



UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Matemática, Estatística e
Computação Científica – IMECC



TRABALHO FINAL

Análise dos dados referentes ao artigo “How computers have changed the wage structure: evidence from microdata, 1984 - 1989”

NOME: Eduardo Vargas Ferreira

Professor: Dr. Maurício Enrique Zevallos Herencia

Campinas, 27 de junho de 2012

SUMÁRIO

1.	REPRODUÇÃO DOS RESULTADOS DO ARTIGO	3
2.	QUESTIONAMENTOS	7
2.1	1º Questionamento	7
2.2	2º Questionamento	8
2.3	3º Questionamento	15
2.4	4º Questionamento	16
3.	PROPOSTA DE MODELO E TESTES DIAGNÓSTICO	18
3.1	Proposta de modelo baseado no banco de dados do ano de 1984	18
3.2	Proposta de modelo baseado no banco de dados do ano de 1989	20
3.3	Conclusão relacionada aos modelos propostos	26

Esse trabalho foi baseado no artigo “How computers have changed the wage structure” de Alan Krueger, publicado em *Quarterly Journal of Economics* (1993) e tem dois objetivos principais: (a) reprodução dos resultados do artigo e (b) propor análises alternativas fazendo o diagnóstico do modelo.

Para que os resultados sejam apresentados de maneira clara e organizada, decidimos dividir o trabalho em três partes. Primeiramente, apresentamos os resultados da reprodução do trabalho. Na sequência, expomos os resultados e conclusões referentes a quatro questionamentos e, por fim, apresentamos todos os procedimentos realizados para seleção de variáveis, escolha do modelo mais adequado e análise de diagnóstico.

1. REPRODUÇÃO DOS RESULTADOS DO ARTIGO

Para a reprodução dos resultados do artigo, foram utilizados dois bancos de dados, os quais contêm informações do “*Current Population Survey*”. Os bancos de dados disponíveis são referentes ao ano de 1984 e 1989 e contêm 165.415 e 161.750 indivíduos, respectivamente.

A fim de reproduzir os resultados da Tabela 1 (“*Percent of workers in various categories Who directly use a computer at work*”), excluimos os indivíduos, os quais não haviam respondido ao questionamento quanto ao uso de computador e com nível educacional negativo (99.848 e 95.298 indivíduos dos anos de 1984 e 1989, respectivamente), bem como selecionamos para a análise somente os indivíduos entre 18 e 65 anos de idade, visto que o autor também o realiza devido ao fato dessa faixa etária representar os indivíduos que estão trabalhando ou têm um trabalho. Conseqüentemente, para a reprodução da tabela foram utilizados para o ano de 1984 61.864 indivíduos e para o ano de 1989 62.979 indivíduos.

Os totais de indivíduos utilizados para a reprodução da tabela não condizem com a quantidade total descrita pelo autor, o qual baseou seu trabalho em 61.712 e 62.748 indivíduos para os anos de 1984 e 1989, respectivamente. Sendo assim, os resultados apresentados não condizem exatamente aos valores obtidos pelo autor, porém próximos (veja Tabela 1).

Tabela 1: Resultados obtidos na reprodução da tabela intitulada por Alan Krueger como “*Percent of workers in various categories Who directly use a computer at work*”.

Grupos	1984	1989
All workers	24,6	37,3
<u>Gender</u>		
Men	21,1	32,1
Women	28,9	43,2
<u>Education</u>		
Less than high school	5,0	7,7
High school	19,2	29,2
Some college	30,5	45,2
College	41,4	58,0
Postcollege	42,6	59,5
<u>Race</u>		
White	28,3	43,7
Black	19,3	27,6
<u>Age</u>		
Age 18 – 24	18,2	27,2
Age 25 – 39	29,1	41,3
Age 40 – 54	23,6	39,0
Age 55 – 65	16,6	25,8
<u>Union status</u>		
Union member	20,0	32,4
Nonunion	27,9	41,1
<u>Hours</u>		
Part-time	14,2	24,3
<u>Region</u>		
Northeast	25,4	37,8
Midwest	23,2	35,8
South	23,1	36,4
West	26,9	39,7

Vale lembrar que a variável *Education* disponível em nosso banco de dados apresenta-se no formato “anos de estudos”, sendo que os resultados apresentados pelo autor estão agrupados em cinco categorias. Para categorizar da mesma forma, consideramos até 11 anos de estudo como *Less than high school*, 12 anos como *High school*, 13 a 15 anos como *Some college*, 16 anos como *College* e 17 e 18 anos *Postcollege*, segundo a classificação Norte Americana.

Além disso, não estava disponível diretamente o resultado da variável *White*, a qual é uma composição entre a variável *Black* e *Other races*. Logo, a partir dessas variáveis, calculamos a percentagem dos indivíduos categorizados como zero em *Black* e zero em *Other races*.

Na sequência, apresentamos os resultados obtidos para a reprodução da Tabela 2 (“*OLS Regression estimates of the effect of computer use on pay (dependent variable: $\ln(\text{hourly wage})$)*”). Para tanto, excluimos os indivíduos que apresentam *real usual hourly wage* menores do que \$1,50 por hora ou

maiores do que \$250 por hora como menciona o autor no apêndice A. Além disso, excluímos os indivíduos os quais não responderam a alguma das covariáveis a serem utilizadas para a geração dos modelos. Assim, foram utilizados 11.411 e 10.835 indivíduos para os anos de 1984 e 1989, respectivamente.

Novamente, os totais de indivíduos utilizados para a reprodução da tabela não condizem com a quantidade total descrita pelo autor (13.335 e 13.379 indivíduos para os anos de 1984 e 1989, respectivamente). Devido a essa discrepância, os resultados apresentados não condizem exatamente com os valores obtidos pelo autor, mas muito próximos (veja Tabela 2).

Tabela 2: Resultados obtidos na reprodução da tabela intitulada por Alan Krueger como “*OLS Regression estimates of the effect of computer use on pay (dependent variable: $\ln(\text{hourly wage})$)*”.

Independent variable	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1,944 (0,005)	0,711 (0,025)	0,714 (0,039)	1,939 (0,006)	0,671 (0,026)	0,689 (0,038)
Uses computer at work (1 = yes)	0,278 (0,011)	0,171 (0,009)	0,140 (0,009)	0,331 (0,010)	0,184 (0,008)	0,156 (0,009)
Years of education	--	0,071 (0,001)	0,049 (0,001)	--	0,076 (0,001)	0,056 (0,001)
Experience	--	0,026 (0,001)	0,025 (0,001)	--	0,026 (0,001)	0,024 (0,001)
Experience-squared/100	--	-0,041 (0,002)	-0,039 (0,002)	--	-0,041 (0,002)	-0,039 (0,002)
Black (1 = yes)	--	-0,098 (0,013)	-0,067 (0,012)	--	-0,128 (0,013)	-0,103 (0,013)
Other race (1 = yes)	--	-0,111 (0,022)	-0,091 (0,021)	--	-0,024 (0,021)	-0,009 (0,021)
Part-time (1 = yes)	--	-0,260 (0,011)	-0,215 (0,010)	--	-0,220 (0,011)	-0,181 (0,011)
Lives in SMSA (1 = yes)	--	0,100 (0,008)	0,096 (0,008)	--	0,170 (0,008)	0,162 (0,008)
Veteran (1 = yes)	--	0,038 (0,011)	0,042 (0,011)	--	0,030 (0,012)	0,035 (0,012)
Female (1 = yes)	--	-0,152 (0,013)	-0,137 (0,013)	--	-0,167 (0,013)	-0,154 (0,013)
Married (1 = yes)	--	0,161 (0,012)	0,138 (0,011)	--	0,161 (0,012)	0,148 (0,011)
Married*Female	--	-0,179 (0,016)	-0,169 (0,015)	--	-0,139 (0,016)	-0,134 (0,016)
Union member (1 = yes)	--	0,180 (0,010)	0,182 (0,009)	--	0,185 (0,011)	0,180 (0,010)
8 Occupation dummies	No	No	Yes	No	No	Yes
R ²	0,050	0,446	0,49	0,083	0,458	0,492

Por fim, os resultados obtidos para a Tabela 3 (“*The return to various uses of computers, october 1989 - dependent variable: ln(hourly wage)*”). Novamente, os resultados não condizem com os obtidos pelo autor, visto que os dados utilizados são os mesmos referenciados para a Tabela 2, isto é, apresentam discrepância entre os valores totais utilizados pelo autor.

Tabela 3: Resultados obtidos na reprodução da tabela intitulada por Alan Krueger como “*The return to various uses of computers, october 1989 (dependent variable: ln(hourly wage))*”.

