Impact of Looks on Wage

Shreyas Nellore,[1] Almitha Veigas,[2] Dhatri Chunchu,[3] Jeevithesh Reddy Narravula Reddy,[4] Rithwik Reddy Koripelly [5], Pidaparthi Murthy,[6] and Tej Sai Pranav Reddy Kagitala [7]

[1] SXN210092, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[2] AXV220022, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[3] DXC220013, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[4] JXN210032, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[5] RXK210087, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[6] PSM220000, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[7] TXK220023, The University of Texas at Dallas, Dallas, 75080, TX, Richardson

Abstract

This paper examines the impact of looks on the wages of the population in Canada. The analysis focuses on interaction effects between looks and other factors like gender, race, marital status, health condition, union membership and city size. The study draws on data that was collected from Canadian surveys in 1978, 1979 and 1981. The findings imply that for below average people there are no significant predictors differentiating wage from one another. Overall, this paper contributes to a better understanding of the complex relationship between wages and the various socio-economic factors.

Keywords: Linear Regression, Ordinary Least Squares, Interaction Effect

1. Introduction

Our research project delves into the intricate relationship between beauty which is measured by facial attractiveness on a scale of 1 to 5 and labor market outcomes specifically a person’s wage. The reason for our interest in this topic is attributed to the fact that all the team members would be graduating with the same degree and therefore we would like to understand how much of an impact do looks have on wage. More importantly, we would also like to understand that apart from looks, what are the significant factors amongst gender, race, marital status, health condition and city. Our initial intuition is that there would be an increase in wage as the looks of the person begins to improve however this would not be a major difference.

2. Literature Review

The original paper, called Beauty and the Labor Market by Daniel S. Hamermesh and Jeff E Biddle, appeared in the December 1994 issue of the American Economic Review (AER). It is concluded in this paper that other things equal, the wages of people with below—average looks are lower than those of average—looking workers; and there is a premium in wages for good—looking people that is slightly smaller than this penalty. The penalty and premium may be higher for men, but these gender differences are not large. There is some evidence that the labor markets sort the best—looking people into occupations where their looks are productive. These results were produced by including controls for Education level, experience, union status, tenure with the firm, and firm or establishment size.

As an extension to this already perfect research, we would like to analyze the interaction effects of the dummy variables below-average and above-average with our factors of interest namely gender, race, marital status, health condition and city to identify which interaction effects are significant and interpret the results.

3. Data

The data set used for this research project is beauty from the Wooldridge package available from R. The original data was collected from Canadian surveys in 1978, 1979 and 1981 and used in a paper published in 1994. The data set contained 1260 observations on 17 variables. All the 17 columns in the dataset are provided in the below screenshot using Stata.

Name	Label
wage	hourly wage
lwage	log(wage)
belavg	=1 if looks <= 2
abvavg	=1 if looks >=4
exper	years of workforce experience
looks	from 1 to 5
union	=1 if union member
goodhlth	=1 if good health
black	=1 if black
female	=1 if female
married	=1 if married
south	=1 if live in south
bigcity	=1 if live in big city
smllcity	=1 if live in small city
service	=1 if service industry
expersq	exper^2
educ	years of schooling

Figure 1. Data Labels

Out of these columns the variables of interest are wage and lwage which represents the hourly wage and log of the wage and these would be our dependent variables. From the other list of variables the independent variables which we use is educ which is duration of school in years, exper which is duration of workforce experience in years, expersq which is the square of the exper column, looks which is a categorical column ranking facial attractiveness on a scale of 1 to 5, belavg and abvavg which classify people who had below average or above average looks on a scale of 1 to 5 and dummy variables for gender, race, marital status, health condition and city denoted by female, black, married, goodhlth and bigcity respectively.

The following table shows summary statistics for the variables.

