Working Hours and Migrants' Settlement Intention: Evidence from China Migrants Dynamic Survey *

By Jingwen Tan

This thesis analyses the impact of the migrants' working hours in the inflow city on their settlement intentions using 2017 China migrants dynamic survey (CMDS) data of 117,152 workers. The results of the linear probability model (LPM) show that for 1% increase in the working hours of the migrant population, their settlement intention decreases by 0.0391%. The results are supported by instrumental variables as well as robustness checks. Mechanism analysis shows that the increase in housing expenditures weakens the effect of labour time on migrants' willingness to settle, while the increase in income-expenditure ratio strengthens the effect of labour time on migrants' willingness to settle. In addition, using the Karlson–Holm–Breen (KHB) method this thesis analyses the mediating effect of the social integration situation. Heterogeneity analysis reveals that the settlement intentions of the female, young person are more likely to be influenced by labour hours. Intention to settle is not sensitive to labour hours for agricultural migrants, but it is sensitive to labour hours for industrial and service migrants.
JEL: J00, J61, J62, J68, J83
Keywords: Migrants, Working Hours, Settlement Intention

I. Methodology

A. Data

This thesis uses data from 2017 China Migrants Dynamics Survey (CMDS), organized by the National Health Planning Commission for empirical analysis. The survey covered 31 provinces (municipalities and autonomous regions) in mainland China, and the sample was selected from the migrants who stayed in the area for more than one month. The survey was conducted according to a stratified, multi-stage PPS (Probability Proportionate to Size Sampling) sampling method proportional to the size of the migrant population.

In this thesis, the data were cleaned before empirical analysis. First, we excluded all samples under the age of 18 as well as age over 60. In China, citizens over the age of 18 are adults, have full capacity for civil conduct, and can independently conduct civil activities. The retirement age is 60 years old, and the settlement policies of most cities are targeted at the working population. Secondly, the population studied in this thesis is the migrants in cities with jobs, so this thesis remove the sample without jobs. Finally, we matched the micro-questionnaire data of the sample with the data from the urban statistical yearbook and retained the successfully matched sample. The total amount of sample before processing is 169989, the processed sample size with a total of 117,152.

* The authors acknowledge the data support provided by the National Health Care Commission of China.

The data relating to the urban control variables were obtained from the China Urban Statistical Yearbook 2017. China Urban Statistical Yearbook is an annual publication sponsored by the Department of Urban Social and Economic Surveys of the National Bureau of Statistics, which comprehensively reflects the economic and social development of Chinese cities. It contains data on population, resources and environment, economic development, science, technology and innovation, people's life, public services and infrastructure of more than 700 urban units in cities at prefecture level and above and county level cities nationwide in the previous year. The per capita GDP, industrial structure, proportion of urban green space, number of teachers and number of doctors among the control variables in this thesis come from the China Urban Statistical Yearbook. The description of all variables are shown in Table 1

Table 1—: Variable Description.

Variables	Description
settle	willingness to settle = 1, no = 0
workhour	logarised hours worked per week
gender	male = 1, female = 0
age	actual age
edu	years of education
hukou	rural household registration = 0, urban household registration = 1
marriage	unmarried = 0, married = 1
hincome	household income
pgdp	logarised gdp per capita
structure	proportion of people employed in tertiary industry
green	proportion of urban green area
pbed	number of hospital beds per 10,000 people
pteacher	number of primary school teachers per 10,000 people

B. Baseline Model

The focus of this thesis is on the impact of migrants' labour time in the inflow city on their willingness to settle. The question "Are you willing to move your household to the inflow city" in the CMDS questionnaire is used to measure migrants' willingness to settle. The indicator is a dummy variable (1=willing, 0=not willing).

Long working hours may lead to physical and mental fatigue and increased psychological stress, while migrants may find it difficult to have time to participate in social activities and integrate into urban society. Long working hours may also lead to reduced communication among family members and reduced family stability, as well as limiting the ability of migrants to pursue other opportunities for personal development. Therefore, excessive labour hours in a city may have a negative impact on migrants' willingness to settle.

