Beauty Makes Wealthy?

Effect of Appearance on Income Based on Quantile Regression

Jianing Cai, Zihe Gao

May 8, 2022

Abstract

We explore the effect of physical appearance on income using nationwide survey data from CFPS. To examine the conditional distribution of income, we use quantile regression strategy instead of classical OLS regression. Different from numerous existing literature, we do not only focus on wage, but also take income beside wage into consideration, since we believe good-looking appearance may bring extra income. The empirical results show that good-looking individuals earn more income relative to those with average and below-average appearance, and there exist disparities of return rates of appearance between different groups. And the return rates with quantile shows decreasing trends. Besides, the results of variables describing individual effort imply that even without innate advantages such as good-looking appearance and high iq, one can still live a better life through acquired effort.
Keywords: beauty premium, income, quantile regression

1 Introduction

With the faster development of and easier access to beauty byproducts, it has come to a time when appearances matter much. Actually, beauty can be considered helpful in many aspects of life, including the labor market.

From another aspect, the exploration for the influencing factors of income has never stopped. Nowadays, it's seen a dramatic increase of the income inequity in all over the world. Scholars have investigated numerous origins about wage inequity in the labor market, the most known ones are that returns to schooling is higher for the most skilled individuals (Martins P S, Pereira P T, 2004), and men are paid more at each of the quantiles

(Montenegro C, 1998). Beyond those factors, labor market outcomes are also influenced by many innate traits, like the character, intelligence quotient, and even appearance. Hamermesh and Biddle (1994) find that physically-attractive workers derive sizable rents from their looks. Tanya S. Rosenblat (2008) suggests beautiful people are good negotiators and explained it by a bargaining game. Some current papers enrich the theory. Karina Doorleya and Eva Sierminska (2015) find the beauty premium of women is concentrated at the bottom of the wage distribution using decomposing technique. One research in China which is similar to us provide that good-looking people can earn roughly 5.4% more than other people and relevant anti-discrimination regulations are needed. (Langchuan Peng, Xi Wang, Shanshan Ying, 2020).

This article explores the impact of appearance on income and goes beyond OLS regression to conditional quantile regression, the less parametric characteristic of which allows us to explore whether the effect of appearance varies across income distribution. We use the data from the 2018 Chinese Family Panel Study (CFPS), which is a nationwide representative micro face-to-face survey data set, to examine the effect of good looking and other variables of interest on individual income. We take both job income and other non-labor income into consideration since we believe appearance may also have an influential effect on income besides wage. Our results show that beauty premium prevails across the whole income distribution for good-looking individuals, besides, we also show that male, people with reading habit, a stable marriage, high iq, a high education level tend to have higher income with significant evidence.

With the advantage of quantile regression, we could explore the disparity of variables of interest in different quantiles and in different groups. For rural-urban comparison, we find that good looking, male, and reading habit have generally higher return rates in rural area, and the decreasing trend of return rate with quantile show disparity among quantiles. For highskill-lowskill comparison, the return rate of low-skill works is generally higher for people with good-looking appearance, male, and reading habit. In contrast, the return rate of high education is higher for high-skill works, which may result from the different demand for labor type.

This article is organized as follows. Section 2 makes a brief introduction to the definition of quantile regression and quantile regression treatment effect, which is the empirical strategy used in our empirical work. Section 3 represents our data, empirical analysis and results, and section 4 concludes.

2 Theory: Quantile regression

2.1 Conditional Quantile and Quantile Regression

Classical Ordinary linear regression cares about mean, assuming the residuals follow independent identical normal distribution and homoscedasticity. However, in realistic applications, these assumptions are difficult to be satisfied. Abnormal values in sample may cause severe bias in estimation. Sometimes, for a specific characteristic and a given sample, people with low characteristic value and those of high characteristic value may initially be different people. Under this condition, we do not focus on the expected mean of the dependent variable, but focus on the global distribution of it, under which condition quantile regression is needed. Quantile regression has advantage over traditional linear regression by measuring the treatment effect separately on different quantiles of the population, which could reveal the real effect on different quantiles and avoid being influenced by the potential heterogeneity of the population.

In quantile regression, quantile τ means an aspect of an individual performs better than the proportion τ of the reference group, and worse than the proportion 1-τ. Median is the 0.5 quantile.

