

**ESTIMANDO RETORNOS DA EDUCAÇÃO NO CEARÁ: Uma aplicação do método
do Lasso para seleção de variáveis instrumentais fracas**

Rafael B. Barbosa

Professor dos Cursos de Ciências Econômicas e Finanças da Universidade Federal do Ceará (UFC),
Campus de Sobral.

rafaelbarrosbarbosa@gmail.com

(85) 999-301-938

Daniel Barboza Guimarães

Professor do Departamento de Administração da Universidade Federal do Ceará (UFC)

barbozadan@hotmail.com

ESTIMANDO RETORNOS DA EDUCAÇÃO NO CEARÁ: Uma aplicação do método do Lasso para seleção de variáveis instrumentais fracas

Resumo

Este trabalho pretende estimar os retornos da educação para o Estado do Ceará utilizando um conjunto grande de instrumentos fracos. Para identificar os instrumentos relevantes será utilizado o método de seleção de variáveis do Lasso (*Least absolute shrinkage and select operator*). Para corrigir a possível presença de viés na estimação de primeiro estágio, será adotado o procedimento do pós-lasso, como proposto por Belloni *et al* (2010, 2012, 2014). Com isso, espera-se obter estimativas consistentes e que não dependem unicamente da teoria econômica para sua determinação. Dos resultados encontrados, verificou-se que os retornos da educação tiveram magnitude significativamente menor do que àqueles encontrados na literatura. Isso pode servir como indicativo de que os métodos anteriores de seleção *ad-hoc* de instrumentos podem gerar resultados viesados.

Palavras Chave: Retornos da educação, Seleção de variáveis, Instrumento fracos, Ceará.

Abstract

The aim of this paper is to estimate the educational returns to Ceará using many weak instruments. To identify relevant instruments it was used the method of variable selection of Lasso (*Least absolute shrinkage and select operator*). The bias in the first stage was corrected by post-lasso method proposed by Belloni *et al* (2010, 2012, 2014). Then, we hope that the educational returns estimates to be consistent and that do not have only dependence by economic theory for its determination. The results show that the magnitude of estimates is significantly lesser than that was estimated by literature. That could indicate that the ad-hoc selection methods of instruments can generate biased estimates.

Key Words: Educational Returns, Variable selection, Weak Instruments, Ceará.

JEL: J24, I21, I25

1. Introdução

O capital humano é um dos principais determinantes da taxa de crescimento e do nível de bem-estar de um país. De acordo com Resende e Wyllie (2006) as fontes de crescimento econômico têm sido crescentemente associadas a investimentos em capital humano e o fenômeno da decolagem (*catching up*) de alguns países asiáticos esteve amplamente relacionado a investimentos expressivos em educação.

O debate em torno do peso da educação como fator explicativo da desigualdade de renda existente no Brasil é intenso. Destaca-se a corrente que defende que a escolaridade é o principal fator explicativo causal da desigualdade, por gerar diferenças de produtividade entre os indivíduos que podem persistir por todo seu ciclo de vida (MENEZES-FILHO (2001)).

As evidências indicam que os retornos econômicos à educação no Brasil estão entre os mais elevados do mundo. No entanto, recentemente, vêm sendo registrado um declínio deste retorno. Esse processo em parte deve-se ao próprio processo de expansão educacional, que aumentou a oferta relativa de pessoas com ensino fundamental e médio (MENEZES-FILHO, (2001)). De fato, o papel da escolaridade vem exercendo forte influência desde o início da década de 1990 em consequência da persistente redução dos retornos médios à educação, como apontado por alguns autores como, por exemplo, Ferreira e Veloso (2006).

Diversos artigos, recentemente publicados na literatura econômica brasileira, são enfáticos ao atribuírem papel importante à educação na redução contínua e acelerada da desigualdade de renda no Brasil, apesar do enorme papel das transferências ao longo dos últimos anos¹.

De acordo com Carvalho, Neri e Silva (2006) o propósito desses trabalhos é avaliar a situação em que indivíduos com atributos produtivos semelhantes, têm salários tão diferenciados. Para que tais propósitos sejam alcançados, ou seja, para capturar os efeitos sobre as diferenças de rendimentos causadas por diferenças de escolaridade formal dos indivíduos, grande parte desses estudos desenvolve versões da equação de salário proposta por Mincer (1974).

No entanto, equações mincerianas são fortemente marcadas pela presença da endogeneidade, isto é, a presença de variáveis omitidas que são fortemente correlacionadas com a educação e que possuem um efeito indireto sobre a renda do indivíduo. Uma das principais estratégias para corrigir o problema da endogeneidade está na utilização de variáveis instrumentais (VI).

A prática comum na utilização de variáveis instrumentais é bastante limitada por dois motivos. Primeiro, as VI's devem ter forte correlação com a variável endógena e baixa correlação com o erro. VI's que possuem esta característica são chamadas de VI's fortes (ANGRIST e PISCHKE (2009)). Tais tipos de VI são bastante difíceis de serem encontradas em base de dados tradicionais (Pesquisas Amostrais e Census), sendo mais comum encontrar um conjunto de VI's fracas, isto é, variáveis que possuem baixa correlação relativa com as variáveis endógenas. Segundo, a escolha das VI é normalmente realizada por meio da intuição teórica dos pesquisadores. Ou seja, poucos critérios estatísticos são utilizados para determinar se uma variável serve ou não como VI.

Diante destes dois tipos de problemas, pesquisadores têm buscado alternativas empíricas para justificar a utilização de um grande conjunto de VI's fracas. Duas estratégias

¹ Soares (2006b), Hoffman (2006), Barros et al. (2006), Barros, Carvalho, Franco e Mendonça (2007), Ferreira et al. (2006).

têm sido tomadas, a redução da dimensão e a seleção de variáveis. Bai e Ng (2009) e Kapetanios e Marcellino (2010) buscam reduzir a dimensão de um conjunto de VI's fracas por meio da estimação de fatores comuns. Groen e Kapetanios (2009), por outro lado, utilizam mínimos quadrados parciais para o mesmo fim.

Mais recentemente, autores como Belloni *et al* (2014) propuseram métodos baseados no *esparçamento* de dados (*Approximate Sparse Models*), com destaque para o método do Lasso (*Least absolute shrinkage and select operator*) e Post-Lasso. A ideia básica consiste em permitir que sejam escolhidas, no primeiro estágio, variáveis instrumentais que têm elevado poder de explicação da variável endógena.

