Journal of Choice Modelling 48 (2023) 100433

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

A cross-sectional exploration of labor supply, gender, and household wealth in urban China

Xuehui Han [a], Tao Zhang [b], John K. Dagsvik [c], Yuan Cheng [d,*]

[a] Fiscal Affair Department, The International Monetary Fund, United States
[b] Ragnar Frisch Centre for Economic Research, Norway
[c] Research Department, Statistics Norway, Norway
[d] School of Social Development and Public Policy, Fudan Development Institute, Fudan University, China

ARTICLE INFO

JEL classification:
C35
J20
J16
R30

Keywords:
Discrete choice theory
Regression tree
Labor supply
Gender
Wealth

ABSTRACT

We propose a modeling framework that uses only cross-sectional data to disentangle labor supply and demand choices simultaneously. This modeling framework extends the labor-market analytical toolkits to adapt to environments where data are limited, flexibility in working hours is lacking, or structural changes are present, as is the case in most emerging and low-income countries. We showcase our model by using the 2011 China Household Finance Survey to decipher labor market choices in urban China. We find that the main discrepancies in labor supply between males and females are driven by the number and age of children, the lower utility of working rather than fewer job opportunities for females, and larger impacts of education and work experience on females' job opportunities. Household wealth in the form of 'cash inflow' incentivizes individuals not to work, while wealth in the form of 'stock' induces higher utility to work for both males and females. The interpretation of empirical findings hinges on particular assumptions that might be disputed.

1. Introduction

Gender differences in labor supply have often been rooted in social norms (traditional view of gender roles), educational attainment, childcare responsibilities, and discrimination in the labor market. These differences are the result of a series of complex social, economic, and cultural factors. In this paper, using the 2011 Chinese Household Finance Survey data, we examine the differences in labor supply choices between females and males by considering both the willingness to work of people of each gender and their job market opportunities through the choices of job offering, with a particular focus on the role of household wealth. We highlight these aspects through a discrete choice model framework that simultaneously models the supply side work participation decision and the demand side job offering decision. We base our reduced-form model on the well-known random utility model, from which we derive the choice model of labor supply and demand. This enables the display of gender inequality and the way in which some key attributes, such as the age of children and wealth type, affect job decisions.

Our research contributes to the literature in three ways. First, we propose a choice-modeling framework that allows us to disentangle labor supply and demand choices simultaneously without needing direct data from the employer side. By observing whether

* Corresponding author.
E-mail addresses: xhan@imf.org (X. Han), tao.zhang@frisch.uio.no (T. Zhang), john.dagsvik@ssb.no (J.K. Dagsvik), chengyuan@fudan.edu.cn (Y. Cheng).

https://doi.org/10.1016/j.jocm.2023.100433
Received 12 October 2021; Received in revised form 21 June 2023; Accepted 6 July 2023

individuals are working, our model can quantify the impacts of factors driving demand and supply decisions simultaneously. To the best of our knowledge, we are among the first to use cross-sectional data to estimate *both* labor supply and demand while emphasizing employees' qualifications in labor demand. The previous literature, e.g., Dagsvik (2000), Menzel (2015), Dagsvik and Jia (2018), and Han et al. (2020), incorporates labor demand by using the size of job choice sets rather than by introducing employees' qualifications explicitly in the modeling framework. In contrast, our approach can be used to estimate the impacts of employees' qualifications on employers' preferences explicitly. This approach could be particularly valuable for analyzing labor supply and demand in emerging and low-income countries, where panel data or long-series repeated cross-sectional data are not easily accessible. Thus, a modeling framework that can be used to estimate both the demand and supply sides of the labor market but for which only cross-sectional data are needed is advantageous for improving the understanding of such labor markets with limited data.

Second, instead of focusing only on the female labor supply, which is often done in the literature, such as Blundell et al. (1987), we also include the labor supply choices of males. This approach allows us to answer questions about whether there are gender-related *inequalities* in terms of expected wages and job opportunities. The motivation for analyzing female and male job participation behavior together is not only rooted in the classic labor supply theories of different preferences in consumption and leisure but also substantiated by the machine learning approach we develop, in which our algorithm chooses "gender" as the decisive factor of labor participation. Moreover, empirical evidence of gender inequality in the labor markets of emerging and low-income countries is not abundant, even for China, i.e., no official unemployment rates by gender have been collected or published.[1] Thus, our modeling framework could hopefully contribute to improving the understanding of gender inequality in such cases.

Finally, we examine the impact of household wealth types on the choice to work. Our definitions of wealth types differ from those in the conventional approach. We regard the income of other household members as the "cash inflow" type of wealth and housing values and financial asset holdings as the "stock" type of wealth. We find that these two types of wealth work in the opposite ways; the "cash inflow" type of wealth tends to discourage individuals from working, whereas the "stock" type of wealth tends to encourage individuals to work.

Historically, China's labor market participation has been relatively high for both males and females. However, in recent years, China has experienced continuous declines in labor participation for both males and females; these levels are getting closer to those of some developed economies, such as Norway and Denmark, but are still well above the levels of Japan and Korea (Fig. 1). Thus, understanding the mechanism behind labor market participation behavior in China would be useful for both China and other countries.

China also provides a welcoming opportunity for analyzing the impact of wealth on labor market behaviors. Since 1978, China has been transitioning toward a more market-oriented economy. With the establishment of stock exchanges in Shenzhen and Shanghai, where the number of listed companies has increased from none to more than one thousand in 2015,[2] financial investment has become an increasing component of household wealth. Similarly, sudden increases in housing wealth from near zero to large amounts occurred in a very short time in China. Before 1999, there was no commercial residential housing market (Chen and Han, 2014). However, in the following decade, housing prices sharply increased despite the equally vast expansion of supply (Han, 2010). Xie and Jin (2015) have documented that in only two years, i.e., from 2010 to 2012, national housing wealth grew 51%. Housing wealth has become an important component in household financial portfolios and a decisive factor in labor supply, which is shown in the later analysis.

We use a cross-sectional data framework for our analysis in this paper. It is argued in the literature that panel or repeated-cross-sectional data can provide more information over time and is thus more suitable for depicting labor market behavior. This argument is certainly true and prevalent. On the other hand, for emerging and less-developed economies, long-term panel data or repeated cross-sectional data are not always available. We demonstrate in this paper that in the absence of applicable panel data, it is possible to conduct labor supply analysis using available cross-sectional data.

Another consideration of cross-sectional data is that in fast-growing and emerging economies such as China, the labor market has experienced rapid structural changes since the late 1990s due to privatization and reform. Caution should be exercised in assuming a constant fundamental causal relationship in labor supply and demand behaviors overtime, which is implied in applications with panel data. Therefore, we choose to focus on using cross-sectional data to depict the facts of the labor market at the time of data collection. Certainly, when long-term panel data or repeated cross-sectional data are available, such analysis of labor supply would benefit from using different types of data sources.

This paper is organized as follows. Section 2 summarizes the related literature. Section 3 lays out our modeling framework. Section 4 outlines the 2011 Chinese Household Finance Survey data and applies the regression tree approach to select variables. Section 5 elaborates on the empirical results, including the estimations of the baseline model and an extension of the model to allow correlations between job seeking and job obtainment. Furthermore, Section 5 presents the robustness checks, including using the method of moments to estimate the baseline model, estimating wage expectations using alternative approaches, and simulating the impact of gender on job offering and job seeking. Section 6 concludes our findings and discusses policy implications.

2. Literature review

Our paper is related to several strands of literature. The first line of research is rooted in the so-called Carnegie conjecture. As discussed in Holtz-Eakin et al. (1993), Andrew Carnegie asserted in 1891 that a large inheritance decreases a person's labor force participation. In the literature that followed, an inheritance was understood as being able take its original form as a true inheritance, as

[1] The labor participation rate using population census constitutes the only by-gender labor statistics that have been published; see ILO (2010).

[2] https://www.statista.com/statistics/225725/number-of-companies-listed-on-the-chinese-stock-exchange/.

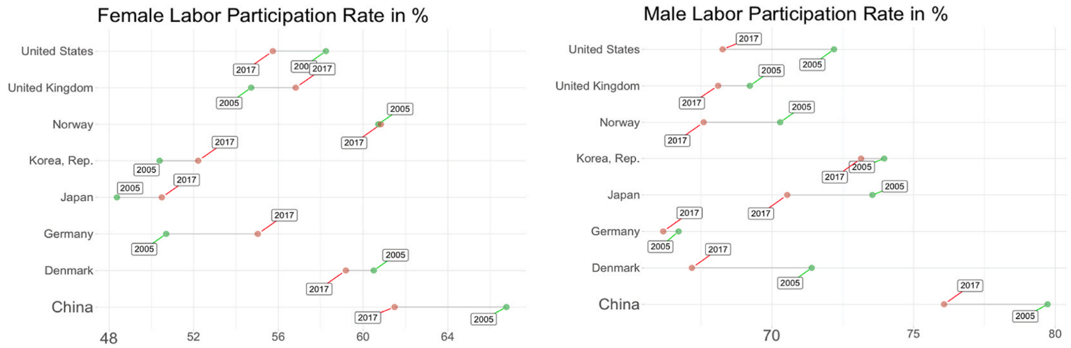

Fig. 1. Labour participation rate changes between 2005 and 2017 (of population ages 15+) by genders. Source: World Bank's World Development Indicators (https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS).

in Holtz-Eakin et al.'s (1993) usage of the federal individual tax returns of a group of people who received inheritances. Alternatively, an inheritance can take the form of more general types of wealth, such as housing wealth (e.g., Zhao and Burge, 2016) or winning the lottery (e.g., Jacob and Ludwig, 2012). The impact of wealth on the analysis of general labor market participation was further extended to a special category of employment, namely, entrepreneurs, e.g., Fairlie and Krashinsky (2012) and Adelino et al. (2015).

In our analysis, we consider wealth in the forms of nonlabor income such as housing wealth and financial assets, and we further differentiate them into the forms of "stock" and "cash inflow". Our analysis shows that only the "cash inflow" form of wealth is consistent with the Carnegie conjecture.

The second line of research focuses on the gender differences in labor supply and wages. The labor market participation of females has been analyzed extensively by researchers in labor economics, for example, Blundell et al. (1987), Ilmakunnas and Pudney (1990), Arellano and Meghir (2006), Blundell et al. (2016), Byker (2016), and Tominey (2016). Some scholars, such as Zhao and Burge (2016), have differentiated the impact of labor factors for females and males. In a recent trend, scholars have tried to attribute the wage gap between females and males to bearing and raising children. For example, Kleven et al. (2018) found that the arrival of children creates a gender gap in earnings of approximately 20% in the long run. Goldin et al. (2017) documented that the wage gap between females and males mainly occurs during child-bearing age. The paid leave law or other income support boosted the employment rate of women after giving birth, e.g., Tominey (2016), Byker (2016), Blundell et al. (2016). Our analysis uses a machine learning approach to search the age spectrum of children and determine whether the age of young children is a key factor affecting females' labor participation behaviors.