User of computer at work	Proportion	Coefficient (std. error)
Uses computer at work for any task ^b	0,399	0,133 (0,011)
<u>Specific Task^c</u>		
Word processing	0,165	0,033 (0,013)
Bookkeeping	0,099	-0,049 (0,013)
Computer-assisted design	0,036	0,007 (0,023)
Electronic mail	0,061	0,127 (0,018)
Inventory control	0,103	-0,056 (0,014)
Programming	0,076	0,048 (0,017)
Desktop publishing or newsletters	0,035	-0,054 (0,023)
Spread sheets	0,093	0,086 (0,016)
Sales	0,062	0,012 (0,017)
Computer games	0,020	-0,089 (0,029)
R ²		0,493

Os resultados da Tabela 3 indicam que indivíduos que utilizam o computador para acessar o e-mail têm um aumento médio no salário de aproximadamente 12,7%, mantendo constantes as outras variáveis do modelo. Analogamente, indivíduos que fazem uso do computador para venda, fixando as demais variáveis, aumenta seu salário, em média, 1,2%. Vale destacar, que o sinal negativo nas estimativas de *Bookkeeping*, *Inventory control*, *Desktop publishing or newsletters* e *Computer games* indicam que a utilização do computador nessas atividades conduz a uma redução no salário médio. É digno de nota, que as variáveis *Computer-assisted design* e *Sales* não foram significativas ao nível de 1%.

2. QUESTIONAMENTOS

2.1 1º Questionamento: Na Tabela II Coluna (6), faça a regressão sem as dummies de região. Mudam bastante os resultados? Relacione os novos resultados com os da Tabela I.

Em relação à Tabela 2, modelo (6), realizamos uma nova regressão sem a variável explanatória “Region”. Os resultados obtidos são apresentados na Tabela 4.

Tabela 4: Estimativas e erros-padrões obtidos para o modelo (6) (proposto pelo autor) e modelo com a nova regressão (desconsiderando a variável explanatória Region).

Independent Variable	Modelo (6)	Modelo Novo
Intercept	0, 689 (0,038)	0,627 (0,037)
Uses computer at work (1 = yes)	0,156 (0,009)	0,157 (0,009)
Years of education	0,056 (0,001)	0,056 (0,001)
Experience	0,024 (0,001)	0,025 (0,001)
Experience-squared/100	-0,039 (0,002)	-0,040 (0,002)
Black (1 = yes)	-0,103 (0,013)	-0,125 (0,013)
Other race (1 = yes)	-0,009 (0,021)	0,017 (0,020)
Part-time (1 = yes)	-0,181 (0,011)	-0,176 (0,011)
Lives in SMSA (1 = yes)	0,162 (0,008)	0,174 (0,008)
Veteran (1 = yes)	0,035 (0,012)	0,033 (0,012)
Female (1 = yes)	-0,154 (0,013)	-0,158 (0,013)
Married (1 = yes)	0,148 (0,011)	0,142 (0,011)
Married*Female	-0,134 (0,016)	-0,132 (0,016)
Union member (1 = yes)	0,180 (0,010)	0,191 (0,010)
8 Occupation dummies	Yes	Yes
R ²	0,492	0,486

De acordo com a Tabela 4, verificamos uma mudança diminuta nas estimativas dos coeficientes. A maior alteração dos resultados ocorreu com a variável raça, decrescendo de -0,103 para -0,125 para variável *Black* e aumentando de -0,009 para 0,017 para *Other race*. Ressaltamos, todavia, que não há evidências suficientes para rejeitar a nulidade referente ao coeficiente de *Other race* (p-valor = 0.6525). É escuso supor, portanto, que ausência da variável região esteja acarretando em um viés por omissão de variável. O R^2 como esperado diminuiu, apesar de muito pouco. Além disso, de acordo com a Tabela 1, a porcentagem de indivíduos que utilizam o computador no trabalho não se modifica expressivamente entre as regiões estudadas, indo ao encontro das conclusões acima relatadas.

2.2 2º Questionamento: Há evidência de correlação serial e/ou heteroscedasticidade? Calcule os erros-padrão robustos das estimativas dos coeficientes de regressão (nas colunas 1 e 4 da Tabela II).

Para todos os modelos presentes na Tabela 2, foram realizados os testes para verificação de correlação serial e heteroscedasticidade, bem como os cálculos de erros-padrão robustos das estimativas dos coeficientes das respectivas regressões (vide Tabela 5).

Tabela 5: Resultados das estimativas de cada modelo proposto por Alan Krueger, juntamente com os erros-padrões robustos de White.

Independent Variable	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1,944 (0,005)	0,711 (0,026)	0,714 (0,040)	1,939 (0,006)	0,671 (0,027)	0,689 (0,038)
Uses computer at work (1 = yes)	0,278 (0,011)	0,171 (0,009)	0,140 (0,009)	0,331 (0,010)	0,184 (0,009)	0,156 (0,009)
Years of education	--	0,071 (0,001)	0,049 (0,001)	--	0,076 (0,001)	0,056 (0,002)
Experience	--	0,026 (0,001)	0,025 (0,001)	--	0,026 (0,001)	0,024 (0,001)
Experience-squared/100	--	-0,041 (0,002)	-0,039 (0,002)	--	-0,041 (0,002)	-0,039 (0,002)
Black (1 = yes)	--	-0,098 (0,012)	-0,067 (0,012)	--	-0,128 (0,013)	-0,103 (0,013)
Other race (1 = yes)	--	-0,111 (0,025)	-0,091 (0,023)	--	-0,024 (0,022)	-0,009 (0,021)
Part-time (1 = yes)	--	-0,260 (0,011)	-0,215 (0,011)	--	-0,220 (0,012)	-0,181 (0,012)
Lives in SMSA (1 = yes)	--	0,100	0,096	--	0,170	0,162

		(0,008)	(0,008)		(0,008)	(0,008)
Veteran (1 = yes)	--	0,038	0,042	--	0,030	0,035
		(0,012)	(0,012)		(0,013)	(0,013)
Female (1 = yes)	--	-0,152	-0,137	--	-0,167	-0,154
		(0,013)	(0,013)		(0,013)	(0,013)
Married (1 = yes)	--	0,161	0,138	--	0,161	0,148
		(0,012)	(0,012)		(0,012)	(0,012)
Married*Female	--	-0,179	-0,169	--	-0,139	-0,134
		(0,016)	(0,015)		(0,016)	(0,016)
Union member (1 = yes)	--	0,180	0,182	--	0,185	0,180
		(0,009)	(0,009)		(0,011)	(0,010)
8 Occupation dummies	No	No	Yes	No	No	Yes
R ²	0,050	0,446	0,49	0,083	0,458	0,492

De acordo com os resultados constantes na Tabela 5, concluímos que os erros-padrão robustos das estimativas dos coeficientes da regressão praticamente não se alteraram (a menos por diferenças de arredondamento) quando comparados com os resultados da Tabela 2. Especialmente, nas colunas (1) e (4) não houve mudanças.

Em relação à verificação da violação da suposição de homoscedasticidade, conduzimos nossa análise considerando os seis modelos descritos na Tabela 2. Primeiramente, realizamos a verificação de homoscedasticidade através de gráficos apropriados. No entanto, tal análise, na maioria das vezes, é subjetiva. Por isso, aliado a essa ferramenta, realizaremos alguns testes de hipótese, a fim de obter resultados mais objetivos a respeito dessa suposição. Para tanto, utilizamos dois testes, Goldfeld-Quandt e Breusch-Pagan, ambos supondo como hipótese nula a homoscedasticidade. A Figura 1 apresenta os gráficos de resíduos versus estimativas para os modelos (1), (2) e (3).

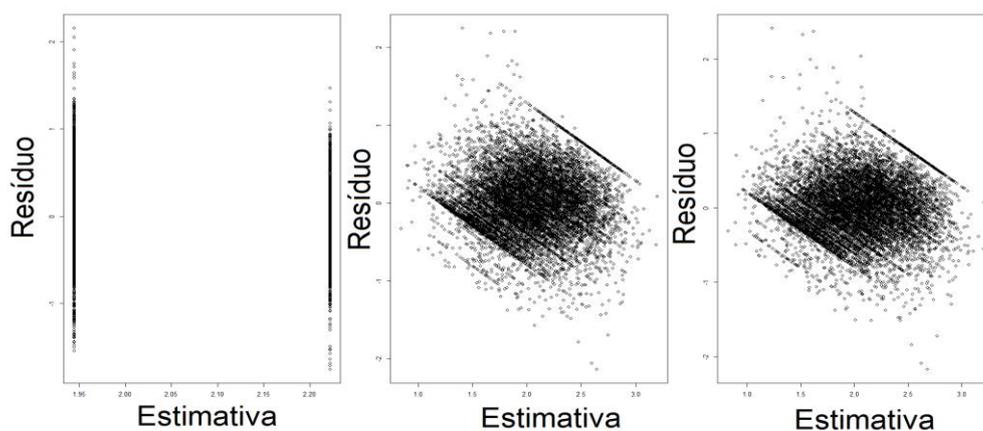


Figura 1: Da direita para a esquerda, os resultados gráficos dos resíduos versus estimativas para os modelos referentes ao ano de 1984, isto é, modelo: (1), (2) e (3) da Tabela 2.

