What's Beauty Worth?
Evidence from the Chinese Labor Market

Jean-Louis Arcand* Xinchen Dai† Minhui Zhou‡

Abstract

We use a Chinese longitudinal survey to examine the association between beauty and individual labor market earnings. We estimate Mincerian earning equations with both cross sectional and longitudinal samples using parental characteristics as instrumental variables for beauty. We include the hazard rate to control for selection into occupations where beauty is perceived as being productive in the cross sectional specification. In the longitudinal specification, we add the inverse Mills ratio in the earning equation and carry out a test of selection bias. We get similar results in both specifications. Estimating over the full sample, we find that beauty is associated with higher earnings for females, that its exogeneity is rejected using the appropriate Hausman test, and that there is no evidence for selection into occupations where beauty is important. Our preferred point estimates indicate that the impact on hourly labor earnings of a one standard deviation increase in beauty is equivalent to a seven year increase in the number of years of schooling for the cross sectional model. For the longitudinal model, a one standard deviation increase in beauty increases women's annual income by 40% and increase men's annual income by more than 20%.

Keywords: Beauty; Wages; Selection; Hazard Rate; China.

JEL Codes: J24; J30; J70

*Department of International Economics, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: jean-louis.arcand@graduateinstitute.ch

†Department of International Economics, Graduate Institute of International and Development Studies, Geneva, Switzerland. Email: xinchen.dai@graduateinstitute.ch

‡Corresponding author. Department of Public Economics, University of International Business and Economics, Beijing, China. Email: zhouminhui@uibe.edu.cn

1 Introduction

In this paper, we study how physical attractiveness affects individual earnings, and attempt to disentangle endogeneity from occupational selection effects using parental characteristics (including their beauty) as a plausibly exogenous source of variation in beauty and occupational selection. We estimate the impact of beauty on individual labor market earnings using a nationally representative household survey —the China Family Panel Studies (CFPS)— in both cross sectional and longitudinal contexts. Our model specifications in the two contexts simultaneously control for occupational selection and the endogeneity of beauty, which is novel in the beauty literature.

In our Mincerian earning equations, beauty, BMI and self-reported health are the key explanatory variables. The beauty data is the score the interviewer attributes to their respondent at the end of the interview. In the cross sectional setting, we also include age, years of schooling, a rural-urban dummy and *hukou* status as additional covariates. Since beauty is potentially endogenous (as are BMI and self-reported health) due both to unobserved heterogeneity and to reverse causality, we adopt an instrumental variables approach in which an individual's beauty is instrumented using his or her parents's beauty.

It is likely that more beautiful individuals will select into occupations where physical attractiveness is advantageous. To control for the occupational selection in the cross sectional model, we use the method popularized by Thomas and Strauss (1997) in which it is shown that an exclusion restriction that is valid in terms of controlling for endogeneity can also be used as an exclusion restriction in a selection procedure. We calculate the underlying hazard rate using a logit model of occupational choice, where occupations are classified ex ante into beauty- and non-beauty related. This hazard rate can then be included as an additional covariate in an earning regression estimated over the subsample of individuals who select into occupations where beauty is likely to be important.

To estimate the longitudinal model, we follow the procedure proposed by Semykina and Wooldridge (2010). First, we estimate the occupational selection equation using the probit model and calculate the inverse Mills ratio. Second, we estimate the earning equation using Fixed Effect-Two Stage Least Squares (FE-2SLS), adding the calculated inverse Mills ratio as an additional control. The hypothesis test of the coefficient of inverse Mills ratio gives us a test of selection bias. If the null hypothesis that the true value of the coefficient of inverse Mills ratio is zero is not rejected, then there is no selection bias. The FE-2SLS estimates of the earning equation are consistent. If the null hypothesis is rejected, we estimate the earning equation by pooled 2SLS, using parental characteristics, the time-specific means of parental characteristics, and inverse Mills ratio as instruments to correct the selection bias.

We find for the cross sectional specification that for females, when occupational selection is not taken into account and we estimate over the full sample, the appropriate Hausman test rejects the null of beauty being exogenous (as well as the exogeneity of BMI and self-reported health), indicating that many of the empirical estimates of the impact of beauty on wages in the existing literature may be biased by endogeneity issues. In contrast, when we estimate over the subsample of individuals in beauty-related occupation and control for selection and endogeneity at the same time, and while the exogeneity null is sometimes not rejected, there is no evidence whatsoever for selection effects. Allowing beauty to be exogenous and controlling solely for selection does nothing to bring the latter into sharper focus. The upshot is that selection into beauty-related occupations is not driving the rejection of the exogeneity null, and that our preferred specification is given by one in which we control for endogeneity but not for selection. In quantitative terms, our estimates of the impact of beauty are large. A one standard deviation increase in a female's beauty score in China is equivalent, in labor earnings terms, to an additional 7 years of schooling. For males, in contrast, we do not find any evidence that beauty affects labor income. Similar pattern holds for the longitudinal specification: there is no strong evidence for the existence of occupational selection for both genders. But even for males, the impact of beauty on individual earning is significantly positive when we control for the endogeneity in the earning equation.

Our paper contributes to the beauty literature by rigorously tackling the endogeneity of beauty and the occupational selection simultaneously for the first time. Hamermesh and Biddle (1994) is the first paper that studies the beauty premium in the labor market. The test for sorting is carried out by testing whether the average rating of people's beauty in beauty-related occupations are higher than people in non-beauty-related occupations. The results show that there does exist selection, but the evidence is not strong. For the endogeneity generated from unobserved heterogeneity, Biddle and Hamermash argue that since the family background does not show noticeable effect in their samples, the relevant covariates are unlikely to bias their results. They realize, though, that it cannot prove that their results are free from omitted variable bias. To argue that simultaneity is not a serious problem, Hamermesh and Biddle (1994) uses the research findings from social-psychological literature which shows that people's physical appearances are quite stable in the adulthood. By focusing on younger workers, they find that there is no strong evidence which shows that the beauty premium is lower for younger workers. However, we think that simultaneity does not necessarily means the beauty premium will accumulate over time, which is the premise of Biddle and Hamermesh's argument. One possibility is that older workers might invest more in maintaining their beauty to neutralize the aging effect.

Biddle and Hamermesh (1998) studies the beauty premium for a particular occupation: lawyer. By dividing the sample into employee and self-employed, and into public sector and private sector, Biddle and Hamermash find that there is no sorting resulted from employer's discrimination, but the lawyers who stay in the private sector 15 years after the graduation are more attractive than those leave after 5 years after of the graduation. The photos used for rating are the photographs taken when the matriculants first enter the law school, so the beauty ratings are exogenous to the labor market earnings. Since the sample used in Biddle and Hamermesh (1998) is drawn from only one law school, Biddle and Hamermash argue that the students are quite homogeneous, thus the variations coming from the correlations between parents' income and children's income is negligible.

The following studies never simultaneously discuss the endogeneity and the occupational selection (Harper, 2000; Pfann et al., 2000; Fletcher, 2009; Johnston, 2010; Salter et al., 2012; Hamermesh and Abrevaya, 2013; Doorley and Sierminska, 2015; Scholz and Sicinski, 2015; Oreffice and Quintana-Domeque, 2016). Some studies get rid of the endogeneity of beauty by using special measures of physical appearance. The beauty score used in Harper (2000) is the rating given to the respondents when they are at 7 and 11 years of old, so as Biddle and Hamermesh (1998), it is an *ex-ante* rating before a person enters the labor market. The *ex-ante* measures purge the simultaneity, but it is arguable whether there is still bias from unobserved heterogeneity. Johnston (2010) studies the natural hair color (blondness), as a representative of beauty. Selection is avoided to some extent by focusing on a specific industry or occupation (Pfann et al., 2000; Salter et al., 2012; Scholz and Sicinski, 2015).