A técnica do Lasso consiste na estimação por Mínimos Quadrados Ordinários aplicando uma penalidade normada do tipo L1. Ou seja, trata-se de um Método de Mínimos Quadrados Ordinários Restrito que faz com que os coeficientes estimados de variáveis que não contribuam fortemente para explicar o comportamento da variável endógena tendam a zero.

O método do Lasso² foi originalmente desenvolvido para selecionar variáveis que possuam um elevado poder preditivo. Devido ao *trade-off* entre variância e viés (HASTIE *et al* (2012)), o Lasso tende a produzir estimativas viesadas. Para reduzir o peso deste viés, Belloni *et al* (2014) propuseram o Post-Lasso, que consiste em estimar o modelo em segundo estágio, utilizando apenas os instrumentos selecionados pelo Lasso, por mínimos quadrados ordinários.

Alguns estudos posteriores têm mostrado relativo sucesso da aplicação de tais técnicas. Belloni *et al* (2014) comparam os resultados apresentados pelo Lasso e o Post-Lasso aplicados a mesma base de dados de Angrist e Krueger (1991) e mostram que o trimestre de nascimento constitui um instrumento forte. Outras aplicações recentes foram Hansen *et al* (2015), Belloni *et al* (2012), Belloni *et al* (2013).

O objetivo deste trabalho consiste em estimar as equações mincerianas para o Estado do Ceará utilizando as técnicas de seleção de variáveis do Lasso e do Post-Lasso. Espera-se que tais métodos forneçam estimativas mais apropriadas dos retornos da educação, visto a dificuldade de selecionar, teoricamente, variáveis instrumentais fortes no contexto brasileiro.

Será utilizada a base de dados da Pesquisa Nacional por Amostragem Domiciliar (PNAD) realizada anualmente pelo IBGE. Para comparar temporalmente os resultados serão estimados os retornos educacionais para os anos de 2002 e 2012. Além disso, a amostra será estratificada entre regiões metropolitanas e não metropolitanas. Para situar o retorno educacional no contexto nordestino, o mesmo exercício será realizado para a região Nordeste e posteriormente comparado com o Ceará.

Portanto, este estudo realiza duas contribuições para a literatura sobre o retorno da educação. Primeiro, contribui para a discussão da seleção de variáveis instrumentais fracas, tendo o retorno da educação como foco. Diante da dificuldade de obtenção de instrumentos fortes, as técnicas do Lasso e do Post-Lasso mostram-se como uma alternativa factível e promissora. Não é do conhecimento dos autores nenhum trabalho que tenha aplicado tais técnicas no Brasil. Segundo, ajuda a compreender o papel da educação na elevação da renda

² Tal técnica faz parte de um conjunto de métodos que buscam selecionar variáveis que reduzam uma função perda, chamados de métodos de redução (*Shrinkage Methods*). Dentre este ainda se incluem, *elastic net*, *ridge regressions*, entre outros.

dos indivíduos no Nordeste e no Ceará. Destaca-se o papel da educação na superação da pobreza entre os indivíduos.

Além desta introdução este trabalho está subdividido nas seguintes seções. Na seção seguinte, será feita uma análise das evidências empíricas da literatura. Na seção três, será apresentada a metodologia do Lasso e do Post-Lasso e discutido a base de dados utilizada. A seção quatro exhibe os resultados e a última seção apresenta e discute as principais conclusões.

2. Revisão da Literatura

O debate sobre o retorno da educação no salário dos indivíduos foi iniciado por Schultz (1961) e Ben-Porath (1967). No entanto, segundo esses autores, tais discussões só foram intensificadas a partir de 1974, quando Mincer formalizou teoricamente a derivação da equação de rendimentos, a qual utiliza apenas o nível educacional como representativo do capital humano. Nesse estudo, o autor constatou que um ano adicional de escolaridade gerava um aumento de 11,5% nos rendimentos.

Desde então, essa formulação passou a ser amplamente utilizada em estudos do mercado de trabalho, os quais procuram obter informações sobre os determinantes do salário dos trabalhadores, a partir de suas características individuais (escolaridade, experiência, cor/raça ou local de moradia) e informações sobre o mercado de trabalho (setor de atividade ou ocupação).

Ashenfelter e Krueger (1994) coletaram dados sobre salários e educação para 198 gêmeos univitelinos, residentes nos Estados Unidos, com diferentes níveis de escolaridade. Os resultados indicaram que mesmo entre os indivíduos nascidos na mesma família e com características genéticas idênticas, aqueles que estudaram mais recebem um salário maior, onde um ano a mais de estudo pode gerar um aumento salarial entre 12% a 16%. Rouse (1999), Arias e McMahon (2001), os quais afirmaram que os retornos da educação variam em torno de 10%, 13% e de 5% a 10%, respectivamente.

Utilizando informações referentes ao nível educacional e a experiência de trabalho, Boudarbat, Lemieux e Riddell (2008) buscaram analisar a evolução dos retornos do capital humano no Canadá sobre as diferenças salariais, no período de 1980 a 2006. Observaram que os retornos à educação permaneceram maiores para as mulheres. No entanto, para o período de 2000-2006 os resultados mostram uma tendência decrescente das diferenças salariais medidas pelo retorno à educação.

Zepeda *et al* (2009) utilizaram variáveis demográficas e socioeconômicas, dentre elas, o nível educacional para identificar os fatores que explicam as alterações observadas na média dos rendimentos em três países da América Latina – Brasil, Chile e México. Os autores constataram que a única mudança mais significativa em todos os países e períodos, foi o impacto do nível educacional sobre a força de trabalho, indicando que a escolaridade é o fator mais importante para o aumento nos rendimentos.

Já Sapelli (2009) verificou as taxas de retornos da educação no Chile, onde afirmou que estas taxas variam de acordo com o nível educacional, divididos em três grupos, educação básica, média e superior, conforme comprovado com os resultados empíricos de 5,6%, 14,3% e 25%, respectivamente. Além disso, o trabalho compara estes resultados com os observados na Argentina, Colômbia, México e Brasil e conclui que apenas os retornos no ensino básico se equiparam com os do Chile.

Um dos primeiros autores a ressaltar a importância da educação como fator explicativo para a desigualdade brasileira foi Langoni (1973). De acordo com esse autor, o aumento da desigualdade no Brasil, entre 1960 e 1970, ocorreu devido ao aumento na demanda por trabalhadores qualificados associados à industrialização.

