste e de 13% no Sudeste. Cabe destacar que para região Nordeste o coeficiente estimado por variável instrumental excedeu a estimativa por MQO, já na região Sudeste isso não ocorreu.

---

1 A TIR corresponde a taxa de desconto que iguala o valor presente dos custos de educação com o valor presente de seus benefícios. Conforme descrito por Moura (2008), devido as hipóteses do modelo de Mincer, o coeficiente minceriano só pode ser considerado como taxa de retorno sob circunstâncias restritas.

Poucos estudos exploram os fatores institucionais na identificação do efeito da escolaridade no rendimento. Nessa linha, Teixeira & Menezes Filho (2012) usando dados das PNADs de 1997 a 2007 construíram um grande conjunto de instrumentos, que incluem a oferta de escolas no estado e no ano de nascimento, o PIB estadual e a Lei 5692 de 1971 que ampliou a escolaridade obrigatória de 4 para 8 anos. As estimativas por variáveis instrumentais variaram entre 5% e 9%, dependendo dos instrumentos utilizados, abaixo do retorno estimado por MQO, de 11%.

Tendo em vista a escassez de estudos que exploram os fatores institucionais relacionados à oferta educacional para identificar o efeito da escolaridade sobre os rendimentos, esse trabalho busca contribuir com a literatura de economia da educação ao empregar medidas de densidade de oferta de escolas para estimar o retorno salarial no ensino médio.

### 3) Bases de dados e descrição das variáveis

Neste artigo, utilizaram-se duas fontes de dados principais: o Censo Escolar, realizado pelo INEP, que traz informações sobre as escolas de ensino médio, e o Censo Demográfico realizado pelo IBGE. Além de trazer informações para o universo da população brasileira, no Censo Demográfico é selecionada uma amostra para responder um questionário mais completo, com perguntas sobre rendimentos, escolaridade, migração, fecundidade, características do emprego e outros temas. A análise empírica utiliza esta riqueza de informações contidas no Censo Demográfico, que fornece o rendimento e o número de horas de trabalho dos indivíduos na data de referência da pesquisa (31/07/2010), e também uma pergunta retrospectiva sobre o município de residência cinco anos antes da data de referência.

O número de escolas de ensino médio em cada município foi dividido pela respectiva população de 15 a 19 anos. Os dados de população foram extraídos do site do DATASUS. Além da oferta de escolas no próprio município, foi construída uma medida de densidade de oferta de escolas nos municípios vizinhos. Para tanto se utilizou a distância geodésica entre as coordenadas centrais dos municípios como fator de ponderação para o número de escolas<sup>2</sup>. As coordenadas dos municípios foram obtidas no site do IBGE, utilizaram-se as malhas territoriais por município de 2005.

#### 3.1) Seleção da amostra

No Censo Demográfico não estão disponíveis as variáveis que identificam o mês e ano de nascimento dos indivíduos. Para construção das coortes, considerou-se a idade dos indivíduos na data de referência, 31/07/2010.

A amostra utilizada neste artigo é composta por indivíduos com nível de escolaridade fundamental ou médio completo, com idade entre 20 e 22 anos em 2010, e ocupados na semana de referência da pesquisa.

Além do município de residência atual, temos informações sobre o tempo de moradia no município e o município em que o jovem morava em 2005. O nosso interesse é na oferta de escolas do município quando o jovem tinha 15 anos. As variáveis instrumentais foram construídas a partir do número de escolas no município onde o jovem residia aos 15 anos, idade de entrada no ensino médio. O número de escolas em cada município foi dividido pela respectiva população de 15 a 19 anos.

---

<sup>2</sup> A distância geodésica é a distância mais curta entre dois pontos em um elipsoide. O cálculo das distâncias foi feito pelo *software R* com pacote Geosphere que produz estimativas de alta precisão.

**Tabela 1: Seleção das Coortes**

Coorte	Data de nascimento	Censo Escolar utilizado	Idade no Censo Demográfico
Coorte 1	01/08/1989 – 31/07/1990	2005	20
Coorte 2	01/08/1988 – 31/07/1989	2004	21
Coorte 3	01/08/1987 – 31/07/1988	2003	22

Para quem sempre morou no mesmo município é possível identificar a oferta de escolas em qualquer instante do tempo. Porém para os jovens que tenham migrado entre o ano em que completou 15 anos e o ano de 2005, não é possível determinar exatamente o município de residência aos 15 anos. Tendo em vista essa limitação, a amostra é restrita aos jovens que estavam com 20 a 22 anos em 2010, considerando que apenas uma pequena proporção da amostra possa ter migrado entre 2003 e 2005, ou entre 2004 e 2005 para as coortes com 22 e 21 anos no censo, respectivamente.

