Salário, educação e experiência: uma abordagem econométrica

Felipe Resende Oliveira[1]
Dieison Lenon Casagrande[2]
Guilherme Studart[3]
Inaldo Bezerra da Silva[4]
Paulo Henrique Monteiro Guimarães[5]

Resumo: Este trabalho busca analisar e identificar as relações entre educação e experiência na fixação dos salários hora dos trabalhadores belgas. O resultado da regressão por MQO ilustra uma forte relação e significante entre o salário, experiência e educação. Existe uma relação positiva entre os níveis de educação e o salário hora, ou seja, o adicional da escolaridade está associado com aumento percentual do salário. O coeficiente da experiência 2 indica uma correlação negativa, o que sugere o retorno decrescente de experiência de trabalho sobre o percentual de aumento de salário.

Palavras-chave: Salário-hora. Modelo de regressão. HC3.

Classificação JEL: I00, C01.

1 Felipe Resende Oliveira felipexresende@gmail.com

2 Dieison Lenon Casagrande dieisonlenon@yahoo.com.br

3 Guilherme Studart guistudart@gmail.com

4 Inaldo Bezerra da Silva inaldobs@gmail.com

5 Paulo Henrique Monteiro Guimarães gmpaulohenrique@hotmail.com

Abstract: This work seeks to analyze and identify the relationship between education and experience in wage setting of Belgian workers. The results of the OLS regression shows a significant and strong relationship between earnings, experience and education. There is a positive relationship between levels of education and the hourly wage, so the additional schooling is associated with percentage salary increase. The experience^2 coefficient indicates a negative correlation, suggesting the diminishing returns of work experience on the salary percentage increase.

Keywords: Hourly wage. Regression model. HC3.

JEL Classification: I00, C01.

1 Introdução

Um dos principais insumos usados na produção dos mais diversos bens da economia é o trabalho, objeto de estudo de várias ciências e vem sendo investigado sob os mais diversos instrumentos teóricos há décadas. Pelo ponto de vista da ciência econômica, nesse caso especifico, viemos através deste trabalho fazer um esforço no sentido de compreender quais fatores estariam determinando o salário de um trabalhador.

No presente estudo, o salário bruto por hora é determinado pela educação e experiência de trabalhadores belgas, num total de 1472 observações para o ano de 1994. Buscamos uma adaptação do modelo de determinação salarial de Mincer para, através de análise de regressão múltipla, identificar a magnitude e relevância de cada variável independente, lembrando que é de extrema importância a correta modelagem do nosso problema bem como a utilização de recursos e testes econométricos para correção de eventuais problemas como: multicolinearidade, heteroscedasticidade ou até mesmo erro na especificação do modelo, só assim tivemos subsídio para escolher um modelo que se aproxima mais fielmente do que realmente ocorre na determinação desses salários.

Esse artigo está divido em sete partes. Logo após a introdução, fazemos uma revisão da literatura do trabalho, uma análise resumida dos principais artigos na área e que nos motivam para fazer esse estudo. Já a terceira parte vem apresentar todo o esforço no sentido de modelar o nosso problema, quais hipóteses foram utilizadas e quais premissas seguidas. Na parte quatro e cinco são feitas a análise descritiva dos dados e a escolha do melhor modelo, respectivamente. Na sexta parte são apresentados os resultados e, logo em seguida, na sétima e última parte expõe-se as conclusões.

2 Revisão de literatura

Este artigo tem como principal interesse explicar como o nível de educação e os anos de experiência de trabalho se relacionam com o salário do indivíduo, esse tema, é de suma importância para qualquer sociedade. O capital humano contribui para determinação dos salários e do próprio crescimento econômico de um país, esse capital humano sendo representado pelo nível de educação do indivíduo. (MINCER, 1984)

Já em 1974, Mincer corroborou com Becker (1964), afirmando que um eventual crescimento do salário de um trabalhador acontece com um aumento de suas habilidades, capital humano, educação ou com mais experiência no trabalho. A literatura também sugere um certo *trade-off* entre os determinantes do salário, e é bastante plausível pensar que em certas circunstâncias é melhor priorizar educação e em outros casos, experiência. Ao longo de sua pesquisa, Mincer, uns dos pais da economia moderna do trabalho, desenvolveu o que viria a ficar conhecida como "equação minceriana" que, justamente, buscava determinar o salário em função dos anos de estudo e da experiência do trabalho do indivíduo.