Variable	Obs	Mean	Std. dev.	Min	Max
wage	1,260	6.30669	4.660639	1.02	77.72
lwage	1,260	1.6588	.5945075	.0198026	4.353113
belavg	1,260	.1230159	.3285858	0	1
abvavg	1,260	.3039683	.4601517	0	1
exper	1,260	18.20635	11.96349	0	48
looks	1,260	3.185714	.6848774	1	5
union	1,260	.2722222	.4452804	0	1
goodhlth	1,260	.9333333	.2495429	0	1
black	1,260	.0738095	.2615645	0	1
female	1,260	.3460317	.4758923	0	1
married	1,260	.6912698	.462153	0	1
south	1,260	.1746032	.3797781	0	1
bigcity	1,260	.2190476	.4137652	0	1
smllcity	1,260	.4666667	.4990857	0	1
service	1,260	.2738095	.4460895	0	1
expersq	1,260	474.4825	534.6454	0	2304
educ	1,260	12.56349	2.624489	5	17
belavg_black	1,260	.0055556	.0743578	0	1

Figure 2. Summary statistics

Some of our interesting observations are:

- The mean wage is $6.30 an hour, which seems very low by our standards. However, considering this data is from 1970's it might make sense.
- The mean for looks is 3.186. Since on a scale of 1 to 5, 3 was "average", this suggests that people were slightly generous in their ratings of other people's attractiveness - or that our sample was slightly above average in looks.
- Only 34% of the people in this data set are female. This is not surprising since, especially back in 1980, women were less likely to participate in the work force.
- Only 7.4% of the people in this data set are black. Again, this isn't surprising since Canada is quite ethnically homogeneous.

The following table shows the correlation between all the variables in the dataset.

	wage	lwage	belavg	abvavg	exper	looks	union	goodhlth
wage	1							
lwage	0.8573	1						
belavg	-0.0834	-0.1067	1					
abvavg	0.0066	-0.0066	-0.2475	1				
exper	0.2346	0.3066	0.0436	-0.1651	1			
looks	0.055	0.0587	-0.6946	0.8339	-0.155	1		
union	0.0945	0.1771	-0.0336	-0.0824	0.0898	-0.0435	1	
goodhlth	0.0068	0.0405	-0.0065	0.0659	-0.1316	0.0586	-0.0224	1
black	-0.0591	-0.1116	-0.041	-0.0018	0.0076	0.0121	0.0319	-0.0341
female	-0.3134	-0.4376	0.0273	0.0416	-0.2472	0.0171	-0.1	-0.0397
married	0.1845	0.2112	0.0097	-0.0738	0.2556	-0.0646	0.0536	-0.0133
south	0.0587	0.0819	0.0314	0.0142	-0.0016	-0.0026	0.0522	-0.0028
bigcity	0.1706	0.1932	-0.0172	-0.0037	0.052	0.0077	0.0813	-0.0046
smllcity	-0.0168	-0.0057	-0.0355	-0.0233	-0.0744	-0.0005	-0.0181	0.0587
service	-0.0537	-0.1175	0.0247	0.0353	-0.0573	0.0102	-0.1036	0.0499
expersq	0.185	0.2434	0.0392	-0.1578	0.9661	-0.1477	0.0797	-0.1368
educ	0.2123	0.2661	-0.0906	0.1323	-0.1862	0.1538	-0.096	0.1071

	black	female	married	south	bigcity	smllcity	service	expersq	educ
wage									
lwage									
belavg									
abvavg									
exper									
looks									
union									
goodhlth									
black	1								
female	0.1137	1							
married	-0.1333	-0.3192	1						
south	-0.0419	-0.0225	-0.0456	1					
bigcity	0.1147	-0.0262	-0.0531	-0.001	1				
smllcity	-0.0633	0.0018	-0.0119	0.0936	-0.4954	1			
service	0.0173	0.268	-0.0751	0.0692	-0.0154	0.0321	1		
expersq	0.0306	-0.2263	0.2026	-0.0007	0.0573	-0.0755	-0.074	1	
educ	-0.1382	0.0091	-0.0477	0.0964	0.1335	0.0501	0.3016	-0.2164	1

Figure 3. Correlation Matrix

Our observations are as following:

- There is significant correlation between wage and lwage which is expected as lwage is derived from wage.
- There is significant correlation between looks and belavg, abvavg which is expected as belavg and abvavg is derived from looks.
- There is significant correlation between smallcity and bigcity which is expected as if the city is not a small city, it is a big city.
- There is significant correlation between exper and expersq which is expected as expersq is derived from exper.

