Dynamic Female Labor Supply: Applications in China*

Heather Chen

MA Program in Computational Social Science
The University of Chicago

Abstract

Keywords: natural language processing, international relations, united nations

1. Methods

The models in this paper are mainly based on Eckstein and Wolpin (1989). This simplified model only takes the subsample of married women and considers their labor market participation decision in the post child-bearing period. After presenting the detailed specification of our model, a brief introduction of estimation methods is provided.

1.1. Simplified Model

In this section, we will consider the simplified model which only takes into account of married women and their labor market choices. In each discrete period t, the model supposes the household maximizes its expected utility over a finite total period T by choosing whether the wife works ($p_t = 1$) or not ($p_t = 0$) and whether to have an additional children (n_t). Under such assumption, the household is targeted to maximize the following equation:

$$E_t\Big[\sum_{k=0}^{zt-t} \delta^j U(p_{t+k},\ N_{t+k,j},\ x_{t+k},\ K_{t+k-1},\ S)\Big] \tag{1}$$

where p_t is a dummy variable that indicates whether the women works or not at period t; K_t indicates the women's former experience which is measured by the number of periods the women has worked before; x_t denotes the consumption at that period; and $N_{t,j}$ indicates the number of children of age group j at period t. The total number of periods considered, T, is determined by a fixed retirement age, which is 50 for urban women in China.

Given the utility maximization problem above, the household faces the following budget constraint:

$$y_t^w p_t + y_t^H = x_t + \sum_{j=1}^{J} c_j N_{t,j} + bp_t \tag{2}$$

*This is the course paper for MACSS perspective research class economics track. The course instructed by Prof. Richard Evans at the University of Chicago.

where the total income of the household at period t is equal to its total consumption. In equation (2), y_t^w and y_t^H denote the wife and husband's annual income at period t, respectively. The household consumption is consisted of three parts: commodity consumption x_t, the aggregate consumption spent on children (where c_j indicates the cost per child of age group j), and a fixed cost of women working b. By demonstrating the household budget constraint as the above, we are implicitly assuming that the wife's labor market decision is exogenous of her husband's income.

The wife's earnings at period t is subject to the standard Mincer earning function, which is given by

$$\ln y_t^w = \beta_1 + \beta_2 K_{t-1} + \beta_3 K_{t-1}^2 + \beta_4 S + \beta_5 R_t + \varepsilon_t \tag{3}$$

where R_t is an indicator of whether the women lives in rural or urban areas, which is an important factors that needed to be taken account of when considering individual income in China. All the other variables except ε_t in equation (3) are the same as previously defined. ε_t is the random effect of the wife's earning that is unaffected by the household decision process. It has a standard normal distribution with zero mean.

The number of children of age group j is determined by the following equation:

$$N_{t,j} = N_{t-1,j} + n_{t,j} - d_{t,j} \tag{4}$$

Equation (4) can be simply understood as the number of children of age group j at time period t is equal to the number of children of age group j at last period plus the number of kids who enter into this age group, $n_{t,j}$, subtracts the number of kids who leaves this age group, $d_{t,j}$.

One's working experience is written as a cumulative function of previous experience:

$$K_t = K_{t-1} + p_t \tag{5}$$

For simplicity, here we assume that the household utility for a single period t has taken a linear form, according to the general form given in equation (1),

$$U_t = \alpha_1 p_t + x_t + \alpha_2 p_t x_t + \alpha_3 p_t K_{t-1} + \sum_{j=1}^{J} \alpha_{4,j} N_{t,j} p_t + \alpha_5 p_t S + f(N_{t,i}) \tag{6}$$

Here $f(N_{t,i})$ is an unspecified term that illustrate how the utility of children enters into the total household utility function. As mentioned by Eckstein and Wolpin in their 1989 paper, there are a few coefficients in equation (6) that needs to be put special emphasis on. α_3 implies the marginal utility of leisure time, with $\alpha_3 < 0$ reflects diminishing marginal utility for non-market leisure time within the life cycle, and $\alpha_3 > 0$ reflects habit persistence that the women who work previously gain utility by continuing working in the current period. α_2 provides a test for wealth maximization. If we cannot reject the hypothesis that $\alpha_2 = 0$, then the individual will be better off with the capital market constraint.