Na tentativa de capturar a influência do nepotismo ou de habilidades não observadas na relação entre educação e salários, Lam e Schoeni (1993) utilizaram informações referentes a educação do pai, da mãe, do sogro e da sogra, além da educação do próprio indivíduo, contidas na PNAD de 1982, como determinantes dos rendimentos dos indivíduos. O efeito da educação sobre os salários dos indivíduos cujos pais, mães, sogros e sogras têm o mesmo nível educacional permanece em torno de 11%.

Barros (1997, 2000) enfatizou que um dos principais problemas sociais no Brasil decorre do baixo nível e da má distribuição da educação na população brasileira. Pessoa e Barbosa Filho (2008) verificaram que as taxas de retorno do ensino primário (10%), ginásial (14%) e universitário (18%) mantiveram-se extremamente elevadas no Brasil. Os autores concluem que o investimento em educação no Brasil é extremamente atrativo, uma vez que oferece elevadas taxas de retorno.

Mariano e Arrais (2011) verificaram que as taxas de retorno da educação variam de acordo com as regiões brasileiras, sendo a região Nordeste a que apresentou o menor valor dentre elas (14,68%). Sendo assim, o diferencial entre os retornos da educação causada pela diferença regional chega a 26%.

Percebe-se que, mesmo em um período onde a desigualdade de renda no Brasil esteve entre as mais baixas nos últimos 30 anos aliada ainda a uma rápida expansão educacional, os resultados encontrados, apesar de não comparáveis com outros estudos semelhantes devido às diferentes variáveis explicativas e base de dados, mostram que o prêmio à escolaridade no Brasil ainda se mantém em patamares elevados (SULIANO; SIQUEIRA, 2012).

Portanto, das diversas variáveis que influenciam o salário de um indivíduo, o nível de escolaridade é um dos que tem maior importância (IOSCHPE, 2004). Além disso, conforme enumerado nos trabalhos supracitados, essa influência apresenta uma variabilidade considerável, ou seja, o impacto da educação no salário apresenta variações significativas quando são realizadas comparações entre os gêneros, raças, países e regiões geográficas, por exemplo.

Esses estudos, ao procurarem evidenciar a importância da educação na explicação dos diferenciais de renda, acabam incorrendo em uma série de problemas. O primeiro problema é o viés de seleção amostral, o qual, de acordo com Heckman (1979), pode surgir praticamente por dois motivos: pela auto-seleção de indivíduos ou unidades amostrais investigadas e pelas decisões de seleção da amostra pelos analistas.

Cameron e Trivedi (2005) indicam que a omissão e/ou a característica endógena de uma variável em uma equação, especificamente de rendimentos, poderá causar um problema grave de viés e inconsistência nos estimadores quando se aplica o usual método de mínimos quadrados ordinários (MQO), em virtude do grave problema de endogeneidade entre educação e salários que as equações mincerianas enfrentam (SULIANO; SIQUEIRA, 2012).

Na tentativa de solucionar o problema da endogeneidade entre educação e salários, diversos autores passaram a utilizar o método de Variáveis Instrumentais (VI). No entanto, Heckman e Urzua (2009) relatam uma série de problemas que podem emergir a partir de uma estimação por VI, como por exemplo, deficientes instrumentos provêm estimativas viesadas;

estimativas de VI advêm de fortes restrições tomadas a priori sobre os dados; em modelos heterogêneos, diferentes instrumentos fornecem estimativas distintas; estimativas de VI, dependendo dos instrumentos utilizados e hipóteses feitas, gerarão diferentes estimativas do retorno da educação, as quais podem levar a incorretas interpretações. Mariano e Arraes (2011) e Suliano e Siqueira (2012) observam que, que o modelo estimado pelo método de VI apresenta um Erro Quadrado Médio (EQM) maior que o modelo endógeno, indicando que, o método de VI apresenta um pior ajuste, o que prejudica a questão da previsão.

3. Metodologia

Um dos problemas recorrentes na estimação de equações de salário mincerianas deve-se à presença de endogeneidade, causada pela existência de correlação entre as variáveis explicativas e o erro. No caso, a educação apresenta forte correlação com o erro devido à existência de variáveis omitidas como habilidade inata, herança familiar entre outras.

Variáveis instrumentais (VI) tem sido frequentemente aplicadas, com relativo sucesso na literatura (ANGRIST e KRUEGER (1991), CARD (1995), ANGRIST (1990), entre outros³), para tentar reduzir o efeito da endogeneidade. Todavia, a identificação de quais variáveis devem ser consideradas instrumentos fortes tem sido, na maioria dos casos, realizada por meio do uso da teoria econômica. Isto é, determinada variável será considerada uma VI ideal desde que possua forte correlação com a variável endógena e nenhuma correlação com o erro. A primeira condição é facilmente testada, porém, a segunda condição não é, uma vez que depende do erro, o qual é não observável.

Dessa forma, para selecionar uma VI os pesquisadores buscam encontrar justificativas teóricas que possibilitem inferir a ausência de correlação com a variável e o erro. Tal estratégia pode não necessariamente resultar em instrumentos capazes de corrigir o problema da endogeneidade, uma vez que tal procedimento pode acabar selecionando variáveis que servem apenas como VI's fracas e, portanto, podem elevar o viés devido à endogeneidade⁴.

Recentemente, algumas estratégias econométricas alternativas têm sido propostas para lidar com a pouca existência de variáveis instrumentais fortes, as quais buscam extrair de um conjunto de variáveis instrumentais fracas aquelas que podem ser utilizadas para estimar os modelos estruturais. Tais alternativas objetivam evitar a total dependência da escolha das VI's pela teoria econômica. O objetivo consiste em aliar a teoria econômica, necessária mesmo que em menor grau, a técnicas estatísticas de eficiência sobre o poder preditivo das candidatas a VI.

Dois métodos têm ganhado destaque: Redução da dimensionalidade e seleção de variáveis. A redução da dimensionalidade consiste em extrair, de um conjunto de variáveis instrumentais fracas, variáveis latentes que resumam o comportamento da relação entre educação e seus instrumentos. Bai e Ng (2008) e Kapetanios e Marcellino (2010) utilizam componentes principais e Groen e Kapetanios (2009) mínimos quadrados parciais.

A redução da dimensionalidade, porém, não se aplica diretamente ao problema aqui estudado, pois, requer um painel de dados, em que são extraídos fatores comuns. Estes

³ Uma recente literatura tem buscado analisar os retornos da educação diante de heterogeneidade provocadas por políticas públicas, mas não captadas pelas variáveis instrumentais. Destacam-se: Heckman et al (2006, 2008), Carneiro et al (2011), entre outros.