Outra restrição do Censo Demográfico 2010 é que no questionário de educação não foi levantada a questão sobre a última série concluída com aprovação para os indivíduos que deixaram de estudar. Com isso não é possível construir a variável contínua de anos de estudo, mais usual em equações de rendimento. No entanto é possível identificar o nível de instrução mais elevado. Para mensurar o efeito diploma do ensino médio compararam-se os rendimentos de jovens com níveis fundamental e médio completo, excluindo do 2º grupo os jovens que frequentavam ou que em algum momento haviam frequentado o ensino superior sem ter concluído. Portanto, em termos de anos de estudo, a comparação foi feita entre os jovens com 11 anos de estudo (ensino médio completo) e aqueles que tinham entre 8 e 10 anos de estudo (fundamental completo e médio incompleto).

### 3.2) Variáveis Instrumentais

Com os Censos Escolares de 2003 a 2005 foram construídos dois tipos de medidas de oferta de escolas por município. A primeira corresponde ao número de escolas do próprio município para cada 1000 habitantes de 15 a 19 anos.

$$(2) \text{Esc}_{mt} = \frac{\text{Escolas}_{mt}}{\text{Populacao 15-19}_{mt}} \times 1000$$

Onde  $m$  indexa o município e  $t$  o ano.

O segundo tipo de instrumento é baseado em Doyle & Skinner (2016) e busca captar a densidade de oferta de escolas próximas ao município, dentro do mesmo Estado. Levando em consideração que a oferta de escolas na região aumenta o conjunto de possibilidades dos jovens de frequentarem o ensino médio, e que os indivíduos irão atribuir um peso maior às instituições que estão mais próximas, foram construídos pesos baseados no inverso da distância geodésica:

$$(3) w_{\rho mk} = \left( \frac{d_{mk}}{\sum_{k=1}^K d_{mk}} \right)^{-\rho}$$

Em que  $d_{mk}$  é a distância geodésica intra-estadual entre o município de residência  $m$  e o município  $k$ , e  $\rho$  é a taxa de influência. Para alterar o peso atribuído à

distância, o termo exponencial,  $\rho$ , pode ser modificado. Em  $\rho = 1$ , os pesos são uma função linear da distância, quando  $\rho > 1$ , os pesos decrescem exponencialmente à medida que a distância aumenta. No limite, com  $\rho \rightarrow \infty$ , todo o peso será atribuído ao município mais próximo. Usamos duas especificações, com  $\rho = 1$  e  $\rho = 2$ .

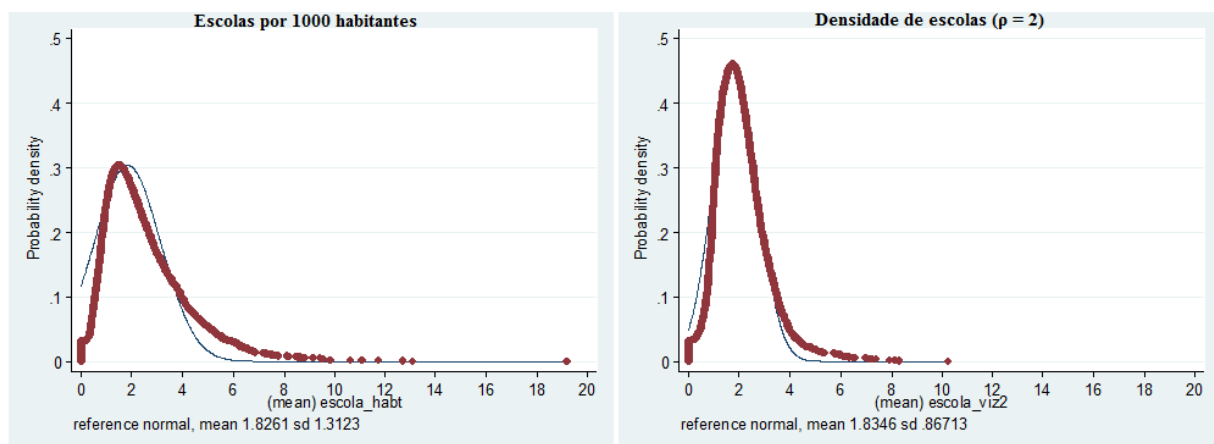
Como o município de residência aos 15 anos é a menor unidade de medida que obtemos no Censo Demográfico<sup>3</sup>, as distâncias geodésicas foram calculadas entre as coordenadas dos centroides do município de residência, e as coordenadas dos centroides dos municípios que estavam a um raio de até 80 quilômetros da distância geodésica, dentro do mesmo Estado. Para cada município a densidade de escolas na vizinhança é dada por:

$$(4) \text{ Densidade\_Esc}_{\rho mt} = \sum_{k=1}^K \frac{w_{\rho mk} \cdot \text{Esc}_{kt}}{\sum_{k=1}^K w_{\rho mk}}$$

### 3.3) Análise descritiva

A figura 1 apresenta as distribuições do número relativo de escolas no município e da medida de densidade de escolas, considerando pesos quadráticos na distância. É possível perceber que há uma grande concentração em torno da média de duas escolas por mil habitantes. A medida de densidade apresenta uma dispersão menor em torno da média.