With the dependent as the dummy variable, models that can be used include linear probability model (LPM) based on ordinary least squares (OLS) estimation and Logit, Probit models based on likelihood estimation. This thesis chooses to use a linear probability model in the baseline regression to estimate the effect of working hours on the intention to settle among the migrants, while the estimation results of the Logit model are reported in the robustness section. The interpretation of the coefficients of the linear probability model is more intuitive than that of the likelihood estimation. And with a reasonable model setup, the direction and significance of the coefficients obtained using least squares estimation are not significantly different from that of the likelihood estimation (Ferrer-i-Carbonell and Frijters, 2004; Angrist and Pischke, 2009). The linear probability model in this thesis can be expressed as

$$settle_{is} = \beta_0 + \beta_1 workhour_{is} + \beta_2 X_{is} + \mu_i + \lambda_s + u_{is} \tag{1}$$

where $settle_{is}$ is the willingness of the migrants to settle, and this variable is a dummy variable. $workhour_{is}$ is the logarithmic weekly labour hours of the migrants i. The labour time used in this paper is the labour time of each individual worker, not the average labour time in the city. X_{it} are other control variables. μ_i is the regional fixed effect, λ_s is the occupation fixed effect. This thesis control for the regional fixed effects at the provincial level while putting in control variables at the local municipality level.

Labor hours as the core explanatory variable is likely to be an endogenous variable. When migrants' willingness to settle in economically developed areas is higher, migrants may work frantically to earn higher incomes so as to have savings to buy a house and settle in the inflow city. In addition, when migrants do not want to settle in the inflow city, their willingness to work may be reduced, which in turn affects the number of working hours. Instrumental variables are widely used in the treatment of endogeneity problems.

This thesis choose "Whether the migrant have been injured or ill in the last two weeks" and "whether the migrant has hypertension or diabetes" as instrumental variables for labour time. Since the questionnaire surveyed the working hours of the migrant population in the last week, "whether they have been ill or injured in the last two weeks" and "whether they have two long-term illnesses" had a significant effect on the number of hours worked by the migrant population, which is in line with the correlation hypothesis of the instrumental variable. In addition, there is no evidence that these two variables have an impact on the willingness of the migrant population to settle. Hypertension and diabetes have well-established treatment programmes and are not directly life-threatening, so the influx of medical resources into the city does not affect the settlement decisions of migrants with these two diseases. This thesis conducts over-identification test on these two instrumental variables, which confirms the exogeneity. Based on the instrumental variables, this thesis conducts a two-stage least squares (TSLS) regression analysis. The expression for the first stage is

$$workhour_{is} = \varphi_0 + \varphi_1 union_{is} + \varphi_2 chronic_{is} + \mu_i + \lambda_s + \psi_{is} \tag{2}$$

where $workhour_{is}$ is the logarithmic weekly working hours of migrants, $union_{is}$ is the migrant whether to join the union or not, $chronic_{is}$ is whether the migrants suffer from two chronic diseases: hypertension or diabetes, μ_i is the regional fixed effect, λ_s is the occupation fixed effect, and ψ_{is} is a random error term.

C. *Variables and Descriptive Statistics*

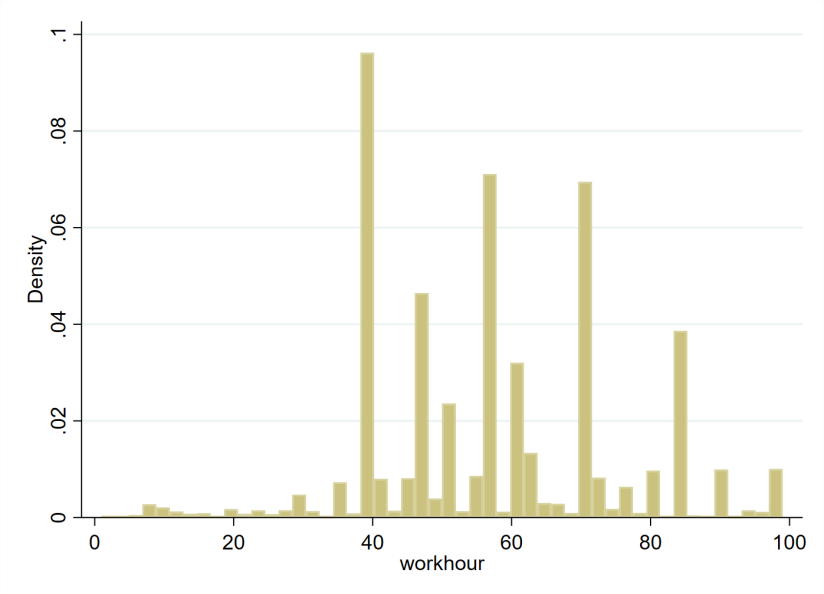

Figure 1. : Distribution of weekly labour hours of the sample.