⁴ Ver Angrist and Pischke (2009).

formam uma combinação multiplicativa entre os fatores temporais e a heterogeneidade dos indivíduos (*fator loadings*).

A seleção de variáveis, por sua vez, busca escolher, dentre um conjunto de variáveis instrumentais fracas, aquelas que têm maior poder de explicação sobre a variável endógena. Tais métodos forçam as variáveis instrumentais fracas, com pouco poder explicativo, a ter seus coeficientes estimados tendendo a zero.

Observe que ambas as técnicas possuem forte apelo para a parcimônia do modelo estrutural. De fato, a seleção de variáveis irá selecionar s , de um conjunto contendo p variáveis instrumentais fracas, onde ($s \ll p$), as quais serão utilizadas como VI's. As demais variáveis não serão utilizadas na estimação.

Existem dois motivos para excluir determinadas variáveis. Primeiro, é que tais variáveis podem não ter significativo poder de explicação do comportamento da variável endógena. Nesse caso, a inclusão desta variável pode resultar em relativa perda de eficiência em comparação ao ganho em poder de explicação. Segundo, duas variáveis podem ter elevada correlação, de forma que, a presença de uma apenas sirva para explicar o comportamento da variável endógena.

Dentre os métodos de seleção de variáveis disponíveis, destaca-se a aplicação do Lasso e do Post-Lasso, como em Belloni *et al* (2012, 2014, 2013) e Hansen *et al* (2015). O método do Lasso foi originalmente desenvolvido por Tibshirani (1996). Este método busca selecionar variáveis que tenham elevado poder preditivo para a variável dependente. Belloni *et al* (2012) adotam este método para estimar equações estruturais, utilizando variáveis instrumentais fracas, as quais foram previamente selecionadas.

Considere a seguinte equação estrutural minceriana:

$$y_i = \alpha + \beta_1 y_{1i} + x' \delta + e_i, \quad i = 1, \dots, n \quad (1)$$

Em que: y_i é o log do salário do indivíduo i ; y_{1i} corresponde a medida de capital humano, no caso deste artigo, será utilizado os anos de estudo; x corresponde a um vetor de variáveis de controle de tamanho ($k \times n$) e e_i corresponde ao erro, que por hipótese tem as seguinte propriedade estatística: $e_i \sim iid N(0,1)$. A endogeneidade é caracterizada pela forte correlação entre y_{1i} e e_i .

Suponha que exista um conjunto Z de tamanho ($p \times n$) de variáveis instrumentais fracas. Baseando-se na estratégia de Mínimos Quadrados Ordinários em Dois Estágios (MQO2E) a equação de primeiro estágio pode ser escrita como:

$$y_{1i} = Z' \gamma + x' \delta + u_i \quad (2)$$

Observe que se p for muito grande, então o método dos mínimos quadrados se torna inviável na estimação do primeiro estágio. Dessa forma, o objetivo do método do Lasso consiste em selecionar s variáveis de Z ($s \ll p$), de modo a tornar a estimação da equação (2) factível. Para tanto, o Lasso estima os parâmetros por meio da otimização convexa restrita, a qual se dá da seguinte forma:

$$\beta_L = \arg \min \hat{Q}(\beta) + \frac{\lambda}{n} \|\beta\|_1 \quad (3)$$

Em que, $\hat{Q}(\beta) = E[(y_{1i} - f'\beta)^2]$, $f = (Z, x)$, $\beta = (\gamma, \delta)$, $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$. Isto é, o Lasso busca minimizar a soma dos quadrados dos resíduos impondo uma restrição aos parâmetros. O termo λ , chamado de *tuning point*, funciona com uma penalidade, uma vez que corresponde ao peso que será dado a restrição. Observe que a equação (3) pode ser rescrita da seguinte forma:

$$\beta_L = \arg \min \sum_{i=1}^n (y_{1i} - f'\beta)^2$$

sujeito a $\sum_{j=1}^p |\beta_j| \leq t$

Esta última forma deixa mais explícita a forma de otimização restrita proposta pelo método do Lasso. Um elemento fundamental da seleção de variáveis via Lasso consiste na escolha correta do *tuning point*. Tibshirani (1996) propôs a escolha de λ via correlação cruzada, isto é, o lasso é aplicado considerando um intervalo para o valor do λ . O λ escolhido é aquele que minimiza o erro quadrático médio.

Tal estratégia, entretanto, não é diretamente aplicável a problemas de estimação de modelos estruturais, por meio de variáveis instrumentais, pois o interesse não é necessariamente realizar previsão. Desse modo, Bickel, Rivot e Tsybakov (2009) propuseram o seguinte valor para λ :

$$\lambda = 2. c\sigma\sqrt{2n \log(2p/\zeta)}$$

Em que, $c > 1$, $1 - \zeta$ refere-se ao intervalo de confiança necessário para ficar próximo de 1, onde σ consiste no desvio padrão do erro. Tal derivação é motivada buscando obter *near-optimal* taxas de convergência dos estimadores após a seleção. Todavia, tal escolha possui duas limitações. Primeiro, foi derivada supondo a homocedasticidade. Segundo, não é factível, tendo em vista que σ não é observável. Portanto, visando resolver esse problema, Belloni *et al* (2010) propuseram um λ factível e aplicável na presença de heteroscedasticidade. Este λ será utilizado para estimar (3) neste artigo.

O Lasso é um caso particular de um conjunto de métodos de seleção de variáveis, conhecidos como *shrinkage methods*. Todavia, o Lasso possui duas vantagens frente às outras técnicas de seleção de variáveis. Primeiro, ele seleciona variáveis ao forçar certos parâmetros irrelevantes a apresentarem valores tendendo à zero. Assim, supondo que existam duas variáveis que possuem alta correlação com a variável explicativa, então uma delas terá seu parâmetro estimado tendendo à zero. Outras técnicas como regressões do tipo *ridge* não produzem esta característica.

Segundo, a forma funcional da equação (3) é convexa. Portanto, é computacionalmente mais fácil a obtenção de ótimos, onde tais ótimos são globais. Estas duas propriedades fazem com que o método do Lasso seja bastante apropriado para selecionar quais variáveis instrumentais fracas são relevantes para explicar o comportamento da variável endógena, como em Belloni *et al* (2010).

Entretanto, o método do Lasso possui também limitações. De fato, o Lasso não lida bem com a presença de multicolinearidade (ZOU e HASTIE, 2005). Isso ocorre

justamente pela forma de seleção de variáveis em que aquelas variáveis que possuem elevada correlação entre si têm seus coeficientes estimados forçados a tenderem a zero.