Formally, for a continuous random variable y, the definition of population quantile τ is: the probability of y smaller than or equal to y(τ) is τ, that is to say,

$$\tau = P(y \leq y(\tau)) = F(y(\tau))$$

In classical linear regression, people do not estimate grouped data separately, and assume the mean fall on a line or some linear surface. The solution of least squared estimation (LSE) is sample mean. Similarly, in quantile regression, the solution of least absolute deviation estimation (LAD) is median. Under LAD, the objective function becomes:

$$\underset{\beta'}{min} \sum_{i=1}^{n} |y_i - \beta' x_i|$$

More specifically, quantile regression is actually giving asymmetric penalties to overprediction and underprediction through giving different weights. The weighted objective function is:

$$\underset{\widehat{\beta'_\tau}}{min} \sum w_\tau |y_i - \widehat{\beta'_\tau} x_i|$$

$$= \underset{\widehat{\beta'_\tau}}{min} \quad \tau \sum_{i: y_i \geq \widehat{\beta'_\tau} x_i} |y_i - \widehat{\beta'_\tau} x_i| + (1-\tau) \sum_{i: y_i \leq \widehat{\beta'_\tau} x_i} |y_i - \widehat{\beta'_\tau} x_i|$$

The objective function is expressed in this way because of the way of minimizing the weighted sum of positive and negative errors. Let's define a function $\rho_\tau(\widehat{e})$, which is also known as "check function":

$$\rho_\tau(\widehat{e}) = \widehat{e}(\tau - I(\widehat{e} < 0))$$

$$\rho_\tau(\widehat{e}) = \begin{cases} \tau|\widehat{e}| & if \quad \widehat{e} > 0 \\ (1-\tau)|\widehat{e}| & if \quad \widehat{e} < 0 \end{cases} \tag{2.1}$$

The weighted residual allows quantile regression to capture distributional impacts of explanatory variables, and makes quantile regression more robust relative to ordinary least sqaure regression, which could be extremely sensitive to outliers or abnormal values.

2.2 Quantile Treatment Effects

After the objective function of quantile regression, it comes to quantile regression treatment effect (QTE). We consider a binary treatment variable D and a continuous outcome variable Y. Y^1 is realized when $D_i = 1$ and Y^0 is realized when $D_i = 0$. Then the treatment effects can be represented as $Y_i \equiv Y_i^1 D_i + Y_i^0 (1 - D_i)$. The estimation of quantile regression treatment effect is powerful and intuitive to discover the effects on the entire distribution. QTE can include conditional QTE and unconditional QTE with exogeneity or endogeneity, here we only focus on conditional QTE.

For conditional exogenous QTE, we can use the quantile regression estimators proposed by Koenker and Bassett (1978). Consider a continuous independent variable Y_d given treatment variables D, we are interested in the τ th quantile of Y_d , represented by $q(d, \tau)$. The QTEs are defined as the changes in the τth quantile of Y_d given a change of D from d_0 to d_1, that is, $q(d_1, \tau) - q(d_0, \tau)$. For continuous D, QTEs can be represented by $\frac{\partial q(d,\tau)}{\partial d}$. The conditional quantile regression using Y_i as the dependent variable can be written as

$$Y_i^d = X_i\beta^\tau + d\delta^\tau + \varepsilon_i, Q_{\varepsilon_i}^\tau = 0, \tag{2.2}$$

where $Q_{\varepsilon_i}^\tau$ refers to the τ-th quantile of the unobserved random variable ε_i, and δ^τ is conditional quantile treatment effect (CQTE). If the covariate X_i and D_i are exogenous, then $\varepsilon_i \perp (D_i, X_i)$, the estimation is defeined by

$$\left(\widehat{\beta}^\tau, \widehat{\delta}^\tau\right) = \underset{\beta,\delta}{\arg\min} \sum \rho_\tau \left(Y_i - X_i\beta - D_i\delta\right) \tag{2.3}$$

For conditional endogenous QTE, the treatment D is self selected and potentially endogenous. Therefore, the covariate X can not be observed. Then instrumental-variable (IV) estimator of Abadie, Angrist, and Imbens (2002) may be applied. In many applications, the treatment D is self-selected and potentially endogenous, in this case, the true effects could be recovered by IV identification strategy.

2.3 Application of Quantile Regression

The charming robustness characteristic of quantile regression is particularly appealing to noisy data such as earnings. In realistic applications, a classical application of quantile regression is the investigation of the impact of various demographic characteristics and maternal behavior on the birthweight of infants. According to quantile regression results, the disparity between boys and girls of infant weight is smaller in lower quantiles of the distribution and considerably larger in the upper tail distribution, while the conventional least square does a poor job to represent this sparsity. The results show that the effects may not be constant across the conditional distribution of the independent variable.

The rapidly expanding empirical quantile regression literature shows the value of "going beyond models for conditional mean" by contributing in many fields including labor economics, applied micro areas, demand analysis, empirical finance, etc. In this article, we make use of quantile regression and focus on the impact of appearance (as known as beauty premium) on the conditional distribution of income to separately explore the effects in different income groups.