Como podemos observar nos gráficos da Figura 1, existe uma evidência de heteroscedasticidade, em especial, para os modelos (2) e (3), no qual observamos uma maior variabilidade na região central do gráfico. Raciocínio análogo para a Figura 2, no qual apresenta os gráficos de resíduos versus estimativas para os modelos (4), (5) e (6), isto é, temos evidências de que a suposição de homoscedasticidade pode estar sendo violada.

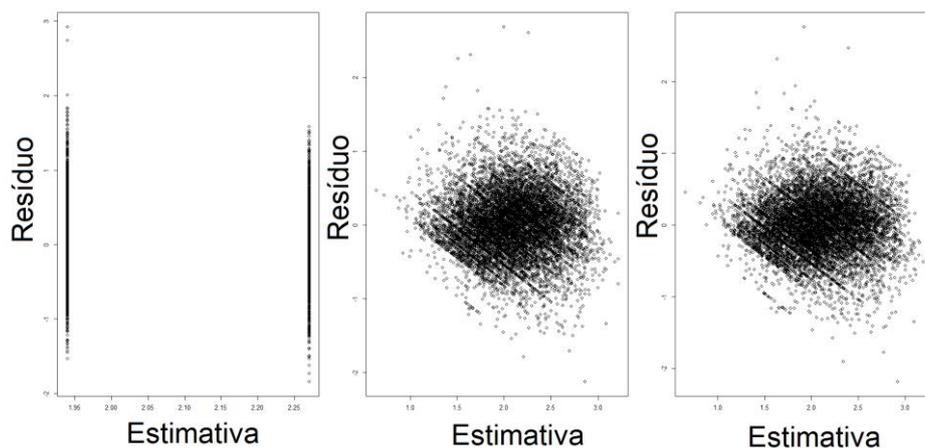


Figura 2: Da direita para a esquerda, os resultados gráficos dos resíduos versus estimativas para os modelos referentes ao ano de 1989, isto é, modelo: (4), (5) e (6) da Tabela 2.

Além disso, como verificamos na Tabela 6, os resultados dos testes concordam em relação a nossas suposições, rejeitando a hipótese de homoscedasticidade do modelo. Nesse caso, devemos corrigir os valores dos erros-padrão das estimativas, e uma alternativa é utilizar os erros-padrão de White (calculados para cada modelo e apresentados na Tabela 5). É digno de nota, que a um nível de 1%, os testes não rejeitam a hipótese nula para os modelos (1) e (4), um dos motivos está no fato de que a causa desta violação, talvez, esteja nas outras variáveis, e não em *Uses computer at work*, como veremos adiante.

Tabela 6: Resultados dos testes de hipótese, Goldfeld-Quandt e Breusch-Pagan, para os modelos propostos pelo autor, relativo aos dados dos anos de 1984 e 1989.

Teste de Hipótese (Heteroscedasticidade)	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Goldfeld-Quandt*</u>						
Estatística de teste	1.302	1.357	1.407	1.319	1.442	1.434
p-valor	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
<u>Breusch-Pagan</u>						
Estatística de teste	10.792	303.154	275.773	4.855	289.751	296.838
p-valor	0.001	< 2.2e-16	< 2.2e-16	0.028	< 2.2e-16	< 2.2e-16

* Para a realização do teste de Goldfeld-Quandt, o banco de dados utilizados na Tabela 2 foi ordenado a partir da variável "School".

As Figuras (3), (4) e (5) mostram os gráficos das variáveis explanatórias versus os resíduos para o modelo (6). Verificamos que, possivelmente, as variáveis que estão mais interferindo na heteroscedasticidade sejam: *Years of education* e *Black*, devido ao comportamento do gráfico em forma de funil crescente (para *Years of education*) e decrescente (para *Black*).

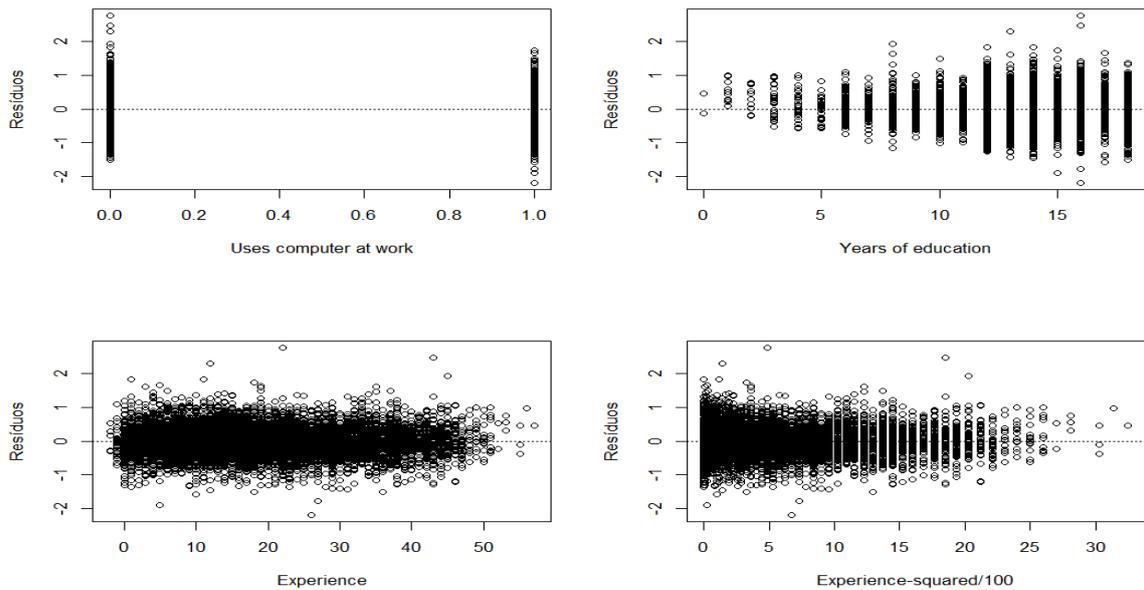


Figura 3: Gráfico das variáveis explanatórias versus resíduos para o modelo (6)

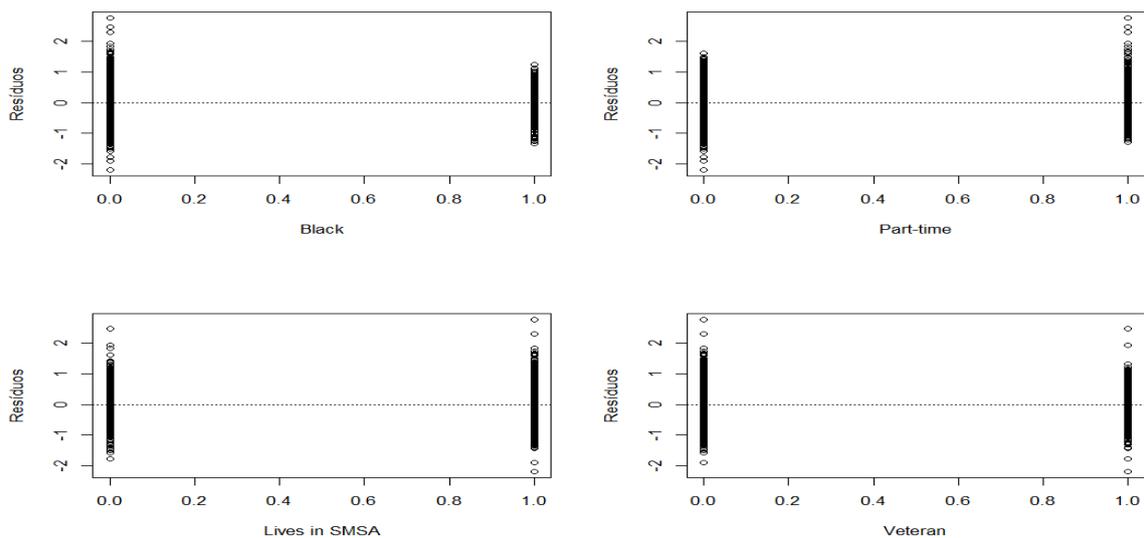


Figura 4: Gráfico das variáveis explanatórias versus resíduos para o modelo (6)