Mais grave, porém, é o fato de o Lasso produzir estimativas viesadas. O método do Lasso foi desenvolvido para selecionar variáveis de forma a aumentar o poder preditivo dos modelos. Isto é, as variáveis são selecionadas buscando minimizar a função perda quadrática. Todavia, à medida que a variância se reduz, o viés se eleva⁵.

Para reduzir o peso do viés, métodos alternativos têm sido propostos. Por exemplo, Zou e Hastie (2002) propuseram o *elastic net*, que busca suavizar a perda com o viés do Lasso sem, no entanto, perder a qualidade na seleção das variáveis. Belloni *et al* (2012, 2013) por sua vez, adotam o Post-Lasso, onde este método consiste em reestimar a equação (2) por mínimos quadrados ordinários (MQO), após a seleção de variáveis.

Dessa forma, primeiro realiza-se a seleção de variáveis, em que s dentre as p possíveis variáveis instrumentais são escolhidas ($s \ll p$). Logo em seguida, a equação (2) é reestimada considerando apenas as variáveis selecionadas. Por fim, utiliza-se a estimativa do primeiro estágio para estimar a equação estrutural (1) de segundo estágio.

Belloni *et al* (2010) mostraram que esta técnica se adapta bem a problemas em que existem poucas (ou nenhuma) variável instrumental forte, porém, existem muitos instrumentos fracos. Este é justamente o caso da maior parte das bases de dados brasileiras de onde se pode extrair dados para a análise do retorno da educação.

Portanto, neste trabalho adota-se a técnica de estimação do Post-Lasso, a qual consiste em aplicar uma seleção *ex-ante* a estimação do modelo estrutural sobre um conjunto grande de variáveis instrumentais fracas.

É importante observar que não necessariamente as variáveis selecionadas precisam ter significado econômico. De fato, estas terão apenas forte poder de explicação sobre a variável endógena.

Todavia, como teoricamente se sabe que algumas variáveis são fundamentais para a estimação das equações mincerianas, estas variáveis não estarão sujeitas a seleção via Lasso e serão consideradas variáveis explicativas durante a estimação da equação estrutural.

Vale ressaltar que, não é de conhecimento dos autores da existência de nenhum outro trabalho que tenha utilizado técnicas de seleção de variáveis para estimação de equações de salário no contexto brasileiro. Assim, ademais a importância social e de planejamento de políticas públicas, este trabalho proporciona um avanço na estimação do retorno da educação no Brasil.

3.1 Base de Dados

Será utilizada a base de dados da Pesquisa Nacional por Amostragem Domiciliar (PNAD) realizada anualmente pelo IBGE. Foram escolhidos dois períodos, com uma diferença de 10 anos, para verificar o comportamento temporal dos retornos da educação. Assim, os anos selecionados foram 2002 e 2012.

As variáveis de controle escolhidas são: experiência (*exp*), experiência ao quadrado (*exp_2*), sexo (feminino como referência), negros e pardos. Tais variáveis serão utilizadas

⁵ De fato, sendo y_0 a variável dependente e $\hat{f}(x_0)$ os previsores estimados então, o erro quadrático médio pode ser decomposto em $E(y_0 - \hat{f}(x_0))^2 = Var(\hat{f}(x_0)) + [Viés(\hat{f}(x_0))]^2 + Var(\epsilon)$, em que: ϵ consiste no erro de aproximar y_0 por $\hat{f}(x_0)$. Para maiores detalhes ver, Hastie et al (2009).

como potenciais instrumentos, entretanto, mesmo que não sejam selecionadas pelo Lasso, figurarão na equação estrutural.

Algumas observações foram eliminadas para gerar resultados plausíveis com a média da população e evitar a presença de *outliers*. Desse modo, foram consideradas pessoas acima dos 22 anos, cuja renda individual é inferior a R\$ 5000,00⁶. A variável dependente é o *log* do salário.

Inicialmente, quarenta e oito (48) variáveis foram selecionadas como potenciais instrumentos fracos. Entretanto, sabe-se que certas transformações geram novas variáveis instrumentais fracas (BELLONI *et al* (2012), CHEN (2007)). Assim, foram realizadas transformações polinomiais (elevar ao quadrado e ao cubo) e transformações por interação entre as VI's fracas. Dessa forma, ao final, foram utilizados oitenta (80) potenciais instrumentos fracos⁷.

A amostra foi estratificada entre região metropolitana e não metropolitana. O objetivo é identificar se existem diferenças no retorno da educação entre capital e interior. Essa possível diferença decorre da existência de estruturas de mercado de trabalho e oferta de serviços educacionais diferenciados entre tais regiões. Por fim, o mesmo exercício foi realizado para a região Nordeste com o objetivo de se verificar como se comportam os resultados encontrados para o Ceará quando comparados à região a qual ele pertence.

A Tabela 01 apresenta os valores médios e os desvios-padrões da renda pessoal e dos anos de estudo para os anos de 2002 e 2012. A amostra foi estratificada entre região metropolitana e não metropolitana.

Tabela 1: Estatísticas Descritivas das Variáveis Renda Pessoal e Anos de Estudo.

Variáveis	Média		Desvio-Padrão	
	2002	2012	2002	2012
Amostra Completa				
Renda Pessoal	352,30	718,30	338,75	385,79
Anos de Estudo	8,29	9,35	4,68	4,2
Região Metropolitana				
Renda Pessoal	372,01	753,54	347,04	379,28
Anos de Estudo	8,63	9,64	4,56	4,07
Região Não Metropolitana				
Renda Pessoal	177,60	463,94	174,59	333,67
Anos de Estudo	5,2	7,24	4,64	4,53

Fonte: Elaboração dos autores a partir dos dados da PNAD.

Depreende-se da Tabela 01 que houve no período em análise um crescimento substancial da renda pessoal, seja para a região metropolitana (102%), não metropolitana (161%) ou considerando a amostra completa (103%). No entanto, mesmo apresentando um crescimento consideravelmente superior, verifica-se que os indivíduos residentes nas regiões

⁶ Valores atualizados em 2002.

⁷ Todas as rotinas utilizadas, sejam para estimar o post-lasso ou para criar e ajustar a base de dados encontram-se disponíveis desde que solicitadas no email: rafael_barbosa@ufc.br.

não metropolitana recebem, em média, quase a metade recebida por uma pessoa que reside na região metropolitana.