3 Empirical Analysis: Beauty premium on Income

3.1 Background and Data

Furthermore, we want to provide evidences on beauty premium on labor market outcome.

The empirical sample we use is drawn from the 2018 Chinese Family Panel Study (CFPS). It is a face-to-face survey which recorded over 18 thousand respondents of all ages in 31 provinces of China, conducted by the Institute of Social Science Survey of Peking University. To estimate the effects of beauty on wage level, our sample is restricted to respondents who have a formal job. Full-time students and retiree are excluded. The questionnaire recorded individuals' earnings from multiple jobs. Given many people have a sufficient income from supplementary occupation or non-labor income, like investment, property income or other extra income, we use the total individual annual income as the dependent variable. The interested independent variable is whether the individual is good-looking. At the same time, we also care about other control variables, like the intelligence quotient, education level, gender, marital status, and reading habit. All these explanatory variables can answer income inequity. The following table 3.1shows the description statistic of these variables.

Table 3.1: Description Statistic

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
age	10,891	36.77	10.80	14	60
ln(income)	10,724	10.25	0.928	0	13.12
iq	9,506	5.379	1.309	1	7
highedu	8,350	0.144	0.351	0	1
male	10,891	0.592	0.491	0	1
marriage	10,891	0.678	0.467	0	1
urban	10,891	0.308	0.462	0	1
goodlooking	7,089	0.681	0.466	0	1
read	10,891	0.339	0.474	0	1

The quantile figure 3.1, also can be seen as Lorenz curve of our sample, shows an extreme inequity in social distribution. To make the conclusion easier to observe, we do a logarithm transform on total income, which is shown in the box graph 3.2. Due to the huge deviation of normal distribution, the assumption of simple OLS regression can not be satisfied. We'll use a quantile regression to obtain a robust conclusion and compare the heterogeneity of different income groups.

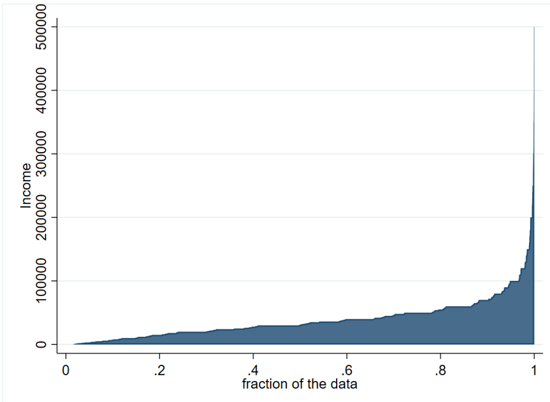

Fig. 3.1. Quantile figure of total income

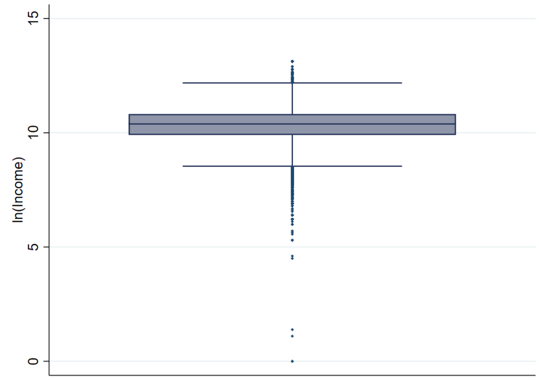

Fig. 3.2. Box graph of ln(income)

The table followed shows a difference test of means between low income group and high income group with respect to appearance, reading habit, age, marital status, intelligence quotient (IQ), and education level. The two groups are separated from a median quantile of the sample. Significant differences could be seen in all aspects. Low income group tends to have a relatively worse physical appearance, less reading, older age, worse marital status, and lower education level relative to high income group. By far, this preliminary research has reflected a strong relationship between our explained variable and explanatory variables.