The dependent variable in this thesis is a dummy variable in the form of willingness to settle of the migrants (settle), and the core explanatory variable in this thesis is the logarithmisation of working hours. The control variables are selected along two dimensions: individual and city. Table III shows the detailed descriptive statistics of the variables.

The demographic characteristics of the sample are often used as control variables. In human capital theory, factors such as education and work experience can help individuals to obtain better jobs and income, thus enabling them to integrate into the inflow cities. In this thesis, gender (gender), years of education (edu), age (age) and marital status (marriage), types of hukou (hukou), and monthly household income (income)) are chosen as individual control variables.

City characteristics may affect the intention of the migrants to stay in the inflow area. The urban control variables selected in this thesis mainly reflect the level of economic development and other infrastructure development in the inflowing city. They mainly include the number of tertiary sector employees (structure), a measure of the city's industrial establishment; logarithmic gross domestic product per capita (lnpgdp), a measure of the city's economic

Table 2—: Percentage of Sample Occupations.

Occupation	Sample	Proportion	Total
Specialised equipment manufacturing	1,714	1.46	1.46
Transport, Warehousing and Postal	4,241	3.62	5.08
Transport Equipment Manufacturing	1,717	1.47	6.55
Instrumentation Manufacturing	365	0.31	6.86
Accommodation Catering	14,849	12.67	19.54
Software and Information Technology Services	1,853	1.58	21.12
Public Administration and Social Organisations	868	0.74	21.86
Other Manufacturing	9,263	7.91	29.76
Agriculture, forestry, animal husbandry and fishery	1,643	1.4	31.17
Chemical Processing	774	0.66	31.83
Pharmaceutical Manufacturing	829	0.71	32.54
Hygiene	1,836	1.57	34.1
Printing Arts & Crafts Office & Recreation	1,242	1.06	35.16
International Organisations	11	0.01	35.17
Residential services, repairs and other services	12,816	10.94	46.11
Construction	8,946	7.64	53.75
Real Estate	3,123	2.67	56.41
Wholesale and retail	24,617	21.01	77.43
Education	2,492	2.13	79.55
Sports and Recreation	1,069	0.91	80.47
Wood Furniture	1,893	1.62	82.08
Water, Environment and Utilities Management	463	0.4	82.48
Electrical Machinery and Manufacturing	2,661	2.27	84.75
Electricity, Coal, Water and Heat Production and Supply	518	0.44	85.19
Social Work	874	0.75	85.94
Scientific research and technical services	359	0.31	86.24
Leasing and Business Services	529	0.45	86.7
Textile and Clothing	5,626	4.8	91.5
Computer and communication electronics manufacturing	3,333	2.85	94.34
Mining	983	0.84	95.18
Finance	1,471	1.26	96.44
Food Processing	4,174	3.56	100

situation; the number of hospital beds per 10,000 population (pbed), a measure of the city's health care resources; the number of primary school teachers per 10,000 population (pteacher), a measure of the city's educational resources; and the proportion of green space in the city (green), used to measure the city's environment.

Table 3—: Descriptive Statistics.

Variables	N	Mean	Median	SD	Min	Max
settle	117152	0.390	0	0.488	0	1
workhour	117152	3.987	4.025	0.375	0	4.595
gender	117152	0.568	1	0.495	0	1
age	117152	35.56	34	9.278	18	60
edu	117152	10.38	9	3.287	0	19
hukou	117152	0.160	0	0.367	0	1
marriage	117152	0.806	1	0.396	0	1
hincome	117152	8.738	8.700	0.585	3.434	12.21
pgdp	117152	11.20	11.28	0.496	9.384	12.28
structure	117152	52.74	52.24	14.71	17.92	87.03
green	117152	41.41	40.90	6.664	3.070	61.58
pbed	117152	66.20	60.61	20.66	14.31	138.6
pteacher	117152	45.04	40.43	20.01	20.54	163.1