O uso da equação de Mincer foi bastante difundido e serve de arcabouço teórico para muitos artigos nos mais diversos países. Katz apontou que a diferença salarial entre os trabalhadores graduados em universidades e empregados sem curso superior cresceu bastante na década de 1980, o mesmo ocorreu entre aqueles que possuíam experiência e os que não tinham. Recentemente, Hotchkiss e Shiferaw (2010) também mostraram que o grande aumento do número de trabalhadores com curso superior contribuiu para intensificar a desigualdade salarial entre os anos de 1980 e 1990.

Esse modelo de determinação salarial desenvolvido por Mincer, apesar de seminal, continua sendo intensamente usado nas pesquisas que buscam entender, do ponto de vista do indivíduo, o retorno da educação e dos anos de experiência no trabalho.

3 O método dos mínimos quadrados ordinários e suas premissas

Nesta seção iremos expor o método de mínimos quadrados ordinários. Este método foi proposto pelo grande matemático/estatístico Carl Friedrich Gauss. Devido sua simplicidade matemática e por ser bastante intuitivo, esse método é um dos mais utilizados para se fazer análise de regressão linear.

O objetivo do modelo de regressão linear é predizer o comportamento da variável y tendo como base o que acontece com a variável x, tendo em vista que y representa a variável dependente e x a independente. Sendo assim, queremos observar como as mudanças nas variáveis independentes estão relacionadas com a variável dependente. O modelo de regressão linear múltipla é dado pela equação (1) abaixo:

$$y_t = \beta_1 + \beta_2 xt2 + \ldots + \beta kxtk + \epsilon t \tag{1}$$

onde β_k com $k = 1,2, \ldots, K$ são parâmetros e $x_{t1}, \ldots, x_{tk}$ são variáveis independentes com $t = 1,2\ldots, T$. O modelo de regressão linear também pode ser escrito da forma matricial, conforme pode ser visto de forma mais compacta na equação (2):

$$y = X\beta + \epsilon \tag{2}$$

e em forma de matriz como

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2t} & X_{3t} & \cdots & X_{kt} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_t \end{bmatrix}$$

em que:

y = vetor coluna Tx1 de observações da variável dependente Y ;

X = matriz *Txk* dando T das *K-1* variáveis de X_2 a X_k, a primeira coluna é de 1 representando o intercepto. Essa matriz é mais conhecida como a matriz de dados;

β = vetor coluna *Kx1* de parâmetros desconhecidos;

ϵ = vetor coluna *Tx1* dos termos de erro.

O objetivo agora é estimar o vetor β, assim poderemos fazer inferências sobre a população a partir de um conjunto de dados. Para isto será utilizado o método de mínimos quadrados ordinários que minimiza a soma dos quadrados dos erros, ou seja, a soma do quadrado da diferença entre o valor estimado e os dados observados.

Para seguir com a estimação, faz-se necessário assumir algumas suposições, a saber:

S_0: O modelo proposto está correto;

S_1: Os erros possuem médias iguais à zero, $E(\epsilon_t) = 0, \forall t$;

S_2: Homoscedasticidade, $Var(\epsilon_t) = \sigma^2 \ (0 < \sigma^2 < \infty) \ \forall t$;

S_3: Os erros são não autocorrelacionados, $Cov(\epsilon_t, \epsilon_t) = 0 \quad \forall t \neq s$;

S_4: A matriz de dados possui posto coluna completo, ou seja, posto (X) = K;

S5: Os erros possuem distribuição Normal, $\epsilon_t \sim N(0, \sigma^2) \quad \forall t$.

Assumindo as suposições expostas, o estimador pelo método dos Mínimos Quadrados Ordinários (MQO) é dado por:

$$b = (X'X)^{-1}X'y \qquad (3)$$

onde b é um vetor de estimadores de β.