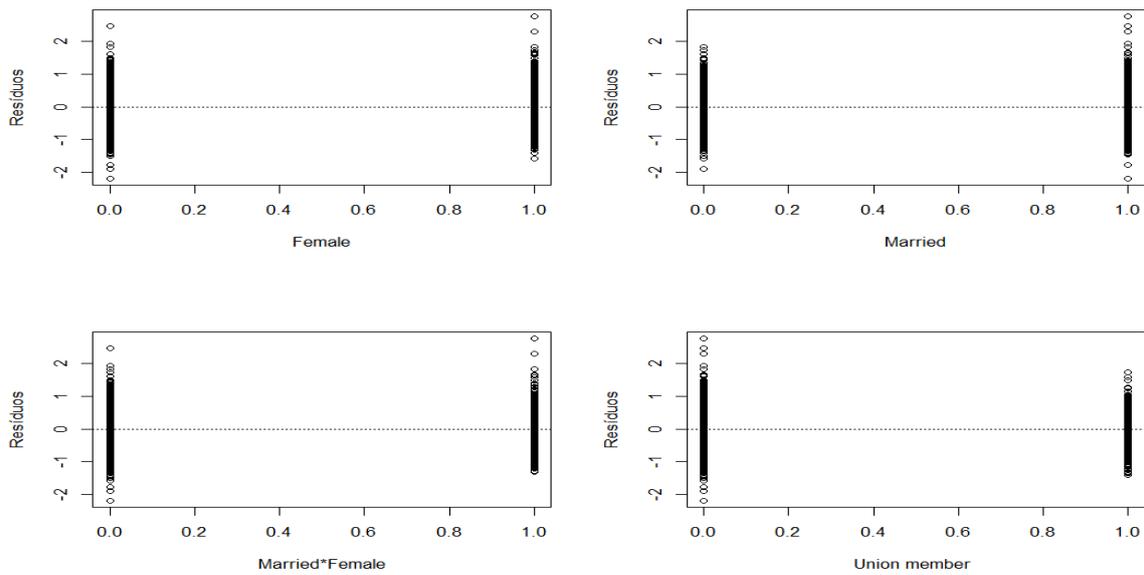


Figura 5: Gráfico das variáveis explanatórias versus resíduos para o modelo (6)

Por fim, realizamos os testes referentes à correlação serial, a fim de verificar a suposição de erros aleatórios independentes. Primeiramente, geramos correlogramas para cada modelo proposto na Tabela 2, com o objetivo de verificar uma possível indicação de que tal suposição está sendo violada. Nas Figuras 6 e 7 apresentamos os resultados gráficos da função de autocorrelação e autocorrelação parcial para os modelos referentes ao ano de 1984.

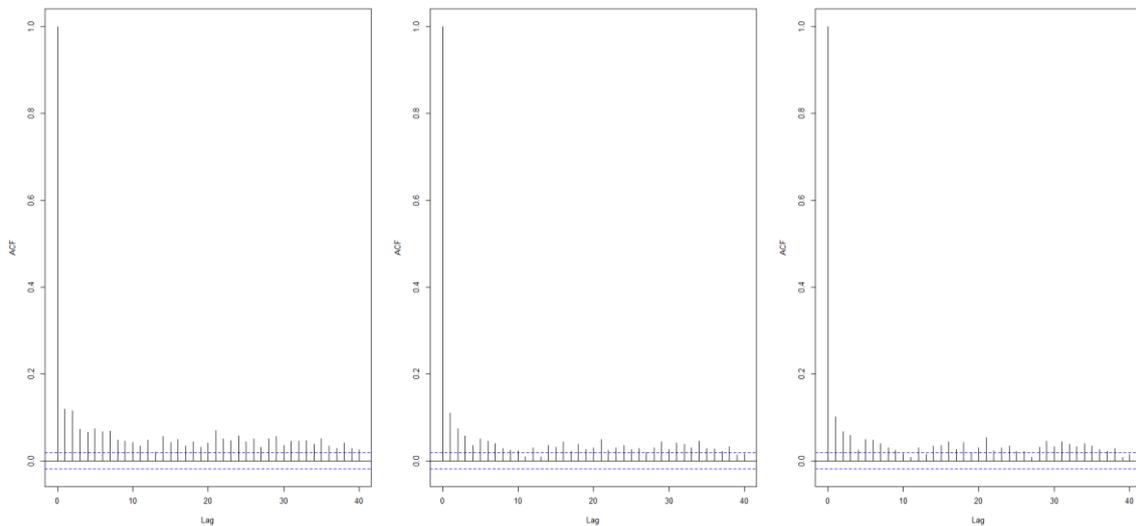


Figura 6: Função de autocorrelação dos resíduos

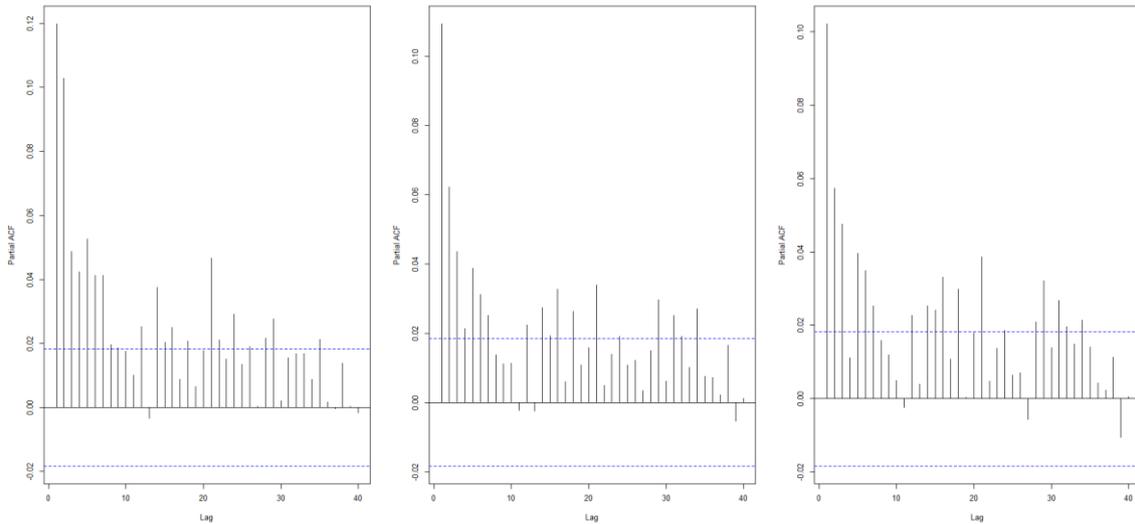


Figura 7: Função de autocorrelação parcial dos resíduos

Como podemos observar na Figura 6, referente aos modelos de 1984, há evidências gráficas que indicam que a suposição de erros aleatórios independentes está sendo violada em todos os modelos propostos para esse ano, isto é, possivelmente existe uma correlação entre os erros. Evidências semelhantes, obtivemos para os modelos referentes ao ano de 1989, como podemos observar nas Figuras 8 e 9.