II. Empirical Analysis and Discussion

A. Baseline Regression Results

In the first part of the empirical study, the thesis reports the results of estimating the effect of labour hours on migrant settlement intention. According to the regression results in Table 4, labour hours have a significant negative effect on migrant settlement intention in all regressions. Specifically, in the absence of control variables, the coefficient of labour hours on migrants' intention to settle is -0.134 (Model 1), implying that for every 1% increase in migrants' labour hours, migrants' intention to settle decreases by 0.134%. The coefficient of labour hours on migrants' intention to settle drops to -0.0589 (Model 4) after the inclusion of urban and individual control variables. The coefficient decreases slightly after adding fixed effects. According to the results of Model 6 after adding control variables and occupation and province fixed effects, the coefficient of labour hours on migrants' intention to settle is -0.0391 (Model 1), implying that for every 1% increase in migrants' labour hours, migrants' willingness to settle will decrease by 0.0391%. All of the above results have a confidence level of 99%.

Table 5 reports the results of the regressions with "membership in union" and " chronic disease" as instrumental variables for hours of labour. The two-stage least squares (TSLS) results of model (1) indicate that the coefficient on labour hours is 0.601 and that potential

Table 4—: Linear Probability Model Regression Results.

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
workhour	-0.134***	-0.090***	-0.092***	-0.058***	-0.054***	-0.039***
	(0.0040)	(0.0039)	(0.0038)	(0.0038)	(0.0039)	(0.0038)
gender		-0.016***		-0.013***	-0.012***	-0.011***
		(0.0028)		(0.0027)	(0.0028)	(0.0028)
age		0.001***		0.001***	0.001***	0.001***
		(0.0002)		(0.0002)	(0.0002)	(0.0002)
edu		0.013***		0.011***	0.011***	0.012***
		(0.0005)		(0.0005)	(0.0005)	(0.0005)
hukou		0.147***		0.140***	0.139***	0.129***
		(0.0042)		(0.0041)	(0.0041)	(0.0040)
marriage		-0.002		0.009**	0.011***	0.013***
		(0.0041)		(0.0041)	(0.0041)	(0.0041)
income		0.075***		0.039***	0.041***	0.026***
		(0.0025)		(0.0025)	(0.0025)	(0.0026)
pgdp			0.185***	0.172***	0.170***	0.088***
			(0.0039)	(0.0039)	(0.0039)	(0.0059)
structure			0.005***	0.005***	0.005***	0.002***
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
green			0.001***	0.001***	0.001***	-0.002***
			(0.0002)	(0.0002)	(0.0002)	(0.0003)
pbed			0.000	0.000	0.000	0.001***
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
pteacher			0.001***	0.001***	0.001***	0.001***
			(0.00008)	(0.00008)	(0.00008)	(0.0001)
constant	0.924***	-0.097***	-1.738***	-2.166***	-2.126***	-0.742***
	(0.0162)	(0.0279)	(0.0456)	(0.0474)	(0.0491)	(0.0714)
Province FE	NO	NO	NO	NO	NO	YES
Occupation FE	NO	NO	NO	NO	YES	YES
N	117152	117152	117152	117152	117152	117152
R^2	0.011	0.051	0.071	0.096	0.101	0.133

Note: $***p < 0.01, **p < 0.05, *p < 0.1$; robust standard errors in parentheses, same table below.

Table 5—: IV Regression Results.

	(1) TSLS	(2) LIML	(3) GMM	(4) IGMM
workhour	-0.165***	-0.165***	-0.165***	-0.165***
	(0.0762)	(0.0762)	(0.0762)	(0.0762)
Control Vars	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES
N	117152	117152	117152	117152
R^2	0.094	0.094	0.094	0.094
Sargan statistic			0.203	
P-value			0.652	

endogeneity issues may biased estimate the effect of labour hours on migrants' settlement intentions. The estimation results of model (2) Limited Information Maximum Likelihood (LIML) are the same as the 2SLS model, proving that the model does not suffer from the problem of weak instrumental variables. The Generalised Moment Estimation (GMM) and Iterated Generalised Moment Estimation (IGMM) estimation results for models (3)(4) are also the same as the 2SLS, proving that the results are robust. Prior to instrumental variables regression, the exogeneity of instrumental variables needs to be justified. There is no research that demonstrates the effect of "whether they have been ill or injured in the last two weeks" and "whether they have two long-term illnesses" on migrants' intention to settle, and over-identification tests for these two instrumental variables are also conducted in this thesis. The results of the over-identification test are reported in the appendix and indicate that the two instrumental variables are exogenous (Sargan statistic = 0.643, p-value = 0.423).