**Figura 1: Distribuição do número de escolas por mil habitantes e da densidade de escolas ( $\rho = 2$ )**



Fonte: Censo Escolar/INEP. Elaboração própria

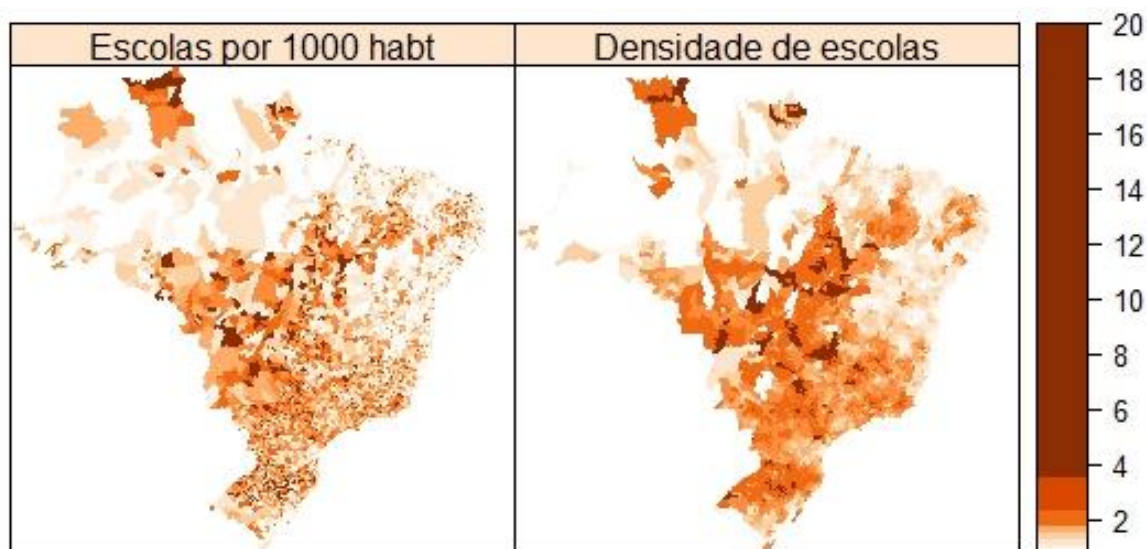
O mapa abaixo apresenta a distribuição regional das duas variáveis. As regiões Centro-Oeste e Sul apresentam as maiores ofertas relativas de escolas, tanto em termos

<sup>3</sup> O setor censitário representa o menor nível de agregação do Censo Demográfico para os domicílios entrevistados na data de referência da pesquisa. Na pergunta retrospectiva sobre o município de residência em 2005 não é possível identificar o setor censitário, portanto o menor nível de agregação é o próprio município.



do número de escolas do próprio município (quadrante do lado esquerdo), quanto em termos de densidade de escolas nos municípios próximos (lado direito). O Sudeste possui o maior número absoluto de escolas, no entanto essa região também concentra a maior parte da população. Em 2005 cerca de 40% dos jovens de 15 a 19 anos moravam na região Sudeste. Nas regiões Norte e Nordeste verifica-se uma menor oferta relativa de escolas.

**Figura 2: Número de escolas de ensino médio por 1000 habitantes de 15-19 anos, média dos anos 2003, 2004 e 2005 escolas ( $\rho = 2$ )**



Fonte: Censo Escolar/INEP. Elaboração própria

A tabela 2 reporta os valores médios das variáveis explicativas usadas no modelo de regressão. Cerca de 60% dos jovens contemplados na amostra tinham completado o ensino médio. Observa-se uma maior predominância de homens (58%), isso se deve a maior participação masculina no mercado de trabalho. Em relação à cor há uma proporção ligeiramente menor de brancos (47%) em comparação a Pretos, Pardos e Índios (53%). Também se verifica uma proporção maior de jovens com 22 anos (35%), o que era esperado visto que a participação no mercado de trabalho tende a aumentar com a idade para os jovens. Por fim, também foram incluídas variáveis regionais, segundo o município de residência em 2005. A maior parte dos jovens morava em municípios da região Sudeste (44%) seguido da região Nordeste (25%).