The methods we adopt in this paper do not rely on the *ex-ante* beauty measures or the data from a group of relatively homogeneous people. All we ask is the identification of "parent / adult children" pairs in the sample and the occupational information. Numerous household surveys allow us to identify enough number of "parent / adult children" pairs for econometrics analysis. The respondent's occupational information is also provided by most household surveys. The information requirement of our methods are lower than the previous studies.

2 Data

The samples we work with are drawn from the China Family Panel Studies (CFPS), which is a nationally representative longitudinal household survey. The baseline survey of CFPS, which was initiated in 2010, covers 14,960 households in 649 urban and rural communities in 25 provinces.[1] In each household, both adult family members and juveniles under 16 years old are surveyed. Family members linked by blood, marriage or adoption ties are followed over time.

For cross sectional estimation, we use the individual survey and the family relationship survey from the 2014's wave of the CFPS. In addition to individual's beauty scores, BMI (Body Mass Index), and self-reported health, this sample contains information on various measures of income, weekly working hours, educational attainment, a rural-urban dummy, *hukou* status, and parental characteristics. We construct the longitudinal sample using three waves of CFPS (2010, 2012 and 2014).[2] The sample contains information on annual gross income, beauty scores, BMI, self-reported health and parental characteristics.

In our Mincerian wage regressions, the dependent variable is the logarithm of labor income (after the usual inverse hyperbolic sine transformation to account for wages near zero). For the cross sectional sample, the measure is net hourly income, which is calculated either by dividing the net monthly salary by 4 times the weekly working hours for employees with wage jobs, or by dividing net annual income by 52 times weekly working hours for individuals with other types of jobs. For the longitudinal sample, the income measure is the annual gross income. [3]

The key right-hand-side variable is the interviewee's beauty score attributed by the interviewer at the end of the interview. In the CFPS, the score is recorded on a Linkert scale of 1 to 7 with higher values representing a higher evaluation of the interviewee's beauty. For the cross sectional model, other covariates include years of schooling, a dummy representing community type, a dummy on whether the individual possesses nonagricultural residential *hukou*, age, and age squared. For the longitudinal model, the right-hand-side variables only include beauty score, BMI and self-reported health as the rest of the covariates used in the cross sectional model are either completely or nearly time invariant.

We construct our instruments for beauty (as well as BMI and self-reported health status) by identifying all possible "parent / adult child" pairs within each household. We implement the identification using the household roster and variables which indicate the relationship of each household member to the household head. This allows us to construct a subset of observations for which we can construct measures of parental beauty, parental BMI and parental self-reported health. Since both parents of the individuals in our sample

[1]Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan are excluded

[2]The latest survey of CFPS is the 2016 wave. However, the occupational information is not available for 2016, so we cannot define the beauty-related job, which is indispensable for studying occupational selection.

[3]We do not calculate the net hourly income for the longitudinal sample because the statistics related to various income measures and working hours are not comparable across the different waves of the CFPS.

are not always interviewed, we use the following coding rule to construct our instruments: when characteristics of both parents are available, we take the mean; otherwise, either paternal or maternal characteristics are used.

Finally, so as to be able to compute the hazard rate and the inverse Mills ratio corresponding to occupational choice, we classify occupations into two categories based on whether physical attractiveness is likely to be productive and potentially reflected in earnings. Typical occupations where beauty is important include sales related jobs and any occupation where frequent face to face interactions with clients are called for. Appendix A provides details concerning our classification. All results are robust to reasonable changes in the underlying dichotomy.

We divide the full sample by gender and restrict the subsamples to adult respondents (18 years old or above) with positive income and whose weekly working hours are between 0 and 150 hours. Table 1 and Table 2 report descriptive statistics. In both cross sectional and panel samples, there is no significant difference in the right-hand-side covariates between females and males, except for income. Men have higher income and the income distribution is more dispersed than women.

3 Cross section evidence

Our Mincerian earning equation in the cross sectional context is given by:

$$w_i = b_i\beta + h_i\theta + x_i\delta + \eta_i, \tag{1}$$

where w_i (after the inverse hyperbolic sine transformation) is hourly labor earnings, b_i is the measure of beauty, h_i represents BMI and self-reported health, x_i is a matrix of covariates which includes age, years of schooling, a rural-urban dummy and *hukou* status, and η_i represents unobservables.

The first problem affecting Equation 1 is that the measure of beauty b_i is likely to be correlated with the unobservables η_i. There are two reasons for this. First, physical appearance is likely to be in part determined by income: a casual conversation on the topic of expenditures on cosmetics and health/esthetic-related products with a typical middle class Chinese woman is usually sufficient to convince anyone of this point. Second, it is likely that unobservables that affect beauty also affect labor market earnings. Concomitantly, a large corpus of literature that has examined the impact on earnings of measures of health, such as BMI and self-reported health status (h_i), has focused on the potential endogeneity of these variables (see e.g. Thomas and Strauss (1997)). As such, an instrumental variable approach is called for. Our three potential IVs, denoted by z_i, are measures of parental beauty, parental BMI and parental self-reported health.

The second problem affecting Equation 1 is that the marginal effect of beauty on labor earnings is likely to differ by occupation. As has been well-documented in the (largely North American) literature, certain occupations, especially those involving significant interactions with customers, will value beauty more than others, and we therefore expect the marginal effect of beauty on wages to be higher in such occupations. The answer in this case is to divide the data into two subsamples and re-estimate over each subsample while controlling for selection effects. Since it is plausible that more beautiful individuals will select into occupations where physical attractiveness is advantageous, we calculate the underlying hazard rate using a logit model of occupational choice, where occupations are classified *ex ante* into beauty- and non-beauty related. This hazard rate can then be included as an additional covariate in an earning regression estimated over the subsample of individuals who select into occupations where beauty is likely to be important.

In formal terms, when b_i and both elements of h_i are jointly endogenous in the earning equation, and if belonging to the subsample of individuals working in "beauty-friendly" occupations is denoted by the dummy variable d_i, the selection equation is given by a standard latent index model:

$$d_i = \Lambda(z_i \pi_z + x_i \phi + \vartheta_i), \tag{2}$$

where $\Lambda(.)$ is the logistic function, and where the earning equation (Equation 1) is estimated over the subsample of individuals for which d_i=1, while including the hazard rate stemming from Equation 2 to control for selection bias.[4]

The appearance of the matrix of instrumental variables z_i (used in the earning equation) in the selection equation is not fortuitous: as first shown by Thomas and Strauss (1997), an exclusion restriction that is valid in terms of controlling for endogeneity can also be used as an exclusion restriction in a selection procedure, on condition of course that it does provide an acceptable degree of identification both in the latter and in the former. This means, again when b_i and both elements of h_i are jointly endogenous in the wage equation, that the three first stage reduced forms are given by:

$$b_i = z_i \xi_b + x_i \varrho_b + \varsigma_{bi}, \tag{3}$$

$$h_{1i} = z_i \xi_{h1} + x_i \varrho_{h1} + \varsigma_{h1i}, \tag{4}$$

$$h_{2i} = z_i \xi_{h2} + x_i \varrho_{h2} + \varsigma_{h2i}, \tag{5}$$

[4] Moving to a probit specification for the selection equation and thus to a two-step Heckman procedure does not appreciably modify the results.

where h_{1i} represents BMI and h_{2i} self-reported health. Our empirical approach is novel in the beauty literature in that it simultaneously allows for occupational selection and the endogeneity of beauty. If elements of (b_i, h_{1i}, h_{2i}) are deemed to be exogenous in the earning equation, the preceding system must be appropriately modified. For example, if BMI (h_{1i}) is assumed exogenous, the modified selection equation is given by:

$$d_i = \Lambda(z_i\pi_z + h_{1i}\pi_{h1} + x_i\phi + \vartheta_i), \tag{6}$$

with the two remaining first-stage reduced forms being given by:

$$b_i = z_i\xi_b + h_{1i}\zeta_b + x_i\varrho_b + \varsigma_{bi}, \tag{7}$$

$$h_{2i} = z_i\xi_{h2} + h_{1i}\zeta_{h2} + x_i\varrho_{h2} + \varsigma_{h2i}. \tag{8}$$

Of course, in such cases where only two variables are considered endogenous in the earning equation, a subset of two elements of z_i may be used while preserving the just-identified nature of the system. This line of reasoning carries over to the case of only one element of (b_i, h_{1i}, h_{2i}) being jointly endogenous in the earning equation.