We will use dynamic programming to solve this life-cycle utility maximization problem. Here we regard the maximum expected life-time utility given by equation (1) as $V_t(K_{t-1},\ \varepsilon_t,\ \Omega_t)$, where ε_t represents the random wage draw given by earnings equation (3), and Ω_t consists of all other factors, including schooling S, the age distribution of children $N_{t,j}$, and her husband's income $\overline{y}_t^H$. At any time period t, the woman makes a discrete choice between participating in the labor market and not participating, that is

$$V_t(K_{t-1},\ \varepsilon_t,\ \Omega_t) = \max\Big[V_t^1(K_{t-1},\ \varepsilon_t,\ \Omega_t),\ V_t^0(K_{t-1},\ \varepsilon_t,\ \Omega_t)\Big] \tag{7}$$

In equation (7), $V_t^1(\cdot)$ implies the household expected utility when the wife chooses to work at period t, and $V_t^0(\cdot)$ implies the expected utility when she decides not to. As a result, equation (7) simply means the household maximizes its expected life-time utility based on whether or not let the wife to work in the current period. Since we are solving this problem using backward induction, we consider the maximization problem at final period T. Given all the specifications above, we may write $V_t^0(\cdot)$ and $V_t^1(\cdot)$ as the following:

$$\begin{aligned} V_T^1(K_{T-1},\ \varepsilon_T,\ \Omega_T) = \alpha_1 + (a+\alpha_2)\Big(\exp\{\beta_1 + \beta_2 K_{T-1} + \beta_3 K_{T-1}^2 + \beta_4 S\ + \beta_5 R_t + \varepsilon_T\} + \overline{y_T^H} - \sum_{j=1}^{J} c^j N_{T,j} - b\Big) \\ + \alpha_3 K_{T-1} + \sum j = 1^J \alpha_{4,j} N_{Tj} + \alpha_5 S\ + f(N_{T,j}) \end{aligned} \tag{8}$$

$$V_T^0(K_{T-1},\ \varepsilon_T,\ \Omega_T) = \overline{y}_T^H - \sum_{j=1}^{J} c_j N_{T,j} + f(N_{T,j}) \tag{9}$$

If the utility the household can get when the wife works $V_T^1(\cdot)$ exceeds that of when she does not work $V_t^0(\cdot)$, then the wife would decide to participate in the labor market at period T. Therefore, knowing her decision logistic, equation (8) and (9) can be simplified as:

$$\begin{aligned} p_T\ = 1 \iff \varepsilon_T \geq \ln\Big[-\alpha_1 - \alpha_2(\overline{y}_T^H - \sum_{j=1}^{J} c_j NT, j + b(1) + \alpha_2) - \alpha_3 K_{T-1} \\ - \sum_{j=1}^{J} \alpha_{4,j} N_{T,j} - \alpha_5 S\ - (\beta_1 + \beta_2 + \beta_3 K_{T-1}^2 + \beta_4 S\ + \beta_5 R_t)\Big] \end{aligned} \tag{10}$$

$$p_T\ = 0 \quad otherwise \tag{11}$$

Similarly, for any period t in between, the wife's utility giving the two choices can be written as:

$$\begin{aligned} V_t^1(K_{t-1},\ \varepsilon_t,\ \Omega_t) = \alpha_1 + (a+\alpha_2)\Big(\exp\{\beta_1 + \beta_2 K_{t-1} + \beta_3 K_{t-1}^2 + \beta_4 S\ + \beta_5 R_t + \varepsilon_t\} + \overline{y_t^H} - \sum_{j=1}^{J} c^j N_{t,j} - b\Big) \\ + \alpha_3 K_{t-1} + \sum_{j=1}^{J} \alpha_{4,j} N_{t,j} + \alpha_5 S\ + f(N_{t,j}) + \delta E V_{t-1}(K_t = K_{t-1} + p_t) \end{aligned} \tag{12}$$

$$V_T^0(K_{t-1},\ \varepsilon_t,\ \Omega_t) = \bar{y}_T^H t - \sum_{j=1}^{J} c_j N_{t,j} + f(N_{t,j}) + \delta EV_{t-1}(K_t = K_{t-1}) \tag{13}$$

which can be simply understood as the estimated utility at current period t add the expected utility the household might get for the future periods if she makes that decision. Given the utility value functions above, we may construct the decision function as below:

$$p_t = 1 \iff \varepsilon_t \geq \ln\Big[-\alpha_1 - \alpha_2(\bar{y}_t^H - \sum_{j=1}^{J} c_j N_{t,j} + b(1+\alpha_2) - \alpha_3 K_{t-1} - \sum_{j=1}^{J} \alpha_{4,j} N_{t,j} - \alpha_5 S + \delta(EV_{t+1}(K_{t-1}) - EV_{t+1}(K_{t-1} + p_t)\Big] - (\beta_1 + \beta_2 + \beta_3 K_{t-1}^2 + \beta_4 S + \beta_5 R_t) \tag{14}$$

$$p_t = 0 \quad otherwise \tag{15}$$

Here, we have listed all the equations needed for estimation.