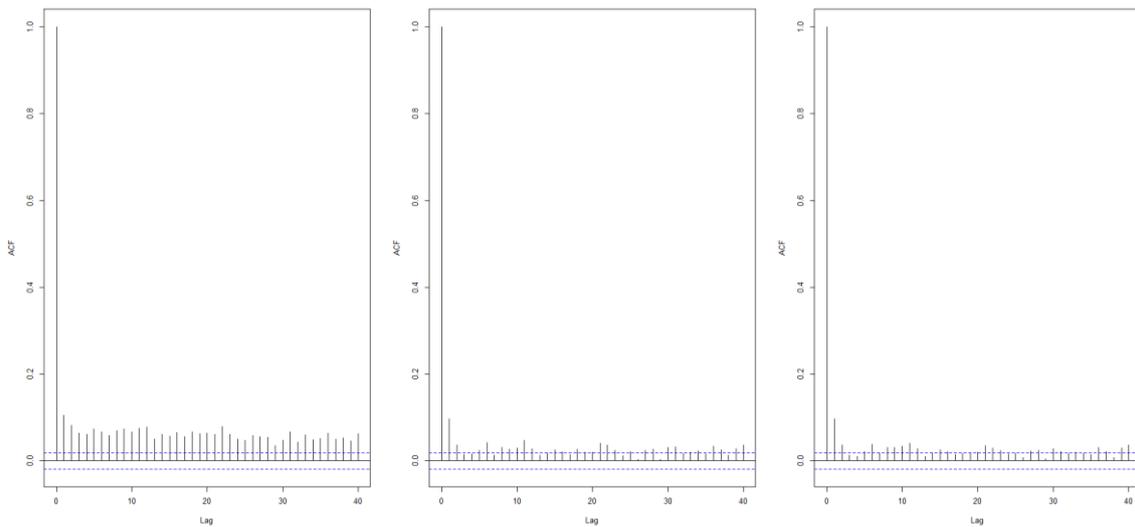


Figura 8: Função de autocorrelação dos resíduos

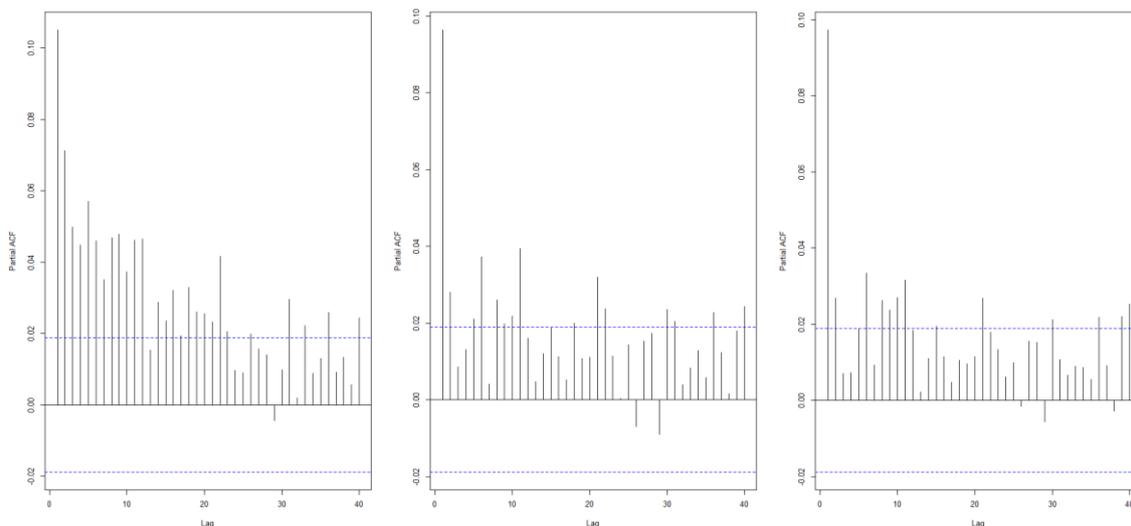


Figura 9: Função de autocorrelação parcial dos resíduos

A fim de obter resultados mais concretos, e confirmar as indicações gráficas, realizamos alguns testes de hipótese. Sendo assim, considere os seguintes testes para a verificação de correlação de resíduos: Durbin-Watson, Breusch-Golfrey e Cochrane. Os resultados obtidos nos testes são apresentados na Tabela 7.

Tabela 7: Resultados dos testes de hipótese, Durbin-Watson, Breusch-Golfrey e Cochrane, para os modelos propostos pelo autor, relativo aos dados dos anos de 1984 e 1989.

Teste de Hipótese (Heteroscedasticidade)	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Durbin-Watson</u>						
Estatística de teste	1.681	1.879	1.881	1.628	1.863	1.880
p-valor	< 2.2e-16	1.29e-11	2.55e-11	< 2.2e-16	3.95e-13	1.31e-10
<u>Breusch-Golfrey</u>						
Estatística de teste	294.817	41.863	40.574	381.419	50.607	39.287
p-valor	< 2.2e-16	9.79e-11	1.89e-10	< 2.2e-16	1.13e-12	3.66e-10
<u>Cochrane</u>						
Estatística de teste	1.7602	1.781	1.7956	1.7898	1.8072	1.8055
p-valor	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

A partir dos resultados obtidos pelos testes de Durbin-Watson, Breusch-Golfrey e Cochrane, os quais foram apresentados na Tabela 7, temos evidência para rejeitar a hipótese nula, isto é, os erros são correlacionados. Novamente, seria muito importante corrigir as estimativas dos erros-padrão obtidos para as estimativas dos parâmetros da regressão.

2.3 3º Questionamento: Baseado nos resultados da coluna 6 da Tabela II, o uso de computador varia segundo union status?. Teste se ambos, uso do computador e efeitos de schooling variam segundo union status. Teste se as interações são significativas.

Com o intuito de investigar se o uso do computador varia segundo *union status*, iremos realizar a regressão representada pelo modelo (6) incluindo a interação das duas variáveis supracitadas (chamaremos de Regressão 1), e a partir do p-valor apresentado, decidiremos sobre sua significância. Dessa forma, caso tal interação seja significativa, poderemos afirmar que existe um efeito significativo da variável *union status* sobre o efeito do uso do computador na resposta. Desenvolvimento semelhante será feito para responder a segunda pergunta – se o uso do computador e efeitos de *schooling* variam segundo *union status* (que chamaremos de Regressão 2), e dessa vez verificaremos a significância conjunta das interações.

Tabela 8: Resultados obtidos para as regressões 1 e 2 propostas para a análise dos efeitos questionados.

Independent variable	Regressão 1			Regressão 2		
	Estimativa	p-valor		Estimativa	p-valor	
Intercept	0.688 (0.038)	< 2e-16	***	0.637 (0.038)	< 2e-16	***
Uses computer at work (1 = yes)	0.184 (0.009)	< 2e-16	***	0.173 (0.009)	< 2e-16	***
Years of education	0.056 (0.001)	< 2e-16	***	0.061 (0.001)	< 2e-16	***
Experience	0.025 (0.001)	< 2e-16	***	0.025 (0.001)	< 2e-16	***
Experience-squared/100	-0.039 (0.002)	< 2e-16	***	-0.040 (0.002)	< 2e-16	***
Black (1 = yes)	-0.101 (0.013)	1.85e-14	***	-0.101 (0.013)	1.95e-14	***
Other race (1 = yes)	-0.010 (0.020)	0.620		-0.012 (0.020)	0.545	
Part-time (1 = yes)	-0.178 (0.011)	< 2e-16	***	-0.178 (0.011)	< 2e-16	***
Lives in SMSA (1 = yes)	0.158 (0.008)	< 2e-16	***	0.155 (0.008)	< 2e-16	***
Veteran (1 = yes)	0.034 (0.012)	0.006		0.032 (0.012)	0.010	
Female (1 = yes)	-0.154 (0.013)	< 2e-16	***	-0.153 (0.013)	< 2e-16	***
Married (1 = yes)	0.145 (0.011)	< 2e-16	***	0.141 (0.011)	< 2e-16	***
Married*Female	-0.132 (0.016)	< 2e-16	***	-0.128 (0.015)	8.55e-16	***
Union member (1 = yes)	0.243 (0.013)	< 2e-16	***	0.666 (0.054)	< 2e-16	***
Uses computer at work* Union	-0.183 (0.022)	< 2e-16	***	-0.122 (0.023)	1.57e-07	***
Years of education*Union	--	--		-0.033 (0.004)	1.74e-15	***
8 Occupation dummies	Yes	--		Yes		
R ²	0.495			0.498		

De acordo com a Tabela 8, verificamos que ambas as interações são significativas. Utilizando o teste t ($t = -8,31$), temos evidências para considerar a interação significativa na Regressão (1); e, utilizando o teste F ($F = 66,95$), temos evidência para considerar as duas interações presentes na Regressão (2) como significativas.

A partir da Regressão (1) e mantendo as demais variáveis fixas, podemos dizer que o fato do indivíduo pertencer a *union* faz com que o uso do computador não influencie no salário, em média. Isto ocorre, porque o coeficiente da interação é praticamente o mesmo coeficiente de *Uses computer at work*, apenas com o sinal trocado.

Da mesma forma, considerando a Regressão (2) e mantendo as demais variáveis fixas, podemos dizer que o fato do indivíduo pertencer a *union* gera um decréscimo de 0.173 para 0,051 no efeito de *Uses computer at work*, ou seja, existe um efeito da variável *union* sobre o efeito de *Uses computer at work* na resposta média. Conclusão análoga, temos sobre os efeitos de *schooling*, isto é, os efeitos de *schooling* variam segundo *union status*.

2.4 4º Questionamento: Os resultados são diferentes de acordo com a raça?

De acordo com o autor, o objetivo principal do trabalho é inferir a influência do uso do computador na média do salário. Portanto, realizaremos uma regressão que contemple as interações entre raça e uso do computador no trabalho. Caso tais variáveis sejam significativas, teremos evidências de que os resultados são diferentes de acordo com raça, isto é, o ganho adicional pelo uso do computador não é o mesmo entre as raças. A Tabela 9 apresenta esse resultado.

Tabela 9: Resultados obtidos para a regressão contemplando as interações entre raça e uso de computador no trabalho.

Independent Variable	Estimativa	p-valor	
Intercept	0,689 (0,038)	< 2e-16	***
Uses computer at work (1 = yes)	0,156 (0,009)	< 2e-16	***

(continua)

Tabela 9: Resultados obtidos para a regressão contemplando as interações entre raça e uso de computador no trabalho.

(conclusão)

Independent variable	Estimativa	p-valor	
Years of education	0,056 (0,001)	< 2e-16	***
Experience	0,024 (0,001)	< 2e-16	***
Experience-squared/100	-0,039 (0,002)	< 2e-16	***
Black (1 = yes)	-0,100 (0,015)	2.95e-10	***
Other race (1 = yes)	-0,021 (0,021)	0.440	
Part-time (1 = yes)	-0,181 (0,011)	< 2e-16	***
Lives in SMSA (1 = yes)	0,162 (0,008)	< 2e-16	***
Veteran (1 = yes)	0,035 (0,012)	0.005	
Female (1 = yes)	-0,154 (0,013)	< 2e-16	***
Married (1 = yes)	0,148 (0,011)	< 2e-16	***
Married*Female	-0,134 (0,016)	< 2e-16	***
Union member (1 = yes)	0,180 (0,010)	< 2e-16	***
8 Occupation dummies	Yes	--	
Uses computer at work* Black	-0,011 (0,027)	0.684	
Uses computer at work*Other race	0,027 (0,041)	0.509	
R ²	0,492		

Podemos afirmar que as interações: *Uses computer at work*Black* e *Uses computer at work*Other race*, não são significativas ao modelo, visto que o R² permaneceu o mesmo até a terceira casa decimal, e, portanto, a partir do teste F, não temos evidência para rejeitar a hipótese nula. Concluimos que os resultados não são diferentes de acordo com raça.