Since the combinations of distinct sets of endogenous variables and admissible subsets of instruments which yield identification or overidentification is large (there are 34 such combinations for each gender) Table 3 and Table 4 provide a flavor of the first stage reduced forms: in general, the parental characteristic for which the null of no effect in the first stage reduced form (conditional on other covariates) is soundly rejected is that corresponding to the same variable for the offspring. That is, parental BMI is a statistically significant determinant of BMI, but not of self-reported health or beauty. The same goes for parental beauty and parental self-reported health. This characteristic of the IVs narrows down admissible specifications considerably, although the results presented in Table 3 and Table 4 display associated partial F-statistics that are all above the usual minimum Rule of Thumb levels.

Our procedure is sequential. First, we consider just-identified systems in which each potentially endogenous RHS variable in the earnings equation is instrumented by the corresponding parental characteristics. For example, if the jointly endogenous RHS variables are beauty and BMI, the corresponding excluded IVs that appear in the two first-stage reduced forms are parental beauty and parental BMI. Second, after performing the corresponding IV procedure, we carry out the Durbin-Wu-Hausman test whose null corresponds to the exogeneity of the potentially endogenous variables. This means, if the p-value associated with the test is low, that the null of exogeneity is rejected. Third, and whether exogeneity is rejected or not, we then re-estimate over the subsample of individuals who have selected into a beauty-related occupation while augmenting the specification with the hazard rate from the appropriate logit specification (which in this running example will include own self-reported health, parental beauty and parental BMI), to control for selection bias. The underlying selection equations are presented in Table 9 and Table 10. Once again, we perform the appropriate Durbin-Wu-Hausman test of the null of exogeneity. Fourth, if the null of exogeneity is not rejected once selection is accounted for, we then re-estimate the selection specification while allowing BMI and Beauty to be exogenous.

Results for this sequential procedure are presented in Table 5 to Table 8. A typical example is given by females in Table 7, in which parental beauty and parental self-reported health are the exclusion restrictions used both to potentially control for endogeneity and selection. In column (2), we estimate the equation by 2SLS for females. The point estimate associated with beauty is equal to 0.185 while that associated with self-reported health is 0.887: both of these are estimated quite precisely. Moreover, the marginal impact of moving from poor to good self-reported health is large, and is equivalent to an 89% increase in hourly labor earnings. We note that the Hausman test rejects the null of exogeneity of beauty and health. In column (4) we re-estimate over the subsample of women working in beauty-related occupations while controlling for selection. The point estimate of the impact of beauty increases to 0.327, and is still marginally significant. The point estimate associated with the impact of good self-reported health is, for its part, not statistically significant. The Hausman test now marginally rejects (at the 6 percent level) while there is no statistically significant evidence of selection bias, leading us to prefer the initial estimates of column (2) which control solely for endogeneity. For argument's sake, we also (in column 6) estimate while controlling for selection but not for endogeneity. The point estimate of the marginal impact of beauty on wages falls to 0.145, while the coefficient associated with self-reported health gains the statistical significance again, but falls to 0.186. This example highlights how failure to appropriately control for the endogeneity of beauty and health can lead to a gross underestimation of the marginal impact of both of these personal attributes on hourly earnings. This pattern holds, roughly speaking, for the three other female specifications considered in Tables Table 5, Table 6 and Table 8. For males, in contrast, we do not find any evidence that beauty affects hourly labor income.

4 Longitudinal evidence

Cross sectional estimates could be biased by time-invariant unobservables, so we construct a longitudinal sample to reestimate the impact of beauty on individual earnings. Our

Mincerian earning equation in the longitudinal context is given by:

$$w_{it} = b_{it}\beta + h_{it}\theta + \mu_i + \eta_{it}, \tag{9}$$

where w_{it} (after the inverse hyperbolic sine transformation) is annual labor earnings, b_{it} is the measure of beauty, h_{it} represents BMI and self-reported health, μ_i is unobserved individual specific heterogeneity, and η_{it} represents idiosyncratic unobservables. The potential IVs, denoted by z_{it}, are still measures of parental beauty, parental BMI and parental self-reported health.

To estimate longitudinal model with both endogenous explanatory variables and sample selection, Semykina and Wooldridge (2010) proposes a test for selection bias and the estimation procedures that correct for the bias while controlling for the endogeneity. The test for selection bias is carried out in three steps:

(1) For each period t, use probit to estimate the selection equation:

$$P(d_{it} = 1|z_i) = \Phi(z_{it}\delta_t + \bar{z}_i\xi_t), \tag{10}$$

where $\bar{z}_i$ represents the time specific means of instrumental variables. Then use the estimates to calculate the inverse Mills ratio $\hat{\lambda}_{it} \equiv \lambda(z_{it}\hat{\delta}_t + \bar{z}_i\hat{\xi}_t)$.

(2) For the selected sample $d_{it} = 1$, use Fixed Effect-Two Stage Least Square (FE-2SLS) to estimate the income equation including $\hat{\lambda}_{it}$ as an additional control:

$$w_{it} = b_{it}\beta + h_{it}\theta + \mu_i + \hat{\lambda}_{it}\rho + \eta_{it}, \tag{11}$$

use $\hat{\lambda}_{it}$ and z_{it1} ($z_{it1} \subset z_{it}$ and the dimension of z_{it1} is not less than the dimension of endogenous explanatory variables) as instruments. One can also add the interactions of $\hat{\lambda}_{it}$ with time dummies to allow ρ to be time varying.

(3) Use the t-statistics to test $H_0 : \rho = 0$. If we cannot reject the null hypothesis, which means that there is no selection issue, the FE-2SLS estimator of Equation 9 is consistent.

If we reject the null hypothesis, we can implement the following procedures to correct the selection bias:

(1) For each period t, use probit to estimate the selection equation $P(d_{it} = 1|z_i) = \Phi(z_{it}\delta_t + \bar{z}_i\xi_t)$. Then use the estimates to calculate the inverse Mills ratio $\hat{\lambda}_{it}$.

(2) For the selected sample $d_{it} = 1$, use pooled 2SLS to estimate the earning equation:

$$w_{it} = b_{it}\beta + h_{it}\theta + \bar{z}_i\varphi + \hat{\lambda}_{it}\gamma + \eta_{it}, \tag{12}$$

use z_{it1}, $\bar{z}_i$, $\hat{\lambda}_{it}$ as instruments. One can also add the interactions of $\hat{\lambda}_{it}$ with time dummies to allow γ to be time varying.

Notice that in the cross sectional model, we can use the same set of identifying instruments for both the selection equation and the income equation, but in the longitudinal model, z_{it1}, the instruments used in the income equation, must be a subset of z_{it}, the instruments used in the selection equation. Therefore, when estimating the selection equation we need at least two variables from z_{it}, and in the income equation we can at most allow two variables to be endogenous. Table 11 and Table 12 report the results of the selection equation estimated by probit model.