B. Robustness

In order to improve the rigour of the findings of this thesis, the robustness test is carried out in the following five ways:(1) Using the robust standard errors of prefecture-level city clustering. Since the perturbation terms of the regression model may have heteroskedasticity and the perturbation terms of the same prefecture-level administrative region may have autocorrelation in different periods, the robust standard errors clustered with prefecture-level administrative regions are used for estimation. (2) The length of working hours without logarithmic treatment is used as the core explanatory variable. (3) Substitution of dependent variables. In order to avoid possible bias in the question of "intention to go to the household", the dependent variable is replaced with the questionnaire's "intention to live for a long time" (whether or not to live in the local city for more than 5 years). (4) Non-linear model setting. In this thesis, the quadratic term of working hours is added to the regression model to avoid the impact of model setting bias on the regression results. (5) Non-linear regression model. In this thesis, the linear probability model based on OLS is replaced in the main regression, and the Logit model based on likelihood estimation is used for robustness testing. As in the

main regression, all five robustness tests include control variables and city with occupation fixed effects.

Table 6—: Robustness Tests

	(1) Clustercity	(2) Nolog	(3) Longstay	(4) Quadratic	(5) Logit	(6) Partiality
workhour	-0.039 ***	-0.001 ***	-0.042***	0.152***	-0.182***	-0.043***
	(0.0061)	(0.0001)	(0.0035)	(0.0240)	(0.0179)	(0.0042)
workhourˆ2				-0.027***		
				(0.0033)		
Control Vars	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES	YES	YES
N	117152	117152	117152	117152	117152	117152
R^2	0.1334	0.1338	0.1410	0.1339	0.1037	-

As can be seen from the results of the robustness regressions using robust standard errors clustered by the prefecture-level city in the model (1) in Table 6, the coefficients and significance of the core explanatory variables remain unchanged. The regression results using hours worked without logarithmic treatment as the core explanatory variable are also significant (Model 2), with the implication being that for every 10-hour increase in migrants' labour hours, the propensity to settle will decrease by 1%. Replacing the dependent variable with the questionnaire's "intention to stay long" resulted in an increase in the model coefficients, but the sign and significance level did not change (Model 3). After adding the quadratic term of working hours (Model 4), the positive coefficient of the primary term and the negative coefficient of the quadratic term prove that the non-linear function is inverted U-shaped, which implies that working hours are negatively correlated with willingness to settle on the right side of the inflection point (log(workhour) = 2.8), which is in line with the finding of the baseline regression. Model (5) uses a logit model based on likelihood estimation, and the direction and significance of the coefficients remain unchanged from the baseline regression. Model 6 reports the marginal effects regression results from the logit model, the significance level and coefficient closely resemble the results of the baseline regression. The above results indicate that the findings of this thesis are robust.

C. Mechanism Analysis

The mechanism of labour hours on migrants' willingness to settle is affected by a variety of factors, which are analysed in this thesis from three perspectives: excessive labour, economic situation and social integration. Firstly, the migrants' excessive labour, this thesis constructs an experimental environment through a matching model, divides the treatment group with excessive labour and the control group without excessive labour according to the working hours, and then analyses the treatment effect of excessive labour on settlement intention.