Em relação aos anos de estudo, verifica-se que também houve um crescimento deste indicador, em termos médios, no período analisado. Além disso, constatou-se que o crescimento dos anos de estudo seguiu o padrão do crescimento verificado na renda, uma vez que este foi mais relevante na região não metropolitana (39%) do que na região metropolitana (12%). Vale salientar que, mesmo com essa disparidade significativa, em média, os residentes na região metropolitana apresentam mais anos de estudo. Diante dos resultados verificados na tabela 1, verifica-se que, em média, indivíduos mais escolarizados recebem maiores salários.

Apesar de analisar o retorno da educação entre as regiões metropolitanas e não metropolitana do Ceará, os resultados da amostra completa serão comparados aos resultados para a região Nordeste. O objetivo deste exercício adicional consiste em entender se o retorno educacional do Ceará está inserido no contexto regional.

4. Resultados

Esta seção se reserva a apresentação e análise dos resultados a partir das estimações da equação minceriana de salários, equação (1), com o propósito de se verificar o impacto dos anos de estudo nas remunerações dos indivíduos no Nordeste e no Ceará. Convém mencionar que foram feitas algumas divisões na amostra referente ao Estado do Ceará, visando homogeneizá-la de forma a dar uma maior robustez aos resultados. Desse modo, as estimativas supracitadas foram realizadas levando-se em consideração a Região Metropolitana e a Região Não Metropolitana. Os resultados podem ser verificados na tabela 2, apresentada a seguir.

Percebe-se que os modelos estão bem ajustados e que quase todos os coeficientes estimados apresentam sinais condizentes com a literatura que aborda a problemática dos retornos da educação.

Em todas as estimações verifica-se que indivíduos do sexo masculino, com um maior nível de escolaridade, brancos e mais experientes recebem, em média, maiores salários. Estes resultados já são esperados na literatura.

Em relação ao retorno à educação observa-se que tanto as regiões Nordeste como o Estado do Ceará apresentaram uma redução neste índice entre os anos de 2002 e 2012. Na região Nordeste, o retorno à educação apresentou uma pequena redução, uma vez que passou de 5,9%, em 2002, para 4,2%, em 2012.

Já o Estado do Ceará, quando comparado ao Nordeste, apresentou uma redução bastante significativa no período analisado. Em 2002, um ano a mais de estudo acarretava em um aumento, em média, de 11% no rendimento do indivíduo, ao passo que em 2012 este aumento passou a ser de apenas 4%. Portanto, atualmente, o Estado do Ceará apresenta um retorno à educação inferior ao da Região Nordeste.

Este resultado pode ser devido ao aumento do crescimento educacional verificado nos últimos anos no Estado. De fato, com a expansão da educação, o mercado de trabalho se torna mais concorrido, reduzindo com isso os retornos médios à educação.

No tocante as análises estratificadas, ou seja, região metropolitana e não metropolitana, verifica-se que nos dois casos manteve-se a trajetória de queda no período analisado. Na região não metropolitana tem-se um maior retorno à educação, 11% em 2002 e 7% em 2012, contra 6% em 2002 e 0,9% em 2012 na região metropolitana.

Tabela 2: Resultados para as Equações Estruturais após a Seleção das Variáveis.

Variáveis	2002				2012			
	Nordeste	Ceará			Nordeste	Ceará		
		Amostra Toda	RM	Não RM		Amostra Toda	RM	Não RM
Educ	0.059* (0.013)	0.1152* (0.0087)	0.0671* (0.0044)	0.1101* (0.0103)	0.042* (0.006)	0.0466* (0.0059)	0.0091* (0.0044)	0.0760* (0.0107)
Exp	-18.306 NaN	0.0772* (0.0069)	0.0862* (0.0082)	0.0471* (0.0068)	0.043* (0.004)	0.0319* (0.0098)	0.0302* (0.0082)	0.0410* (0.0160)
Exp_2	-0.001* (0.000)	-0.0009* (0.0001)	-0.0011* (0.0002)	-0.0005* (0.0001)	-0.001* (0.000)	-0.0004* (0.0002)	-0.0004* (0.0002)	-0.0005* (0.0003)
Sexo	0.224* (0.007)	0.2020* (0.0208)	0.1915* (0.0197)	0.2452* (0.0279)	0.323* (0.021)	0.2325* (0.0237)	0.2459* (0.0197)	0.2935* (0.0436)
Negro	-0.122* (0.013)	-0.1141* (0.0643)	-0.1872* (0.0718)	-0.1026* (0.0804)	-0.043* (0.017)	0.0045 (0.0378)	-0.0303* (0.0718)	-0.0180* (0.0799)
Const	-80252 (1113385)	3.1700* (0.0978)	3.4489* (0.1150)	3.3089* (0.1221)	4.965* (0.111)	5.1259* (0.1168)	5.5538* (0.1150)	4.4009* (0.2179)
Pardo	-0.111* (0.013)	-0.0918* (0.0129)	-0.1188* (0.0139)	-0.0834* (0.0294)	-0.090* (0.010)	-0.0899* (0.0230)	-0.0788* (0.0139)	-0.0576* (0.0534)
R2	0.985	0.9846	0.9882	0.9792	0.991	0.9912	0.9948	0.9855

Fonte: Elaboração própria a partir dos dados das PNADs de 2002 e 2012.

Nota: Erros- Padrões entre parênteses, calculados de acordo com Belloni et al (2014). Coeficientes com asteriscos indicam que as estimativas foram significativas, considerando nível de significância de 5%. NaN é uma abreviação para *Not a Number*. A inferência no lasso requer técnicas especiais e em alguns casos, este resultado pode aparecer. Para maiores detalhe, ver Belloni *et al* (2014).

Tal resultado pode ser explicado pelo fato da incipiente oferta de trabalhadores escolarizados nos municípios pertencentes à região não metropolitana cearense. Além disso, vale destacar a expressiva redução do retorno à escolaridade na região metropolitana de Fortaleza, onde tal fato pode ser explicado pelo expressivo crescimento das instituições de ensino em tais municípios.

Ao se contrastar os resultados encontrados neste artigo com os de outros autores, verifica-se que a magnitude dos retornos à educação se mostraram bem inferiores quando comparados com os obtidos por Mariano e Arraes (2011) e Suliano e Siqueira (2012), os quais também realizaram estimações da equação minceriana de salários para medir o retorno à educação no Estado do Ceará.

Este resultado pode decorrer da melhor escolha de variáveis instrumentais realizadas pela técnica do lasso e do post-lasso. De fato, Teixeira e Menezes-Filho (2012), propuseram nova abordagem para a estimação de retorno da educação utilizando a legislação brasileira como instrumento. Os resultados encontrados por esses autores caminham no mesmo sentido dos evidenciados aqui, onde a escolha apropriada de instrumentos reduz significativamente a magnitude dos retornos da educação.