The final line of research is the modeling framework differentiating demand and supply. The random utility framework that we use was pioneered by McFadden (1978, 1984). Dagsvik (1994), van Soest (1995), van Soest et al. (2002), and Dagsvik and Strøm (2006) extended the application to investigate labor supply choices. Some previous research that addressed the issue of demand constraints in labor supply models includes that of Blundell et al. (1987), Ilmakunnas and Pudney (1990), and Dagsvik et al. (2013). Starting from the standard labor supply model as suggested by Heckman (1974), Blundell et al. (1987) relaxed the assumptions that an individual can find a job with a probability equal to one conditional on her or his willingness to work. The probability of finding a job is then affected by labor market conditions, such as involuntary unemployment. Using a dataset from Finland, Ilmakunnas and Pudney (1990) identified the effect of constraints on working hours on labor market behaviors. Dagsvik et al. (2013) used data from Norway and formulated demand constraints in the form of the discouraged workers' effect through the searching cost. In recent efforts to introduce the impact of labor market demand, scholars have viewed the labor market equilibrium as a two-sided matching market in which firms are also looking for a suitable match with a worker (e.g., Dagsvik (2000), Menzel (2015), and Dagsvik and Jia (2018), and Han et al. (2020)).

Our framework is closer in nature to the original modeling framework suggested by Blundell et al. (1987). While they focused on the number of working hours chosen, we focus on the decision to work or not work. Using our modeling specification, we can measure the impact of demographic characteristics and that of the local job market condition in an explicit way.

3. The model

In our analysis, an observation of an individual working is the result of fulfilling two conditions: first, the utility of that individual working is larger than the utility of not working. Second, the utility of the firm hiring that individual is larger than that of not hiring. In other words, observed labor participation is a result of the positive net utility of these two conditions.

We depart from the conventional labor supply choice model that is similar to that of Dagsvik et al. (2014) by defining the agent's

Table 1
Summary statistics for main variables.

	JobDummy	Age	Gender	Married	HHSize	YSchool	YWork	MedExp (CNY)	NumChild
Female									
Minimum	0	16	0	0	1	0	0	0	0
1st quartile	0	29	0	1	3	9	10	0	0
Median	1	37	0	1	3	12	20	0	1
Mean	0.73	36.7	0	0.84	3.8	11.1	19.6	122.1	0.9
3rd quartile	1	44	0	1	5	15	28	100	1
Maximum	1	55	0	1	13	19	49	21000	5
Male									
Minimum	0	16	1	0	1	0	0	0	0
1st quartile	1	30	1	1	3	9	12	0	0
Median	1	39	1	1	3	12	22	0	1
Mean	0.90	39.2	1	0.83	3.8	11.4	21.8	123.5	0.88
3rd quartile	1	48	1	1	5	15	31	0	1
Maximum	1	60	1	1	13	19	54	20,000	5

	AgeChild	**LabourInc** (CNY)	**NonlabInc** (CNY)	**HousingValue** (10,000 CNY)	**DepositCash** (10,000 CNY)	**FinAssetValue** (10,000 CNY)	**Rental** (10,000 CNY)	**Mortgage** (10,000 CNY)	**FinProfit** (10,000 CNY)
Female									
Minimum	0	2400	0	−9.4	0	0	0	0	−5
1st quartile	6	12,100	5774	10	0.2	0	0	0	0
Median	12	20,000	13,744	25	1	0	0	0	0
Mean	12.7	28,078	27,821	57.1	4.8	2.2	0.24	2.7	0.05
3rd quartile	19	32,035	24,703	60	3.7	0	0	0	0
Maximum	36	460,000	2,886,751	1050	700	420	120	570	26
Male									
Minimum	0	2085	0	−9.4	0	0	0	0	−10
1st quartile	6	16,736	1061	10	0.2	0	0	0	0
Median	13	25,298	9192	25.8	1	0	0	0	0
Mean	13.8	37,608	24,090	57.8	5.0	2.1	0.2	2.8	0.08
3rd quartile	21	40,000	20,294	64.5	3.8	0	0	0	0
Maximum	40	2,000,000	2,886,751	1050	700	204	120	570	40

Note: Labour income is conditional statistics for individuals reported salary higher than 2000. Non labour income is standardized by the household size.

utility function U of real disposable income and hours of work as follows:

$$U(C,h) = v(C,h) + \varepsilon, \tag{1}$$

where $v(C,h)$ is a positive deterministic term that represents the mean utility across individuals. The random error term ε is i.i.d with c.d.f. $\exp(-exp(-x))$.[3]

We assume there is only one option of working hours – a full-time job. There is no continuous choice of working hours. This is a main difference from the previous literature. The reason for this assumption is that in China, the common employment form is salary employment with monthly payment instead of hourly wage employment. The usual terminology "wage rate" is in this case not an hourly rate but a monthly rate. Second, Article 36 of the labor law requires that working hours cannot exceed 8 per day and 44 h per week with overtime. Among the 7984 individuals who reported job status in our data, 6549 individuals were working. A total of 3531 individuals reported their working hours. Among these 3531 individuals, 2010 reported exactly 40 working hours per week. Only 414 individuals had fewer than 40 working hours per week, and 211 had fewer than 35 working hours per week. The number of observations with fewer than 35 working hours is very small and clustered.[4] Thus, it is reasonable to take China's unique labor market into account, and in our model, a full-time job is the only option for working.[5]

Furthermore, for the sake of simplicity, we assume that the wage only depends on individual qualifications and does not vary across jobs. This might not be an innocuous assumption because the wage differences among various occupations and industries are influenced by a variety of factors (training and skills needed, competition levels, etc.) and can have a significant impact on an individual's labor participation choices. Krueger and Summers (1984) and subsequent research have shown that industry does truly matter in explanations of wage differences. The literature on job polarization (e.g., Autor and Dorn 2009, 2010, and 2013) shows that occupation also has an impact on wage differences, even after controlling for education.[6] Nevertheless, we credit the wage differences across occupation and industries mainly to the corresponding requirements for quality of labor, e.g., education and experience. In addition, our focus in this paper is on dichotomous labor market participation and job offer decisions instead of what type of jobs to choose. When potential job seekers decide to search or not search for a job and form their expected wages, they rely on their own education and experience instead of what jobs they might get. Therefore, we do not include variables on occupations and industries. This is the common approach in the literature in such cases, e.g., Dagsvik et al. (2014). Both job seekers and employers form their expectations of wages offered/desired based on individual qualifications. This assumption is rather stylized.

As there are only full-time jobs, the utility function is simplified such that the disposable income C captures all impact through wages in this stylized model:

$$U(C) = v(C) + \varepsilon. \tag{2}$$

We can further expand Equation (2) to differentiate labor and nonlabor income in disposable income and allow for the individual characteristics that affect the utility of working and not working. Let

$$U(W,Z) = v(W,Z) + \varepsilon, \tag{3}$$

where W is labor income and Z represents all other characteristics that affect willingness to work, such as age, marriage status, number of children, nonlabor income, and housing wealth. Non-labor income is included in the model as one element of Z. The functional form of $v(.)$ can be approximated by a linear transformation of W and Z through Taylor extensions.[7] Building on Equation (3), we separately formulate the utility function of working and not working to make the modeling more intuitive and easier to understand. A unified framework of this model (simultaneous decisions of working and not working) can be found in Appendix B.

Let U_1 be the agent's utility of working and let U_0 be the agent's utility of not working. Assume that the utility of working depends solely on the labor wage W as follows[8]:

[3] In the terminology of Resnick (1987) this c.d.f. is called the type III (standard) extreme value distribution, or Gumbel distribution. The reason why we assume the error term takes type III extreme value distribution is that the difference of two such error terms follows logistic distribution and yields computable choice probabilities function form such as Equation (8).

[4] The working hour tabulation for hours less than 35 per week.

Hours Worked	0.5	5	7.5	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5
No. of Obs.	3	3	1	7	2	6	3	26	5	27	7	96	25

[5] Even working hours varied, and since the data are cross-sectional rather than panel, there is a fundamentally different relationship between extensive margin (labor market participation) and the intensive margin of work hours. With panel data, we can examine changes over time in the intensity of work hours for the same individual (i.e., controlling for many unobservable influences), but this is not possible with cross-sectional data.

[6] Thanks to an anonymous reviewer for pointing this out.

[7] Han (2010) has presented a detailed Taylor extension approximation of choice utility.

[8] This assumption ignores other nonwage benefit related to working experience, such as vacation days associated with seniority. As one remedy, the age and age squared entering the utility of not working can help to capture the unobservable impact of seniority. The reason we do not introduce the nonlabor income in the utility of working is that we introduce the nonlabor income in the utility of not working. For the identification purpose, we can only introduce the nonlabor income into one utility function of either working or not working. We choose the latter to be consistent with the literature.

$$U_1 = \theta logW + \varepsilon_1^*, \tag{4}$$

where θ is an unknown positive parameter and ε_1^* is a random term that represents the unobservable preference associated with working. The utility of not working is assumed to have the following structure:

$$U_0 = Z\gamma + \varepsilon_0, \tag{5}$$

where Z is the vector containing the variables that motivate those individuals not to work, γ is the set of parameters to be estimated, and ε_0 is a random error term.

We need to explain our assumption of a linear form for Equations (4) and (5). Our choice of linear utilities of working and not-working is a rather simplified attempt to ease the mathematical derivations of reduced form solutions and later likelihood functions. A generalized nonlinear formulation would certainly be beneficial for the applicability of our model but come at the expense of a complex modeling framework. It is also a general approach in choice models to assume linearity in decision equations, e.g., Dagsvik and Strøm (2006) and Dagsvik et al. (2013).

Variables that have an impact on the decision *not to work* are own age, own age squared, the number of children, age of the youngest child, nonlabor income (labor income of other household members), and status of health. In particular, we also include variables of interest, such as the net value of housing, rental income, net value of financial assets, and profit earned from financial investments.

One might argue that the variables in Z (such as nonlabor income) would possibly affect both the utility of working and not working, with different effects (coefficients). As shown later, the decision to participate in the labor market is determined by the utility differences between U_1 and U_0 , in which only the net effect of variables in Z are necessary for our model. Thus, we adopt the simplified modeling framework, as shown in Equations (4) and (5).