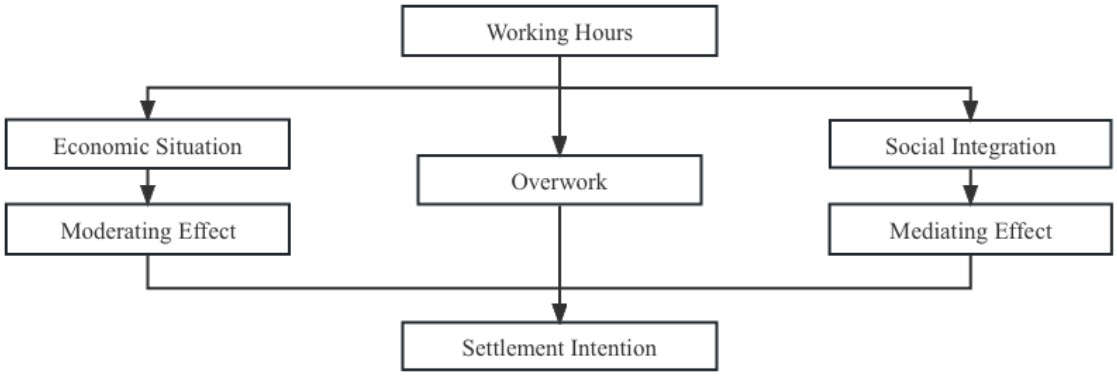

Figure 2. : Mechanism analysis

Second, the economic situation of migrants also has an impact on the regression mechanism in this thesis. This thesis puts the interaction terms of the three variables with working hours: housing expenditure, personal income and household income-expenditure ratio, into the model to examine the moderating effect of migrants' economic situation. Thirdly, this thesis considers the effect of working hours on social integration and uses the KHB model to analyse the mediating effect of working hours-social integration-intention to settle.

Overwork and Settlement Intention

In order to measure the impact of "excessive labour" on migrants' willingness to settle, this thesis sets the length of working hours as a dummy variable of "excessive labour" at four points: 44h, 48h, 56h and 70h. According to the Labour Law of the People's Republic of China, the weekly working hours cannot exceed 44h, so we define weekly working hours exceeding 44h as excessive labour. Considering the distribution of the data, this thesis adds 48h, 56h, and 70h constructed dummy variables for robustness testing. When the core independent variable is a dummy variable, the data structure is suitable for using a matching model to statistically construct a "counterfactual" environment to obtain more accurate estimates, and this thesis uses the PSM model.

Figures 3 and 4 report the before and after matching and density plots, demonstrating better results for the matched data. Table 7 shows the results of the PSM estimation, and it can be seen that overwork has a significant negative effect on migrants' willingness to settle. The results of the dummy variable constructed with 44 hours yield a treatment effect of -0.0519, and the regressions of the dummy variables constructed with 48h, 56h, and 70h yield similar coefficients.

Economic Situation and Settlement Intention

Table 8 reports the results of regressions of working hours on the interaction terms of housing expenditures (house), monthly income (income), and household expenditures-to-income ratio

Table 7—: Overwork and Settlement Intention

	PSM_1:1	PSM_1:2	PSM_radius	PSM_kernel
workhour44	-0.052	-0.052	-0.052	-0.052
	(0.0049)	(0.0048)	(0.0048)	(0.0048)
workhour48	-0.050	-0.053	-0.054	-0.049
	(0.0048)	(0.0044)	(0.0048)	(0.0048)
workhour56	-0.046	-0.046	-0.047	-0.046
	(0.0042)	(0.0042)	(0.0042)	(0.0039)
workhour70	-0.048	-0.048	-0.048	-0.048
	(0.0042)	(0.0041)	(0.0042)	(0.0042)

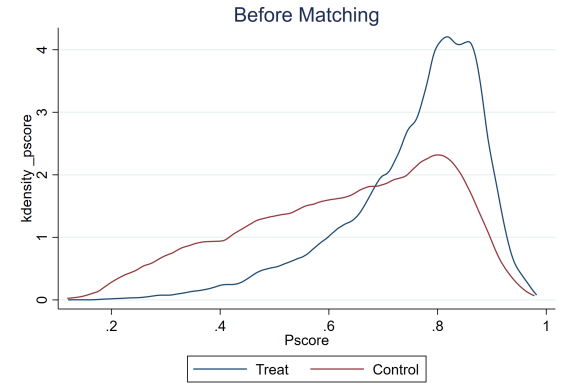

Figure 3. : Pre-matching kernel density

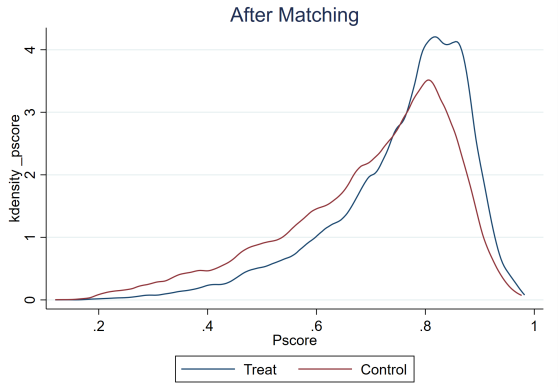

Figure 4. : Post-matching kernel density

(ratio), respectively. All three interaction terms are statistically significant. The results show that an increase in housing expenditures weakens the effect of labour time on migrants' willingness to settle, while an increase in income and the income-expenditure ratio strengthens the effect of labour time on migrants' willingness to settle.