<b>Tabela 2: Características da amostra</b>	
<b>Rendimento/ hora</b>	5.02
<b>Escolaridade</b>	
Ensino médio completo	59.3%
Ensino Fundamental completo	40.7%
<b>Sexo</b>	
Homens	58.0%
Mulheres	42.0%
<b>Cor</b>	
Branços	47.2%
Pretos, Pardos e Índios	52.8%
<b>Idade no Censo 2010</b>	
20 anos	31.3%
21 anos	33.5%
22 anos	35.2%
<b>Características regionais</b>	
Metropolitana	47.6%
Não metropolitana	52.4%
Nordeste	24.8%
Sudeste	44.2%
Sul	16.6%
Centro Oeste	7.3%
Norte	7.0%
Observações	351,062

Fonte: Censo Demográfico 2010. Elaboração própria.

#### 4) Modelo Empírico

Para estimar o retorno salarial associado a cursar o nível médio de ensino, partimos de uma equação de rendimentos adaptada:

$$(5) \ln(y_i) = \alpha + \beta EM_{completo_i} + \delta X_i + \mu_i$$

Em que  $EM_{completo}$  representa uma variável binária com valor igual a 1 para indivíduos que tenham completado o Ensino Médio,  $X_i$  é um vetor de características do indivíduo que inclui sexo, cor, faixa etária dos jovens, e região geográfica,  $\ln(y_i)$  é o logaritmo do rendimento por hora no trabalho principal.

Uma especificação como (5) é problemática visto que há fatores não observados que podem estar afetando simultaneamente a probabilidade de completar o Ensino Médio e a capacidade de auferir renda. Na ausência de dados experimentais, buscamos encontrar características do município em que o jovem morava aos 15 anos que possam estar razoavelmente relacionadas ao seu nível de escolaridade, mas que estejam relacionadas aos rendimentos apenas através do mecanismo de escolaridade.

Para lidar com o problema de endogeneidade, aplicamos o método de mínimos quadrados em dois estágios (2SLS). Um problema adicional com essa especificação é que a variável de escolaridade é uma dummy. Convencionalmente, o modelo em dois estágios é dado por:

**1º Estágio:**

$$(6) EM_{completo_i} = \pi_0 + \pi_1 Z_i + \pi_2 X_i + u_i$$

**2º Estágio:**

$$(7) Ln(y_i) = \tau + \theta EM_{completo_i} + \gamma X_i + \eta_i$$

Em que  $Z_i$  é um vetor de instrumentos que inclui a medida de escolas por município (2) e a medida de densidade de municípios vizinhos dada por (4).

No entanto, como  $EM_{completo_i}$  é uma *dummy*, a função de expectativa condicional  $E(EM_{completo_i} | Z_i, X_i)$  provavelmente é não linear. Com isso, o 1º estágio do MQO é apenas uma aproximação da verdadeira função de expectativa condicional. Segundo Angrist & Pische (2008), os pesquisadores podem ficar tentados a substituir o 1º estágio por um modelo não linear como um Probit ou Logit, e depois utilizar os valores preditos diretamente no 2º estágio. Essa é a chamada “*Forbidden Regression*”, que foi considerada proibida por Hausman em 1975. A justificativa é que somente na estimativa por MQO é que se pode garantir que os resíduos do 1º estágio sejam não correlacionados com os valores preditos e com as covariáveis.

Uma alternativa a “*Forbidden Regression*” proposta por Angrist & Pische (2008) é usar o valor predito por um modelo não linear como instrumento num modelo convencional de mínimos quadrados em dois estágios. Sendo assim teremos um modelo em três etapas. Primeiro estimamos o valor predito da *dummy* de ensino médio num modelo Probit:

$$(8) p_i^{\wedge} = EM_{completo_i}^{\wedge} = \Phi[\pi_1 Z_i + \pi_2 X_i]$$

Depois o valor predito  $p_i^{\wedge}$  entra na regressão de 1º estágio como um instrumento:  $Z'_i = Z_i + p_i^{\wedge}$

**1º Estágio:**

$$(9) EM_{completo_i} = \phi_0 + \phi_1 Z'_i + \phi_2 X_i + e_i$$

**2º Estágio:**

$$(10) Ln(y_i) = \lambda + \psi EM_{completo_i}^{\wedge} + \varphi X_i + n_i$$

A mesma estratégia empírica foi usada por Zhang (2013) para estimar o efeito de ter frequentado classes particulares (tutorias) no desempenho de alunos. Serão reportados os resultados do 2º estágio com e sem correção de não linearidade para a variável de ensino médio.