3. PROPOSTA DE MODELO E TESTES DIAGNÓSTICO

Decidimos realizar os procedimentos para a elaboração dos modelos a partir dos dois bancos de dados.

3.1 Proposta de modelo baseado no banco de dados do ano de 1984

Primeiramente, realizamos a etapa de seleção de covariáveis utilizando o método de *backward*, isto é, incluímos inicialmente todas as covariáveis disponíveis no banco de dados e a cada passo retiramos apenas a covariável que apresentou menor significância, até que somente as covariáveis significativas permanecessem no modelo. Na Tabela 10, apresentamos os resultados obtidos pelo processo de *backward* realizado (manualmente).

Tabela 10: Etapas e resultados da seleção de covariáveis pelo método de backward para banco de dados referente ao ano de 1984 e com variável dependente igual ao $\ln(\text{hourly wage})$.

	Parâmetro	Valor z	p-valor associado	
Passo 1				
Uses computer at work (1 = yes)	0.140382	15.239	< 2e-16	***
Years of education	0.049518	26.615	< 2e-16	***
Experience	0.025090	22.590	< 2e-16	***
Experience-squared/100	-0.039441	-16.181	< 2e-16	***
Black (1 = yes)	-0.067394	-5.408	6.51e-08	***
Other race (1 = yes)	-0.091321	-4.165	3.14e-05	***
Part-time (1 = yes)	-0.215989	-19.766	< 2e-16	***
Lives in SMSA (1 = yes)	0.096343	11.674	< 2e-16	***
Veteran (1 = yes)	0.042882	3.727	0.000195	***
Female (1 = yes)	-0.137780	-10.417	< 2e-16	***
Married (1 = yes)	0.138782	11.885	< 2e-16	***
Married*Female	-0.169151	-10.710	< 2e-16	***
Union member (1 = yes)	0.182696	18.489	< 2e-16	***
Occupation 1	0.493907	15.149	< 2e-16	***
Occupation 2	0.283878	8.910	< 2e-16	***
Occupation 3	0.084442	2.601	0.009311	**
Occupation 4	0.403208	12.496	< 2e-16	***
Occupation 5	0.240985	7.586	3.55e-14	***
Occupation 6	NA	NA	NA	
Occupation 7	NA	NA	NA	
Occupation 8	NA	NA	NA	
Region 2	-0.010609	-1.006	0.314400	
Region 3	-0.022795	-2.227	0.025983	*
Region 4	0.109251	9.428	< 2e-16	***

Para o banco de dados referente ao ano de 1984, verificamos que somente a variável *dummy Region 2* não apresentou efeito significativo. Contudo, como as demais variáveis *dummy* referentes à *Region* apresentaram efeito significativo, principalmente, a *Region 4*, decidimos por mantê-la no modelo, sendo assim, permaneceremos com o modelo (3) proposto pelo autor e reproduzido na Tabela 2.

Na sequência realizamos os seguintes testes de diagnóstico: RESET, CUSUM e CUSUMSQ. Realizando o RESET (utilizando a library *lmtest*) obtivemos evidências de que alguma suposição está sendo violada nesse modelo, pois a estatística de teste obtida foi de 30,3297 e o p-valor associado é de 7.293e-14. A Figura 10 e 11 apresentam os resultados gráficos obtidos para o CUSUM e o CUSUMSQ, respectivamente. Lembrando que para a realização desses testes é necessário que possamos assumir que os dados podem ser ordenados de maneira natural, sendo assim, ordenamo-los a partir da covariável *School*.

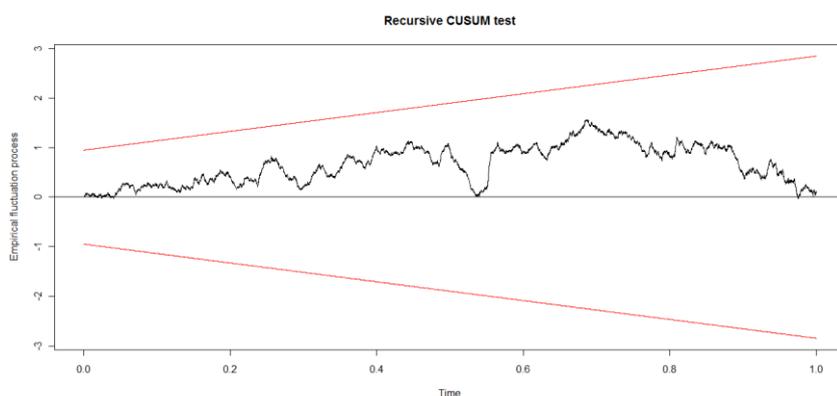


Figura 10: Resultado gráfico referente ao CUSUM teste do modelo (3) referente ao ano de 1984.

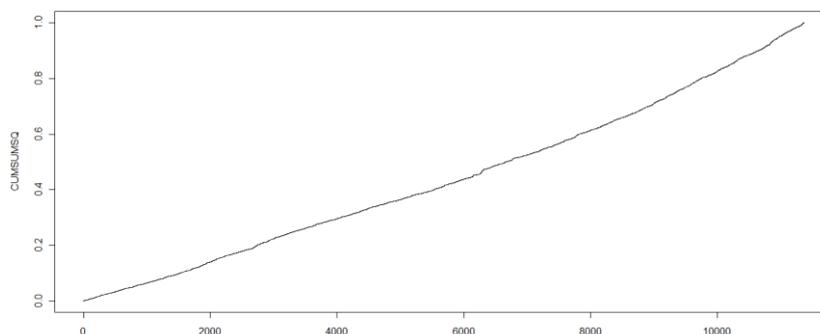


Figura 11: Resultado gráfico referente ao CUSUMSQ teste do modelo (3) referente ao ano de 1984.

Tanto o gráfico de CUSUM como de CUSUMSQ, não nos trazem evidências de que alguma suposição estaria sendo violada, visto que o comportamento de ambos é aparentemente adequado. Contudo, decidimos realizar testes mais específicos, como teste de hipótese para a suposição de homoscedasticidade e correlação serial a fim de identificar se as suposições estão ou não sendo violadas. Visto que o modelo proposto é idêntico ao modelo (3) proposto pelo autor, os resultados e conclusões referentes aos testes de hipótese já foram apresentados anteriormente, e por tal motivo serão omitidos nessa etapa. A Tabela 11 apresenta o modelo proposto para os dados referentes ao ano de 1984.

Tabela 11: Resultados das estimativas e erros-padrões robustos de White para o modelo selecionado para o ano de 1984.

Independent Variable	October 1984 Modelo Proposto	
	Estimativa	Erro-Padrão
Intercept	0.714749	0.040493
Uses computer at work (1 = yes)	0.140382	0.009493
Years of education	0.049518	0.001998
Experience	0.025090	0.001106
Experience-squared/100	-0.039441	0.002418
Black (1 = yes)	-0.067394	0.012045
Other race (1 = yes)	-0.091321	0.023590
Part-time (1 = yes)	-0.215989	0.011600
Lives in SMSA (1 = yes)	0.096343	0.008254
Veteran (1 = yes)	0.042882	0.012001
Female (1 = yes)	-0.137780	0.013406
Married (1 = yes)	0.138782	0.012267
Married*Female	-0.169151	0.015830
Union member (1 = yes)	0.182696	0.009459
Occupation 1	0.493907	0.033302
Occupation 2	0.283878	0.032258
Occupation 3	0.084442	0.032594
Occupation 4	0.403208	0.032385
Occupation 5	0.240985	0.031892
Region 2	-0.010609	0.010378
Region 3	-0.022795	0.010248
Region 4	0.109251	0.011762
R ²	0.4900	

3.2 Proposta de modelo baseado no banco de dados do ano de 1989

Seguindo o procedimento realizado para os dados de 1984, realizamos agora a etapa de seleção de covariáveis utilizando o método de *backward* para o banco de dados referente ao ano de 1989. Vale lembrar, que para o ano de

1989, há mais covariáveis disponíveis no banco de dados, sendo assim, iniciaremos o processo incluindo todas as covariáveis presentes na Tabela 2 juntamente com as covariáveis relacionadas à Tabela 3. Na Tabela 12, apresentamos os resultados obtidos pelo processo de *backward* realizado (manualmente).