1.2. Estimation Methods

As long as we have stated all the methods and provide a brief introduction of data in the previous sections, now we came to present the methods for estimation. Since the Mincer earning equation (3) of the wife's income does not involve decision at the current period, we may easily obtain the wage parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, and the random effect ε_t for each individual. Also, for simplicity, using the data in the final period T, other parameters of interest, α_2, α_3, α_5, $\alpha_1 + b(1+\alpha_2)$, and $\alpha_2 c_j - \alpha_{4,j}$ $(j = 1, ..., J)$ can be further identified. Since the parameters in the budget constraint, b and c_j cannot be directly estimated and is implicitly expressed in the parameters of utility maximization, we will here assume $b = 0$ and $c_j = 0$ without loss of generality.

Generalized Method of Moments (GMM) is employed in all estimations in this analysis.

2. Data

In this paper, we utilize 10 waves of data from China Health and Nutrition Survey (CHNS) in 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, and 2015. CHNS dataset is the longest longitudinal survey that contains detailed household and individual data in China. Its baseline survey in 1989 covers 15,907 indivuduals within about 3,795 households. The number of participated households increases to over 7,000 in 2015. In all, there are 11,130 households and 42,829 individuals who ever participated in CHNS survey. Individual-level data of our interest includes information about the interviewee's current working status, marital status, annual income, and education level. Household-level data of interest includes how many children and their ages the household has and its living area (urban or rural). The reason why we choose this dataset is rather simple direct, since it is the only household-level survey dataset that contains longitudinal data for more than 10 years (The population census data conducted by Chinese government are not available for public academic use). Also, its sampling communities ranges across 15 provinces in China, making the dataset rich in regional context for analysis. Last but not least, CHNS dataset has an unique "ever-married women" dataset that includes all participated women that have ever been married. Using this dataset as the basis, we are able to link each mother to her children as well as the child's basic information.

However, CHNS dataset also has it inherent drawbacks. Firstly, The survey is not conducted at an annual level, which implies we cannot obtain an accurate year-based working experience data of women. Also, the survey asks about the participant's current working status as "*Are you presently working?*". The survey question does not touch on the participant's employment status during the survey interval. As a result, we can only impute the experience variable in the following way, if the interviewee reports that she is currently working and she has worked before, then we will treat all the years during the gap of surveys as her "experience". If she reports not working currently but has work in the previous survey, we will assign half of the survey gap as her experience. If she does not report working for two consecutive surveys, then no experience gained. Secondly, the survey does not record individual schooling in a cumulative way, which means the survey in all does not have a continuous variable that represents the individual's years of schooling. Therefore, the schooling variable S using our data takes the form of categorical variable: no education, graduated from primary school, graduated from junior high school, graduated from senior high school, technical or vocational degree, college degree and above.

In the following two subsections, we shall first present sample selection and descriptive statistics for the data, then a brief introduction for estimation strategies is followed.

2.1. Sample Selection and Descriptive Statistics

Using 10 waves of CHNS data as introduced above, we first calculate the labor market participation rate for all married adults.

As we can see from Figure 1, the labor market participation rate for both married men and women has declined significantly from 1989 to 2015. The labor market participation rate of married women has decreased from over 80% in 1989 to a little above 40% in 2015. The overall trend corresponds to what we have seen in introduction part using data from the World Bank, but the decline is larger in magnitude using CHNS data.