Since wages are only observed for those who work, for those who do not work, we have no observations of potential wages. To disentangle this problem, we define a wage equation, which serves as an instrument variable, as follows:

$$logW = X\beta + \eta \tag{6}$$

where X is the vector containing an individual's qualifications (years of schooling, experience, experience squared), and β is a vector of parameters to be estimated. Following common practice, we define labor market experience as age minus years of schooling and minus 7.[9] However, with such treatment, there is the caveat that absence from the labor market, e.g., unemployment, may be excluded. Equation (6) is the wage equation used to approximate individuals' wage expectations if they choose to work[10]. Furthermore, our empirical findings on the impact of expected wages on the utility of working might be overestimated if the seniority or experience with the company brings substantial non-wage related benefits. Without including such non-wage benefit, the expected wage approximates but understates the overall expected benefit of working. Thus, the coefficient of the should-be larger benefit would be smaller. When Equation (6) is inserted into Equation (4), we obtain the following reduced-form expression of the utility of working:

$$U_1 = \theta X\beta + \varepsilon_1, \tag{7}$$

where $\varepsilon_1 = \varepsilon_1^* + \theta\eta$. Suppose that when $U_1 > U_0$, the individual wishes to work, and vice versa. Assume further that the random error terms ε_0 and ε_1 are independent and follow the extreme value distribution with a c.d.f. of $(-e^{-x})$. It is straightforward that the probability $\lambda(X,Z)$ that the individual is willing to work can be expressed as follows:

$$\lambda(X,Z) = P(U_1 > U_0|X,Z) = P(\theta X\beta - Z\gamma \geq \varepsilon_0 - \varepsilon_1|X,Z) = \frac{1}{1 + exp(Z\gamma - \theta X\beta)}. \tag{8}$$

As elaborated earlier, for an individual to obtain a job, her net utility of working must be positive, and the utility of employer to hire this individual must be positive as well. Originating from the random utility model proposed by McFadden (1978, 1984), the hiring decision of the employer is a choice decision based on comparisons between the revenue and cost of hiring. Let V_1 be the utility of an employer of offering a job to the individual and let V_0 be the employer's utility of not offering the job to this individual. V_1 can be thought of as the productivity/revenue the worker provides, while V_0 is the labor cost. Thus, the employer will only offer employment if $V_1 > V_0$. For simplicity, we assume that the utility of the employer for hiring is derived solely from the productivity measured by the characteristics of the employee (in our case, this is the same as the wage determination factors education and working experience). We assume that this utility can be specified as follows:

$$V_1 - V_0 = X\delta + \xi \tag{9}$$

where δ is a parameter vector to be estimated, and ξ is a random term that follows the logistic distribution.[11] From Equation (9), it follows that the probability $q(X)$ that the individual is offered a job is given by the following:

[9] The entrance age for the preliminary school is 7. We do not have observations of labor market movement, such as possible change of participation due to marriage, childbirth, etc.; therefore, we apply a crude assessment of labor market experience. This can be amended in future research with more detailed data.

[10] We also examine the impacts of alternative approaches to formulating the expectation of wages in Section 5.

[11] When V_1 and V_0 follows type I extreme value distribution, their difference follows logistic distribution.

$$q(X) = P(V_1 - V_0 > 0) = \frac{1}{1 + exp(-X\delta)} \tag{10}$$

Assuming that job participation decisions and hiring decisions are independent, we thus have the probability of an individual being observed working as follows:

$$P(X, Z) = \lambda(X, Z)q(X) \tag{11}$$

That is, the observed employment probability of one individual is jointly determined by the labor supply decision of this individual and the demand side decision in terms of an employment offer probability. In Appendix C, we demonstrate that the model is identifiable.

An alternative approach is to introduce the expected wage, such as Equation (6), into the labor demand Equation (10). However, from the identification perspective, the impacts of individual characteristics on the expected wage (labor cost) and on productivity (revenue to the employer) are not separable. In our baseline modeling framework, the payoff for the employer ($V_1 - V_0$) is the net of productivity/revenue over labor cost (wage), where both are dependent on the individual characteristics X. Accordingly, we choose to model the unified payoff as a function of X. A detailed discussion can be found in Appendix B.

One caveat of our modeling framework is the lack of formal treatment of searching costs. However, the way we formulate the labor market supply and demand could be regarded as the result of searching with searching cost implied. As shown in Dagsvik et al. (2013), the searching cost is positively proportional to the search time, which is the inverse of the arrival intensity of an acceptable job offer following a Poisson process. When the rate of flow out of unemployment to employment is equal to the rate of flow from employment into unemployment, the searching cost can be explicitly expressed as the inverse of the probability of obtaining an acceptable job minus one. Their probability of obtaining an acceptable job takes a functional form that is similar to ours (Equation (10)), which is determined by education and work experiences. Even without being explicitly introduced into the modeling framework, the implied searching cost can be calculated as $1/q(x) - 1$.

Furthermore, in our model, employees make their job-seeking decision purely based on expected wages and other factors such as age, nonlabor income, and housing wealth, as reflected in Equation (8), but not on the probability of being offered a job. The job offers are addressed by the labor market demand, as reflected in Equation (10). Our modeling framework assumes an equilibrium labor market, implying that the productivity/revenue generated by the employee and the cost of the employee (wage) warrant the equilibrium.

Our analysis also introduces males' labor market behavior using a similar modeling framework, which allows us to examine whether employers treat females differently from males in terms of offering jobs. By estimating female and male labor market choices, we deviate from the traditionally assumed Stackelberg assertion in the previous literature, i.e., the male decides first that he will work, regardless of the wife's preferences, and subsequently, the wife makes her choice, taking the husband's participation and wage as given (exogenous). In our analysis, we assume that the Stackelberg assertion (e.g., Bjorn and Vuong, 1985) applies symmetrically to males. We do not implement a joint decision-modeling framework between females and males since the main purpose of this analysis is to compare whether the same variables affect females and males in a different way.

4. Data and variable selection using the regression tree approach

The data we used are from the 2011 Chinese Household Finance Survey.[12] The survey has the advantage of providing detailed housing wealth information, which is an important part of our investigation. After preliminary cleaning,[13] we included 8438 households and 29,324 individuals in our analysis. Geographically, the survey covered 22 provinces/municipalities. We included only males between the ages of 15 and 60 years old and females between the ages of 15 and 55 years old, as in China, the typical retirement ages are 60 and 55 years for males and females, respectively. Furthermore, we included only those individuals living in urban areas, and we excluded individuals for which there was no information on labor and nonlabor income and those who were full-time students or retirees. The final dataset included 8128 individuals; among these, 3726 females and 4195 males had complete information on the explanatory variables.

Table 1 presents the summary statistics of the main variables included in the analysis. They are labor market participation (JobDummy), age, gender, marital status (Married), household size (HHSize), years of schooling (YSchool), years of working experience (YWork), medical expenses (MedExp), number of children (NumChild), age of the youngest child (AgeChild), labor income (LaborInc), nonlabor income (NonlabInc), net value of housing (HousingValue), value of time deposit and cash (DepositCash), value of financial assets (FinAssetValue), rental income (Rental), mortgage, and profit earned on financial assets (FinProfit). The job dummy variable is equal to one if the individual reported that she or he was working and 0 otherwise. The variable of gender is equal to 1 if the respondent

[12] More details of the survey can be found in https://chfs.swufe.edu.cn/datacenter/apply.html.

[13] We excluded households with a suspicious age gap among family members such as (a) an age gap between two generations (i.e., parents and children) of less than 16 years; (b) a child being "older" than either the parent or grandparent; (c) an age gap between spouses that is too large; (d) an age gap between a child and grandparent of less than 32 years; (e) an age gap between in-laws (e.g., respondent and daughter-in-law); that is too small; and (f) an age gap between siblings that is too large (e.g., approximately 1 or more generations apart). We also removed the households with more than one household member tagged as the spouse of the respondent, where more than two household members are tagged as parents, where the "spouse" of the child is suspiciously young or where there is no indicator for the respondent, so that it is difficult to distinguish the family members.

Table 2
Estimations of the wage equation.

	Females	Males
Intercept	1.00**	0.38**
	(0.12)	(0.09)
Years of Education	0.23**	0.19**
	(0.02)	(0.02)
Years of Working Experience	0.27**	0.28**
	(0.06)	(0.06)
Years of Working Experience squared	−0.02**	−0.01**
	(0.003)	(0.002)
Fixed effects for provinces	Yes	Yes
Adjusted R^2	0.145	0.08
No. of Obs.	2735	3763

Note: The years of education and years of working experience are standardized values by minus their means and divided by their standard deviations. The years of working experience squared are years of working experience divided by 5 and squared.

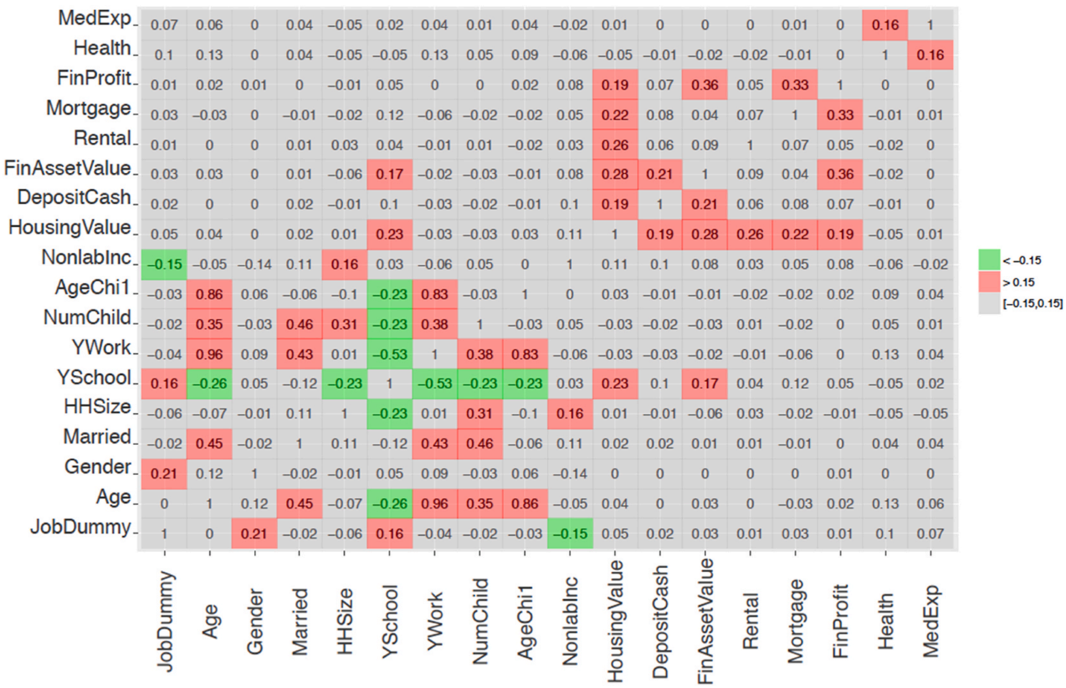

Fig. 2. Pair-wise correlations between variables.

is male and 0 otherwise.

Labor income is the annual after-tax income reported in the survey for 2010 and includes salary, bonus, and any other forms of compensation related to the job. The nonlabor income includes all labor income of other household members. The net value of housing corresponds to the market value minus the cost of acquisition reported in the survey. Rental income and profit earned on financial assets correspond to the annual amount reported in the survey for 2010. The value of time deposits and cash, the value of financial assets, and mortgages are directly reported in the survey.

As shown in Table 1, the average value of labor market participation for males (0.9) is higher than that of females (0.73). Due to the higher retirement age for males, the average age, years of experience, and age of the youngest child of males are higher than those of females. The average labor income of males is higher than that of females, while the average nonlabor income (income of other family members) of males is lower. The average values of marriage status, household size, years of schooling, medical expenses, number of children, housing values, value of time deposit and cash, value of financial assets, rental income, and mortgage for females and males are similar. These values serve as the basis for the estimates in our model so that we can further investigate whether males have a higher labor participation rate because they have a higher willingness to work or more job market opportunities and whether they have higher labor incomes because they have more work experience/higher education or because of the effect of gender discrimination.