Diante do exposto, verifica-se que os resultados corroboram com a literatura no sentido de que se verifica uma redução no prêmio à escolaridade ao longo dos últimos anos. Além disso, este artigo apresenta um benefício considerável para a literatura que se dedica ao estudo dos retornos da educação, uma vez que apresenta uma metodologia inovadora que busca corrigir o problema da falta de instrumentos fortes para corrigir a endogeneidade entre educação e salários.

5. Conclusões

Este artigo buscou reexaminar os retornos da educação no Ceará utilizando técnicas robustas de correção da endogeneidade. Utilizou-se os métodos de seleção do Lasso e técnicas de estimação pós seleção (Post-Lasso) para lidar com tal problema.

Os métodos do Lasso e do Post-Lasso estabelecem critérios estatísticos para selecionar, dentre um conjunto de instrumentos fracos, aqueles que possuem o maior potencial preditivo da variável endógena. Dessa forma, evita-se, assim, a escolha de instrumentos *ad-hoc*, por meio, exclusivamente, da teoria econômica, a qual pode estar sujeita a enganos.

Os resultados foram estimados para dois anos diferentes, 2002 e 2012. Além disso, estratificou-se a amostra, em cada um dos anos, entre região metropolitana e não metropolitana. Esta separação deve-se somente a análise em ambientes diferenciados de mercado de trabalho e oferta educacional.

Além disso, o mesmo exercício foi realizado para a região Nordeste. Dessa forma, foi possível identificar se os retornos da educação no Ceará resultam de um contexto regional mais amplo.

Dos resultados encontrados, verificou-se o mesmo que já havia sido observado na literatura. Primeiro, que os retornos educacionais vêm se reduzindo em magnitude ao longo do tempo. Este resultado foi robusto a amostra completa, estratificada do Ceará e ao Nordeste.

Segundo, a região metropolitana apresenta menor retorno a educação do que a região não metropolitana. Como possível explicação para essa evidência destaca-se a carência, ou a falta de concorrência de pessoas qualificadas na região não metropolitana, como observado na análise descritiva dos dados.

Terceiro, em comparação com o Nordeste, embora em termos de magnitude os retornos da educação aparentemente estejam próximos, indicando a inserção no contexto regional, o Ceará reduziu muito mais fortemente os retornos da educação ao longo do tempo.

Por fim, e mais importante, a magnitude dos retornos da educação estimados neste trabalho foi significativamente menor do que os encontrados em outros trabalhos semelhantes, como, por exemplos, os de Mariano e Arraes (2011) e Suliano e Siqueira (2012). Este resultado pode indicar que a utilização de instrumentos utilizando apenas a teoria econômica não é recomendável sob a pena de selecionar instrumentos fracos que elevam o viés causado pela endogeneidade.

REFERÊNCIAS BIBLIOGRÁFICAS

ANGRIST, Joshua. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Record. **American Economic Review**, June 1990.

- ANGRIST, J. and PISCHKE, Jorn-Steffen. **Mostly Harmless Econometrics**. An empiricist's companion. Princeton University Press, New Jersey, 2009.
- ANGRIST, Joshua and KRUEGER, Alan. Does Compulsory School Attendance Affect Schooling and Earnings? **Quarterly Journal of Economics**. November, 1991.
- ARIAS, O; MCMAHON, W. Dynamic Rates of Return to Education in the U.S. **Economics of Education Review**. v.20, n.1, p. 121-138, 2001.
- ASHENFELTER, O.; KRUEGER, A. Estimates of the Economic Return to Schooling from a New Sample of Twins. **American Economic Review**. v. 84, n. 5, p. 1157-1173, 1994.
- BAI, Jushuan and NG, Serena. Selecting instrument variables in a data rich environment. **Journal of Time Series Econometrics**, 1, Nº 4, 2009
- BARROS, R.P. (1997). Os Determinantes da Desigualdade no Brasil, Seminário 22/97, IPE-USP.
- BARROS, R.P., HENRIQUES, R. e MENDONÇA, R.(2000), Education and Equitable Economic Development, **Economia**, Volume 1, No1.
- BARROS, R. P.; CARVALHO, M.; FRANCO, S.; MEDONÇA, R. Uma Análise das Principais Causas da Queda Recente na Desigualdade de Renda Brasileira. **Econômica**, Rio de Janeiro, v.8, n.1, p.117-147, jun. 2006.
- BARROS, R. P.; CARVALHO, M.; FRANCO, S.; MEDONÇA, R. A Queda Recente da Desigualdade de Renda no Brasil. Rio de Janeiro: IPEA, jan. 2007. (Texto para Discussão, 1258).
- BELLONI, A; HANSEN, C. and CHERNOZHUKOV, V. Inference Methods for High Dimensional Sparse Econometric Models. **Advances in Economics and Econometrics**, 10th World Congress of the Econometric Society, 2013.
- _____. High Dimensional Methods and Inference on Structural and Treatment Effects. **Journal of Economic Perspectives**, 28(2), pg. 29-50, 2014.
- _____. LASSO Methods for Gaussian Instrumental Variable Models. Working paper, **Duke University**, 2010.
- BELLONI, A; HANSEN, C.; CHEN, D.; and CHERNOZHUKOV, V. Sparse Models and Methods for Optimal Instruments with Application to Eminent Domain. **Econometrica**, 80(6), pg. 2369-2429, 2012.
- BEN-PORATH, Y. The Production of Human Capital and the Life Cycle of Earnings. **Journal of Political Economy**, v. 75, n.4, p. 352-365, 1967.
- BICKEL, P. J.; RITOV, Y. and TSYBAKOV, A. B. Simultaneous analysis of Lasso and Dantzig selector, **Annals of Statistics**, 37(4), pg. 1705-1732, 2009.
- BOUDARBAT. B; LEMIEUX, T.; RIDELL, W. C. **The Evolution of the returns to Human Capital in Canada, 1980-2006**. (Working Paper n.1) University of British Columbia, 2008.
- CARD, David. Earnings, Schooling and Ability Revisited. In Solomon Polachek, editor, **Research in Labor Economics**, vol. 14, 1995.
- CARNEIRO, P.; HECKMAN, J. and VYTLACIL, E. Estimating Marginal Returns to Education. **American Economic Review**, 101(6), 2011.
- CARVALHO, A. P.; NERI, M. C. ; SILVA, D. B. do N. . Diferenciais de Salários por Raça e Gênero: Aplicação dos procedimentos de Oaxaca e Heckman em Pesquisas Amostrais

Complexas. In: **XV Encontro Nacional de Estudos Populacionais**, 2006, Caxambu - MG. Anais do XV Encontro Nacional de Estudos Populacionais, 2006.