Tabela 12: Etapas e resultados da seleção de covariáveis pelo método de *backward* para banco de dados referente ao ano de 1989 e com variável dependente igual ao $\ln(\text{hourly wage})$.

	Parâmetro	Valor z	p-valor associado	
Passo 1				
Uses computer at work (1 = yes)	0.135055	11.957	< 2e-16	***
Years of education	0.054890	29.029	< 2e-16	***
Experience	0.024427	21.075	< 2e-16	***
Experience-squared/100	-0.038393	-14.930	< 2e-16	***
Black (1 = yes)	-0.102490	-7.749	1.01e-14	***
Other race (1 = yes)	-0.009197	-0.441	0.659223	
Part-time (1 = yes)	-0.170010	-14.029	< 2e-16	***
Lives in SMSA (1 = yes)	0.157316	18.369	< 2e-16	***
Veteran (1 = yes)	0.039579	3.165	0.001556	**
Female (1 = yes)	-0.151169	-11.585	< 2e-16	***
Married (1 = yes)	0.147773	12.459	< 2e-16	***
Married*Female	-0.132916	-8.327	< 2e-16	***
Union member (1 = yes)	0.186713	17.152	< 2e-16	***
Occupation 1	0.440759	13.735	< 2e-16	***
Occupation 2	0.223930	7.184	7.22e-13	***
Occupation 3	0.108150	3.423	0.000622	***
Occupation 4	0.354352	11.207	< 2e-16	***
Occupation 5	0.220243	7.132	1.05e-12	***
Region 2	-0.078762	-7.079	1.54e-12	***
Region 3	-0.090904	-8.169	3.44e-16	***
Region 4	0.007824	0.654	0.513324	
Working	0.030538	1.751	0.079957	.
Word processing	0.029447	2.154	0.031261	*
Bookkeeping	-0.050627	-3.391	0.000698	***
Computer-assisted design	0.002847	0.124	0.901124	
Electronic mail	0.130790	7.105	1.28e-12	***
Inventory control	-0.057941	-3.889	0.000101	***
Programming	0.051302	3.022	0.002520	**
Desktop publishing or newsletters	-0.055366	-2.373	0.017659	*
Spread sheets	0.084822	5.167	2.42e-07	***
Sales	0.012547	0.704	0.481167	
Computer games	-0.088774	-3.048	0.002312	**
Passo 2				
Uses computer at work (1 = yes)	0.135122	11.977	< 2e-16	***
Years of education	0.054889	29.030	< 2e-16	***
Experience	0.024425	21.076	< 2e-16	***
Experience-squared/100	-0.038391	-14.931	< 2e-16	***
Black (1 = yes)	-0.102493	-7.749	1.01e-14	***
Other race (1 = yes)	-0.009213	-0.442	0.658622	
Part-time (1 = yes)	-0.170015	-14.030	< 2e-16	***
Lives in SMSA (1 = yes)	0.157314	18.370	< 2e-16	***
Veteran (1 = yes)	0.039581	3.165	0.001554	**
Female (1 = yes)	-0.151213	-11.593	< 2e-16	***
Married (1 = yes)	0.147789	12.462	< 2e-16	***

Married*Female	-0.132949	-8.330	< 2e-16	***
Union member (1 = yes)	0.186717	17.153	< 2e-16	***
Occupation 1	0.440816	13.739	< 2e-16	***
Occupation 2	0.223940	7.185	7.18e-13	***
Occupation 3	0.108179	3.424	0.000619	***
Occupation 4	0.354380	11.208	< 2e-16	***
Occupation 5	0.220262	7.133	1.05e-12	***
Region 2	-0.078791	-7.083	1.49e-12	***
Region 3	-0.090923	-8.172	3.36e-16	***
Region 4	0.007817	0.653	0.513665	
Working	0.030533	1.751	0.079991	.
Word processing	0.029507	2.160	0.030802	*
Bookkeeping	-0.050629	-3.391	0.000698	***
Electronic mail	0.130975	7.139	1.00e-12	***
Inventory control	-0.057947	-3.890	0.000101	***
Programming	0.051673	3.092	0.001995	**
Desktop publishing or newsletters	-0.054953	-2.380	0.017342	*
Spread sheets	0.084960	5.187	2.17e-07	***
Sales	0.012457	0.700	0.483919	
Computer games	-0.088521	-3.047	0.002320	**

Passo 2

Uses computer at work (1 = yes)	0.135079	11.974	< 2e-16	***
Years of education	0.054849	29.043	< 2e-16	***
Experience	0.024439	21.097	< 2e-16	***
Experience-squared/100	-0.038423	-14.950	< 2e-16	***
Black (1 = yes)	-0.102178	-7.737	1.11e-14	***
Part-time (1 = yes)	-0.169960	-14.027	< 2e-16	***
Lives in SMSA (1 = yes)	0.157220	18.365	< 2e-16	***
Veteran (1 = yes)	0.039776	3.183	0.001463	**
Female (1 = yes)	-0.151199	-11.592	< 2e-16	***
Married (1 = yes)	0.147764	12.461	< 2e-16	***
Married*Female	-0.132987	-8.333	< 2e-16	***
Union member (1 = yes)	0.186594	17.148	< 2e-16	***
Occupation 1	0.440910	13.742	< 2e-16	***
Occupation 2	0.223977	7.186	7.10e-13	***
Occupation 3	0.108070	3.421	0.000627	***
Occupation 4	0.354456	11.211	< 2e-16	***
Occupation 5	0.220229	7.132	1.05e-12	***
Region 2	-0.078702	-7.077	1.57e-12	***
Region 3	-0.090946	-8.175	3.29e-16	***
Region 4	0.007081	0.598	0.550186	
Working	0.030534	1.751	0.079967	.
Word processing	0.029545	2.163	0.030579	*
Bookkeeping	-0.050573	-3.388	0.000707	***
Electronic mail	0.131082	7.145	9.57e-13	***
Inventory control	-0.057894	-3.886	0.000102	***
Programming	0.051482	3.081	0.002065	**
Desktop publishing or newsletters	-0.054943	-2.379	0.017357	*
Spread sheets	0.084981	5.189	2.15e-07	***
Sales	0.012580	0.707	0.479548	
Computer games	-0.088509	-3.046	0.002323	**

Passo 3

Uses computer at work (1 = yes)	0.135982	12.133	< 2e-16	***
Years of education	0.054823	29.035	< 2e-16	***
Experience	0.024421	21.087	< 2e-16	***
Experience-squared/100	-0.038392	-14.940	< 2e-16	***
Black (1 = yes)	-0.102351	-7.752	9.89e-15	***
Part-time (1 = yes)	-0.170120	-14.043	< 2e-16	***
Lives in SMSA (1 = yes)	0.157193	18.362	< 2e-16	***
Veteran (1 = yes)	0.039727	3.179	0.001482	**

Female (1 = yes)	-0.151441	-11.615	< 2e-16	***
Married (1 = yes)	0.147884	12.472	< 2e-16	***
Married*Female	-0.133069	-8.339	< 2e-16	***
Union member (1 = yes)	0.186233	17.134	< 2e-16	***
Occupation 1	0.441040	13.747	< 2e-16	***
Occupation 2	0.224634	7.211	5.94e-13	***
Occupation 3	0.108329	3.429	0.000607	***
Occupation 4	0.354378	11.209	< 2e-16	***
Occupation 5	0.220282	7.134	1.04e-12	***
Region 2	-0.078634	-7.071	1.63e-12	***
Region 3	-0.090934	-8.174	3.31e-16	***
Region 4	0.007089	0.598	0.549686	
Working	0.030563	1.753	0.079675	.
Word processing	0.028924	2.122	0.033882	*
Bookkeeping	-0.049971	-3.353	0.000802	***
Electronic mail	0.131661	7.184	7.21e-13	***
Inventory control	-0.055350	-3.829	0.000129	***
Programming	0.051853	3.105	0.001906	**
Desktop publishing or newsletters	-0.054269	-2.352	0.018675	*
Spread sheets	0.085691	5.242	1.62e-07	***
Computer games	-0.088043	-3.031	0.002443	**