We only consider the just-identification case when individual beauty is the only endogenous explanatory variable in the income equation. The over-identification specification does not give us any results that are fundamentally different. Table 13 presents the results for the female sample when individual beauty is instrumented by parental beauty in the income equation and both parental beauty and parental BMI are used in the selection equation. BMI is excluded in the income equation in column (1), column (3) and column (5). In column (2), column (4) and column (6), we include BMI and take it to be exogenous in the earning equation. In column (1) - Column (4), we test the existence of selection bias by estimating Equation 11. z_{it} is parental beauty and BMI and z_{it1} is parental beauty. The coefficients of inverse Mills ratio (IMR) are statistically indistinguishable from zero, whether we take it time invariant or not. This implies that there is no selection bias and the FE-2SLS estimates of Equation 9 in column (5) and column (6) are consistent. The effect of beauty is huge: one standard deviation increase in beauty score will increase annual income by 40 percent ($0.95 \times 0.422 = 0.401$). BMI has no significant impact on annual income, though. The Hausman test in column (5) and column (6) rejects the null of exogeneity of beauty.

Table 14 reports the results for the male sample. We see the similar pattern hold for the males. None of the coefficients of inverse Mills ratio is statistically significant, so we estimate the earning equation without controlling for the occupational selection. Compared to the results from the female sample, the effect of beauty on income is also statistically significant for males, but the economic significance is smaller. In column (5), one standard deviation increase in beauty score will raise annual income by $1.02 \times 0.256 = 0.261 = 26.1\%$. In column (6) we add BMI to the right-hand-side of the earning equation. The coefficient of BMI is statistically significant and positive: one standard deviation increase in BMI will increase men's annual income by 28%. The economic magnitude of beauty drops slightly from 0.256 to 2.242 when BMI is added. Hausman test still rejects the exogeneity of beauty in both columns.

Table 15 and Table 16 report the results when parental BMI is replaced by parental self-reported health. For females, the pattern from Table 13 still holds: the coefficients of inverse Mills ratio are not statistically significant, so there is no selection bias. Then we estimate the earning equation taking self-reported health to be exogenous and instrumenting individual beauty with parental beauty. For males, the results are different from what we have in Table 14. In column (1) and column (2), when inverse Mills ratio is time invariant, its coefficients are marginally significant, which implies that there exists some extent of selection bias, Therefore, we carry out the procedures for correcting the selection bias in column (5) and column (6). We estimates Equation 12, using pooled 2SLS and controlling for the inverse Mills ratio and the time-specific means of parental characteristics. The effect of beauty on men's annual income now becomes negative once we control for the occupational selection. Since the coefficients of inverse Mills ratio are only marginally statistically significant, in column (7) and column (8) we still estimate the earning equation using FE-2SLS and do not control for the occupational selection. The economic significance of the impact of beauty on annual income is similar to that in Table 14 and the impact of self-reported health is highly significant: moving from poor to good self-reported health will increase annual income by $0.42 \times 0.437 = 0.184 = 18.4\%$.

Finally, we consider the case when all three parental characteristics are used in the selection equation. Table 17 and Table 18 report the results. Now for both subsamples, we cannot reject the null hypothesis of the test of selection bias: the coefficients of inverse Mills ratio are all statistically indistinguishable from zero. Hence, we just need to estimate the earning equation controlling solely for the endogeneity of individual beauty. The results are again similar from what we have in the previous tables. For both females and males, the impact of beauty is significantly positive. The impact for males is smaller than for females. When we add BMI and self-reported health in the list of covariates, one standard deviation increase in the individual beauty will increase annual income by $0.95 \times 0.374 = 0.355 = 35.5\%$ for female and by $1.02 \times 0.233 = 0.238 = 23.8\%$. BMI does not have any significant impact on annual income for females but does for males. One standard deviation increase in BMI will increase men's annual income by $2.96 \times 0.076 = 0.225 = 22.5\%$. The impacts of self-reported health on annual income are statistically significant for both females and males.

5 Conclusion

We estimate the impact of beauty on individual earnings in China in both cross sectional and longitudinal contexts. We tackle simultaneously the endogeneity of beauty and the occupational selection for the first time in the beauty literature. Using the 2010, 2012 and 2014 waves of CFPS, we match all the "parent-adult child" pairs and use parental characteristics as potential instrumental variables. For the cross sectional specification, we control for occupational selection by adding the hazard rate as additional control. In longitudinal model, we carry out the test of selection bias and correcting procedures proposed by Semykina and Wooldridge (2010). We do not find strong evidence of the existence of selection bias in both cross sectional and longitudinal specifications. The exogeneity of beauty is rejected using the appropriate Hausman test. The cross sectional results show that beauty is associated with higher earning for females: one standard deviation increase in beauty is equivalent to a seven year increase in the number of years schooling. Higher beauty score will bring higher earning for both genders in the longitudinal estimation: one standard deviation increase in beauty will lead to roughly 40% increase in annual income for female and at least 20% increase for males.

References

Biddle, Jeff E. and Daniel S. Hamermesh, "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, 1998, *16* (1), 172–201.

Doorley, Karina and Eva Sierminska, "Myth or Fact? the Beauty Premium across the Wage Distribution in Germany," *Economics Letters*, 2015, *129*, 29–34.

Fletcher, Jason M, "Beauty Vs. Brains: Early Labor Market Outcomes of High School Graduates," *Economics Letters*, 2009, *105* (3), 321–325.

Hamermesh, Daniel S. and Jason Abrevaya, "Beauty Is the Promise of Happiness?," *European Economic Review*, 2013, *64*, 351–368.

_ **and Jeff E. Biddle**, "Beauty and the Labour Market," *American Economic Review*, 1994, *84* (5), 1174–1194.

Harper, Barry, "Beauty, Stature and the Labour Market: a British Cohort Study," *Oxford Bulletin of Economics and Statistics*, 2000, *62* (s1), 771–800.

Johnston, David W, "Physical Appearance and Wages: Do Blondes Have More Fun?," *Economics Letters*, 2010, *108* (1), 10–12.

Oreffice, Sonia and Climent Quintana-Domeque, "Beauty, Body Size and Wages: Evidence from a Unique Data Set," *Economics and Human Biology*, 2016, *22* (1), 24–34.

Pfann, Gerard A, Jeff E Biddle, Daniel S Hamermesh, and Ciska M Bosman, "Business Success and Businesses Beauty Capital," *Economics Letters*, 2000, *67* (2), 201–207.

Salter, Sean P., Franklin G. Mixon Jr., and Ernest W. King, "Broker Beauty and Boon: A Study of Physical Attractiveness and Its Effect on Real Estate Brokers Income and Productivity," *Applied Financial Economics*, 2012, *22* (10), 811–825.

Scholz, John Karl and Kamil Sicinski, "Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study," *Review of Economics and Statistics*, 2015, *97* (1), 14–28.

Semykina, Anastasia and Jeffrey M. Wooldridge, "Estimating Panel Data Models in the Presence of Endogeneity and Selection," *Journal of Econometrics*, 2010, *157* (2), 375–380.

Thomas, Duncan and John Strauss, "Health and Wages: Evidence on Men and Women in Urban Brazil," *Journal of Econometrics*, 1997, *77* (1), 159–185.