## 5) Resultados

Nesta seção são apresentadas as estimativas do prêmio salarial de ensino médio pelos métodos de variável instrumental e de mínimos quadrados ordinários. Os resultados do 1º estágio e os efeitos marginais obtidos no modelo probit para a probabilidade de completar o ensino médio se encontram no apêndice.

Na regressão de 1º estágio foi aplicado o teste F para verificar se conjuntamente os coeficientes associados aos instrumentos excluídos da equação de rendimento são estatisticamente significativos. Em todas as especificações os valores obtidos foram bem altos, mostrando que eles são significativos ao nível de 1%.

No 2º estágio, as variáveis explicativas se mostraram significativas e com sinal esperado na equação de rendimentos. As diferenças salariais entre homens e mulheres e brancos e negros são expressivas. O rendimento é crescente com a idade, em termos regionais se verificam maiores salários nas regiões metropolitanas.

Observa-se um efeito positivo e significativo da conclusão do ensino médio nos rendimentos. Os modelos com variável instrumental revelam retornos bastante elevados, que variam de 20% na especificação com apenas a oferta relativa de escolas do município, a 34% quando se inclui a medida de densidade de escolas ponderada pelas distâncias com peso linear. Com a medida de densidade de escolas e peso não linear, o retorno ficou em 28%. Essas estimativas ficaram bem acima do retorno estimado por mínimos quadrados.

Considerando um cenário de retornos heterogêneos à escolaridade, as estimativas de variável instrumental fornecem um efeito médio de tratamento local para o subconjunto da amostra que é induzida a obter mais escolaridade por esse conjunto particular de instrumentos (Doyle & Skineer 2015 apud Card 2001).

É possível que o retorno esteja sendo superestimado devido a erros de mensuração na variável de ensino médio. Tendo isso em vista, a tabela 4 reporta os resultados do 2º estágio considerando os valores preditos da probabilidade de conclusão do ensino médio num modelo probit como instrumento no 1º estágio, ou seja, consideramos explicitamente o problema destacado de “*Forbidden regression*” e adotamos a estratégia de Angrist & Pischke (2008). Na 1ª coluna, com apenas a oferta de escolas do próprio município como instrumento, o coeficiente da escolaridade se mostrou não significativo. Com a inclusão da medida de densidade de escolas o retorno do ensino médio varia entre 15% a 20%, uma diferença considerável em relação às estimativas da tabela 3.

**Tabela 3: Estimativa do logaritmo do rendimento/hora**

	(0) MQO	(1) IV	(2) IV	(3) IV
Médio completo	0.100*** (0.00267)	0.207*** (0.0762)	0.340*** (0.0716)	0.283*** (0.0717)
Homem	0.167*** (0.00261)	0.180*** (0.00912)	0.195*** (0.00860)	0.189*** (0.00861)
Branco	0.0943*** (0.00270)	0.0879*** (0.00529)	0.0801*** (0.00510)	0.0835*** (0.00509)
21 anos	0.0334*** (0.00314)	0.0300*** (0.00393)	0.0259*** (0.00389)	0.0277*** (0.00387)
22 anos	0.0684*** (0.00314)	0.0633*** (0.00475)	0.0569*** (0.00462)	0.0597*** (0.00460)
Região Metropolitana 2005	0.185*** (0.00260)	0.182*** (0.00314)	0.178*** (0.00304)	0.179*** (0.00302)
Nordeste 2005	-0.187*** (0.00624)	-0.188*** (0.00627)	-0.189*** (0.00630)	-0.189*** (0.00627)
Sudeste 2005	0.128*** (0.00590)	0.123*** (0.00678)	0.117*** (0.00667)	0.120*** (0.00667)
Sul 2005	0.120*** (0.00624)	0.124*** (0.00683)	0.129*** (0.00686)	0.126*** (0.00681)
Centro Oeste 2005	0.135*** (0.00732)	0.134*** (0.00735)	0.132*** (0.00744)	0.132*** (0.00739)
Constante	0.962*** (0.00640)	0.900*** (0.0446)	0.823*** (0.0421)	0.856*** (0.0421)
Instrumentos:				
Oferta de escolas no município		✓	✓	✓
Densidade de escolas ( $\rho = 1$ )			✓	
Densidade de escolas ( $\rho = 2$ )				✓
Observações		351,062	349,108	349,108
R <sup>2</sup>		0.091	0.067	0.080

Fonte: Elaboração própria. Os coeficientes são relatados com os erros-padrão em parênteses. \*\*\*, \*\* e \* indicam os níveis de significância de 1, 5 e 10%