FERREIRA, E. A.; VELOSO, F. A. Intergeneration Mobility of Wages in Brazil.

Brazilian Review of Econometrics, v. 26, p. 181-211, 2006.

FERREIRA, F. H. G.; LEITE, P. G.; LITCHFIELD, J.; ULYSSEA, G. A. Ascensão e Queda da Desigualdade de Renda no Brasil. **Econômica**, Rio de Janeiro, v.8, n.1, p.147-169, jun., 2006.

HASTIE, Trevor, TIBSHIRANI, Robert and FRIEDMAN, Jerome. **The Elements of Statistical Learnings: Data Mining, Inference and Prediction**. Springer Series in Statistics, Second Edition, 2012.

HANSEN, C.; CHERNOZHUKOV, V. and SPLINDER, M. Post-Selection and Post-Regularization inference in linear models with very many controls and instruments.

American Economic Review: Papers and Proceedings, 105(5), pg. 486-490, 2015.

HECKMAN, J.J. Sample Selection Bias as a Specification Error. **Econometrica**, v. 47, n. 1, p. 153-161, Jan. 1979.

HECKMAN, James; URZUA, S., and VYTLACIL, E. Understanding Instrumental Variables in Models with Essential Heterogeneity. **Review of Economics and Statistics**, 88(3). Pg 389–432, 2006.

_____. Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case. **Les Annales d’Economie et de Statistique**, N° 91, pg 151-174, 2008.

_____. **Comparing IV with Structural Models**: What Simple IV Can and Cannot Identify. NBER Working Paper, n. 14.706, 33 p., 2009.

HOFFMANN, R. Transferências de Renda e a Redução da Desigualdade no Brasil e cinco Regiões entre 1997 e 2004. **Econômica**, Rio de Janeiro, v.8, n.1, p.55-81, jun., 2006.

IOSCHPE, Gustavo. A ignorância custa um mundo: o valor da educação no desenvolvimento do Brasil. São Paulo: Francis, 2004. 324p.

KAPETANIOS, George and MARCELLINO, M. Factor-GMM Estimation with Large Sets of Possibly Weak Instruments. **Computational Statistics and Data Analysis**, 54(11), pg. 2655–2675, 2010.

KAPETANIOS, George and GROEN, J. J. Parsimonious Estimation with Many Instruments. **Federal Reserve Bank of New York**. Staff Report n° 386, 2009.

LANGONI, C. G. **Distribuição da renda e desenvolvimento econômico do Brasil**. Rio de Janeiro: Expressão e Cultura, 1973.

LAM, D.; SCHOENI, R. Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil. **Journal of Political Economy**, v.101, n.4, p.710-739, aug., 1993.

MARIANO, F. Z. ; ARRAES, R. A . Endogeneidade da Educação na Previsão da Taxa de Retorno: Avaliação Metodológica e Aplicação para Regiões Brasileiras e Estados Selecionados. In: **VII Encontro da Economia do Ceará em Debate**, 2011, Fortaleza.

MENEZES-FILHO, N. A. A Evolução da Educação no Brasil e seu Impacto no Mercado de Trabalho. Instituto Futuro Brasil, 2001a.

MENEZES-FILHO, N. A. Equações de Rendimentos: Questões Metodológicas. In: CORSEUIL, C. H. (EDITOR). Estrutura Salarial: aspectos conceituais e novos resultados para o Brasil. Disponível em:

http://www.ipea.gov.br/agencia/images/stories/PDFs/livros/capitulo2_equacoes.pdf. Acesso em: 5 jun. 2002

MINCER, J. B. Schooling, experience and earnings. New York: NBER, 1974. 152 p.

PESSÔA, S. A.; BARBOSA FILHO, Fernando de Holanda. Retorno da Educação no Brasil. **Pesquisa e Planejamento Econômico** (Rio de Janeiro), v. 38, p. 97-125, 2008.

RESENDE, M. ; WYLLIE, R. . Retornos para Educação no Brasil: Evidências Empíricas Adicionais. **Economia Aplicada**, FIPE-USP, v. 10, p. 349-365, 2006.

ROUSE, C. Further Estimates of the Economic Return to Schooling from a New Sample of Twins. **Economics of Education Review**. v.18, n.2, p. 149-157,1999.

SACHSIDA, A., LOUREIRO, P.R.A., MENDONÇA, M.J.C. (2004), Um Estudo sobre Retornos em Escolaridade no Brasil, **Revista Brasileira de Economia**, 58, 249-265

SAPELLI, C. **Los Retornos a la Educación en Chile**: Estimaciones por Corte Transversal y por Cohortes. Documento de Trabajo, n. 349. ISSN: 0717-7593. 2009

SCHULTZ, T. W. Investment in Human Capital. **The American Economic Review**, vol 51, nº 1, Mar 1961, p. 1-17.

SOARES, S. S. D. Análise do Bem-Estar e Decomposição por Fatores da Queda na Desigualdade entre 1995 e 2004. **Econômica**, Rio de Janeiro, v.8, n.1, p.83-115, jun., 2006b.

SULIANO, D. C., SIQUEIRA, M. L. Retornos da educação no Brasil em âmbito regional considerando um ambiente de menor desigualdade. **Economia Aplicada**, vol.16, nº 1, Ribeirão Preto Jan./Mar. 2012.

TEIXEIRA, W. M., MENEZES-FILHO, N. A. Estimando o Retorno à Educação do Brasil Considerando a Legislação Educacional Brasileira como um Instrumento. **Revista de Economia Política**, vol. 32, nº 3 (128), pp. 479-496, julho-setembro/2012.

TIBSHIRANI, R. Regression shrinkage and selection via lasso. **Journal of Royal Statistics Society B**, v. 58 (1), pg. 267-288, 1996.

ZEPEDA et al. **Changes in Earnings in Brazil, Chile And Mexico**: Disentangling The Forces Behind Pro-Poor Change In Labour Markets. (Working Paper, n.51). International Policy Centre for Inclusive Growth (IPC - IG). Mar. 2009

ZOU, H. and HASTIE, T. Regularization and variable selection via elastic net. **Journal of Royal Statistics Society B**, v. 67 (2), pg. 301-320, 2005.