O estimador b detém algumas características interessantes:

C_1: b é um estimador linear, logo, pode ser escrito como um múltiplo de y;

C_2: Se y tiver distribuição normal, b também terá, pois, b é um múltiplo de y;

C_3: b é não-viesado para β sob S0 e S1. Isso significa que em média b será igual a β; C_4: $Cov(b) = \sigma^2 (X'X)^{-1}$. A matriz de variância e covariância é utilizada para fazer testes de hipótese e estimação intervalar;

C_5: b é consistente. Este estimador é assintoticamente não viesado e sua variância tende a zero quando $T \rightarrow \infty$;

C6: Sob S0, S1, S2 e S3, o Teorema de Gauss-Markov garante que b é o melhor estimador linear e não-viesado de β.

É preciso de uma medida da qualidade global do ajuste do modelo estimado. Pode-se tomar os desvios de y, variável dependente, em relação a sua média. O coeficiente de determinação ajustado ($\overline{R}^2$). O $\overline{R}^2$ é dado pela equação a seguir:

$$\bar{R}^2 = 1 - \frac{SSE/(T-K)}{SST/(T-1)} \qquad (4)$$

em que SSE é soma dos quadrados dos resíduos e SST é a soma dos quadrados totais.

3.1 Violação das suposições

Abordaremos agora alguns testes para detectar possíveis violações das suposições assumidas para a estimação dos parâmetros via MQO. Não apresentaremos teste para checar problemas de autocorrelação, uma vez que nossos dados não são de séries temporais.

Sob (S0) assumimos que o modelo está correto. Todavia, os erros de especificação dos modelos são relativamente frequentes. Tais erros são oriundos da ausência de variáveis importantes no modelo, erros de medida em variáveis e erros de simultaneidade.

Em 1969, Ramsey propôs um teste para checar se o modelo está especificado corretamente. O teste Reset vai constatar se, de fato, o modelo é linear. Quando não sabemos a forma de não linearidade, a solução proposta por Ramsey é a seguinte:

$$y=X\beta+Z\gamma+\epsilon \qquad (5)$$

onde y é o vetor de variável dependente, X é a matriz dados, Z é a matriz de regressores artificiais, β e γ são vetores de parâmetros. Na matriz de regressores artificiais estão inclusos regressores não lineares.

Hipóteses de teste: $\begin{cases} H_0: \gamma = 0 \text{ especificação correta} \\ H_1: \gamma \neq 0 \text{ especificação incorreta} \end{cases}$

O teste consiste em verificarmos se γ é um vetor nulo. Caso este fato se confirme o modelo está especificado da forma certa. Há outros testes que visam checar se o modelo está corretamente especificado, como por exemplo, o teste do arco-íris proposto por Utts em 1982. O objetivo do teste é comprovar linearidade. Há ainda o teste de *outliers*, observações que destoam de forma acentuada da amostra, que podem causar problemas na especificação do modelo. Contudo, o teste mais geral é o de Ramsey.

Algo que acontece com uma certa frequência nos trabalhos empíricos é a violação de (S2). A heteroscedasticidade dos erros afetam as inferências sobre o modelo, sendo assim, os estimadores de MQO deixam de ser eficiente, logo, o Teorema de Gauss-Markov não pode ser atendido.

Breusch-Pagan, em 1979, propuseram uma maneira para testar a presença de heteroscedasticidade. A ideia é modelar a variância a partir de uma função cedástica que seja duas vezes continuamente diferenciável e não muda de observação para observação. A função (6) representa uma função cedástica:

$$\sigma_t^2 = h(\alpha_1 + \alpha_2 Z_{t2} + \ldots + \alpha_s Z_{ts}) \qquad (6)$$

Iremos testar a seguinte hipótese nula:

Hipóteses de teste: $\begin{cases} H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_s & \text{(homoscedasticidade)} \\ H_1: \text{caso contrário} & \text{(heterocedasticidade)} \end{cases}$