**Tabela 4: Estimativa do logaritmo do rendimento/hora sem a Forbidden Regression**

	(4) IV*	(5) IV*	(6) IV*
Médio completo	0.0846 (0.0697)	0.201*** (0.0656)	0.151** (0.0658)
Homem	0.166*** (0.00846)	0.180*** (0.00799)	0.174*** (0.00801)
Branco	0.0952*** (0.00502)	0.0884*** (0.00484)	0.0914*** (0.00484)
21 anos	0.0338*** (0.00382)	0.0303*** (0.00376)	0.0319*** (0.00376)
22 anos	0.0691*** (0.00456)	0.0635*** (0.00443)	0.0659*** (0.00443)
Região Metropolitana 2005	0.185*** (0.00307)	0.182*** (0.00297)	0.183*** (0.00296)
Nordeste 2005	-0.187*** (0.00627)	-0.188*** (0.00625)	-0.187*** (0.00625)
Sudeste 2005	0.129*** (0.00666)	0.124*** (0.00654)	0.126*** (0.00655)
Sul 2005	0.119*** (0.00674)	0.123*** (0.00672)	0.122*** (0.00670)
Centro Oeste 2005	0.135*** (0.00733)	0.132*** (0.00735)	0.131*** (0.00733)
Constante	0.971*** (0.0408)	0.903*** (0.0386)	0.932*** (0.0386)
Instrumentos:			
Oferta de escolas no município	✓	✓	✓
Densidade de escolas ( $\rho = 1$ )		✓	
Densidade de escolas ( $\rho = 2$ )			✓
Observações	351,062	349,108	349,108
R <sup>2</sup>	0.097	0.092	0.096

Fonte: Elaboração própria. Os coeficientes são relatados com os erros-padrão em parênteses. \*\*\*, \*\* e \* indicam os níveis de significância de 1, 5 e 10%

## **6) Considerações finais**

Esse artigo teve como objetivo estimar o retorno educacional de nível médio para jovens brasileiros no ano de 2010. Além de atualizar as estimativas encontradas na literatura para períodos anteriores, o presente estudo propôs um conjunto de variáveis instrumentais que buscam captar a densidade de escolas quando o jovem tinha 15 anos, idade de entrada no ensino médio. O trabalho também inova ao usar dados censitários e no enfoque para população de jovens, recém-formados no ensino médio.

Os resultados apontam que a densidade de escolas possui uma relação positiva e significativa com a probabilidade de o jovem ter o ensino médio completo em 2010. As estimativas do retorno do ensino médio obtidas pelo método de variável instrumental (IV) excederam consideravelmente a estimativa pelo método de mínimos quadrados ordinários (MQO). Esse resultado é condizente com a literatura empírica que utiliza fatores institucionais relacionados ao sistema educacional como instrumento para educação. Se os retornos educacionais não foram constantes, e se a proximidade de escolas afetar mais as escolhas de indivíduos que, de outra forma, teriam uma escolaridade relativamente mais baixa, é esperado que o estimador IV superestime o retorno educacional

Diversos desdobramentos do exercício empírico aqui apresentado podem ser feitos. Os mais evidentes são a estimação dos retornos separadamente por sexo e a aplicação da mesma metodologia para edição 2000 do Censo Demográfico, e edições de 1993 a 1995 do Censo Escolar, de modo a possibilitar a comparação dos retornos de ensino médio entre os anos de 2000 e 2010. Outra possibilidade é estender a análise para as coortes mais velhas para uma amostra de jovens não migrantes, que residam no mesmo município desde os 15 anos.

## **7) Referências Bibliográficas**

ANGRIST, J.; KRUEGER, A. (1991) “Does compulsory school attendance affect schooling and earnings?” June 1990, Quarterly Journal of Economics Vol. 106, No. 4, November, p.979-1014

ANGRIST, J.; PISCHKE, J.-S. (2008). Mostly harmless economics: An empiricist’s companion. Princeton, NJ: Princeton University Press.

ASHENFELTER, O.; HARMON, C.; OSTERBEEK, H. (1999). A review of estimates of the schooling/earnings relationship with tests of publication bias. Labour Economics, 6, p. 453-470.

BARBOSA FILHO, F. H. & PÊSSOA, S. A. (2008) Retorno da Educação no Brasil. Pesquisa e Planejamento Econômico, Rio de Janeiro, v. 38, n. 1, p. 97-125.

BECKER, G. S. (1975). Human capital: a theoretical and empirical analysis. New York: Columbia University Press.

CARD D. (1995). Earnings, Schooling and Ability Revisited, in: Research in Labor Economics, Volume 14. Solomon W. Polachek, ed. Greenwich, Conn. and London: JAI Press, pp. 23–48.