Moreover, if we emphasize on the labor market participation rates for women of different marital status, we may generate similar trends as follows: As illustrated in Figure 2, even though the labor market participation rate of never married women and married women demonstrate overall trend of declining, the rate of married women has decreased more drastically than that of never married women. What have shown in the two figures above has accentuated the

Figure 1: Labor market participation rate for married participants

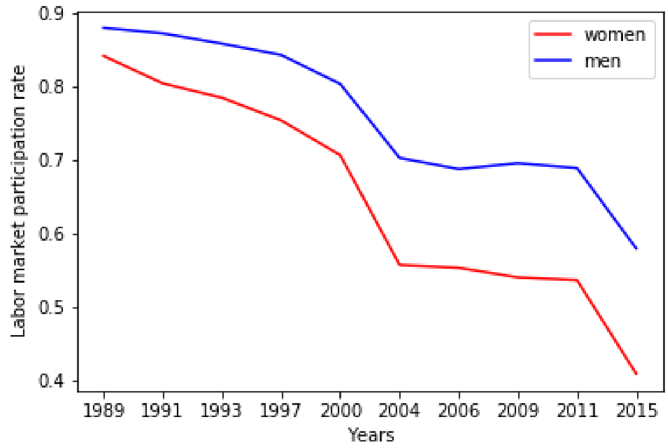
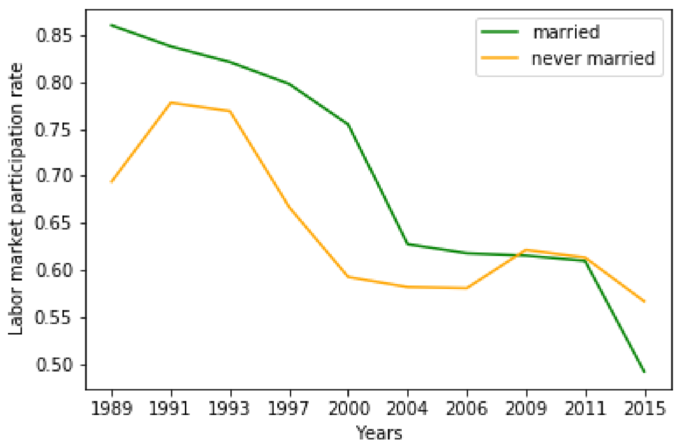


Figure 2: Labor market participation rate for women of different marital status



importance of our analysis, as we are aiming at using dynamic programming to unmask the reason behind decreasing female labor participation in China, especially for married women.

The paper of Eckstein and Wolpin (1989) selects a subsample of 318 women who age between 39 to 44 (post child-bearing) in 1967 and have at least 4 consecutive year of employment information. The horizon T is set to 60. In our sample, since we have longer time range, and the usual retirement age for female workers in China is 50. Thus our sample selection criterion is listed as follows: we include ever-married women who age between 35 to 40 in any sample year expect 2009, 2011, and 2015 with at least 3 waves of consecutive information about their employment status and income information. If a women has not participated one of the surveys in between, her survey data after the gap will be taken into account. The selecting criterion also makes sure that any new survey interviewees who meet the standard will also be included in the sample. Also, given the unique urban-rural stratification in China, [1], we set the fixed retirement age (when a final decision about whether to work or not has to be made) as 55 for all sample, and 60 for the subsample of rural women. The distribution of our data is listed as in Table 1.

Table 1: Distribution of participants by number of consecutive surveys participated

Number of surveys participated	3	4	5	6	7	8	9	10
Number of participants	352	293	220	122	102	35	47	38
Cumulative proportion of participants (%)	28.95	53.04	71.13	81.17	89.56	92.43	96.30	100.00

As we can see, there are in all 1216 female participants that meet our sample selection criterion in CHNS dataset, which constitutes of 5861 observations in all. The sample size is far more larger than that of Eckstein and Wolpin (1989), which only has a sample size of 318 women and 3020 observations. Of all the qualified participants, approximately half of them has consecutive employment and income data for more than 5 surveys. Also, 38 participants have engaged in all 10 waves of surveys. Table 2 presents the descriptive statistics for all the observations. The descriptive statistics are obtained from the 5861 observations. For most

Table 2: Descriptive Statistics

	Variable name	Sample Mean	Standard deviation	Sample mean for rural areas	Sample mean for urban areas	t-statistics
0	Experience	5.74	5.37	5.90	5.31	3.66***
1	Age	43.18	6.07	43.27	42.94	1.83*
2	Husband Income	12243.04	18224.16	11457.95	14477.17	-5.55***
3	Degree	1.35	1.31	1.15	1.92	NaN
4	Number of Children under 6	0.12	0.41	0.13	0.09	3.56***
5	Number of Children from 6 to 18	1.05	1.02	1.10	0.93	5.50***
6	Annual Income	9219.29	13228.88	8583.70	11028.02	-6.20***

of our variables of interest, there exist significant difference between rural women and urban women. Generally speaking, women in rural areas works more often, thus have more working experience than urban women. However, their income, including their own annual income and husbands' income, are significantly lower than those of urban women. Also, from the data regarding to children we can see that, urban women give birth to children relatively later than rural women, thus making them have more children age under 6 years old and less children age between 6 to 18 years old. Most of the women have a degree of junior high school.