Passo 4

Uses computer at work (1 = yes)	0.135996	12.133	< 2e-16	***
Years of education	0.054791	29.017	< 2e-16	***
Experience	0.024691	21.508	< 2e-16	***
Experience-squared/100	-0.038870	-15.210	< 2e-16	***
Black (1 = yes)	-0.101590	-7.697	1.51e-14	***
Part-time (1 = yes)	-0.177813	-15.747	< 2e-16	***
Lives in SMSA (1 = yes)	0.157211	18.363	< 2e-16	***
Veteran (1 = yes)	0.039430	3.155	0.001608	**
Female (1 = yes)	-0.151391	-11.610	< 2e-16	***
Married (1 = yes)	0.148039	12.484	< 2e-16	***
Married*Female	-0.133002	-8.334	< 2e-16	***
Union member (1 = yes)	0.186019	17.114	< 2e-16	***
Occupation 1	0.441555	13.762	< 2e-16	***
Occupation 2	0.225026	7.223	5.44e-13	***
Occupation 3	0.108915	3.448	0.000567	***
Occupation 4	0.354886	11.225	< 2e-16	***
Occupation 5	0.220916	7.154	8.96e-13	***
Region 2	-0.078745	-7.081	1.52e-12	***
Region 3	-0.091099	-8.188	2.95e-16	***
Region 4	0.007001	0.591	0.554709	
Word processing	0.028890	2.119	0.034109	*
Bookkeeping	-0.049774	-3.340	0.000842	***
Electronic mail	0.131608	7.181	7.40e-13	***
Inventory control	-0.055265	-3.823	0.000133	***
Programming	0.051679	3.095	0.001976	**
Desktop publishing or newsletters	-0.054398	-2.358	0.018409	*
Spread sheets	0.085963	5.259	1.48e-07	***
Computer games	-0.087520	-3.013	0.002593	**

A partir dos resultados apresentados na Tabela 12 (a partir da seleção de covariáveis para o modelo referente ao ano de 1989) verificamos que as variáveis explicativas *Computer-assisted design*, *Other race*, *Sale* e *Working* não apresentaram efeito significativo e por tal motivo foram retiradas do modelo. Além disso, apesar da variável *Region 4* não apresentar efeito

significativo, como as demais variáveis referentes a *Region* apresentaram efeito significativo, decidimos por mantê-la no modelo.

A fim de verificarmos se o modelo ajusta-se adequadamente aos dados, realizamos os seguintes testes de diagnóstico: RESET, CUSUM e CUSUMSQ (para a realização dos dois últimos, os dados foram ordenados a partir da variável explanatória *School*). Realizando o RESET teste, utilizando a library *lmtest* obtivemos evidências que alguma suposição está sendo violada, pois a estatística de teste obtida é de 33.0255 e o p-valor associado é de 5.022e-15.

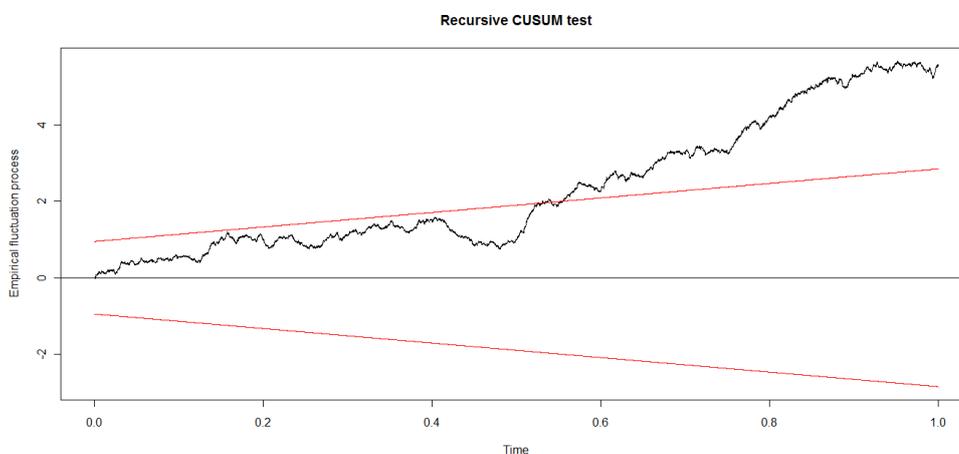


Figura 12: Resultado gráfico referente ao CUSUM teste do modelo proposto, após a seleção das covariáveis, referente ao ano de 1989.

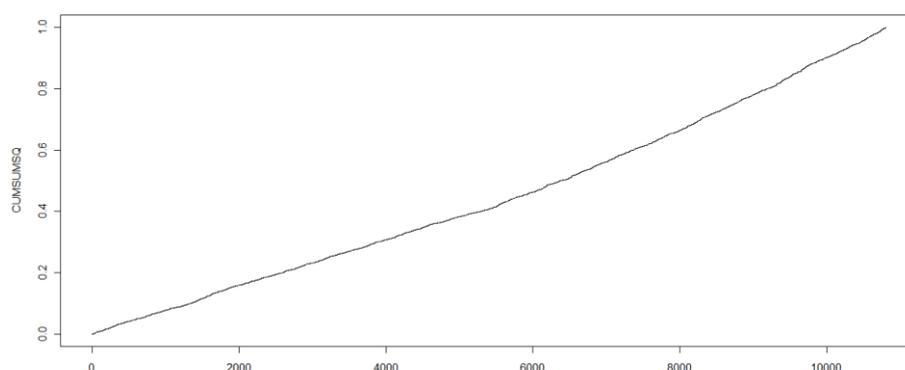


Figura 13: Resultado gráfico referente ao CUSUMSQ teste do modelo (3) referente ao ano de 1989.

Analisando os resultados gráficos apresentados para os testes de CUSUM e CUSUMSQ, apresentados nas Figuras 12 e 13, respectivamente; notamos que o teste de CUSUM nos trouxe evidências a favor da violação de

alguma das suposições assumidas no modelo, enquanto o teste CUSUMSQ não trouxe tais evidências. Pelo fato, da análise gráfica ser subjetiva e de que não encontramos um comportamento adequado para o teste de CUSUM, então optamos por realizar testes mais específicos a fim de verificar quais suposições estão sendo violadas.

Após a realização de uma análise subjetiva, partimos para a realização de testes de hipótese como Goldfeld-Quandt, Breusch-Pagan, Breusch-Golfrey e Durbin-Watson, sendo os dois primeiros para testar a suposição de homoscedasticidade e os dois últimos para testar a suposição de erros não-correlacionados. Os resultados dos testes foram apresentados na Tabela 13.

Tabela 13: Resultados dos testes de hipótese referentes as suposições de homoscedasticidade e autocorrelação dos resíduos para o modelo proposto para o ano de 1989.

Teste de Hipótese	Estatística de Teste	p-valor associado
<u>Goldfeld-Quandt</u>	1.420	< 2.2e-16
<u>Breusch-Pagan</u>	301.422	< 2.2e-16
<u>Durbin-Watson</u>	1.883	4.147e-10
<u>Breusch-Golfrey</u>	37.057	1.147e-09

Como podemos observar, temos evidências de que as suposições de homoscedasticidade e erros não correlacionados estão sendo violadas. Logo, as estimativas de erro-padrão das estimativas do modelo, obtidas através do uso do estimador de mínimos quadrados, estão subestimadas, levando-nos a realizar o procedimento de cálculo dos erros-padrão de White. O modelo proposto para ajuste dos dados referentes ao ano de 1989 é apresentado na Tabela 14, já com as estimativas robustas dos erros-padrão.

Tabela 14: Resultados das estimativas e erros-padrões robustos de White para o modelo selecionado para o ano de 1989.

Independent Variable	October 1984 Modelo Proposto	
	Estimativa	Erro-Padrão
Intercept	0.708410	0.037819
Uses computer at work (1 = yes)	0.135996	0.011440
Years of education	0.054791	0.002035
Experience	0.024691	0.001152
Experience-squared/100	-0.038870	0.002622
Black (1 = yes)	-0.101590	0.012878
Part-time (1 = yes)	-0.177813	0.012439
Lives in SMSA (1 = yes)	0.157211	0.008419
Veteran (1 = yes)	0.039430	0.013106
Female (1 = yes)	-0.151391	0.012979
Married (1 = yes)	0.148039	0.012206
Married*Female	-0.133002	0.015723
Union member (1 = yes)	0.186019	0.010748
Occupation 1	0.441555	0.030456

Occupation 2	0.225026	0.028972
Occupation 3	0.108915	0.029257
Occupation 4	0.354886	0.029216
Occupation 5	0.220916	0.028314
Region 2	-0.078745	0.010813
Region 3	-0.091099	0.011048
Region 4	0.007001	0.012089
Word processing	0.028890	0.013451
Bookkeeping	-0.049774	0.014153
Electronic mail	0.131608	0.017745
Inventory control	-0.055265	0.013893
Programming	0.051679	0.016301
Desktop publishing or newsletters	-0.054398	0.022927
Spread sheets	0.085963	0.015920
Computer games	-0.087520	0.027841
R ²	0.4995	

3.3 Conclusão relacionada aos modelos propostos

Como observamos, as covariáveis selecionadas para os modelos de 1984 foram similares ao do ano de 1989, demonstrando que tais variáveis realmente têm efeito significativo sobre a variável resposta (log-salário) na década de 80. Em relação ao ajuste dos modelos, notamos que ambos apresentaram violação das suposições de homoscedasticidade e de erros não correlacionados. Logo, deveríamos realizar algum procedimento para a correção do modelo, como, por exemplo, realizar o cálculo dos erros-padrão robustos (de White).