Com os resultados de uma regressão auxiliar, é possível mostrar que a estatística de teste será dada por:

$$LM_{BP} = \frac{SSR_A}{2} \qquad (7)$$

Sob H_0, a estatística de teste apresenta distribuição assintótica χ^2_{s-1}, logo, rejeita-se H_0 quando $LM_{BP} > \chi^2_{1-\alpha,\ s-1}$. Não é preciso assumir normalidade (S_5), todavia, quando o número de observações é pequeno o teste não funciona muito bem. Isso eleva a probabilidade do erro tipo II. Koenker (1981) propôs um teste que é mais poderoso que o teste Breusch-Pagan quando não há normalidade. Sem assumir normalidade o teste de Koenker apresentará uma probabilidade do erro tipo II bem menor que o revelado pelo teste de Breusch-Pagan.

A violação de (S_2) torna o estimador de MQO ineficiente. É possível demonstrar, neste caso, que a matriz de covariância do vetor de estimadores toma a seguinte forma:

$$cov(b) = (X'X)^{-1} X'\phi X (X'X)^{-1} \qquad (8)$$

onde $\phi = E(e'\,e)$. Com heterocedasticidade $\phi = \sigma^2\,\Omega$, em que Ω é uma matriz positiva definida e, pela decomposição de Cholesky pode ser expressa como $\Omega = P'P$.

Tomando o modelo original e multiplicando-o pela matriz P, pode-se mostrar que o estimador de MQO será dado por:

$$b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y \qquad (9)$$

Este estimador é conhecido como o estimador de Mínimos quadrados generalizados e corrige o problema da heteroscedasticidade. Mas a matriz Ω^{-1} depende das variâncias de σ_t^2, que são desconhecidas. Estima-se a equação (8), que é a matriz de covariâncias de b. O problema consiste na forma como a matriz será estimada. White (1980) mostrou que para que a *cov(b)* seja estimada de forma consistente, basta que $X'\phi X$ seja estimada de forma consistente. White propôs o seguinte estimador:

$$\hat{\phi} = diag\ (\hat{u}_1^2, \dots, \hat{u}_T^2) \qquad (10)$$

Esse estimador ficou conhecido como HC0.

MacKinnon e White (1985) corrigiram o estimado HC0 pelos seus graus de liberdade, num estimador que ficou sendo chamado de HC1. Uma outra sugestão dos autores foi a de corrigir cada resíduo por um termo que leve em conta a alavancagem daquele ponto. Essa sugestão é chamada de HC2. Outro estimador é o HC3, que aumenta o peso da medida de alavancagem em cada termo. Cribari-Neto (2004) propôs o HC4 que atribui um novo peso para a medida de alavancagem.

De modo geral, todas essas formas consistentes de estimar a matriz tem como objetivo corrigir o problema da heteroscedasticidade. Contudo, salientamos que, ao utilizarmos esses métodos, a forma dos testes t de significância dos regressores têm que sofrer algumas alterações. O erro padrão dos estimadores passam a ser representados em termos de , de modo que não se pode mais utilizar os valores críticos de uma distribuição t de Student. No caso do estimador de White, é possível mostrar que a estatística t converge em distribuição para uma normal padrão, permitindo usar seus valores críticos assintóticos. Tais testes de hipóteses são chamados de testes quase-t.

Checaremos, agora, a validade de (S_5). O teste mais utilizado foi proposto por Bera e Jarque (1987). Já que os erros não são observáveis, usa-se o resíduo a fim de realizar o teste. O teste checa as seguintes hipóteses:

Hipóteses de teste: $\begin{cases} H_0: \text{Normalidade} \\ H_1: N\tilde{a}o - \text{normalidade na família de } pearson \end{cases}$

A estatística de teste é dada por:

$$BJ = (\frac{assimetria^2}{6} + \frac{excessodecurtose^2}{24}) \qquad (11)$$

Entretanto este teste apresenta o seguinte problema: a distribuição nula do teste é desconhecida. Dessa forma o valor crítico dessa distribuição é desconhecido. A solução é fazer um teste aproximado em que, sob H_0, BJ converge em distribuição para χ^2_2. Dessa forma, rejeita-se H0 quando $BJ > \chi^2_{1-\alpha,2}$, em que α é a probabilidade de erro tipo I. Todavia, essa distribuição assintótica não funciona muito bem quando o tamanho da amostra é pequeno. Quando isso acontece, uma sugestão é usar a técnica do Bootstrap para estimar a distribuição nula de BJ e, assim, encontrar o valor crítico.