Tables

Table 1: Summary Statistics: CFPS 2014

	Standard deviation	Min	Median	Max
Female				
Hourly wage	16.92	0.002	11.69	210.00
Beauty	0.97	3	6	7
BMI	2.81	10.16	20.31	32.06
Self-reported health	0.50	0	0	1
Community type (urban=1)	0.47	0	1	1
Years of schooling	3.40	0	12	19
Hukou type (Nonagriculture=1)	0.49	0	0	1
Age	7.27	18	25	68
Parental beauty	1.05	2	5.50	7
Parental BMI	2.61	13.06	23.44	36.73
Parental self-reported health	0.37	0	0	1
Male				
Hourly wage	100.35	0.002	13.75	4230.77
Beauty	1.02	1	6	7
BMI	3.41	11.25	22.86	48.31
Self-reported health	0.50	0	1	1
Community type (urban=1)	0.49	0	1	1
Years of schooling	3.51	0	9	19
Hukou type (Nonagriculture=1)	0.47	0	0	1
Age	8.87	18	30	67
Parental beauty	1.16	1	5.50	7
Parental BMI	3.05	11.72	22.94	38.46
Parental self-reported health	0.35	0	0	1

Note: This table reports the descriptive statistics for the female and the male sample of 2014's wave of CFPS. We restrict our sample according to the following rules: age *ge* 18, weekly working hours ≤ 150, and income > 0. 37,147 individuals are surveyed in the original sample. We are able to identify "parent / adult child" pairs for 2,951 individuals, among whom there are 806 women and 2,145 men. The unit for wage is Chinese Yuan.

Table 2: Summary Statistics: CFPS (2010/2012/2014)

	S.D	S.D (between)	S.D (within)	Min	Median	Max
Female						
Annual income	16129.53	12741.40	9925.86	10	20000	120000
Beauty	0.95	0.66	0.68	3	6	7
BMI	2.96	2.67	1.28	15.33	20.70	33.87
Self-reported health	0.43	0.29	0.32	0	1	1
Parental beauty	1.06	0.79	0.70	2	5	7
Parental BMI	2.56	2.31	1.11	13.33	23.44	31.18
Parental self-reported health	0.41	0.30	0.27	0	0.5	1
Male						
Annual income	22502.12	17978.54	13548.64	1	23000	230000
Beauty	1.02	0.70	0.74	2	6	7
BMI	3.42	3.15	1.35	10	22.86	43.25
Self-reported health	0.42	0.29	0.31	0	1	1
Parental beauty	1.15	0.84	0.78	1	5	7
Parental BMI	2.79	2.52	1.21	14.60	22.89	37.28
Parental self-reported health	0.42	0.30	0.29	0	0.5	1

Note: This table reports the descriptive statistics for the female and the male sample of the longitudinal sample (CFPS 2010/2012/2014). We restrict our sample according to the following rules: age ≥ 18 and annual gross income > 0. 635 individuals (154 women and 481 men) and at least one of their parents are successfully followed across all three waves. The unit for wage is Chinese Yuan.

Table 3: First Stage: CFPS 2014 (Female)

	(1) Beauty	(2) Beauty	(3) Beauty	(4) Beauty	(5) BMI	(6) BMI	(7) Health	(8) Health
Parental beauty	0.500***	0.500***	0.499***	0.499***	-0.022	-0.018	-0.008	-0.008
	(0.028)	(0.028)	(0.028)	(0.028)	(0.088)	(0.089)	(0.017)	(0.017)
Parental BMI		0.001		0.001	0.304***	0.303***		-0.004
		(0.011)		(0.011)	(0.036)	(0.036)		(0.007)
Parental health			0.029	0.029		-0.100	0.276***	0.274***
			(0.079)	(0.079)		(0.251)	(0.047)	(0.047)
Community type (urban = 1)	-0.084	-0.084	-0.083	-0.084	0.432*	0.430*	-0.041	-0.039
	(0.068)	(0.069)	(0.068)	(0.069)	(0.217)	(0.217)	(0.041)	(0.041)
Years of schooling	0.013	0.013	0.013	0.013	-0.129***	-0.130***	-0.001	-0.001
	(0.010)	(0.010)	(0.010)	(0.010)	(0.031)	(0.031)	(0.006)	(0.006)
Residence type (nonagriculture = 1)	0.049	0.049	0.050	0.050	0.035	0.032	0.004	0.004
	(0.072)	(0.072)	(0.072)	(0.072)	(0.227)	(0.227)	(0.043)	(0.043)
Age	0.014	0.013	0.014	0.014	0.187**	0.186**	-0.026	-0.025
	(0.022)	(0.022)	(0.022)	(0.022)	(0.071)	(0.071)	(0.013)	(0.013)
Square of age	-0.000	-0.000	-0.000	-0.000	-0.001	-0.001	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	806	806	806	806	806	806	806	806
R^2	0.305	0.305	0.305	0.305	0.167	0.167	0.076	0.076
F-statistics	58.471	50.057	50.082	43.769	22.824	19.969	9.351	8.230

Note: This table reports results of the first stage regressions for the female sample from 2014's wave of CFPS. Column (1) is the just identification case where the parental beauty is the unique exclusion restriction. Column (2) and column (5) represent the case where parental beauty score and BMI are used as instruments. In column (3) and column (7) parental beauty score and parental self-reported health are the instruments. Column (4), column (6), and column (8) are the results when all three are used as instruments. If the two parents of certain individual are both surveyed, we take the mean for parental characteristics; otherwise, either father's or mother's characteristics are used. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: First Stage: CFPS 2014 (Male)

	(1) beauty	(2) beauty	(3) beauty	(4) beauty	(5) bmi	(6) bmi	(7) healthy	(8) healthy
Parental beauty	0.428***	0.428***	0.429***	0.429***	0.064	0.058	0.008	0.008
	(0.017)	(0.017)	(0.017)	(0.017)	(0.061)	(0.062)	(0.009)	(0.009)
Parental BMI		-0.005		-0.005	0.288***	0.288***		-0.001
		(0.006)		(0.006)	(0.023)	(0.023)		(0.004)
Parental health			-0.034	-0.033		0.214	0.178***	0.178***
			(0.056)	(0.056)		(0.203)	(0.030)	(0.030)
Community type (urban = 1)	0.020	0.021	0.019	0.020	0.239	0.248	-0.075**	-0.075**
	(0.043)	(0.043)	(0.043)	(0.043)	(0.156)	(0.156)	(0.023)	(0.023)
Years of schooling	0.032***	0.032***	0.032***	0.032***	0.023	0.023	0.002	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.023)	(0.023)	(0.003)	(0.003)
Residence type (nonagriculture = 1)	-0.016	-0.012	-0.015	-0.012	0.368*	0.366*	-0.077**	-0.076**
	(0.048)	(0.049)	(0.048)	(0.049)	(0.178)	(0.178)	(0.026)	(0.027)
Age	0.010	0.010	0.009	0.009	0.304***	0.310***	-0.027***	-0.027***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.048)	(0.048)	(0.007)	(0.007)
Square of age	-0.000	-0.000	-0.000	-0.000	-0.003***	-0.003***	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	2145	2145	2145	2145	2145	2145	2145	2145
R^2	0.267	0.267	0.267	0.267	0.118	0.119	0.086	0.086
F-statistics	129.714	111.234	111.204	97.344	40.920	35.946	28.572	24.999

Note: This table reports results of the first stage regressions for the male sample from 2014's wave of CFPS. Column (1) is the just identification case where the parental beauty is the unique exclusion restriction. Column (2) and column (5) represent the case where parental beauty score and BMI are used as instruments. In column (3) and column (7) parental beauty score and parental self-reported health are the instruments. Column (4), column (6), and column (8) are the results when all three are used as instruments. If the two parents of certain individual are both surveyed, we take the mean for parental characteristics; otherwise, either father's or mother's characteristics are used. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5: Wage equation: IV (parental beauty)