The following image shows a plot of average wage and average lwage by looks. It can be clearly seen from the visual that there is a larger increase in wage between looks 1 to 3 when compared to increase in wage between looks 3 to 5.

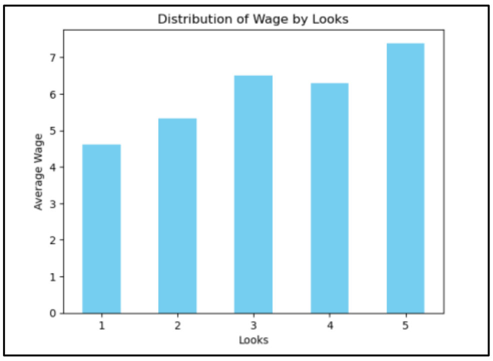

Figure 4. Wage by Looks

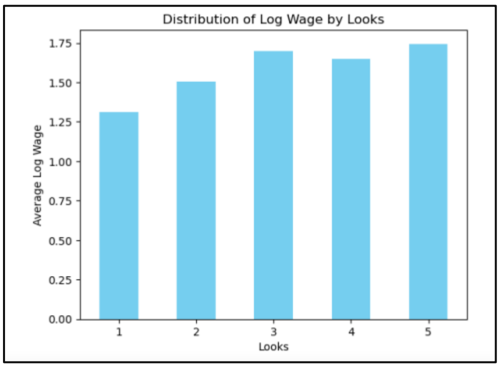

Figure 5. Log wage by Looks

The following image shows a plot of average wage and average lwage by gender and marital status. The following observations can be made from the visual.

- It can be clearly seen that men on average earn more than women. This can be attributed to the wage disparities between genders that existed in 1970's.
- Married men earn more than unmarried men. This could be simply because married men are usually older than unmarried men and this wage gap is because of the experience difference and not just marital status.
- Married women earn equal or less than unmarried women. This could be because married women in the 1970's usually quit their jobs to take care of the kids and family post marriage.

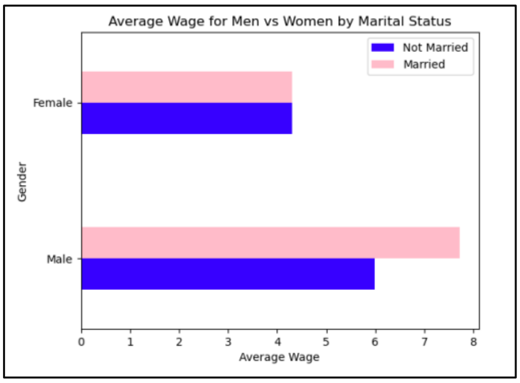

Figure 6. Wage by Gender and Marital Status

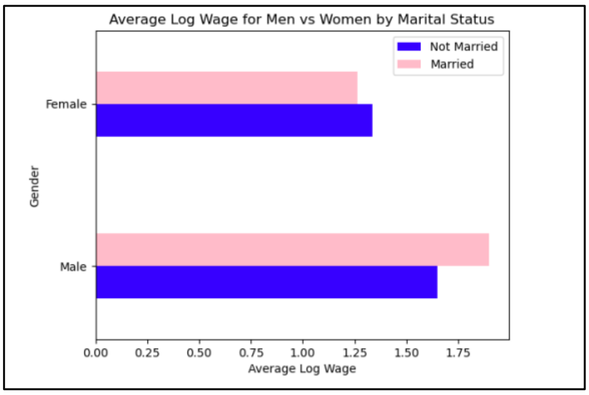

Figure 7. Log wage by Gender and Marital Status

4. Empirical Method

Initially, we start with building 2 regression equations with wage and lwage as dependent variables respectively and educ, exper, expersq and looks as the independent variables. These models can be mathematically represented using the following equations.