- CARD D. (1999) The Causal Effect of Education on Earnings, in: Ashenfelter O., Card D. (Eds.) Handbook of Labor Economics, New York, North-Holland.
- CARD D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69 (5), 1127–1160.
- CRESPO, A; REIS, M. (2009). Sheepskin effects and the relationship between earnings and education: Analyzing their evolution over time in Brazil. *Revista Brasileira de Economia*, 63(3):209–231.
- DOYLE, W. R.; SKINNER, B. T. (2016). Estimating the education-earnings equation using geographic variation. *Economics of Education Review*, 53:254–267
- DUFLO, E. (2001) Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment”, *American Economic Review*. Vol 91, No 4, 795-813.
- HECKMAN J.J, LOCHNER L.J., TODD P.E. (2003) Fifty Years of Mincer Earnings Regressions, NBER Working Papers, no 9732, National Bureau of Economic Research, May.
- KASSOUF, A. L. (1994), ‘The wage rate estimation using heckman procedure’, *Revista de Econometria* 14(1), 89–107.
- KASSOUF, A. L. (1998), ‘Wage gender discrimination and segmentation in the brazilian labor market.’, *Brazilian Journal of Applied Economics* 2(2), 243–269.
- LAM, D.; SHOENI, R. (1993), ‘Effects of family background on earnings and returns to schooling: Evidence from brazil’, *Journal of Political Economy* 101, 710–739.
- LANGONI, C. G. (1974) *As causas do crescimento econômico do Brasil*. Rio de Janeiro: Apec.
- MINCER, J. (1974) *Schooling, experience e earnings*. Columbia University Press.
- MOURA, R. L. (2008). Testando as Hipóteses do Modelo de Mincer para o Brasil. *Revista Brasileira de Economia*, Rio de Janeiro, v. 64, n. 4, p. 407-449.
- RESENDE, M.; WYLLIE, R. (2006) Retornos para educação no Brasil: evidências empíricas adicionais. *Economia Aplicada*, São Paulo, v. 10, n. 3, p. 349-365, jul./set.
- PSACHAROPOULOS, G. (1994). Returns to investment in education: a global update. *World Development*, Oxford, v. 22, n. 9, p. 1325-1343, Sept. 1994.
- PSACHAROPOULOS, G.; PATRINOS, H. A. (2004) Returns to investment in education: a further update. *Education Economics*, Athens, v. 12, n. 2, p. 111-134.
- SACHSIDA, A; LOUREIRO, P. R. A. & MENDONÇA, M. J. C. (2004) Um Estudo Sobre Retorno em Escolaridade no Brasil. *Revista Brasileira de Economia*, Rio de Janeiro, v. 58, n. 2, p. 249-265.



SULIANO, D. C.; SIQUEIRA, M. L.(2012) Retornos da educação no Brasil em âmbito regional considerando um ambiente de menor desigualdade. *Economia Aplicada*, São Paulo, v. 16, n. 1, p. 137–165, jan./mar.

TEIXEIRA, W. M.; MENEZES-FILHO, N. A. (2012). Estimando o retorno à educação do Brasil considerando a legislação educacional brasileira como um instrumento. *Revista de Economia Política*, v. 32, n. 3, p. 479-496, São Paulo.

UEDA, E. M.; HOFFMANN, R. (2002) Estimando o retorno da educação no Brasil. *Economia Aplicada*, São Paulo, v. 6, n. 2, p. 209-238, abr./jun.

ZHANG, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance? A case study from Jinan, China. *Economics of Education Review*, 32, 1-28.

## 8) Apêndice: Resultados do 1º Estágio

**Tabela 5: Primeiro estágio do modelo de variável instrumental**

	(1)	(2)	(3)
	MQO	MQO	MQO
Oferta de escolas no município	0.0232*** (0.00136)	0.0222*** (0.00142)	0.0226*** (0.00141)
Densidade de escolas ( $\rho = 1$ )		0.00949*** (0.00185)	
Densidade de escolas ( $\rho = 2$ )			0.00679*** (0.00160)
Homem	-0.115*** (0.00195)	-0.115*** (0.00195)	-0.115*** (0.00195)
Branco	0.0584*** (0.00205)	0.0586*** (0.00206)	0.0586*** (0.00206)
21 anos	0.0326*** (0.00241)	0.0332*** (0.00241)	0.0331*** (0.00241)
22 anos	0.0504*** (0.00238)	0.0519*** (0.00240)	0.0515*** (0.00240)
Região Metropolitana 2005	0.0324*** (0.00203)	0.0338*** (0.00210)	0.0332*** (0.00210)
Nordeste 2005	0.00932** (0.00415)	0.0111*** (0.00417)	0.0110*** (0.00417)
Sudeste 2005	0.0346*** (0.00407)	0.0306*** (0.00411)	0.0321*** (0.00409)
Sul 2005	-0.0464*** (0.00448)	-0.0522*** (0.00461)	-0.0498*** (0.00455)
Centro Oeste 2005	-0.00862 (0.00533)	-0.0271*** (0.00560)	-0.0237*** (0.00553)
Constante	0.548*** (0.00448)	0.536*** (0.00505)	0.539*** (0.00491)
Observations	351,062	349,108	349,108
R-squared	0.023	0.023	0.023
F	622.46	570	568.73
Prob > F	0	0	0