	Dependent variable: hourly wage					
	2SLS (no selection)		2SLS (with hazard rate)		OLS (with hazard rate)	
	male	female	male	female	male	female
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.026	0.192***	-0.067	0.408*	-0.026	0.121**
	(0.037)	(0.058)	(0.100)	(0.171)	(0.047)	(0.040)
Community type	0.060	0.006	0.185	0.204*	0.183	0.159
(urban = 1)	(0.049)	(0.061)	(0.150)	(0.091)	(0.151)	(0.081)
Years of schooling	0.034***	0.027**	0.064**	0.034**	0.064**	0.039***
	(0.006)	(0.009)	(0.024)	(0.013)	(0.025)	(0.011)
Residence type	0.036	0.011	-0.052	0.658	-0.051	0.080
(nonagriculture = 1)	(0.048)	(0.060)	(0.165)	(0.353)	(0.165)	(0.187)
Age	0.099***	0.135***	0.045	0.142*	0.042	0.104
	(0.016)	(0.034)	(0.049)	(0.057)	(0.049)	(0.054)
Square of age	-0.001***	-0.002***	-0.000	-0.003**	-0.000	-0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hazard rate			-0.040	-1.606*	-0.054	-0.277
			(0.696)	(0.798)	(0.693)	(0.400)
Test of exogeneity	0.043	6.262	0.218	2.981		
(p-value)	(0.836)	(0.013)	(0.641)	(0.085)		
Observations	2145	806	287	373	287	373

Note: This table reports results of the 2SLS estimation of wage equation using the sample from 2014's wave of CFPS. Parental beauty is the only instrument variable. Column (1) and column (2) are the results of 2SLS estimations without controlling the hazard rate. In column (3) and column (4), we control hazard rate and the sample only contains individuals who choose the occupations where beauty could be productive. Result in column (3) and column (4) shows that we cannot reject the hypothesis that beauty is exogenous, we estimate the wage equation again by OLS in column (5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6: Wage equation: IV (parental beauty, parental BMI)

	Dependent variable: hourly wage					
	2SLS (no selection)		2SLS (with hazard rate)		OLS (with hazard rate)	
	male	female	male	female	male	female
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.029	0.189***	-0.043	0.395*	-0.026	0.120**
	(0.038)	(0.057)	(0.106)	(0.182)	(0.046)	(0.043)
BMI	-0.009	-0.036	-0.033	0.014	-0.003	-0.008
	(0.021)	(0.036)	(0.049)	(0.062)	(0.013)	(0.013)
Community type	0.062	0.027	0.214	0.188*	0.183	0.160
(urban = 1)	(0.049)	(0.062)	(0.149)	(0.086)	(0.131)	(0.083)
Years of schooling	0.034***	0.023*	0.073*	0.035**	0.064**	0.039**
	(0.006)	(0.010)	(0.029)	(0.012)	(0.022)	(0.013)
Residence type	0.041	0.012	-0.009	0.633	-0.052	0.082
(nonagriculture = 1)	(0.049)	(0.060)	(0.171)	(0.380)	(0.165)	(0.196)
Age	0.102***	0.143***	0.042	0.136*	0.042	0.105*
	(0.017)	(0.033)	(0.050)	(0.056)	(0.051)	(0.043)
Square of age	-0.001***	-0.002***	-0.000	-0.003*	-0.000	-0.002*
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hazard rate			-0.238	-1.535	-0.045	-0.279
			(0.728)	(0.872)	(0.684)	(0.385)
Test of exogeneity	0.119	3.245	0.277	1.277		
(p-value)	(0.888)	(0.040)	(0.758)	(0.281)		
Observations	2145	806	287	373	287	373

Note: This table reports results of the 2SLS estimation of wage equation using the sample from 2014's wave of CFPS. We add BMI in the equation and use parental beauty and parental BMI as instrument variables. Column (1) and column (2) are the results of 2SLS estimations without controlling the hazard rate. In column (3) and column (4), we control hazard rate and the sample only contains individuals who choose the occupations where beauty could be productive. Result in column (3) and column (4) shows that we cannot reject the hypothesis that beauty is exogenous, we estimate the wage equation again by OLS in column (5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7: Wage equation: IV (parental beauty, parental health)

	Dependent variable: hourly wage					
	2SLS (no selection)		2SLS (with hazard rate)		OLS (with hazard rate)	
	male	female	male	female	male	female
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.032	0.185**	-0.070	0.327*	-0.031	0.145***
	(0.038)	(0.061)	(0.100)	(0.154)	(0.047)	(0.035)
Self-report health	-0.198	0.887**	-0.172	0.481	0.113	0.186**
	(0.312)	(0.296)	(0.725)	(0.470)	(0.091)	(0.070)
Community type	0.043	0.048	0.173	0.220*	0.168	0.179*
(urban = 1)	(0.054)	(0.071)	(0.151)	(0.093)	(0.149)	(0.081)
Years of schooling	0.034***	0.029**	0.060*	0.044**	0.060*	0.041***
	(0.006)	(0.010)	(0.024)	(0.014)	(0.024)	(0.011)
Residence type	0.021	0.015	-0.086	0.464	-0.076	0.259
(nonagriculture = 1)	(0.054)	(0.068)	(0.166)	(0.309)	(0.162)	(0.151)
Age	0.093***	0.160***	0.031	0.124*	0.048	0.115*
	(0.018)	(0.032)	(0.069)	(0.056)	(0.049)	(0.053)
Square of age	-0.001***	-0.002***	-0.000	-0.002*	-0.000	-0.002*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Hazard rate			0.124	-1.179	0.092	-0.675*
			(0.678)	(0.675)	(0.670)	(0.322)
Test of exogeneity	0.497	7.338	0.181	2.741		
(p-value)	(0.608)	(0.000)	(0.835)	(0.066)		
Observations	2145	806	287	373	287	373

Note: This table reports results of the 2SLS estimation of wage equation using the sample from 2014's wave of CFPS. We use the dummy on self-reported health parental in addtion to beauty, so parental health is added in the instrument variables. Column (1) and column (2) are the results of 2SLS estimations without controlling the hazard rate. In column (3) and column (4), we control hazard rate and the sample only contains individuals who choose the occupations where beauty could be productive. Result in column (3) and column (4) shows that we cannot reject the hypothesis that beauty is exogenous, we estimate the wage equation again by OLS in column (5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8: Wage equation: IV (parental beauty, parental BMI, parental health)

	Dependent variable: hourly wage					
	2SLS (no selection)		2SLS (with hazard rate)		OLS (with hazard rate)	
	male	female	male	female	male	female
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.034	0.183**	-0.049	0.336*	-0.031	0.142***
	(0.038)	(0.060)	(0.107)	(0.152)	(0.047)	(0.036)
BMI	-0.009	-0.018	-0.029	0.024	-0.002	-0.006
	(0.021)	(0.038)	(0.056)	(0.053)	(0.013)	(0.012)
Self-reported health	-0.183	0.874**	-0.105	0.431	0.113	0.185**
	(0.312)	(0.304)	(0.779)	(0.454)	(0.091)	(0.071)
Community type (urban = 1)	0.047	0.058	0.200	0.201*	0.168	0.177*
	(0.055)	(0.070)	(0.151)	(0.088)	(0.150)	(0.080)
Years of schooling	0.034***	0.027**	0.068*	0.045**	0.061*	0.041***
	(0.006)	(0.011)	(0.029)	(0.014)	(0.024)	(0.011)
Residence type (nonagriculture = 1)	0.028	0.015	-0.043	0.483	-0.077	0.249
	(0.055)	(0.068)	(0.177)	(0.303)	(0.168)	(0.147)
Age	0.096***	0.164***	0.034	0.121*	0.048	0.115*
	(0.019)	(0.032)	(0.070)	(0.055)	(0.049)	(0.052)
Square of age	-0.001***	-0.002***	-0.000	-0.002*	-0.000	-0.002*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Hazard rate			-0.070	-1.210	0.097	-0.649*
			(0.725)	(0.669)	(0.687)	(0.304)
Test of exogeneity (p-value)	0.389	4.818	0.256	1.824		
	(0.761)	(0.003)	(0.857)	(0.143)		
Observations	2145	806	287	373	287	373