Se o teste indicar que não há normalidade nos resíduos, uma alternativa para corrigir o problema é usar a transformação proposta por Box e Cox (1964). A vantagem desse método é que essa transformação reduz os desvios de (S_2) e (S_5).

4 Análise descritiva dos dados

Essa parte do trabalho dedica-se a descrever a massa de dados. Aqui faz-se uma explanação do salário por hora, medido em euros, de 1472 trabalhadores belgas no ano de 1994.

Tabela 1: Descrição das variáveis

	Wage	Educ	Exper
Mínimo	2.191	1.000	0.00
Mediana	10.127	3.000	16.50
Média	11.051	3.378	17.22
Máximo	47.576	5.000	47.00
Desvio-Padrão	4.450513	1.204522	10.16667

Fonte: Elaboração própria.

O salário médio por hora, variável dependente, é de cerca de 11 euros, variando de pouco mais de dois até quase 48 unidades da moeda europeia.

É importante chamar atenção para o fato de que uma das variáveis independentes, a variável Educ, que mede o nível de educação do trabalhador, é categórica, ou seja, vai de um a cinco proporcionalmente, posteriormente levaremos a hipótese de que ela possa ser posta em forma de dummy, pois é plausível pensar que os aumentos com a educação não sejam perfeitamente proporcionais.

Por fim, mas não menos importante, temos as características descritivas da variável independente Exper que nos informa os anos de experiência do trabalhador, observe que no extremo inferior temos o caso do trabalhador completamente inexperiente e do outro lado um indivíduo com 47 anos de experiência, a média fica em torno de 17 anos.

Analisa-se neste momento a correlação entre as variáveis. Para isto utiliza-se a matriz de correlação, conforme a Tabela 2. Com base nesses valores é pouco provável que exista multicolinearidade entre os dados da amostra.

Tabela 2: Matriz de correlação

	Wage	Educ	Exper
Wage	1.00	0.39	0.31
Educ	0.39	1.00	-0.29
Exper	0.31	-0.29	1.00

Fonte: Elaboração própria

5 Escolha do modelo

A primeira etapa numa análise de regressão é encontrar a forma funcional mais adequada para explicar o processo gerador de dados. Tal processo de busca pelo melhor ajuste torna-se melhor fundamentado quando há uma base teórica para o problema em questão. Com isso, analisaremos as formas funcionais contidas na Tabela 3.

Tabela 3: Modelos teóricos

	Modelos
1	$Wage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Educa_t$
2	$Wage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Educ_t$
3	$logWage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Educ_t$
4	$logWage_t = \beta_1 + \beta_{t2}Expp_t + \beta_{t3}Educ_t$
5	$logWage_t = \beta_1 + \beta_{t2}Expp_t + \beta_{t3}Educa_t$
6	$logWage_t = \beta_1 + \beta_{t2}Exper2_t + \beta_{t3}Educa_t$
7	$logWage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Educa_t$
8	$logWage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Exper2_t + \beta_{t4}Educa_t$
9	$logWage_t = \beta_1 + \beta_{t2}Exper_t + \beta_{t3}Exper2_t + \beta_{t4}Educa_t + \beta_{t5}Wage * Exper$

Fonte: Elaboração própria.

De acordo com a Tabela 4, foi realizado testes Bera-Jarque em cada modelo e detectamos a ausência da normalidade dos erros. Através dos testes RESET, vimos que os modelos 3, 4, 5, 6, 7 estão bem especificados. Dos modelos que passaram no teste, apenas o modelo 6 apresentou evidência de problema de multicolinearidade quase exata, uma vez que seu número de condição foi superior a 30.