Table 3.2: Means and Standard Deviations by wage groups

Variables	Low income		High income		Mean Difference
	N	Mean	N	Mean	
goodlooking	3765	0.638	3210	0.737	-0.100***
read	5388	0.268	5336	0.413	-0.145***
age	5388	37.326	5336	36.367	0.959***
marriage	5388	0.646	5336	0.718	-0.071***
iq	4573	5.224	4803	5.533	-0.309***
highedu	3953	0.065	4255	0.217	-0.152***

3.2 Quantile Regression

We formulate a quantile regression model to account for the effect of beauty premium on personal income, the estimation function is formulated as follows:

$$ln(Income)_i = \beta_0 + \beta_1 Goodlooking_i + C_i + \epsilon_i, \tag{3.1}$$

in which subscript i represents the i-th individual, C_i is the vector of control variables ($C_i = \beta_2 male_i + \beta_3 read_i + \beta_4 marriage_i + \beta_5 iq_i + \beta_6 highedu_i$). $Goodlooking_i$ equals 1 if appearance score (range 1-7) from investigators is larger than 4, and equals 0 otherwise; $read_i$ equals 1 if the individual has reading habit, equals 0 otherwise; $marriage_i$ equals 1 if the individual has a stable marriage, equals 0 otherwise; iq_i is a continuous score from investigators; $highedu_i$ equals 1 if the individual has a degree above (including) bachelor, equals 0 otherwise.

Traditional OLS estimators aim at predicting the mean effects, whereas the quantile regression is able to provide information for the conditional distribution, and τ-th quantile of the conditional distribution of explained variable ln(income) given all covariates X_i can be represented as

$$Q_{\tau(Y_i|X_i)} = X_i B_\tau \tag{3.2}$$

The parameter vector of the τ-th quantile of the conditional distribution is estimated by

$$\widehat{\beta}_\tau = \operatorname{argmin} E|Y_i - X_i'\beta| \tag{3.3}$$

We choose a combination of variables with significant effects, and our main regression results are presented in Table 3.3. We apply a simple OLS regression on column (1), and 0.1, 0.25, 0.5, 0.75, 0.9 quantile regression on column (2)-(6), respectively. To better demonstrate the change of coefficients of variables with quantile, we plot them all in Figure 3.3.

Table 3.3: Quantile Regression and OLS Regression on Annual Income

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS			QR		
VARIABLES		0.1	0.25	0.5	0.75	0.9
male	0.427***	0.640***	0.448***	0.371***	0.381***	0.326***
	(15.08)	(7.05)	(10.26)	(18.41)	(21.28)	(9.23)
goodlooking	0.100**	0.182	0.102	0.056	0.066**	0.081**
	(2.13)	(1.12)	(1.42)	(1.00)	(2.02)	(1.97)
read	0.156***	0.267***	0.160***	0.154***	0.133***	0.154***
	(5.24)	(3.17)	(2.95)	(5.29)	(5.33)	(4.67)
marriage	0.329***	0.905***	0.511***	0.167***	0.090**	0.112***
	(9.73)	(10.28)	(9.35)	(5.47)	(2.51)	(3.08)
iq	0.043***	0.069*	0.058***	0.028**	0.025***	0.030*
	(3.24)	(1.66)	(2.89)	(2.20)	(3.50)	(1.90)
highedu	0.632***	0.827***	0.577***	0.476***	0.563***	0.651***
	(15.39)	(7.35)	(11.73)	(10.95)	(16.63)	(13.70)
Constant	9.287***	7.320***	8.712***	9.736***	10.184***	10.443***
	(132.06)	(44.22)	(64.64)	(162.87)	(205.21)	(137.33)
Observations	3,741	3,741	3,741	3,741	3,741	3,741
R-squared	0.145					

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A mechanism for good-looking appearance causes higher income might be that good-looking people are believed to have better social skills, communication skills and self-confidence, which is preferred in competition. More critically, returns rate to a good looking is higher at the bottom of the income distribution. A good-looking people can earn 18.2% more than others on 0.1 quantile of income as shown in column (2). Besides, it's seen that the value is fluctuating around 10% from Figure 3.3. The main reason could be that, as wages increase, personal faculty is more valued while appearance becomes less powerful.

Furthermore, we also find some other determinants of the individual income. The most influential factor that increases income is education. High education can bring returns of more than 40% in all quantiles. This is consistent with our expectation that knowledge changes destiny. And the education level still plays a such important role in low income group. Besides, the second important contributor is gender. The income of a man could be above 30% higher than a woman. Except that women's social division of labor focuses more on the family, gender discrimination in the labor market is still a problem that cannot be ignored. Moreover, returns rate to reading habit is roughly 15%, implying the importance of reading. One interesting finding is that a stable marriage can make extremely higher return in 0.1 quantile. It may suggest that people who are loyal to marriage and good at handling intimacy are trusted by their employers and consumers, and then likely to earn more money. Another potential channel is that supported and motivated by family members, individuals may perform more active and energetic in work especially in tough situation. However, we haven't expected that the intelligence quotient is relatively less important than other factors.

Generally speaking, acquired effort is more important than innate factors, but good-looking people and man are born to earn more than the rest. Our income data includes the non-labor income since we believe appearance may also have influential effect on extra non-labor income. Thus, the income gap might be lager and result in larger coefficients. However, we still think the results reliable. The return rates of each factor decrease as income increases, showing a diminishing marginal trend.