Social Integration and Settlement Intention

Excessive labour hours can take up migrants' social time, affecting their social integration and thus reducing their willingness to settle. In order to test this hypothesis, this thesis conducted a regression using social integration indicators as the dependent variable and labour hours as the independent variable. In this thesis, the social integration indicators are divided into two categories, Perceptual Integration and Community Integration, and the detailed descriptions of the indicators are shown in Table 9. According to Table 10, it can be seen that the effect of labour hours on the six social integration indicators is negative and significant.

This study measured the mediating effect of social integration using the KHB method. The

Table 8—: Economic Situation Interaction Term Regression Results

	(1) OLS	(2) OLS	(3) OLS
workhour	-0.051***	-0.004	-0.053***
	(0.0071)	(0.0222)	(0.0104)
house	-0.005		
	(0.0048)		
workhour*house	0.002*		
	(0.0012)		
income		0.017***	
		(0.0112)	
workhour*income		-0.004**	
		(0.0027)	
ratio			0.228***
			(0.0977)
workhour*ration			-0.046**
			(0.0246)
Control Vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
N	117152	117152	117152
R^2	0.134	0.134	0.134

Table 9—: Social Integration Indicators Description.

Perceptual Integration	
Indicator 1	I like the city/place where I live now
Indicator 2	I am concerned about changes in the city/place where I live now
Indicator 3	The customs of my current place of residence are more important to me.
Community Integration	
Indicator 1	I'd love to fit in with the locals and be a part of it
Indicator 2	I feel that the locals would like to accept me as one of them
Indicator 3	I feel like I'm already a local

Table 10—: Social Integration and Settlement Intention

Panel A. Dependent Variable: Perceptual Integration Indicator			
	(1)	(2)	(3)
	Indicator 1	Indicator 2	Indicator 3
workhour	-0.006***	-0.014***	-0.003*
	(0.0015)	(0.0018)	(0.0040)
Control vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
N	117152	117152	117152
R2	0.004	0.015	0.044
Panel B. Dependent Variable: Community Integration Indicator			
	(4)	(5)	(6)
	Indicator 1	Indicator 2	Indicator 3
workhour	-0.017***	-0.017***	-0.049***
	(0.0022)	(0.0023)	(0.0034)
Control vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
N	117152	117152	117152
R2	0.028	0.030	0.089

Table 11—: KHB Analysis Results

	(1)	(2)	(3)
	Perceptual Integration	Community Integration	Perceptual+Community
Reduced	-0.039***	-0.039***	-0.039***
	(0.0037)	(0.0037)	(0.0037)
Full	-0.037***	-0.030***	-0.030***
	(0.0037)	(0.0037)	(0.0042)
Diff	-0.002***	-0.008***	-0.009***
	(0.0004)	(0.0007)	(0.0007)

coefficients in the simplified model represent the direct effect of the key independent variable (Hours of Labour) on the dependent variable (Willingness to Settle), while the coefficients in the full model represent the effect of the key independent variable (Hours of Labour) on the dependent variable through the mediating variable (Social Integration), and the coefficients in the difference model measure the effect of the mediating variable. Table 8 shows the mediating effects of Perceptual Integration, Community Integration, and Perceptual and Community Integration respectively. According to the results in Table 11, the 1% increase in the working hours of the migrants will reduce the willingness to settle by 0. 039%, where the effect of perceived integration is 0.002% (model 1), the effect of community integration is 0.008% (model 2), and the sum of the two effects is 0.009% (model 3).

D. Heterogeneity Analysis

The heterogeneity analysis in this thesis is divided into two parts, group regressions according to the sample's personal characteristics and work industry, respectively. Characteristics such as age, gender, and marital status are of importance in social science research. They may be associated with many key factors such as health, economic status, social status, etc., and heterogeneity analyses for these variables are necessary. In this thesis, regressions are first conducted according to gender (male, female), age (age¡=35, age¿35), and marital status (unmarried, married). The results in Table 12 find that the settlement intentions of migrants who are female and age¡=35 are more sensitive to labour hours. The regression results grouped by marital status did not pass the SUR subgroup difference test. Young people and female migrants are likely to place greater emphasis on health and life, where overwork may have a negative impact on physical and mental health, while they may be more focused on balancing life. In addition, young people are more focused on diversifying their career goals, and overwork can prevent them from pursuing other career or educational opportunities, which can negatively affect future career development.