In estimating the baseline model, to reduce the impacts of different scales of variables, we rescaled the variables. As the nonlabor

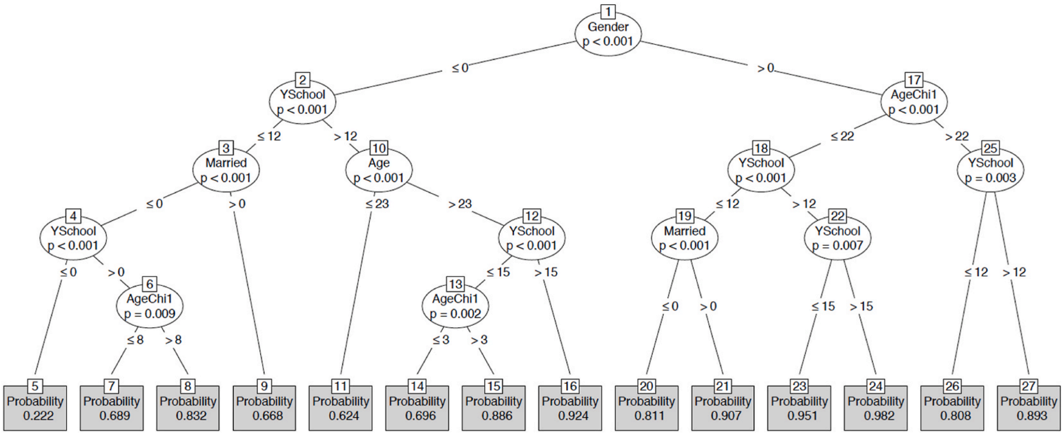

Fig. 3. Regression tree analysis for labour participation probability.

income varies with household size, we first rescaled it by dividing the square root of household size and then taking logarithm transformation. For labor income, we only took logarithm transformation. Age, years of schooling, and years of work experience were divided into 5 groups. Before we began the likelihood estimations, we further standardized all variables, except for dummy variables and the years of work experience squared, by subtracting their means and dividing by their standard deviations. Thus, we needed to rescale the coefficient estimates as in the wage equation in Table 2 back to the original value of years of education and years of working experience to obtain expected wages.

In examining and selecting suitable variables that would affect the decision processes, we first plot the pairwise correlations between the selected variables to check the unconditional correlations. Fig. 2 shows highlighted values that are higher than positive 0.15 or lower than negative 0.15. Unconditionally, the job dummy variable is only positively correlated with gender (males) and with education, i.e., those with more years of schooling tend to have higher job participation, with correlations stronger than 0.15. Conversely, the job participation dummy is only negatively correlated with nonlabor income, i.e., a higher nonlabor income tends to be associated with lower participation, with a correlation coefficient of −0.15. Among the other variables with weaker correlations, it is surprising to observe that individuals with a higher net housing value, a higher value of their financial assets, or worse health status tend to have higher participation levels, while those with longer periods of working experience tend to have lower participation levels.

Based on the unconditional and linear correlations revealed above, we further apply a machine learning approach—regression tree—to guide us in selecting and redefining the variables for the structural model. The specific regression tree approach that we use is called the conditional inference tree (CIF). The main purpose of the CIF is to search among all the potential explanatory variables and select the most powerful ones in terms of predicting job participation tendency. For each explanatory variable, the algorithm searches for the value to split all the observations into two groups, with one group having an average participation tendency that is significantly different from that of the other group. The breakpoint that can best differentiate the participation behavior of the included individuals is the threshold by which to perform the splitting. The explanatory variable with the threshold that can best explain the participation behavior is regarded as the most powerful variable and is thus selected to perform the splitting. Such searching and splitting mechanisms are repeated until a specific ending criterion is reached. Each individual is ultimately grouped into one of the end-result groups, whose members share similar participation tendencies and the same path on the tree (similar characteristics). A methodological description of CIF can be found in Hothorn et al. (2006) and Hothorn and Zeileis (2015), as well as in Han and Wei (2017).

As shown in Fig. 3, the leading pick of the variables is gender, which supports the common claim that the labor market participation behaviors of females and males are different. Based on this evidence, we will estimate the models for males and females separately.

The right arm extending from the top node labeled as gender is for males, and the left arm is for females. For females, the second variable picked for the splitting is years of schooling; females with more than 12 years of schooling go further along the right arm, and those with less than 12 years of schooling go to the left arm.

Furthermore, if females are older than 23 years and have more than 15 years of schooling (above a college education, ending Node 16), they have an average participation rate of 92.4%. However, with less than 15 years of schooling, the participation rate depends on whether the female has children younger than 3 years old. If they do, their participation rate is 69.6%, but if not, their participation rate is higher at 88.6%. On the other hand, when a female has fewer than 12 total years of schooling is married, her average participation rate is 66.8% (Node 9). If a female is not married and has no education at all, her participation rate is only 22.2%. If she is not married but has some education, her participation rate depends on the age of her youngest child; if the child is older than 8 years, the mother's participation is 83.2%, while if the child is younger than 8 years old, the mother's participation is 68.9%.

The right arm of the tree is for males. The age of the youngest child works very differently for males than for females. When the youngest child is younger than 22 years old, which is the age at which most people graduate from university, on average, their fathers' participation rate is higher. If a father has a college education, he has a participation rate of 98.2% (Node 24), which is the highest participation rate among their peers. For those who have fewer than 12 years of education, those who are married have higher participation rates than those who are unmarried.

The CIF analysis revealed several interesting results. First, there are significant differences in labor market participation behaviors

Table 3
Estimations of labour supply and demand probabilities.

	Independent model		Dependent model	
	Females	Males	Females	Males
Probability of supplying labour $\lambda(X,Z)$: Coefficient for variables (except for expected wage) contribute positively to utility of not working (U_0)				
Constant	−96.78**	−53.14**	−135.75**	−53.34**
	'(0.17)	'(0.22)	'(0.17)	'(0.51)
Age	0.07	0.19**	0.1	0.19**
	'(0.05)	'(0.07)	'(0.05)	'(0.08)
Age squared	0.17**	0.25**	0.2**	0.29**
	'(0.05)	'(0.05)	'(0.05)	'(0.08)
Age of the youngest child	0.2**	−0.12	0.2**	−0.12
	'(0.04)	'(0.08)	'(0.04)	'(0.08)
Married	0.6**	−0.29	0.62**	−0.31
	'(0.16)	'(0.2)	'(0.16)	'(0.21)
Non-labour income	0.54**	0.51**	0.51**	0.51**
	'(0.07)	'(0.07)	'(0.07)	'(0.13)
Value of housing	−0.13**	−0.21**	−0.18**	−0.25**
	'(0.05)	'(0.07)	'(0.05)	'(0.09)
Value of financial investment	−0.03	−0.08	−0.05	−0.1
	'(0.05)	'(0.08)	'(0.06)	'(0.09)
Medical expenditures	−647.27**	−348.04**	−914.45**	−348.89**
	'(0.09)	'(0.59)	'(0.03)	'(1.37)
Expected wage	1.09**	0.75**	0.68**	0.33
	'(0.14)	'(0.22)	'(0.14)	'(0.24)
Probability of getting a job offer $q(X)$:				
Constant	133.97**	104.2	189.4**	106.51
	'(2.46)	'(119.35)	'(2.42)	'(129.66)
Education	42.32**	17.06	58.43**	17.94
	'(7.84)	'(27.79)	'(7.6)	'(33.33)
Work experience	−13.31	−34.22	−43.08	−34.78
	'(65.94)	'(36.25)	'(43.79)	'(37.42)
Work experience squared	67.08	115.34**	105.33**	114.93**
	'(74.71)	'(39.02)	'(49.92)	'(32.85)
Provincial employment rate	−19.81**	−16.15	−32.19**	−15.78
	'(3.78)	'(25.87)	'(1.7)	'(26.51)
Afa[a]	–	–	1.00**	0.984**
	–	–	'(0.00)	'(0.00)
Log-Likelihood at Maximum	−1770.74	−1189.61	−1790.74	−1194.59
Log-Likelihood at Reference	−4076.61	−5340.09	−4076.61	−5340.09
McFadden's Rho	0.57	0.78	0.56	0.78
No. of Obs.	3726	4195	3726	4195

[a] In the estimation, we transform Afa to 1/(1+exp(x)), where x can be searched continuously. The estimated x for females is −26.21 with the corresponding Afa as 1 and the estimated x for males is −4.12 with the corresponding Afa as 1.

between females and males in urban China. Second, the age of the children matters differently for the job participation of males and females. For females, caring for young children lowers their participation rate, while for males, supporting the child until university graduation increases their participation rate. Third, among the competing explanatory variables, years of schooling, marriage, age, and age of the youngest child are among the most informative variables.

5. Empirical results

We organize our results in the following way. We first present the results from the baseline model, where we assume independence between job seeking and job obtainment. We then relax this assumption and allow correlation between job seeking and obtainment. In addition, we conduct some robustness checks of our results with respect to endogeneity. Additionally, we explore some alternative approaches to modeling the wage equation. Finally, we present a policy simulation.

5.1. Assuming independence between job seeking and job obtainment

We first consider the case where job seeking and obtainment are independent, that is, assuming in Equation (11) that $\lambda(X, Z)$ and $q(X)$ are independent of each other.

We carry out a two-stage estimation strategy. In stage one, we estimate the wage equation in Equation (6) by OLS for those participating in the job market with observed wages.[14] Let $j = 1$ if the person is working and $j = 0$ if not. Equation (6) can be further written as follows:

$$\log W_{i,j=1} = X_{i,j=1}\beta + \eta_{i,j=1}, \textit{where } j = 1 \textit{ working}. \tag{12}$$

After estimating $\widehat{\beta}$, we can then perform an out-of-sample prediction of wages for those who do not participate by using $log\ \widehat{W}_{i,j=0} = X_{i,j=0}\widehat{\beta}$. In stage two, we use the predicted wages for all individuals to approximate their wage expectations. We then employ the maximum likelihood method to estimate θ, γ, and δ. The corresponding log-likelihood function can be expressed as follows:

$$logL = \sum_i Y_i \log P(X_i, Z_i) + \sum_i (1 - Y_i)\log(1 - P(X_i - Z_i)). \tag{13}$$

where $Y_i = 1$ if individual i works and $Y_i = 0$ otherwise. $P(X_i, Z_i)$ is for the full sample and is defined in Equation (11).

Table 2 reports the coefficient estimations for wage Equation (12). As expected, the coefficient of education is significant and positive. The coefficient of working experience is positive, while the coefficients for working experience squared are negative, which shows that there is a concave function of working experience and “optimal” working experience in terms of bringing the highest wage. The goodness of fit is better for females than for males, as suggested by the adjusted R^2 (0.145 vs. 0.08).

Guided by the CIT results, for the age of the youngest child, we define a dummy variable that is equal to one when the child is younger than three years old for females and when the child is younger than twenty-two years old for males in the baseline model.

Since we standardize all the variables (except for the dummy variables) by first deducting the mean and then dividing by one standard deviation for each variable, the coefficient estimates of the continuous variables are the impacts of one standard deviation from their means.

In Table 3, in the first and second columns, we report the parameter estimations for the model given in Equation (13) for both females and males, respectively. Apart from the value of financial investments and the age of the females, most of the parameters are precisely estimated. Notably, the parameter estimators presented in Table 3 under the title of the “probability of supplying labor $\lambda(X, Z)$” are for the variables contributing to the utility of not working, except the coefficient for the log wage variable. This is due to the model specifications in Equations (4) and (5). Therefore, any positive estimations for the parameters will have a negative contribution to the probability of willingness to work.

From the estimation results, we note that the utility of not working for both females and males takes a convex function regarding age, which means that the utility of not working initially decreases by age until it reaches a minimum and then subsequently increases after that. The age at the minimum utility of not working is 34.5 years old for females[15] and 31 for males.[16] Our results indicate that for females (males) younger than 34.5 (31) years old, the willingness to work decreases with age. However, for people older than the threshold age, the willingness to work increases with age. That is, the ages of 34.5 years old for females and 31 for males are the age thresholds at which the willingness to work is at its lowest.