Note: This table reports results of the 2SLS estimation of wage equation using the sample from 2014's wave of CFPS. Both BMI and self-reported health are added to the covariates. Column (1) and column (2) are the results of 2SLS estimations without controlling the hazard rate. In column (3) and column (4), we control hazard rate and the sample only contains individuals who choose the occupations where beauty could be productive. Result in column (3) and column (4) shows that we cannot reject the hypothesis that beauty is exogenous, we estimate the wage equation again by OLS in column (5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 9: Selection equation: CFPS 2014 (Female)

	Dependent variable: "beauty" occupation			
	(1)	(2)	(3)	(4)
Parental beauty	0.100	0.100		0.108
	(0.070)	(0.070)		(0.070)
Parental BMI		0.011	0.009	0.010
		(0.028)	(0.028)	(0.028)
Parental health			-0.133	-0.170
			(0.197)	(0.198)
Community type	0.005	-0.001	0.008	-0.004
(urban = 1)	(0.171)	(0.171)	(0.171)	(0.172)
Years of schooling	-0.002	-0.003	-0.003	-0.004
	(0.024)	(0.024)	(0.024)	(0.024)
Residence type	0.453*	0.453*	0.473**	0.449*
(nonagriculture = 1)	(0.179)	(0.179)	(0.179)	(0.180)
Age	0.060	0.058	0.053	0.057
	(0.072)	(0.072)	(0.072)	(0.072)
Square of age	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	806	806	806	806

Note: This table reports results of the selection equation estimated by Logit model using the female sample from 2014's wave of CFPS. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10: Selection equation: CFPS 2014 (Male)

	Dependent variable: "beauty" occupation			
	(1)	(2)	(3)	(4)
Parental beauty	0.015	0.018		0.022
	(0.060)	(0.060)		(0.061)
Parental BMI		-0.015	-0.014	-0.015
		(0.024)	(0.023)	(0.024)
Parental health			-0.115	-0.123
			(0.196)	(0.197)
Community type	0.536**	0.539**	0.536**	0.533**
(urban = 1)	(0.165)	(0.165)	(0.165)	(0.166)
Years of schooling	0.103***	0.103***	0.103***	0.103***
	(0.023)	(0.023)	(0.023)	(0.023)
Residence type	0.777***	0.788***	0.793***	0.789***
(nonagriculture = 1)	(0.160)	(0.161)	(0.161)	(0.161)
Age	0.085	0.085	0.082	0.082
	(0.064)	(0.065)	(0.065)	(0.065)
Square of age	-0.002	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	2145	2145	2145	2145

Note: This table reports results of the selection equation estimated by Logit model using the male sample from 2014's wave of CFPS. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 11: Selection equation: CFPS Panel (Female)

	Dependent variable: "beauty" occupation		
	(1)	(2)	(3)
Parental beauty	0.109	0.115	0.110
	(0.088)	(0.088)	(0.089)
Parental BMI	0.076		0.075
	(0.056)		(0.056)
Parental health		0.073	0.026
		(0.224)	(0.227)
Parental beauty (mean)	-0.095	-0.119	-0.107
	(0.117)	(0.118)	(0.118)
Parental BMI (mean)	-0.050		-0.053
	(0.062)		(0.063)
Parental health (mean)		0.243	0.276
		(0.301)	(0.304)
Individuals	154	154	154

Note: This table reports results of the selection equation estimated by Probit model using the female sample from CFPS 2010/2012/2014. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 12: Selection equation: CFPS Panel (Male)

	Dependent variable: "beauty" occupation		
	(1)	(2)	(3)
Parental beauty	0.101	0.097	0.098
	(0.059)	(0.059)	(0.059)
Parental BMI	0.030		0.029
	(0.038)		(0.038)
Parental health		0.101	0.095
		(0.161)	(0.161)
Parental beauty (mean)	0.082	0.100	0.093
	(0.081)	(0.082)	(0.082)
Parental BMI (mean)	0.003		0.005
	(0.042)		(0.042)
Parental health (mean)		-0.167	-0.180
		(0.224)	(0.224)
Individuals	481	481	481

Note: This table reports results of the selection equation estimated by Probit model using the male sample from CFPS 2010/2012/2014. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 13: Wage equation: CFPS Female Panel (z_{it}: parental beauty, parental BMI; z_{it1}: parental beauty)

	Dependent variable: annual income					
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.018	0.017	-0.091	-0.092	0.422***	0.419***
	(0.908)	(0.913)	(0.825)	(0.834)	(0.122)	(0.124)
BMI		0.030		-0.009		0.006
		(0.073)		(0.070)		(0.033)
IMR	-0.120	-0.026				
	(1.936)	(2.015)				
IMR1			0.206	0.186		
			(1.734)	(1.787)		
IMR2			0.729	0.710		
			(1.858)	(1.911)		
IMR3			0.719	0.705		
			(1.801)	(1.842)		
Test of exogeneity	0.020	0.019	0.002	0.002	10.063	9.823
(p-value)	(0.886)	(0.889)	(0.962)	(0.962)	(0.002)	(0.002)
Individuals	51	51	51	51	154	154

Note: This table reports the tests of selection bias for female sample of CFPS 2010/2012/2014 using FE-2SLS. z_{it} used in the selection equation is parental beauty and parental BMI. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add BMI in the covariates, taking it exogenous. Since the coefficients of inverse Mills ratio are statistically insignificant in column (1)-column (4), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity in column(5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 14: Wage equation: CFPS Male Panel (z_{it}: parental beauty, parental BMI; z_{it1}: parental beauty)

	Dependent variable: annual income					
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.672	0.632	0.410	0.425	0.256***	0.242***
	(0.761)	(0.716)	(0.555)	(0.567)	(0.058)	(0.057)
BMI		0.161		0.058		0.082***
		(0.098)		(0.081)		(0.017)
IMR	1.142	1.363				
	(5.304)	(5.162)				
IMR1			0.946	1.090		
			(4.241)	(4.340)		
IMR2			1.479	1.599		
			(4.306)	(4.394)		
IMR3			1.655	1.744		
			(4.378)	(4.451)		
Test of exogeneity	3.829	3.345	2.204	2.158	13.256	11.756
(p-value)	(0.050)	(0.067)	(0.138)	(0.142)	(0.000)	(0.001)
Individuals	37	37	37	37	481	481

Note: This table reports the tests of selection bias for male sample of CFPS 2010/2012/2014 using FE-2SLS. z_{it} used in the selection equation is parental beauty and parental BMI. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add BMI in the covariates, taking it exogenous. Since the coefficients of inverse Mills ratio are statistically insignificant in column (1)-column (4), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity in column(5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 15: Wage equation: CFPS Female Panel (z_{it}: parental beauty, parental self-reported health; z_{it1}: parental beauty)