Uma das possibilidades para não aceitação dos modelos e a rejeição da normalidade é a existência de observações influentes. Sendo assim, adotamos a estratégia de identificar observações atípicas e retirá-las da base de dados, já que tais observações podem alterar o processo gerador de dados e criar problemas na estimação dos nossos modelos de regressão. Para tanto, utiliza-se como critério grau de alavancagem, Dfbeta, Dffit e Distância de Cook. Em seguida, retiramos as observações atípicas de cada modelo e depois procedemos com os mesmos testes.

A próxima tabela mostra a quantidade de pontos influentes retirados, valor do $\overline{R}^2$, teste RESET, o valor da estatística Bera-Jarque e o número de condição.

Tabela 4: Modelos

	$\overline{R}^2$	RESET	Bera-Jarque	Nº de condição
Modelo 1	0,3454	15,6067	4894,364	18,35943
Modelo 2	0,3436	18,0878	5143,985	9,516519
Modelo 3	0,3506	1,577	659,6538	9,516519
Modelo 4	0,3506	1,577	659,6538	9,738397
Modelo 5	0,3494	1,3864	664,1139	14,74904
Modelo 6	0,2948	2,7377	583,2558	79,34022
Modelo 7	0,3494	1,3864	664,1139	14,49692
Modelo 8	0,3741	3,4153	466,9689	79,26727
Modelo 9	0,9237	7095,776	69570,45	97,07162

Fonte: Elaboração própria.

Após a retirada das observações influentes, de acordo com a tabela acima, os modelos 3, 4, 5, 6, 7, 8 passaram no teste RESET com nível de 1% de significância, ou seja, apresentaram boa especificação. Nesta etapa da análise, mesmo que o modelo 8 não tenha passado no Teste Número de condição, iremos selecionar este modelo por três motivos. i) Sua forma funcional é simples, ii) Alto iii) Os erros do modelo (8) seguem distribuição normal na família de Pearson com 5% de significância. O modelo escolhido consegue captar o efeito da idade através da variável experiência 2 (EXPER2), pois ela absorve o efeito do indivíduo ficar velho.

Tabela 5: Modelos de observações influentes

	Observações influentes	$\overline{R}^2$	RESET	Bera-Jarque	Nº de condição
Modelo 1	103	0,3588	16,699	5168,868	31,94015
Modelo 2	58	0,3752	21,6574	7292,086	9,491484
Modelo 3	77	0,4146	3,7015	4,0976	9,462502
Modelo 4	77	0,4146	3,7015	4,0976	9,690942
Modelo 5	121	0,4276	0,9655	5,1702	21,00495
Modelo 6	134	0,3679	0,1324	5,9816	119,5205
Modelo 7	121	0,4276	0,9655	5,1702	26,90958
Modelo 8	134	0,4524	3,6809	3,1901	101,3718
Modelo 9	80	0,9697	21036,47	1513,887	83,73932

Fonte: Elaboração própria.

Sabemos que experiência (EXPER) e experiência 2 (EXPER2) são correlacionadas (0.96), porém, essa correlação não é mais linear. É crível que o viés da variável omitida idade pode causar um problema de estimação, uma vez que é um regressor bastante importante para a estimação, e essa variável está sendo captada pela experiência 2 (EXPER2). A ideia é que a partir do momento que o indivíduo fica mais experiente, ele também fica mais velho, dessa forma, é natural esperar que em média os indivíduos com muita experiência tenham uma idade mais elevada, o que pode fazer com que seu salário comece a se reduzir, pois uma idade avançada pode influenciar negativamente no salário hora.

O modelo (9) apresentou um $\overline{R}^2$ ajustado bem elevado, porém a sua especificação não foi considerada boa com 5% de confiança. A variável wage*experiência não se mostra um bom regressor. Sendo assim, o $\overline{R}^2$ do modelo (8) foi o mais elevado, e este é um critério bastante utilizado na literatura para seleção de dois ou mais modelos. Portanto, a análise será feita com base no modelo (8), que possui 1.338 observações.

Deve-se verificar a violação da hipótese (S2) de heteroscedasticidade. Para isso, foi utilizado dois testes: teste de Breuch-Pagan e teste de Koenker. Em ambos os testes, detectou-se a presença de heteroscedasticidade, a estatística do teste foi LMBP=35.7252, rejeitando a hipótese nula, de que se tem homoscedasticidade. Portanto, o modelo aceita as hipóteses S_0, S_1, S_3, S_4 e S_5, porém tem a violação de S2.