Our findings are at odds with those of previous studies. In previous literature dedicated to the analysis of female labor, the typical

[14] OLS estimate may be subjected to bias caused by the self-selection. First, related empirical analysis shows no evidence that supports the selection bias (Dagsvik and Strøm, 2006). Second, in the next section, we carry out a robustness check with Heckman's selection corrections for wage estimation. However, we have demonstrated that the selection corrected wage estimate shows negative correlation with the labor supply, which is counter intuitive. Therefore, we choose to use the OLS as the baseline framework.

[15] We rescaled the original data by dividing it by 5 first. The mean and standard error for the rescaled age for females is 7.3 and 1.9, respectively. Therefore, the age with the lowest utility will fulfill the condition that the first-order derivative of utility at this age is equal to zero. That is, ((x/5–7.3)/1.9) = -0.07/2*0.17. Therefore, x = (7.3-0.07/0.34*1.9) *5 = 34.5.

[16] The mean and standard error for the rescaled age for males is 7.8 and 2.14, respectively. Therefore, x = (7.8-0.19/0.25*2.14) *5 = 31.

relationship between the utility of not working and age usually takes the opposite shape, that is, the utility of not working increases with age and then decreases. For example, Lundberg et al. (1994) found that in Sweden, the age at which females are most reluctant to work is approximately 35. The possible reason for this disparity might lie in the other variables that we controlled for in the estimation, including the age of the youngest child, marital status, nonlabor income, and wealth in housing.

Our findings suggest that having a child younger than three years old reduces the utility of supplying labor for married females but not for males. Additionally, being married significantly reduces the willingness to work for females (0.6) but not for males. Having a young child or being married tends to encourage males to work, albeit not in a significant way.

With higher nonlabor income, both females and males are more reluctant to work. The impact of wealth in the form of housing is surprising – with a higher net value in housing, people are more willing to work. This outcome holds for both females and males and is significant. Wealth in financial investments works in the same direction but is insignificant. This result is contrary to the common belief associated with the Carnegie conjecture that higher wealth discourages work. Our results show that not all forms of wealth work in the same way. Nonlabor income tends to discourage work, but wealth in housing and financial investments tends to encourage work, at least in China. We further check the robustness of the results in Section 5.3.

Another surprising result is related to the impacts of health on the willingness not to work. The coefficient estimate of medical expenses is significantly negative, which means that with poor health status implied by higher medical expenses, individuals have a strong willingness to work. We also tried to replace medical expenses with self-reported subjective health status, which yielded the same results. We noted that in the data, all the individuals who reported their health status as being very poor also reported being employed, which explains why poorer health is associated with a higher willingness to work. This might be because in China, the medical care system is closely tied to employment. To be able to access sufficient medical care, one needs to be employed to qualify for reimbursement of medical expenditures. There is clearly an indication of self-selection in this respect.

An alternative unified explanation could be that nonlabor income represents cash inflow, while the value of housing does not directly result in cash inflow. It is often just value on paper, unless realized in sales. Medical expenses constitute a need for cash, which induces willingness to work to cover the expenses. Cashable assets reduce the willingness to work.

The coefficient estimate for the wage equations is significant for both females and males, with a higher explanatory power for females than for males, which means that the elasticity of the utility of working with respect to wages is larger for females than for males.

The lower panel of Table 3 presents the estimation of $q(X)$, i.e., the probability of an employer offering a job, conditional on education, work experiences, work experiences squared, and regional employment rate. The regional employment rate approximates the tightness of the local labor market. A higher regional employment rate implies a tighter local labor market, in which it would be more difficult to get a job. The estimates are consistent with our expectations but are only significant for females. Education attainment significantly increases job opportunities for both males and females. Work experience positively contributes to the probability of obtaining a job but only after reaching a certain year of experience, which is 22.7 years for females and 26.3 years for males. Combined with the estimates for age and age squared in the supply function, it is shown that before their mid-thirties, individuals face some unfavorable job opportunities but are more willing to work. In contrast, after reaching their mid-forties, individuals are more valued in the job market but are less keen to work. Therefore, the labor market tends to favor younger and less-experienced individuals less.

With regard to the goodness-of-fit, we find that McFadden's ρ^2 for our model is equal to 0.57 for females and 0.78 for males, which indicates a good fit.

5.2. Allowing correlations between job seeking and job obtainment

We extend our analysis by allowing the utility of working and the utility of obtaining a job to be correlated with the entrance of unobservable factors that jointly affect utility. An example of such an unobservable factor is a booming period that boosts an increase in both labor demand and supply. To include these factors, we include a random effect factor represented by α. We assume that there is a correlation between the error terms of willing to work $(\varepsilon_1 - \varepsilon_0)$ and the error terms of being offered a job (ξ), as $corr(\varepsilon_1 - \varepsilon_0, \xi) = 1 - \alpha^2$. In addition, α is a parameter that is restricted to $0 < \alpha \le 1$..

The joint probability takes the following form:

$$\overline{P}(V_i) = E(\lambda(V, \nu_1) q(X, \nu_2)) = \frac{1}{1 - \exp(Z\gamma - \theta X\beta - X\delta)} \left(\frac{1}{1 + \exp(Z\gamma\alpha - \alpha\theta X\beta)} + \frac{1}{1 + \exp(-X\delta\alpha)} - 1 \right) \tag{14}$$

and

$$\lambda(V) = E\lambda(V, \upsilon_1) \frac{1}{1 + \exp(Z\gamma\alpha - \theta\alpha X\beta)} \tag{15}$$

Table 4
Robustness checks for labour supply and demand probabilities.

Females	(1)	(2)	(3)	(4)
Probability of supplying labour $\lambda(X,Z)$: Coefficient for variables (except for expected wage) contribute positively to utility of not working (U_0)				
Constant	−74.62**	−86.85**	−112.98**	−112.99**
	'(0.18)	'(0.18)	'(0.17)	'(0.17)
Age	0.03	0.04	0.1	0.1
	'(0.05)	'(0.05)	'(0.05)	'(0.05)
Age squared	0.15**	0.16**	0.2**	0.2**
	'(0.05)	'(0.05)	'(0.05)	'(0.05)
Age of the youngest child	0.21**	0.2**	0.2**	0.2**
	'(0.04)	'(0.04)	'(0.04)	'(0.04)
Dummy for married	0.61**	0.63**	0.62**	0.62**
	'(0.16)	'(0.16)	'(0.16)	'(0.16)
Non-labour income	0.43**	0.45**	0.51**	0.52**
	'(0.06)	'(0.06)	'(0.07)	'(0.07)
Value of housing	–	–	−0.19**	−0.17**
	–	–	'(0.05)	'(0.05)
No. of housing owned	–	−0.3**	–	–
	–	'(0.07)	–	–
Regional housing appreciation* no. of housing owned	−0.17	–	–	–
	'(1.31)	–	–	–
Value of financial investment	−0.11	−0.09	−0.05	−0.05
	'(0.06)	'(0.06)	'(0.06)	'(0.06)
Housing rental	–	–	0.01	–
	–	–	(0.04)	–
Mortgage	–	–	–	−0.07
	–	–	–	(0.07)
Medical expenditures	−491.85**	−575.81**	−759.37**	−759.37**
	'(0.03)	'(0.03)	'(0.03)	'(0.03)
Expected wage	0.81**	0.79**	0.68**	0.67**
	'(0.14)	'(0.14)	'(0.14)	'(0.14)
Probability of getting a job offer $q(X)$:				
Constant	1535.62**	120.88**	155.41**	155.41**
	'(0.36)	'(7.03)	'(2.28)	'(2.27)
Education	572.94**	40.13**	48.09**	48.09**
	'(0.5)	'(2.26)	'(7.31)	'(7.32)
Work experience	217.34**	−72.46**	−40.01	−40.02
	'(0.44)	'(13.29)	'(42.07)	'(42.08)
Work experience squared	461.4**	135.78**	101.05**	101.04**
	'(0.36)	'(12.51)	'(47.98)	'(48.03)
Provincial employment rate	−9.98**	−47.83**	−22.84**	−22.84**
	'(0.01)	'(0.18)	'(2.43)	'(2.42)
Log-Likelihood at Maximum	−1798.16	−1787.51	−1790.69	−1790.14
Log-Likelihood at Reference	−4076.61	−4076.61	−4076.61	−4076.61
McFadden's Rho	0.56	0.56	0.56	0.56
No. of Obs.	3726	3726	3726	3726
Males	**(1)**	**(2)**	**(3)**	**(4)**
Probability of supplying labour $\lambda(X,Z)$: Coefficient for variables (except for expected wage) contribute positively to utility of not working (U_0)				
Constant	−29.21	−32.64	−53.17**	−53.22**
	'(17.17)	'(20.86)	'(0.27)	'(0.21)
Age	0.21**	0.21**	0.19**	0.19**
	'(0.08)	'(0.08)	'(0.07)	'(0.07)
Age squared	0.36**	0.35**	0.29**	0.29**
	'(0.06)	'(0.06)	'(0.05)	'(0.05)
Age of the youngest child	−0.18	−0.18	−0.12	−0.12
	'(0.1)	'(0.1)	'(0.08)	'(0.08)
Dummy for married	−0.3	−0.29	−0.3	−0.3
	'(0.23)	'(0.22)	'(0.2)	'(0.2)
Non-labour income	0.54**	0.55**	0.5**	0.51**
	'(0.08)	'(0.08)	'(0.07)	'(0.07)
Value of Housing	–	–	−0.24**	−0.21**
	–	–	'(0.08)	'(0.07)
No. of housing owned	–	−0.49**	–	–
	–	'(0.1)	–	–
Regional housing appreciation* no. of housing owned	−5.6**	–	–	–
	'(1.45)	–	–	–
Value of financial investment	−0.14	−0.12	−0.09	−0.09
	'(0.09)	'(0.09)	'(0.09)	'(0.08)
Housing rental	–	–	−0.05	–

(continued on next page)

Table 4 (*continued*)

Males	(1)	(2)	(3)	(4)
	–	–	’(0.15)	–
Mortgage	–	–	–	−0.28
	–	–	–	’(0.17)
Medical expenditures	−185.31	−210.41	−348**	−348.15**
	’(116.96)	’(142.1)	’(1.13)	’(0.15)
Expected wage	0.22	0.27	0.33	0.31
	’(0.24)	’(0.24)	’(0.22)	’(0.22)
Probability of getting a job offer $q(X)$:				
Constant	11.37**	10.29**	104.21	104.76
	’(3.26)	’(2.9)	’(119.95)	’(120.05)
Education	1.43**	1.64**	17.07	17.37
	’(0.4)	’(0.44)	’(27.86)	’(29.25)
Work experience	−7.28	−6.94	−34.23	−34.71
	’(4.14)	’(4.31)	’(35.66)	’(37.42)
Work experience squared	9.71**	9.62	115.33**	115.01**
	’(4.77)	’(4.97)	’(40.12)	’(34.03)
Provincial employment rate	−126.55	−89.77	−16.15	−16.08
	’(79.9)	’(69.47)	’(25.01)	’(24.69)
Log-Likelihood at Maximum	−1189.84	−1185.62	−1194.40	−1192.17
Log-Likelihood at Reference	−5340.09	−5340.09	−5340.09	−5340.09
McFadden’s Rho	0.78	0.78	0.78	0.78
No. of Obs.	4195	4195	4195	4195

Table 5
Estimations of labour supply and demand using Method of Moments.