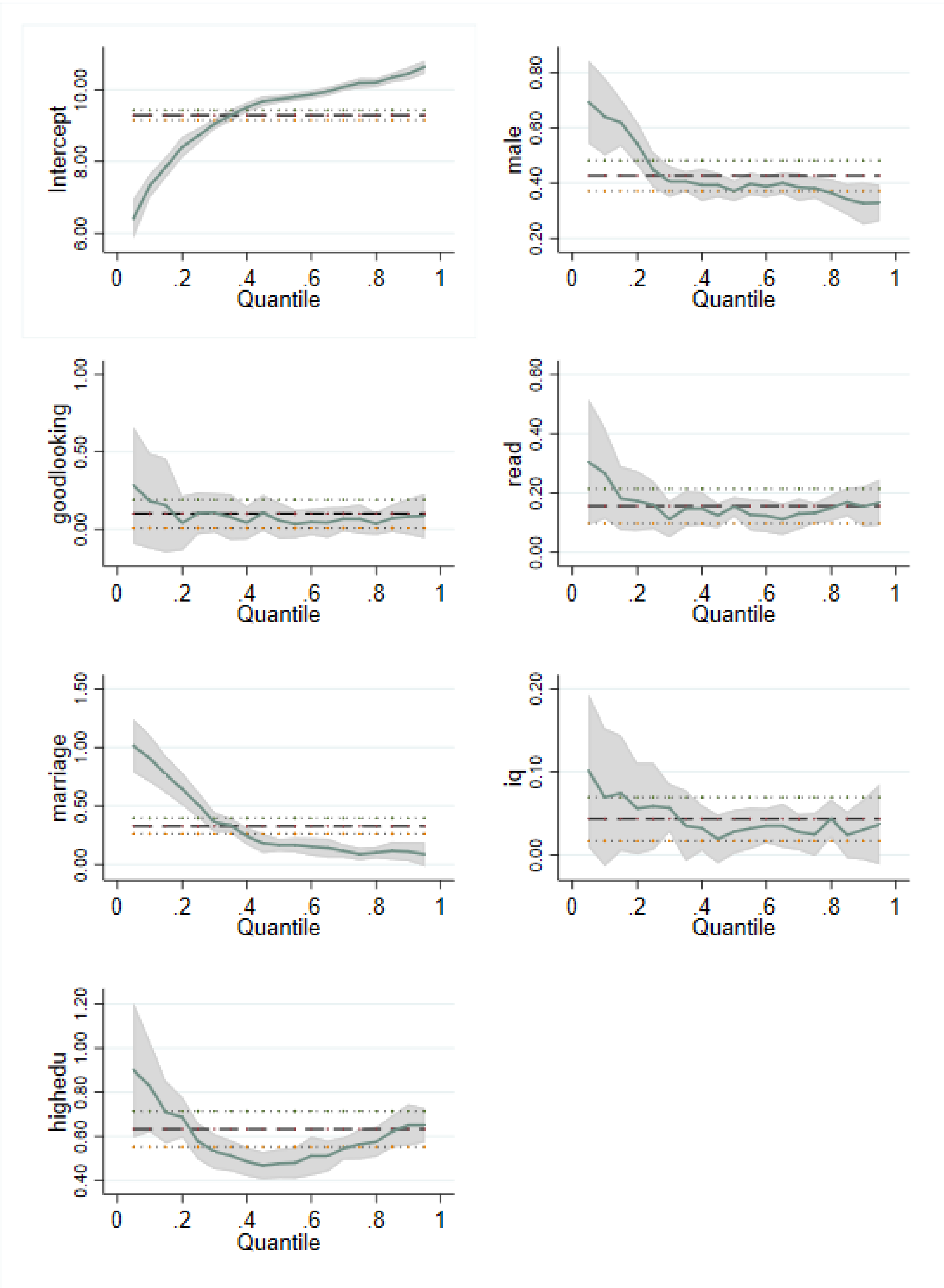

Fig. 3.3. Quantile Regression Estimates

3.3 Rural-urban Disparity

By far, we have argued that good-looking, gender, and reading habit have influential effects on income through quantile regression empirical strategy, then, what about the change trend of the effects in different quantiles? Does there exist any differences of the effects between different groups divided by characteristics? To answer these two questions, the following sections show rural-urban disparity and skill disparity of return rates of some explanatory variables we're interested in, respectively.

The following curves show the "return rates" of explanatory variables reflected by the quantile regression coefficients in different quantiles, which numerically reflect the effects mentioned above.