Na presença de heteroscedasticidade os estimadores de mínimos quadrados ordinários continuam não viesados, lineares, consistentes e assintoticamente normais, porém não são eficientes. Existem diversas maneiras de estimação da matriz de variância de forma consistente, foi utilizado a matriz de covariância robusta HC3 para obter a estimação do modelo final.

6 Resultados

Tabela 6: Resultados da estimação por MQO com HC3

Variável	Estimativa	Desvio-Padrão	Valor t	Pr(> |t|)
(Intercept)	1.5989e+00	6.0753e-02	26.3181	< 2.2e-16 ***
educa2	1.1912e-01	5.9293e-02	2.0091	0.04473 *
educa3	2.6327e-01	5.8672e-02	4.4871	7.842e-06 ***
educa4	4.1871e-01	5.8791e-02	7.1220	1.736e-12 ***
educa5	6.0573e-01	5.9638e-02	10.1568	< 2.2e-16 ***
exper	3.4922e-02	2.6099e-03	13.3803	< 2.2e-16 ***
exper2	-4.8336e-04	7.5259e-05	-6.4227	1.859e-10 ***

Fonte: Elaboração própria.

De acordo com a tabela, todas as variáveis explicativas foram estatisticamente significantes e maiores do que zero a 95% de confiança. Note que o coeficiente da experiência 2 é negativo e a experiência é positivo. O coeficiente da experiência 2 revela que o retorno da experiência com o número de anos vai diminuindo com o passar do tempo.

Nosso modelo é do tipo log-lin. Os coeficientes das dummies de educação foram positivos. Um nível maior de escolaridade faz com que o trabalhador em média tenha maior aumento percentual em termos de salários. Note que, quanto maior for a escolaridade do trabalhador, mantendo a experiência constante, maior será, em termos percentuais, o salário hora do indivíduo.

Em relação à experiência, temos que um ano de experiência aumenta cerca de 3,9% o salário hora. Para um adicional de um ano de experiência, o efeito do declínio de salário é de cerca de 0,042%. O modelo escolhido indica que 45,48% das variações do salário hora podem ser explicadas pelos regressores utilizados.

7 Conclusões

Este estudo fornece evidência empírica, com base em MQO *cross--section*, de que os regressores utilizados têm alto poder de explicação sobre a variável dependente analisada. O modelo escolhido possui a mesma forma funcional do modelo encontrado por Wannakrairoj (2012). Este modelo baseia-se nas equações mincerianas com grande volume de dados cross-section, Mincer (1974).

Os resultados indicam claramente uma forte relação significante entre o salário, experiência e educação. Existe uma relação positiva entre os níveis de educação e o salário hora, ou seja, o adicional da escolaridade está associado com aumento percentual do salário. O coeficiente da experiência 2 indica uma correlação negativa, o que sugere o retorno decrescente de experiência de trabalho sobre o percentual de aumento de salário.

Referências

BECKER, G. S. *Human capital*: a theoretical and empirical analysis, with special reference to education. Chicago, Ill: University of Chicago Press, 2009.

CRIBARI-NETO, F. Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis,* Amsterdam, v. 45, n. 2, p. 215-233, mar. 2004.

KOENKER, R. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, Amsterdam, v. 17, n. 1, p. 107-112, Sept. 1981.

MINCER, J. Schooling and earnings. In: MINCER, J. *Schooling, experience, and earnings*. New York: Columbia University, 1974. p. 41-63.

MINCER, J. Human capital and economic growth. *Economics of Education Review*, Oxford, UK, v. 3, n. 3, p. 195-205, 1984.

WANNAKRAIROJ, W. The effect of education and experience on wages: the case study of Thailand in 2012. *Southeast Asian Journal of Economics*, Bangkok, v. 1, n. 1, p. 27-48, Dec. 2013.

WHITE, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, Menasha, Wis., v. 48, n. 4, p. 817-838, May 1980.