	Females	Males
Probability of supplying labour $\lambda(X,Z)$: Coefficient for variables (except for expected wage) contribute positively to utility of not working (U_0)		
Constant	−34.77	−33.50
	(19.50)	(19.17)
Age	0.14**	0.20*
	(0.07)	(0.11)
Age squared	0.22**	0.43**
	(0.07)	(0.08)
Age of the youngest child	0.21**	−13.22**
	(0.05)	(2.99)
Dummy for married	0.80**	−0.24
	(0.20)	(0.32)
Non-labour income	0.49**	0.56**
	(0.08)	(0.11)
Value of housing	−0.18**	−0.22**
	(0.07)	(0.1)
Value of financial investment	−0.02	−0.07
	(0.09)	(0.12)
Medical expenditures	−225.87	−180.30
	(132.82)	(137.59)
Expected wage	0.15	−0.28
	(0.22)	(0.42)
Probability of getting a job offer $q(X)$:		
Constant	43.74**	8.33**
	(0.13)	(1.71)
Education	4.82**	0.90**
	(0.27)	(0.36)
Work experience	−34.28**	−4.99**
	(0.07)	(2.56)
Work experience squared	33.92**	7.18**
	(0.07)	(3.28)
Provincial employment rate	−63.21**	−86.49**
	(0.04)	(40.24)
Log-Likelihood at Maximum	−334	−387
Log-Likelihood at Reference	−1496	−2342
McFadden’s Rho	0.78	0.83
No. of Obs.	3726	4195

Table 6
Estimations of labour supply and demand using different approaches for wage expectations - Females.

	Using Regression Tree Approach
Probability of supplying labour $\lambda(X,Z)$: Coefficient for variables (except for expected wage) contribute positively to utility of not working (U_0)	
Constant	−66.24**
	(0.17)
Age	0.12**
	(0.05)
Age squared	0.24**
	(0.05)
Age of the youngest child	0.19**
	(0.04)
Dummy for married	0.64**
	(0.16)
Non-labour income	0.50**
	(0.07)
Value of housing	−0.21**
	(0.05)
Value of financial investment	−0.06
	(0.06)
Medical expenditures	−440.56**
	(0.02)
Expected wage	0.50**
	(0.12)
Probability of getting a job offer $q(X)$:	
Constant	649.35**
	(0.94)
Education	215.95**
	(3.01)
Work experience	−29.77**
	(0.43)
Work experience squared	287.40**
	(0.20)
Provincial employment rate	−8.12**
	(0.03)
Log-Likelihood at Maximum	−1794.32
Log-Likelihood at Reference	−4076.61
McFadden's Rho	0.56
No. of Obs.	3726

$$q(X) = Eq(X, \upsilon_2) = \frac{1}{1 + \exp(-X\delta\alpha)}, \tag{16}$$

The log-likelihood function, in this case, is equal to the following:

$$logL = \sum_i Y_i \, log\overline{P}(V_i) + \sum_i (1 - Y_i \log(1 - \overline{P}(V_i)). \tag{17}$$

whereas β is estimated in the first stage wage equation, the parameters α, γ and θ are estimated in the second stage with maximum likelihood.

The results are presented in Columns (3)–(4) of Table 3. To estimate α continuously without artificially imposing the limit, we transform it to $1/1 + exp(\chi)$. As shown in Table 3, the estimated α is close to one, which indicates that the potential correlation between willingness to work and being offered a job is near zero. We find that our data do not support the hypothesis that labor supply decisions and labor demand are correlated. This means that the variables determining job seeking and job obtainment are most likely included in the modeling specifications. No significant variable that drives both events is missing.

5.3. Robustness check of wealth effect

In this section, we check the robustness of the wealth effect. The negative coefficient estimates of housing wealth and the positive coefficient estimates of nonlabor income have tentatively revealed that the "stock" form of wealth and the "cash inflows" form of wealth have opposite effects on the utility of not working.

As discussed in the Introduction section, housing wealth rose from nothing to a significant value in a very short time in China. Housing wealth grew 51% between 2010 and 2012. Such a sharp and sudden increase serves as an exogenous shock. This is helpful for addressing the endogeneity concern. That is, it is highly impossible for the labor supply decision to lead to housing wealth changes. Nevertheless, we carried out several robustness checks.

First, we replace the value of housing with the value of inherited housing reported in the survey. Individual labor supply choices

Table 7
Scenarios simulations for females and males.

	Females		Males	
	Simulated with estimates of females	Simulated with estimates of males	Simulated with estimates of males	Simulated with estimates of females
Probability of willing to work	73.53%	76.71% (76.3%)	90.39%	89.42%
Probability of getting an offer	99.79%	99.30%	99.24%	99.83%

cannot affect the inherited housing value. Only the inherited housing value can affect individual labor supply choices. Among the 7157 individuals who reported housing values, only 466 reported housing as inherited. The coefficient of the inherited housing value for females is estimated as −0.013 but is insignificant, with a standard error of 0.04.[17] Even though it is insignificant, the negative sign is consistent with our baseline results in Table 3 and supports our conclusion.

Second, we replace the net house value with the value approximated by multiplying the number of houses owned with regional housing appreciations (to present the house wealth increase) (Column (1) of Table 4) or the number of houses owned (Column (2) of Table 4) to delink housing wealth from individual idiosyncratic house choices/values. That is, individual labor market choices cannot affect regional housing prices as strongly as their own housing values do. As shown in Columns (1) and (2) of Table 4, the coefficients remain negative and significant and only insignificant for females when interacting the number of houses owned with regional housing appreciation. Thus, the positive impacts of wealth in stock form are robust across the different approaches of measuring.

Next, we check the impact of other forms of wealth on cash flows by including housing rentals and mortgages. Column (3) of Table 4 shows that housing rental (as cash inflows) is positive for females and shares the same sign as nonlabor income but is negative for males. Both are insignificant. Column (4) shows that mortgages (as cash outflows) are negative for both females and males and have an opposite sign of cash inflow, as expected.[18] Both are insignificant. Together with the fact that coefficient estimates of nonlabor income remain positive and significant for both females and males across all four specifications, the results provide evidence supporting that cash inflows work in the opposite way as stock wealth.

The opposite effects of the cash inflow and stock forms of wealth are interesting and unique. Although our findings are at odds with the popular results in the literature where stock wealth diminishes the incentive to work, we believe this might be a particular phenomenon pertaining to China. China has a strict residential permit system (Hukou) in which households cannot freely migrate to cities or between housing markets. In addition, housing booming and development are highly regulated by authorities. An individual can accumulate vast housing wealth simply because he or she is a legal resident of a booming market (contrary to residents of rural areas or less developed areas). Therefore, individual housing wealth development is purely exogenous in these respects.

The cash flow wealth enters the contemporary budget constraint directly, which determines the "affordability" of not working. In contrast, as discussed above, the sharp increase in the real estate market in China has boosted individuals' stock wealth, but this only exists on paper unless realized through sale. The wealth in stock thus does not change the budget constraints for job decision-making directly. Furthermore, it is obvious that in cities with booming housing markets, the living cost increases considerably when housing prices rise. Even with higher wealth values of housing, unless they are liquidated, individuals would need to face the rising prices of commodities and living costs. Therefore, the stock form of wealth in housing would induce the incentive to work in this context. Stock wealth may also signal the potential "opportunity benefit" of working by accumulating wealth or signal the better ability of people who choose to work to accumulate wealth, e.g., buy more houses.

5.4. Estimating the model with the method of moments

The method of moments is used to estimate dynamic structural models, for which constructing the likelihood function is impossible or computationally intractable (see Low and Meghir, 2017; Blundell et al., 2016; Low et al., 2010). We carry out the method of moments in this section to perform robustness checks for our static structural model.

In our analysis, guided by the regression tree results, we divide our samples into four education groups: noncompletion of preliminary school, preliminary school to high school completion, university/college education, and postgraduate education. We also use marriage status as another dimension by which to group, i.e., married, and unmarried. First, for the education grouping, within each group, we sum up the absolute difference between the model-predicted probability of the individuals being observed on the labor market and the corresponding education-group-sample-average market participation rate. Second, across the education groups, we again sum up the group sums after weighing their sample shares. Third, for marriage status, we carry out procedures that are similar to those used for the education grouping. Finally, we minimize the target function, i.e., the total of the summed absolute differences across the education group and the marriage status group.

The results are presented in Table 5. Compared with the first two columns in Table 3, the coefficient estimates are similar,

[17] The limited observations of inherited housing value made it difficult to estimate the standard errors for the labor demand variables. The likelihood at maximum is estimated as −1777.

[18] We also tried to only include houses that no longer have mortgages. The result still holds.

especially for our variables of focus: nonlabor income, housing values, and value of financial investments. All coefficients for the job offering functions are significant for both genders. Unlike in Table 3, the coefficients for medical expenditures and wage equations are smaller and insignificant in Table 5. The log-likelihood improvement measured by McFadden's ρ is better in Table 5, especially for females – 0.56 in Tables 3 and 0.78 in Table 5. However, since McFadden's ρ is not a mathematically approved measure, we use it here only as an indication. We conclude that the coefficient estimates of our focus are quite consistent between the two methods.

5.5. Forming the wage expectation using the regression tree approach

To explore the alternative approach of wage Equation (3), we again resort to using the regression tree approach (CIF). This approach is used to relax the functional form assumption of the wage equation. In addition to the variables already in the linear wage equation, we include dummy variables for province, years of schooling, age, marital status, number of children, age of the youngest child, and health status as the independent variables. Instead of working experience (as used in the linear wage equation), in the CIF analysis, we include age, which has an equivalent power in terms of explaining the wage but also a more intuitive explanation. Because the CIF can incorporate the nonlinear relationship between variables, there is no need to include the squared age. We only use females as an example to examine the impacts of different wage equation approaches. The results are presented in Figure A1 in Appendix A. The ending nodes are the predicted average wages for the group of individuals with the shared characteristics shaped by the branches leading to the ending nodes. Years of schooling is the leading variable in determining the wages of females. Additionally, age, number of children, and province all affect females' wages. For comparison with the linear wage equation, we use the standard absolute residuals of the predictions: 0.7498 for the linear wage equation and 0.73 for the CIF. A smaller standard absolute residual of the CIF indicates that the prediction of the CIF is slightly better. Using the estimate results presented in Figure A1, we predict the wage for each individual (both working and not working) and use it to estimate the probability of labor supply. As shown in the first column of Table 6, the expected wage has a significant positive impact on the probability of working but at a smaller value than that of the baseline case in Table 3. All the other estimates are similar to those in the baseline model. Additionally, the log-likelihood value at the maximum is −1794.32, which is smaller than that of the baseline at −1770.74, indicating that the baseline model outperforms the regression tree-based wage expectation approach in terms of explaining power.