Figure 3.4 shows that the return rate of good looking increases with quantile in urban area, while has a decrease trend with quantile in rural area. And generally, the beauty premium in rural area is more obvious than that in urban area.

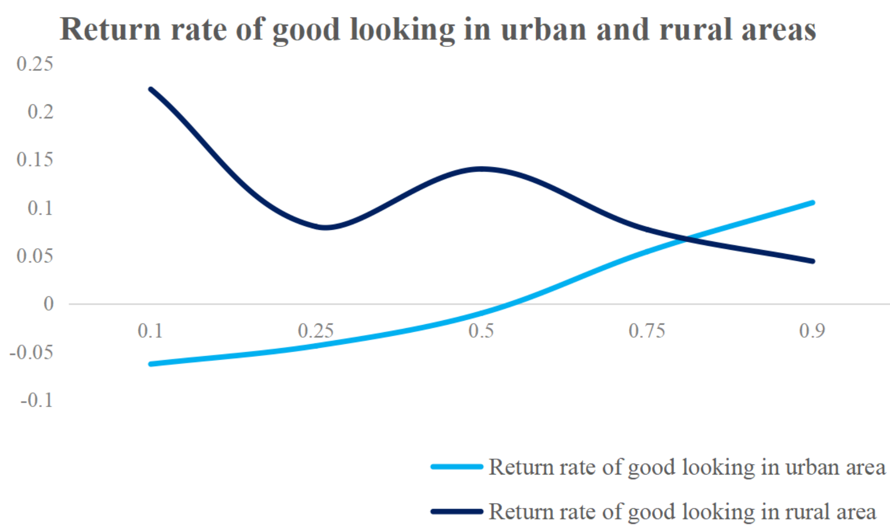

Fig. 3.4. Good looks

As mentioned in previous section, the channel through which physically attractive individuals could earn higher income might be that they have higher level of social skills, communication skills and self-confidence. Thus, an explanatory mechanism for different beauty premium effects between rural and urban areas is that, the competition in labor market in urban area is always more fierce, which requires many other experience and individual skills beside communication skills or social skills. This makes contribution of good looking lower in competition since the "weight" of communication or social skills are made lower by addition of requirements for other skills.

Figure3.5 reflects the change trend of return rate of male in urban and rural areas. The return rate of male in rural area is obviously higher than that in urban area, which is consistent with the intuition that gender discrimination is more severe in rural area. The reason of this discrimination might be that there is more demand for manual labor in rural labor market, in which field male physically have advantages over female. The decreasing trend of return rate of male with quantile in both urban and rural areas also supports this explanation, since low income quantiles may typically be manual laborers.

From the aspect of marital status, figure 3.6 shows that return rate of stable marriage decrease in both urban and rural areas and the return rate is positive in all quantiles, implying stable marriage has positive effect on income. The channel might be that family helps to make individuals more efficient and energetic through support and encouragement among family members and providing motivation for hardworking. The decreasing trend implies that the positive influence brought by family is larger in poorer families, consistent with the intuition that individuals does need support from intimate person especially in tough conditions. The return rate is obviously higher in rural area in low to medium quantiles, which may result from the faster living pace in urban area and traditionally highlighted family ties reserved more in rural area.

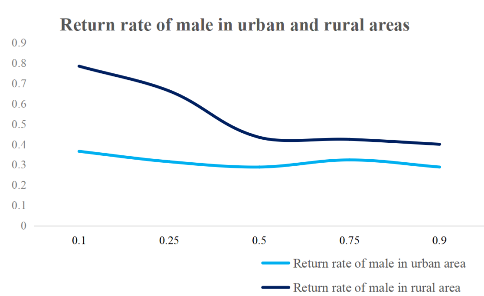

Fig. 3.5. Male

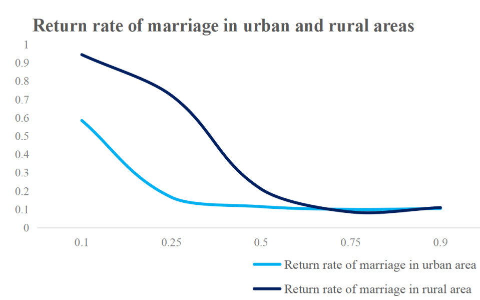

Fig. 3.6. Reading

3.4 Skill Disparity

Since the requirements for high-skill works and low-skill works can be essentially different, we concern about the difference of return rates of explanatory variables of interest between high-skill group and low-skill group. Here, "high-skill works" mean jobs which require minimum education level to be bachelor's degree.