[1]Female workers in urban China conform to strict policies of retirement at age 50 or 55 (based on their position at work), while rural women in China usually engage in agricultural work in their household's own land until very late in life.

3. Results

3.1. Parameter Estimation Results

Table 3: Generalized Method of Moments Estimation Results

Parameters	Logistic Regression (1)	Bayes Logistic Regression (2)
α_1	1.266×10^{15}	1.058×10^{15}
α_2	-3.896×10^{8}	-3.065×10^{9}
α_3	2.154×10^{13}	2.777×10^{13}
α_{41}	-1.305×10^{13}	3.764×10^{14}
α_{42}	-3.980×10^{14}	-9.879×10^{13}
α_5	3.871×10^{13}	-3.730×10^{13}
β_1	8.758	-
β_2	-0.004	-
β_3	0.001	-
β_4	0.083	-
β_5	0.013	-

Parameter estimation results are presented in Table 3. The relative number of these estimated parameters are not of our interest and also do not have explicit meanings, especially for those parameters in logistic model. We mainly focus on interpreting the signs of parameters in the table above. The t-tests that the α parameters are equal to zero are all rejected at 1% level.

Looking at the first column – logistic regression – of the table above, $\alpha_1 > 0$, which implies in China, labor market participation actually increases women's utility. The result is different from conclusions obtained from American data, where women gain disutility from the labor market. $\alpha_2 < 0$ indicates that the increase in husband's income (or household total consumption) generally discourage women from participating in the labor market. Then we have $\alpha_3 > 0$, another variable that is distinct from American data. Chinese woman's utility in working rises when she has more previous experience. Having children, regardless of their age, generally reduces mother's utility from work as $\alpha_{41} < 0$ and $\alpha_{42} < 0$. From the result we also have $\alpha_5 > 0$, woman with higher education are more willing to work, which correspond to the basic understanding.

While using R to gain these estimation results, we have encountered a problem that the data somewhat suffers perfect separation issues. We try to alleviate this problem by using Bayes logistic regression, the results are presented in column (2) of Table 3. For consistency in explanation, we will still use parameters obtained in column (1) for goodness of fit test as well as simulations.

3.2. *Goodness of Fit Test for Estimation Results*

In this section we present a goodness of fit test for the estimation parameters. Given the

Table 4: Actual and predicted labor participation rate by age group and experience

Experience	All ages		39-45		46-50		50+		$\chi^2(row)$
	A	P	A	P	A	P	A	P	
0-5	0.847	0.804	0.863	0.800	0.434	1.000	0.571	1.000	1.06
6-10	0.870	0.811	0.967	0.737	0.794	0.920	0.312	1.000	1.60
11-15	0.862	0.949	-	-	0.960	0.931	0.710	0.977	0.083
15+	0.768	0.983	-	-	-	-	0.768	0.983	-

results in Table 4, the method of dynamic life-cycle modeling generally does not provide a good prediction of labor market participation of different age groups and experience, even though the χ^2 test does not reject the hypothesis that the actual rates and predicted rates are the same.

If taken a closer look at predicted data, we can see that the prediction results are the worse when the women is beyond 50 years old, since there are few data lie within this dataset. And the actual labor market participation in China for elder women experienced a non-linear trend – it declines as the experience increases from 0-5 years to 6-8 years, and increases as experience further goes up to over 10 years. We anticipate that it might be the reason that the linear household utility model in this dataset cannot well capture the features in Chinese data. The prediction results are the best when the woman is between 39-45 years old with 0-5 years of experience, and when she is 46-50 years old with 11-15 years of experience. The following section will mainly use these two subsamples to obtain simulation results.

3.3. *Simulations*

4. Reflections and Extensions

References

Eckstein, Z., Wolpin, K.I., 1989. Dynamic labour force participation of married women and endogenous work experience. The Review of Economic Studies 56, 375–390.