Another commonly raised concern associated with the wage equation in labor economics is self-selection bias. However, we believe that carrying out the self-selection correction in forming the wage expectation in our analysis is inappropriate. First, let us review the commonly used Heckman's two-stage approach (Heckman, 1979). The first stage is to run a probit model with the potential job market participation decision variables. The second stage is to introduce the estimated probability of working into the error term for the wage equation to correct the selection bias. As a result, the wage in the wage equation is observed to be conditional on individuals choosing to work. In our context, if we employ Heckman's two-stage approach to form the wage expectation, the labor supply function in our modeling framework would overlap with the first-stage job participation regression. As an analogy, doing so is similar to estimating the impacts of different factors on the job participation decision with those exact impacts already estimated in the first place. The self-selection corrected wage would reflect the equilibrium of labor supply and demand already. Therefore, we believe that wage expectations can only be formed based on the observed wages of those who chose to work and stick to our baseline approach without correcting the self-selection.

5.6. Simulating the effect of gender on labor demand and supply

In this section, we use the coefficient estimates from the baseline model in Table 3 to simulate the probabilities of females being willing to work and being offered a job by using the coefficient estimates for both females and males, respectively. That is, we treat a female in the labor market as if she were a male by artificially changing the gender of females to males. Since the other characteristics are unchanged for each individual, the differences between these two scenarios can provide an assessment of the gender differences in the job market.

As shown in the first two cells of Table 7, if the females were males, their probability of willing to work would increase from 73.5% to 76.7%, which is the net effect of the positive contributions from the age of the youngest child, marriage, and wealth in housing and the negative contributions from lower work incentives associated with the wage expectations and medical expenses of males. For the probability of getting an offer, surprisingly, we found that if the females were males, the probability of getting an offer would decrease from 99.8% to 99.3%, which might be due to the much lower coefficient estimate for education for the males (17.06) compared to that for the females (42.32). The pattern revealed by the simulations for the males is consistent with that for the females; if the males were females, they would be less willing to work but only have a marginal decrease in probability from 90.4% to 89.4%, whereas they would be offered a job with a higher probability, i.e., from 99.2% to 99.8%. Such findings support Stevenson's (2016) recent argument about the United States that "manly men need to do more girly jobs" to encourage more men to take service jobs in occupations that have been traditionally dominated by women.

Furthermore, as demonstrated by the high job offer probabilities for both females and males, all above 99%, the job opportunities in urban China are ample. A probability of 99% means that an individual will immediately receive a job offer after searching. The lower labor market participation rates were mainly driven by the willingness to work due to family considerations or wage expectations. The tight labor market might help to explain why there are no gender differences in the probability of getting a job.

The relatively high average probability of obtaining an offer shown in Table 7 does not necessarily mean that people would be offered a job without their education and work experience being considered. For example, based on our coefficient estimates of the model, for a female with no education and 20 years of working experience who is in Jilin Province, which has an unemployment rate of 3.8% (at the middle level compared to the other provinces), her simulated probability of getting an offer is 46%, which is far from the 99.79% presented in Table 7. The high average probability of obtaining an offer across the sample is reflected in the individuals who have favorable levels of education and working experience.

We revisit the higher unconditional labor income of males revealed in summary statistics in Table 1. Part of the reason behind this is that labor regulations in China regulate retirement age differently for females and males: 55 years old for females and 60 years old for males. The regulation-induced seniority difference naturally leads to higher wages for males, as revealed by the labor income statistics in Table 1. However, when we simulate the wages for females and males with the same characteristics based on the wage equation results in Table 2, the result is surprising. For instance, a female with 12 years of education and 15 years of work experience in Anhui province is estimated to earn 6468 CNY, while a male with the same education and work experience is estimated to earn 5091 CNY. Therefore, if we control for education and experience, females' wages are not lower than those of males and may be higher. Thus, the apparent gender wage gap in China has different roots than that in other countries.

In contrast to the discrimination and statistical discrimination documented in the previous literature (Phelps, 1972; Arrow, 1973, Aigner and Cain, 1977; Coate and Loury, 1993), our findings provide no evidence of "unobserved" discrimination. As we have shown above, simulated by the wage equation, females in urban China would have higher wages compared to males with the same qualifications and a slightly higher probability of being hired. The observable discrimination is induced by regulation: the retirement age for males is 60 years old, whereas that for females is 55 years old. The resulting difference in seniority between genders leads to differences in labor income.

The results we presented in Section 5.1, including Tables 2 and 3, are the baseline results for our analysis. Sections 5.2 to 5.5 serve as robustness checks of the baseline results by allowing correlations between job seeking and job obtainment, replacing the maximum likelihood method with the method of moments to estimate the baseline model, using alternative housing values to represent housing wealth, and replacing the OLS with the regression tree to form wage expectations. The consistency between the robustness checks and our baseline results provides strong support for our findings. In Section 5.6, we employ the baseline model to simulate the gender impacts on the probability of working and the probability of getting a job, which helps to clarify that the labor income gaps between females and males observed in the statistics are not due to unobserved discrimination but to differences in retirement age regulation.

6. Conclusion

In this paper, we propose a probabilistic cross-sectional structural model to disentangle both labor supply choices and labor demand choices simultaneously. This model enables us to identify and estimate labor supply preferences, as well as the labor demand preference of employers, by only using cross-sectional microdata, which extends the labor-market analytical toolkits to adapt to environments where data are limited, flexibility in working hours is lacking, and structural changes are present, as is the case in most emerging and low-income countries. Thus, our model can potentially be used for estimating model-based unemployment rates in countries without formal and systematic labor market survey systems, in countries with limited panel data or repeated cross-sectional data, or in countries experiencing structural changes.

As an empirical application of the model, we re-examine the differences in labor market supply and demand preference between males and females and the impact of different types of household wealth on work preferences by using the cross-sectional 2011 China Household Finance Survey data.

The following types of inquiries can be explored using our modeling framework and dedicated to the data we found. First, the number and age of children affect the labor supply decisions of females but not those of males. The lower participation rate of females is mainly driven by lower utilities of working rather than by fewer job opportunities. Second, the education and work experience of females affect employers' job offering decisions more than those of males do. However, the unconditional wage gap observed is mainly driven by the regulated higher retirement age for males. In contrast to common belief, neither the simulated wages nor the probabilities of obtaining a job offer for females are lower than those of males with the same attributes in China. Third, different types of household wealth affect labor supply decisions in different ways. Household wealth in the 'flow' form, especially labor income from other family members, tends to discourage individuals from working, while household wealth in the 'stock' form, such as housing and investment in financial assets, motivates both males and females to choose to work.

Finally, individuals with poor health status implied by higher medical expenses, which could be understood as a special form of 'outflow' wealth, have a strong willingness to work.

One concern with regards to the third finding is that the result might be driven by the fact that the 'stock' form of wealth represents a higher cost of consuming housing. The higher cost of consuming housing is commonly found to increase individuals' incentives to work. To alleviate this concern, we include housing rentals and mortgages as an additional control variable to control the cost of consuming housing in Columns (3) and (4) in Table 4. The coefficient estimate of the value of housing remains negative and significant.

We do subscribe to the prevalent view that stock wealth diminishes the incentives of labor supply. In this case, our results might indicate some nuanced effects of stock wealth. This could be a reflection of the limitations of cross-sectional data and the absence of housing prices in the model. On the other hand, we offer some possible explanations of our findings by suggesting that stock wealth might be an indication of a higher cost of housing consumption, as well as a spillover of a general increase in living costs in a booming housing market. It might also involve the income effects and investment incentives of households, seeing that housing is not only a consumption of living but also an object-based investment. In such cases, the stock wealth of housing might incentivize workers to work more so that they can further invest in the housing market. Of course, it is important to have richer data and extensive modeling to capture such effects, if any such effects exist.

Furthermore, as we use the cross-sectional approach to examine the current housing wealth, current housing tenure choice (i.e., renting versus living in self-owned housing), and the number of properties owned, despite the strict residential permit system that enhances the exogeneity or likelihood of legal residents of a booming market accumulating vast housing wealth, the labor market supply choices and housing choices might be driven directly by the same wealth and income profiles. In such cases, our empirical findings would be subjected to the caveat of endogeneity. Again, in this paper, our main purpose is to propose a theoretical model that enables us to use cross-sectional data to disentangle labor supply and demand choices simultaneously. The empirical analysis showcases the application of the theoretical model. With better data, such endogeneity would be of less concern.

We suggest a possible explanation via the institutional limitation of mobility such that—at least seen by households—housing wealth accumulation might be exogenous (to an extent). This could certainly be a unique phenomenon for China. Before the housing reform, all housing was provided by the government based on the hukou system. In the early stage of housing market reform in the late 1990s, residents could buy the housing they resided in for a symbolic sum based on residential permits. Thus, we suggest that the initial housing wealth and later housing wealth accumulations are somewhat exogenous in this context. It is certainly true that when considering the whole housing market and institutions, wealth accumulation and labor supply are endogenous. We hope our contributions will shed light not only on the unique situation in China but also perhaps the situations present in other emerging economies.

Our analysis is helpful for understanding the ever-changing labor markets, such as that seen in China, through the choices of both employees and employers, where labor forces have evolved from low skilled to high skilled and experienced, and the wages/labor costs have increased rapidly in recent years. With an aging population, decreasing labor market participation rates, and increasing personal/household wealth, understanding labor market behaviors is a challenging task. Our analysis hopefully contributes to the literature by providing a tool to better face such challenges. Finally, as mentioned several times in the paper, our choice modeling framework provides a useful approach for simultaneously measuring the preferences of both employees and employers, particularly when the data are limited.

CRediT authorship contribution statement

Xuehui Han: Methodology, Software, Writing – original draft. **Tao Zhang:** Writing – review & editing. **John K. Dagsvik:** Conceptualization, Writing – review & editing. **Yuan Cheng:** Supervision, Software, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A

Appendix B. Introduction of wage into labour demand and supply

Instead of separating utility and disutility for labour supply, and productivity vs cost for demand function, we can define a unified utility for worker, and unified payoff for employer.