3.4 Skill Disparity

Figure 3.7 shows significant difference of return rates of good looking in all quantiles between high-skill works and low-skill works. The obviously higher return rate in low-skill works group could be explained with the similar channel mentioned in the previous section that high-skill works may require performances in many other fields beside appearance and related social communication skills.

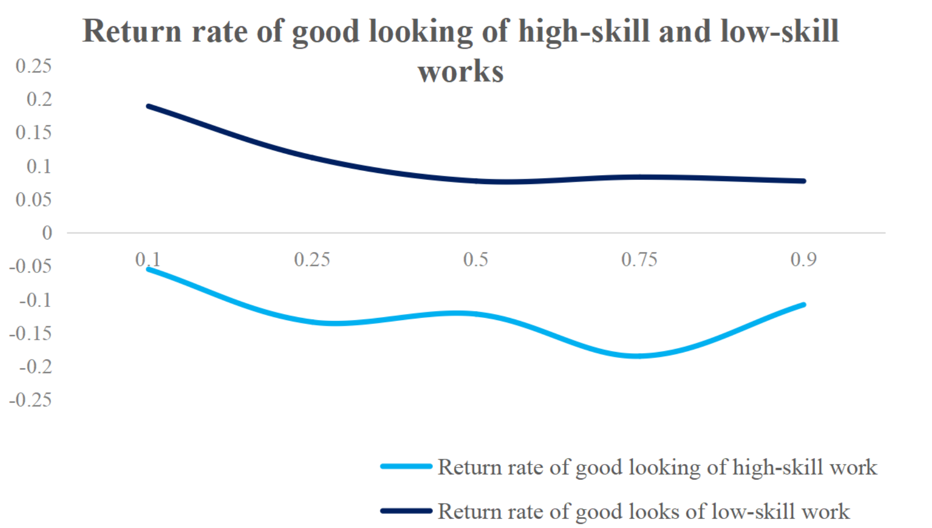

Fig. 3.7. Good looks

Then, what about gender? Consistent with intuition and also consistent with the mentioned manual labor demand issue, Figure 3.8 shows that return rate of male with low-skill works is higher than that with high-skill works. Low-skill works typically demands more manual labor, which has preference for strong male over female.

From figure 3.9, it's shown that return rate of reading is higher with low-skill works than that with high-skill works for all quantiles. Given similar education level and skill level, reading helps to learn additional skills. By intuition, a student with low grades can easily overrun his low-grade peers through hardworking, but it could be very difficult for an initially high-grade student to overrun his high-grade peers although working hard since nearly everyone works hard in high-grade group. Specifically, the marginal improvement of ability diminishes from low-skill people to high-skill people, which may result in the difference of return rate.

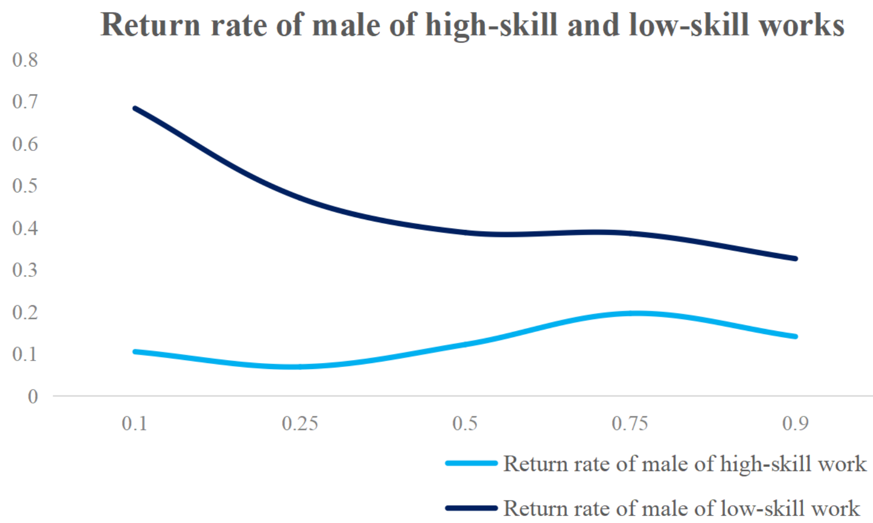

Fig. 3.8. Male

From the perspective of education level, 3.10 shows that the return rate of high education is generally higher in high-skill works than that in low-skill works. This is consistent with the basic requirement of high-skill works: bachelor degree at least. Another potential channel might be that education level may not appear to be that important in low-skill works, which may typically demand more manual labor. And the decreasing trend of both curves with quantile imply that high education substantially improves individuals' competitiveness among his low-income peers.