Model1: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 looks + u$

Model2: $\log{(wage)} = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 looks + u$

The results of these 2 models are as following:

. reg wage educ exper expersq looks

Source	SS	df	MS			
				Number of obs	=	1,260
				F(4, 1255)	=	52.17
Model	3898.86609	4	974.716523	Prob > F	=	0.0000
Residual	23448.5731	1,255	18.684122	R-squared	=	0.1426
				Adj R-squared	=	0.1398
Total	27347.4392	1,259	21.7215561	Root MSE	=	4.3225

wage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.4229037	.0481563	8.78	0.000	.3284279	.5173795
exper	.2993902	.0396703	7.55	0.000	.2215628	.3772176
expersq	-.0043274	.000892	-4.85	0.000	-.0060774	-.0025774
looks	.4368066	.1816278	2.40	0.016	.0804791	.7931342
_cons	-3.79552	.872656	-4.35	0.000	-5.507545	-2.083494

Figure 8. Model 1 result

reg lwage educ exper expersq looks

Source	SS	df	MS			
				Number of obs	=	1,260
				F(4, 1255)	=	94.91
Model	103.347555	4	25.8368886	Prob > F	=	0.0000
Residual	341.632418	1,255	.272217066	R-squared	=	0.2323
				Adj R-squared	=	0.2298
Total	444.979972	1,259	.353439215	Root MSE	=	.52174

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0683689	.0058127	11.76	0.000	.0569653	.0797725
exper	.0487326	.0047884	10.18	0.000	.0393386	.0581267
expersq	-.0006985	.0001077	-6.49	0.000	-.0009097	-.0004872
looks	.0620681	.0219232	2.83	0.005	.019058	.1050782
_cons	.0462774	.105333	0.44	0.660	-.1603708	.2529255

Figure 9. Model 2 result

From the above two models, the model with lwage has a better R-squared value (0.23) and therefore explains the variation in wage better. However, the coefficient of looks cannot be interpreted easily, therefore, we would leverage the dummy variables belavg and abvavg in the next model.

Model3: $\log{(wage)} = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 belavg + \beta_5 abvavg + u$

The results are the following:

reg lwage educ exper expersq belavg abvavg

Source	SS	df	MS			
				Number of obs	=	1,260
				F(5, 1254)	=	77.81
Model	105.369105	5	21.0738209	Prob > F	=	0.0000
Residual	339.610868	1,254	.270822063	R-squared	=	0.2368
				Adj R-squared	=	0.2338
Total	444.979972	1,259	.353439215	Root MSE	=	.52041

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0687694	.0057921	11.87	0.000	.0574061	.0801327
exper	.0483835	.0047764	10.13	0.000	.0390129	.0577542
expersq	-.0006996	.0001074	-6.51	0.000	-.0009104	-.0004889
belavg	-.1800631	.0461625	-3.90	0.000	-.2706273	-.0894988
abvavg	-.012913	.0334551	-0.39	0.700	-.0785471	.0527212
_cons	.2719752	.0844519	3.22	0.001	.1062927	.4376577

Figure 10. Model 3 result

The following inferences can be made from the above model:

- People with below average looks earn around 18% wage less than people with average looks.
- Testing for H_0: $\beta_4 = 0$, this is significant at 1% level as well (p-value = 0). Therefore, we reject the null hypothesis.
- Testing for H_0 : $\beta_5 = 0$, It is shocking to see abavg coefficient is negative. However, this is not significant even at 10% level (p-value = 0.7). Therefore, we do not reject the null hypothesis.
- Being Average vs above average in looks does not really matter.