Fonte: Elaboração própria. Os coeficientes são relatados com os erros-padrão em parênteses. \*\*\*, \*\* e \* indicam os níveis de significância de 1, 5 e 10%

**Tabela 6: Efeitos marginais sobre a probabilidade de completar o ensino médio**

	(1) Probit	(2) Probit	(3) Probit
Oferta de escolas no município	0.0237*** (0.00141)	0.0227*** (0.00147)	0.0231*** (0.00147)
Densidade de escolas ( $\rho = 1$ )		0.00946*** (0.00187)	
Densidade de escolas ( $\rho = 2$ )			0.00677*** (0.00163)
Homem	-0.116*** (0.00196)	-0.115*** (0.00196)	-0.115*** (0.00196)
Branco	0.0593*** (0.00209)	0.0595*** (0.00210)	0.0595*** (0.00210)
21 anos	0.0328*** (0.00242)	0.0334*** (0.00242)	0.0333*** (0.00242)
22 anos	0.0508*** (0.00239)	0.0523*** (0.00241)	0.0519*** (0.00241)
Região Metropolitana 2005	0.0328*** (0.00206)	0.0340*** (0.00214)	0.0335*** (0.00214)
Nordeste 2005	0.00917** (0.00414)	0.0110*** (0.00416)	0.0108*** (0.00416)
Sudeste 2005	0.0348*** (0.00407)	0.0309*** (0.00412)	0.0324*** (0.00410)
Sul 2005	-0.0479*** (0.00456)	-0.0538*** (0.00470)	-0.0514*** (0.00464)
Centro Oeste 2005	-0.00906* (0.00537)	-0.0277*** (0.00568)	-0.0243*** (0.00560)
Constante			
Observations	351,062	349,108	349,108

Fonte: Elaboração própria. Os coeficientes são relatados com os erros-padrão em parênteses. \*\*\*, \*\* e \* indicam os níveis de significância de 1, 5 e 10%

**Tabela 7: Primeiro estágio do modelo de variável instrumental, sem Forbidden Regres**

	(1) MQO*	(2) MQO*	(3) MQO*
Oferta de escolas no município	-0.0713*** (0.0101)	-0.0660*** (0.00964)	-0.0690*** (0.00981)
Densidade de escolas ( $\rho = 1$ )		-0.0283*** (0.00450)	
Densidade de escolas ( $\rho = 2$ )			-0.0206*** (0.00334)
Homem	0.353*** (0.0499)	0.342*** (0.0496)	0.350*** (0.0496)
Branco	-0.179*** (0.0254)	-0.174*** (0.0254)	-0.179*** (0.0254)
21 anos	-0.100*** (0.0144)	-0.0990*** (0.0146)	-0.101*** (0.0145)
22 anos	-0.155*** (0.0220)	-0.155*** (0.0226)	-0.157*** (0.0224)
Região Metropolitana 2005	-0.0991*** (0.0141)	-0.100*** (0.0147)	-0.101*** (0.0144)
Nordeste 2005	-0.0289*** (0.00583)	-0.0334***	-0.0337***
Sudeste 2005	-0.106*** (0.0156)	-0.0913*** (0.0139)	(0.0145)
Sul 2005	0.143*** (0.0206)	0.156*** (0.0230)	0.152*** (0.0220)
Centro Oeste 2005	0.0261*** (0.00647)	0.0806*** (0.0129)	0.0719*** (0.0116)
$p_1^{\wedge}$	4.079*** (0.434)		
$p_2^{\wedge}$		3.986*** (0.433)	
$p_3^{\wedge}$			4.060*** (0.432)
Constante	-1.688*** (0.238)	-1.600*** (0.232)	-1.649*** (0.233)
Observations	351,062	349,108	349,108
R-squared	0.023	0.024	0.024
F	578.7	533.4	532.9
Prob > F	0.00	0.00	0.00

Fonte: Elaboração própria. Os coeficientes são relatados com os erros-padrão em parênteses. \*\*\*, \*\* e \* indicam os níveis de significância de 1, 5 e 10%