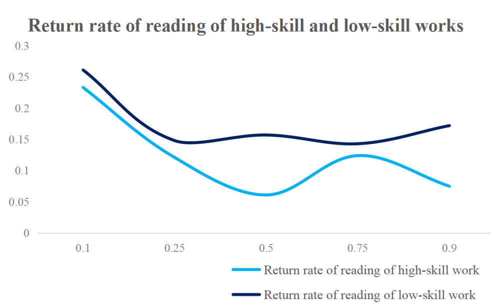

Fig. 3.9. Reading

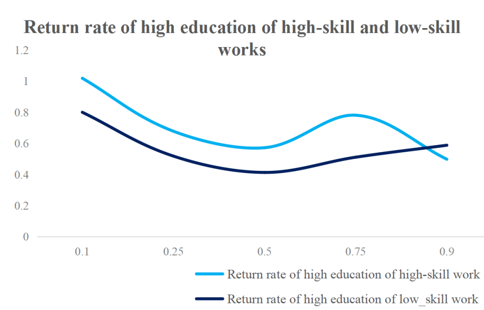

Fig. 3.10. High education

4 Conclusion

In this article, we use the method of quantile regression to estimate the effect of appearance on income. Besides focusing on appearance, we also include variables describing innate factors such as gender and iq, and variables describing individual effort such as reading habit and high education level. Generally, we find that individual effort has larger effect on income relative to innate factors although good-looking person and male initially earn more than others. This finding contributes to the expanding literature about beauty premium and income inequality by comparing the effect of innate factors and individual effort, which is less highlighted in existing literature.

Besides, our comparison of return rates of variables of interest between different groups with rural-area division and highskill-lowskill division shows disparity between groups. The comparison of return rates of appearance between rural and urban areas is less focused on by existing literature.

One potential point which may cause critics comes from the quality of data. Our explanatory variable "good-looking" comes from the ordinal feature "appearance" score ranging from 1 to 7, which is judged by the investigators. The investigators' subjective evaluation of appearance might not be accurate enough, which may cause the numerical value of the coefficients not that convincing. But we believe the sign of coefficients, trend with quantiles, and differences between groups reliable, since the criterion of investigators should be stable for different levels.

However, there exist some limitations of our research. Although we have controlled several dependent variables, there still exist potential confounders not included, such as age, working experience, measure of individual ability, health condition, etc. Failure to control these confounders may cause bias in our estimation. Besides, there may exist multicollinearity among our dependent variables. For example, marital status might be correlated with good looking, and iq might be correlated with education level. These points still need further exploration.

References

[1] Koenker R, Hallock K F. Quantile regression[J]. Journal of economic perspectives, 2001, 15(4): 143-156.

[2] Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas[J]. Journal of the Royal Statistical Society: Series D (The Statistician), 2003, 52(3): 331-350.

[3] 程名望,Jin Yanhong,盖庆恩,史清华.农村减贫:应该更关注教育还是健康?——基于收入增长和差距缩小双重视角的实证[J].经济研究,2014,49(11):130-144.

[4] 刘生龙.教育和经验对中国居民收入的影响——基于分位数回归和审查分位数回归的实证研究[J].数量经济技术经济研究,2008(04):75-85.

[5] Montenegro C. Wage distribution in Chile: Does gender matter? A quantile regression approach[M]. World Bank, Development Research Group/Poverty Reduction and Economic Management Network, 2001.

[6] [1] Hamermesh D , Biddle J . Beauty and the Labor Market[J]. Jeff Biddle.

[7] Martins P S, Pereira P T. Does education reduce wage inequality? Quantile regression evidence from 16 countries[J]. Labour economics, 2004, 11(3): 355-371.

[8] Powell, David. (2019). Quantile Treatment Effects in the Presence of Covariates. The Review of Economics and Statistics. 102. 1-39. 10.1162

[9] 方超,黄斌.教育扩张与农村劳动力的教育收益率——基于分位数处理效应的异质性估计[J].经济评论,2020(04):81-96.DOI:10.19361/j.er.2020.04.10.

[10] Mobius M M, Rosenblat T S. Why beauty matters[J]. American Economic Review, 2006, 96(1): 222-235.

[11] Rosenblat T S. The beauty premium: Physical attractiveness and gender in dictator games[J]. Negotiation Journal, 2008, 24(4): 465-481.

[12] Peng L, Wang X, Ying S. The heterogeneity of beauty premium in China: Evidence from CFPS[J]. Economic Modelling, 2020, 90: 386-396.

[13] Doorley K, Sierminska E. Myth or fact? The beauty premium across the wage distribution in Germany[J]. Economics Letters, 2015, 129: 29-34.