	Dependent variable: annual income					
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	-3.508	-3.785	-3.042	-3.155	0.422***	0.379**
	(2.422)	(2.896)	(2.160)	(2.331)	(0.122)	(0.120)
Self-reported health		-0.566		-0.477		0.205
		(1.026)		(0.818)		(0.113)
IMR	-10.642	-11.825				
	(7.435)	(9.098)				
IMR1			-8.720	-9.268		
			(6.775)	(7.405)		
IMR2			-8.832	-9.325		
			(7.164)	(7.772)		
IMR3			-8.525	-9.005		
			(7.015)	(7.602)		
Test of exogeneity	17.387	16.859	15.563	15.510	10.063	8.071
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.005)
Individuals	51	51	51	51	154	154

Note: This table reports the tests of selection bias for female sample of CFPS 2010/2012/2014. z_{it} used in the selection equation is parental beauty and parental self-reported health. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add self-reported health in the covariates, taking it exogenous. Since the coefficients of inverse Mills ratio are statistically insignificant in column (1)-column (4), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity in column(5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 16: Wage equation: CFPS Male Panel (z_{it}: parental beauty, parental self-reported health; z_{it1}: parental beauty)

	Dependent variable: annual income							
	FE-2SLS (1)	FE-2SLS (2)	FE-2SLS (3)	FE-2SLS (4)	2SLS (5)	2SLS (6)	FE-2SLS (7)	FE-2SLS (8)
Beauty	-0.569	-0.567	0.052	0.073	-1.396*	-1.344*	0.256***	0.245***
	(0.557)	(0.558)	(0.799)	(0.793)	(0.545)	(0.553)	(0.058)	(0.057)
Self-reported health		0.122		-0.119		0.150		0.437***
		(0.403)		(0.394)		(0.272)		(0.076)
IMR	-8.597*	-8.578*			-11.182**	-10.762**		
	(4.189)	(4.200)			(3.474)	(3.558)		
IMR1			-2.149	-1.930				
			(6.817)	(6.766)				
IMR2			-1.672	-1.435				
			(6.958)	(6.904)				
IMR3			-1.558	-1.318				
			(7.120)	(7.065)				
Parental beauty (mean)					-1.042**	-1.008**		
					(0.379)	(0.381)		
Parental health (mean)					0.315	0.270		
					(0.424)	(0.427)		
Test of exogeneity	0.000	0.000	0.303	0.345	5.131	4.408	13.256	13.742
(p-value)	(0.982)	(0.991)	(0.582)	(0.557)	(0.025)	(0.038)	(0.000)	(0.000)
Individuals	37	37	37	37	150	150	481	481

Note: This table reports the tests of selection bias for male sample of CFPS 2010/2012/2014. z_{it} used in the selection equation is parental beauty and parental self-reported health. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add self-reported in the covariates, taking it exogenous. In column (1) and column (2), the coefficients of inverse Mills ratio are statistically significant, which means that there exists selection bias. To correct the selection bias, in column(5) and column(6) we estimate wage equation $w_{it} = b_{it}\beta + h_{it}\theta + \bar{z}_i\varphi + \hat{\lambda}_{it}\gamma + \eta_{it}$ by 2SLS, using $z_{it1}, \bar{z}_i$ and $\hat{\lambda}_{it}$ as instruments. In column (7) and column (8), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity as if there was no selection bias. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 17: Wage equation: CFPS Female Panel (z_{it}: parental beauty, parental BMI, parental self-reported health; z_{it1}: parental beauty)

	Dependent variable: annual income					
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	-0.194	-0.156	-0.234	-0.229	0.422***	0.374**
	(0.847)	(0.867)	(0.773)	(0.797)	(0.122)	(0.121)
BMI		0.032		-0.011		0.010
		(0.074)		(0.071)		(0.032)
Self-reported health		0.290		0.054		0.295*
		(0.340)		(0.307)		(0.129)
IMR	-0.783	-0.375				
	(1.769)	(2.047)				
IMR1			-0.346	-0.323		
			(1.609)	(1.801)		
IMR2			0.150	0.166		
			(1.726)	(1.896)		
IMR3			0.141	0.163		
			(1.684)	(1.836)		
Test of exogeneity	0.008	0.007	0.055	0.052	10.063	7.793
(p-value)	(0.927)	(0.935)	(0.814)	(0.820)	(0.002)	(0.005)
Individuals	51	51	51	51	154	154

Note: This table reports the tests of selection bias for female sample of CFPS 2010/2012/2014. z_{it} used in the selection equation is parental beauty, parental BMI and parental self-reported health. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add BMI and self-reported health in the covariates, taking them exogenous. Since the coefficients of inverse Mills ratio are statistically insignificant in column (1)-column (4), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity in column(5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 18: Wage equation: CFPS Male Panel (z_{it}: parental beauty, parental BMI, parental self-reported health; z_{it1}: parental beauty)

	Dependent variable: annual income					
	(1)	(2)	(3)	(4)	(5)	(6)
Beauty	0.129	0.089	0.304	0.299	0.256***	0.233***
	(0.487)	(0.458)	(0.464)	(0.450)	(0.058)	(0.056)
BMI		0.112		0.061		0.076***
		(0.080)		(0.075)		(0.017)
Self-reported health		0.053		-0.224		0.416***
		(0.439)		(0.425)		(0.075)
IMR	-3.011	-2.948				
	(3.337)	(3.245)				
IMR1			0.113	0.145		
			(3.548)	(3.472)		
IMR2			0.640	0.673		
			(3.629)	(3.550)		
IMR3			0.803	0.796		
			(3.700)	(3.609)		
Test of exogeneity	2.037	1.515	1.932	1.900	13.256	12.297
(p-value)	(0.154)	(0.218)	(0.165)	(0.168)	(0.000)	(0.001)
Individuals	37	37	37	37	481	481

Note: This table reports the tests of selection bias for male sample of CFPS 2010/2012/2014. z_{it} used in the selection equation is parental beauty, parental BMI and parental self-reported health. The IV (z_{it1}) used in the wage equation is parental beauty. In column (1) and column (2), inverse Mills ratio (IMR) is time invariant. In column (3) and column (4), we add three interaction terms of inverse Mills ratio and time dummies (2010, 2012, 2014). In column (2) and column (4), we add BMI and self-reported health in the covariates, taking them exogenous. Since the coefficients of inverse Mills ratio are statistically insignificant in column (1)-column (4), we reestimate the wage equation using FE-2SLS, controlling only the endogeneity in column(5) and column(6). Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix A. Occupation Classification

All the occupations in the CFPS are coded using National Classification and Codes of Occupations (GB/T 6565 - 1999). The following occupations are classified as occupations for which physical attractiveness could be productive.

- 10548 Self-employed persons in sports and recreation services
- 10549 Stall owners in the marketplace
- 10550 Food stall owners on the street
- 10551 Other stall owners on the street
- 20600 Transactors
- 20605 International businessmen
- 20700 Financial personnel
- 20701 Banking personnel
- 20702 Insurance business personnel
- 20703 Security personnel
- 20709 Other financial personnel
- 21003 Actors
- 21005 Practitioners in film, television, radio and stage industry
- 21006 Fine art professionals
- 21007 Crafts and arts personnel
- 21204 Announcers and Presenters
- 21301 Religious professionals
- 30100 Administrative office staff
- 30101 Administrative personnel
- 30102 Administrative affairs personnel
- 40102 Promotional staff, exhibition staff and hostesses
- 40103 Purchasing staff
- 40104 Auctioneers, Pawnbrokers and Leasing Agents
- 40109 Other Salesman And Purchaser

- 40303 Bartenders and tea specialists
- 40305 Attendants in restaurants and dining halls
- 40400 Staff in hotels, tourist sites, sports and recreation services
- 40401 Attendants in inns and hotels
- 40402 Attendants in public places and tourist attractions
- 40403 Attendants in fitness centers and entertainment venues
- 40409 Other staff in hotels, tourist sites, sports and recreation services
- 40704 Staff in the hair and beauty industry