For the final model where we try to estimate the interaction effects, we create the interaction variables as shown in the following screenshot.

```
generate belavg_black = belavg*black

generate belavg_female = belavg*female

generate belavg_married = belavg*married

generate belavg_goodhlth = belavg*goodhlth

generate belavg_bigcity = belavg*bigcity

generate abvavg_black = abvavg *black

generate abvavg_female = abvavg *female

generate abvavg_married = abvavg *married

generate abvavg_goodhlth = abvavg *goodhlth

generate abvavg_bigcity = abvavg *bigcity
```

Figure 11. Creating interaction variables

The final model is:

Model4: $\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 belavg + \beta_5 abvavg + \beta_6 black + \beta_7 female + \beta_8 married + \beta_9 goodhlth + \beta_{10} bigcity + \beta_{11} belavg * black + \beta_{12} belavg * female + \beta_{13} belavg * married + \beta_{14} belavg * goodhlth + \beta_{15} belavg * bigcity + \beta_{16} abvavg * black + \beta_{17} abvavg * female + \beta_{18} abvavg * married + \beta_{19} abvavg * goodhlth + \beta_{20} abvavg * bigcity + u$

5. Results

```
. reg lwage educ exper expersq belavg abvavg black female married goodhlth bigcity belavg_
> black belavg_female belavg_married belavg_goodhlth belavg_bigcity abvavg_black abvavg_fe
> male abvavg_married abvavg_goodhlth abvavg_bigcity
```

Source	SS	df	MS			
				Number of obs	=	1,260
				F(20, 1239)	=	38.75
Model	171.233192	20	8.5616596	Prob > F	=	0.0000
Residual	273.74678	1,239	.220941711	R-squared	=	0.3848
				Adj R-squared	=	0.3749
Total	444.979972	1,259	.353439215	Root MSE	=	.47004

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0595127	.0053986	11.02	0.000	.0489212	.0701041
exper	.0413745	.0045104	9.17	0.000	.0325256	.0502234
expersq	-.0006554	.0001001	-6.55	0.000	-.0008518	-.0004591
belavg	-.0533545	.1832284	-0.29	0.771	-.4128267	.3061178
abvavg	.18001	.149398	1.20	0.228	-.1130909	.473111
black	-.106124	.0670554	-1.58	0.114	-.2376788	.0254307
female	-.4616475	.04038	-11.43	0.000	-.5408684	-.3824267
married	.044806	.0420399	1.07	0.287	-.0376713	.1272834
goodhlth	.1204346	.0654291	1.84	0.066	-.0079294	.2487987
bigcity	.2082872	.0433	4.81	0.000	.1233379	.2932366
belavg_black	-.1617266	.2003792	-0.81	0.420	-.5548467	.2313935
belavg_female	.0227159	.0904494	0.25	0.802	-.154735	.2001669
belavg_married	-.0074369	.0954778	-0.08	0.938	-.194753	.1798792
belavg_goodhlth	-.1165617	.1625025	-0.72	0.473	-.4353721	.2022488
belavg_bigcity	.0468345	.1047375	0.45	0.655	-.158648	.252317
abvavg_black	.1000533	.1148228	0.87	0.384	-.1252153	.325322
abvavg_female	.0781582	.066329	1.18	0.239	-.0519713	.2082877
abvavg_married	-.0500709	.0678177	-0.74	0.460	-.1831212	.0829793
abvavg_goodhlth	-.1900512	.1372143	-1.39	0.166	-.4592494	.0791469
abvavg_bigcity	-.0312491	.0728529	-0.43	0.668	-.1741777	.1116795
_cons	.4678803	.1043979	4.48	0.000	.263064	.6726965

Figure 12. Model 4 result

The results are as follows:

- All the variables except belavg, abavg, black, married and goodhealth are individually significant at 1% level.
- None of the interaction effects are significant at the 1% level.
- Adding one more year of education results in 5.9% increase in wage holding everything else fixed.
- Women earn 46% less than men holding everything else fixed.
- Living in a big city results in 20.8% increase in wage holding everything else fixed.

6. Conclusion

Our model shows that below average people earn less than average people who earn less than above average people. Education, experience, gender, and city are all significant factors in determining the wage of an individual. On the contrary looks, marital status and health condition do not play a significant role at 1% level. Looking at the p-values of all the interaction variables it can be concluded that none of these variables are significant at 1% level. This can be interpreted as among belavg or abvavg people, gender, race, marital status health condition and city does not play a significant role in determining the wage of an individual.