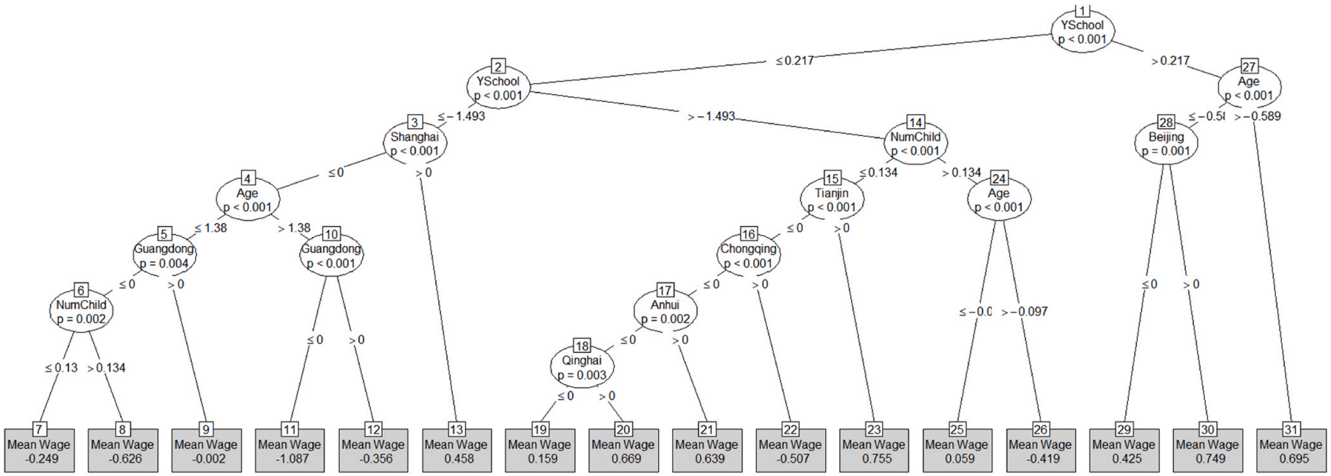

Fig. A1. Regression Tree Analysis for Wage Expectation - Female
Note: all the variable values are standardized.

$$U^* = U(W^*, Z) + \varepsilon^*$$

$$V^* = V(P, W^*) + \xi^*$$

where U^* is the utility of worker, and V^* is the utility of employer. W^* is the market wage which can be modelled as:

$$W^* = logW = X\beta + \eta$$

Then

$$U^* = U(W^*, Z) + \varepsilon^* = \theta logW + \gamma Z + \varepsilon^* = \theta X\beta + \gamma Z + \varepsilon^* + \eta$$

$$= \theta X\beta + \gamma Z + \varepsilon, where\ \varepsilon = \varepsilon^* + \eta$$

$$V^* = V(P, W^*) + \xi^* = X\delta - \tau \log W + \xi^* = X\delta - \tau(X\beta + \eta) + \xi^*$$

$$= (\delta - \tau\beta)X + \xi^* - \tau\eta = \lambda X + \upsilon, where$$

$$\lambda = \delta - \tau\beta, and\ \upsilon = \xi^* - \tau\eta$$

The worker participate the labour market when

$$\lambda(X, Z) = P(U^* > 0|X, Z) = P(\theta X\beta + Z\gamma \geq \varepsilon|X, Z) = \frac{1}{1 + exp(-Z\gamma - \theta X\beta)}$$

And the employer is willing to offer a job when

$$q(X) = P(V^* > 0) = \frac{1}{1 + \exp(-X\lambda)}$$

The identification is the same as in the baseline case, except for a minor change of sign of parameters. But (δ, τ) can not be identified.

Appendix C. Identification of the model

Let

(1) $P\ (X,Z)\ = \lambda(X,Z)q(X),$ .where

$$\lambda(X, Z) = \frac{1}{1 + \exp(Z\gamma - \theta X\beta)}, q(X) = \frac{1}{1 + \exp(-X\delta)}.$$

Recall that identification in this context means that when $P(X, Z)$ is given, as a function of (X, Z) , and Eq. (1) holds then the parameters $(\gamma, \delta, \theta\beta)$ are uniquely determined. Let Z_1 denote non-labor income (which is part of Z).

We have that

(2) $\frac{\partial \log\ P(X,Z)}{\partial Z_1}\ =\ -\ \gamma \frac{\exp(Z\gamma - \theta X\beta)}{1 + \exp(Z\gamma - \theta X\beta)}\ =\ -\ (1\ -\ \lambda(X, Z))\gamma.$

and

(3) $\frac{\partial^2 logP(X,Z)}{\partial Z_1^2}\ = \gamma \frac{\partial (X,Z)}{\partial Z_1}\ =\ -\ \gamma^2 \frac{\exp(Z\gamma - \theta X\beta)}{(1 + \exp(Z\gamma - \theta X\beta))^2}\ =\ -\ \lambda(X,Z)(1\ -\ \lambda(X,Z))\gamma^2.$

Define

$$g(X, Z) = \log\left(\frac{\{\partial \log\ P(X, Z)/\partial Z\}^2}{-\partial^2\ \log\ P(X, Z)/\partial Z}\right).$$

From (2) and (3) we obtain that

(4) $g(X, Z)\ = \log\left(\frac{1 - \lambda(X,Z)}{\lambda(X,Z)}\right)\ = Z\gamma\ -\ \theta X\beta..$

Since $P(X, Z)$ is assumed to be known then the respective partial derivatives of $P(X, Z)$ are also known. Therefore, the function $g(X, Z)$ defined above is also known. Since (4) is linear in γ and $\theta\beta$ it follows readily that $(\gamma, \theta\beta)$ is uniquely determined by (4). Recall that β is determined from the wage equation, and therefore θ is identified. Accordingly, $\lambda(X, Z)$ is identified and from (1) we therefore obtain that

$$(5)\ log\left(\frac{P(X,Z)}{\lambda(X,Z)-P(X,Z)}\right) = log\left(\frac{q(X)}{1-q(X)}\right) = X\delta.$$

which demonstrates that δ is identified. Thus, we have demonstrated that the model (together with the wage equation) is identified.

References

Arellano, Manuel, Meghir, Costas, 2006. Female labour supply and on-the-job search: an empirical model estimated using complementary data sets. Rev. Econ. Stud. https://doi.org/10.2307/2297863.
Arrow, K.J., 1973. The theory of discrimination. In: Ashenfelter, Orley, Rees, Albert (Eds.), Discrimination in Labor Markets. Princeton University Press, Princeton, NJ, pp. 3–33.
Autor, David, Dorn, David, 2009. This job is getting old: measuring changes in job opportunities using occupational age structure. Am. Econ. Rev.: Papers and Proceedings 99 (2), 45–51.
Autor, David, Dorn, David, 2010. Inequality and Specialization: the Growth of Low-Skilled Service Employment in the United States. MIT Working Paper.
Autor, David H., Dorn, David, 2013. The growth of low-skill service jobs and the polarization of the US labor market. Am. Econ. Rev. 103 (5), 1553–1597.
Bjorn, P.A., Vuong, Q.H., 1985. Econometric Modeling of a Stackelberg Game with an Application to Labor Force Participation. California Institute of Technology, Division of the Humanities and Social Sciences. Working Papers 577.
Blundell, Richard, Ham, John, Meghir, Costas, 1987. Unemployment and female labour supply. Econ. J. https://doi.org/10.2307/3038229.
Blundell, Richard, Costa Dias, Monica, Meghir, Costas, Shaw, Jonathan, 2016. Female labour supply, human capital, and welfare reform. Econometrica 84 (5), 1705–1753.
Byker, Tanya S., 2016. Paid parental leave laws in the United States: does short-duration leave affect women's labour-force attachment? Am. Econ. Rev. 106 (5), 242–246.
Chen, J., Han, X., 2014. The evolution of the housing market and its socioeconomic impacts in the post-reform people's Republic of China: a survey of the literature. J. Econ. Surv. 28, 652–670.
Coate, S., Loury, G.C., 1993. Will affirmative-action policies eliminate negative stereotypes? Am. Econ. Rev. 83, 1220–1240.
Dagsvik, J.K., 1994. Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. Econometrica 62, 1179–1205.
Dagsvik, J.K., 2000. Aggregation in matching markets. Int. Econ. Rev. 41, 27–57.
Dagsvik, J.K., Jia, Z., 2018. Aggregate Behavior in Matching Markets with Flexible Contracts and Non-transferable Representations of Preferences (No. 875). Discussion Papers.
Dagsvik, J.K., Strøm, S., 2006. Sectoral labour supply, choice restrictions and functional form. J. Appl. Econom. 21 (6), 803–826.
Dagsvik, John K., Kornstad, Tom, Terje Skjerpen, 2013. Labour force participation and the discouraged worker effect. Empir. Econ. https://doi.org/10.1007/s00181-012-0598-9.
Dagsvik, J.K., Jia, Z., Kornstad, T., Thoresen, T.O., 2014. Theoretical and practical arguments for modeling labor supply as a choice among latent jobs. J. Econ. Surv. 28 (1), 134–151.
Fairlie, Robert W., Krashinsky, Harry A., 2012. Liquidity constraints, household wealth, and entrepreneurship revisited. Rev. Income Wealth. https://doi.org/10.1111/j.1475-4991.2011.00491.x.
Goldin, C., Kerr, S.P., Olivetti, C., Barth, E., 2017. The expanding gender earnings gap: evidence from the LEHD-2000 census. Am. Econ. Rev.: Papers and Proceedings 107 (5), 110–114.
Han, X., 2010. Housing demand in Shanghai: a discreet choice approach. China Econ. Rev. 21, 355–376.
Han, X., Wei, S.J., 2017. Re-examining the middle-income trap hypothesis (MITH): what to reject and what to revive?. In: Journal of International Money and Finance. Elsevier, pp. 41–61, 73(PA).
Han, X., Dagsvik, J., Cheng, Y., 2020. Disability and job constraint in post civil war Cambodia. J. Dev. Stud. 56 (12), 2293–2307.
Heckman, J.J., 1974. Shadow prices, market wages and labour supply. Econometrica 42, 679–694.
Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.
Holtz-Eakin, D., Joulfaian, D., Rosen, H., 1993. The Carnegie conjecture: some empirical evidence. Q. J. Econ. 108 (2), 413–435.
Hothorn, T., Zeileis, A., 2015. Partykit: a modular toolkit for recursive partitioning in R. J. Mach. Learn. Res. 16, 3905–3909.
Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph Stat. 15 (3), 651–674.
Ilmakunnas, S., Pudney, S., 1990. A model of female labour supply in the presence of hours restrictions. J. Publ. Econ. 41, 183–210.
Jacob, Brian A., Ludwig, Jens, 2012. The effects of housing assistance on labour supply: evidence from a voucher lottery. Am. Econ. Rev. 102 (1), 272–304.
Kleven, Henrik, Landais, Camille, Egholt Søgaard, Jakob, 2018. Children and Gender Inequality: Evidence from Denmark. NBER Working Paper No. 24219.
Low, H., Meghir, C., 2017. The use of structural models in econometrics. J. Econ. Perspect. 31, 33–58.
Low, H., Meghir, C., Pistaferri, L., 2010. Wage risk and employment risk over the life cycle. Am. Econ. Rev. 100, 1432–1467.
Lundberg, U., Mardberg, B., Frankenhaeuser, M., 1994. The total workload of male and female white-collar workers as related to age, occupational level, and number of children. Scand. J. Psychol. 35, 315–327.
McFadden, D., 1978. Modeling the choice of residential location. Transport. Res. Rec. 672, 72–77.
McFadden, D., 1984. Econometric analysis of qualitative response models. In: Grilliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. II. Elsevier Science Publishers BV, New York, pp. 1395–1457.
Menzel, K., 2015. Large matching markets as two-sided demand systems. Econometrica 83, 897–941.
Phelps, E.S., 1972. The statistical theory of racism and sexism. Am. Econ. Rev. 62, 659–661.
Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer-Verlag, Berlin.
Stevenson, Betsey, 2016. Manly men need to do more girly jobs. Bloomberg View December 7, 2016. https://www.bloomberg.com/view/articles/2016-12-07/manly-men-need-to-do-more-girly-jobs.
Tominey, Emma, 2016. Female labour supply and household employment shocks: maternity leave as an insurance mechanism. Eur. Econ. Rev. 87, 256–271.
Van Soest, A., 1995. Structural models of family labor supply. A discrete choice approach. J. Hum. Resour. 30, 63–88.
Van Soest, A., Das, M., Gong, X., 2002. A structural labour supply model with flexible preferences. J. Econom. 107, 345–374.
Xie, Y., Jin, Y., 2015. Household wealth in China. Chinese Sociol. Rev. 47 (3), 203–229.
Zhao, Lingxiao, Burge, Gregory, 2016. Housing wealth, property taxes, and labour supply among the elderly. J. Labour Econ. 35 (1), 227–263.