Table 12—: Individual Heterogeneity Analysis Results

	(1)	(2)	(3)	(4)	(5)	(6)
	female	male	age<=35	age>35	unmarried	married
lnworkhour	-0.066***	-0.052***	-0.069***	-0.045***	-0.055***	-0.058***
	(0.0055)	(0.0050)	(0.0053)	(0.0052)	(0.0092)	(0.0041)
Control	YES	YES	YES	YES	YES	YES
N	50654	66498	63620	53532	22751	94401
R^2	0.109	0.086	0.100	0.093	0.074	0.105
Difference	-0.014		-0.024		0.003	
χ^2	3.42**		10.20***		0.10	

This thesis also conducted regressions by group according to the industry of migrants, and Table 13 shows that the willingness to settle of migrants working in agriculture is not sensitive to labour hours. In China, self-employment is still the main form of agricultural production,

Table 13—: Sector Heterogeneity Analysis Results

	(1) agriculture	(2) industrial	(3) services
workhour	0.013	-0.036***	-0.037***
	(0.0254)	(0.0077)	(0.0044)
Control Vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
N	1643	34574	80935
R^2	0.185	0.132	0.138

and there is no formal employment relationship. Agricultural workers usually decide their own working hours and have a high degree of freedom, so their willingness to settle is not sensitive to their working hours. Migrants working in the service sector are the most sensitive to the length of their working hours. This may be due to the fact that excessive overtime is more widespread in the service sector, where work is more flexible.

III. Conclusion

This thesis analyses the impact of the migrants' working hours in the inflow city on their settlement intentions using 2017 China migrants dynamic survey (CMDS) data of 117,152 workers. The results of the linear probability model (LPM) show that for 1% increase in the working hours of the migrant population, their settlement intention decreases by 0.0391%. The results are supported by instrumental variables as well as robustness checks. Mechanism analysis shows that the increase in housing expenditures weakens the effect of labour time on migrants' willingness to settle, while the increase in income-expenditure ratio strengthens the effect of labour time on migrants' willingness to settle. In addition, using the Karlson–Holm–Breen (KHB) method this thesis analyses the mediating effect of the social integration situation. Heterogeneity analysis reveals that the settlement intentions of the female, young person are more likely to be influenced by labour hours. Intention to settle is not sensitive to labour hours for agricultural migrants, but it is sensitive to labour hours for industrial and service migrants.

Excessive labour and labour protection have always been important issues that have aroused widespread concern. This paper takes a unique approach, using the impact of overwork on the willingness of migrants to settle as an entry point, and deeply discusses the urgency and significance of avoiding overwork. This new perspective can help policy makers to better understand the impact of overwork on individuals and society. Excessive labour not only puts a heavy workload on individuals, leading to physical and mental health problems, but also reduces their desire to settle in the inflow area in the long term. This is a direct call to improve labour conditions and reduce work stress to ensure the well-being and quality of life of workers. Migrants are often an important labour resource for a society, but if they are unwilling to settle in a place, then communities and regional economies are missing out

on opportunities. By attracting mobile people to stay, governments can increase investment in talent and promote business and community growth, leading to more robust economic development.

Understanding how overwork affects willingness to settle is also valuable information for policymakers. Governments and employers can use this research to design more humane policies, including enacting minimum wage legislation, setting limits on working hours, and providing better labour protection and welfare systems. This will help build a fairer and more inclusive labour market, maintain social stability and reduce social inequality. This study will benefit not only individuals but also society as a whole. It will help build a more just and prosperous labour environment, improve the quality of life and promote sustainable social development.

REFERENCES

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ferrer-i-Carbonell, A. and Frijters, P. (2004). How Important is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal*, 114(497):641–659.

Appendix

Table A1—: IV Test Results.

Underidentification test (Anderson canon. corr. LM statistic):	296.726
Chi-sq	0
Weak identification test (Cragg-Donald Wald F statistic):	148.683
Stock-Yogo weak ID test critical values: 10% maximal IV size	19.93
15% maximal IV size	11.59
20% maximal IV size	8.75
25% maximal IV size	7.25
Sargan statistic (overidentification test of all instruments):	0.203
Chi-sq	0.6524

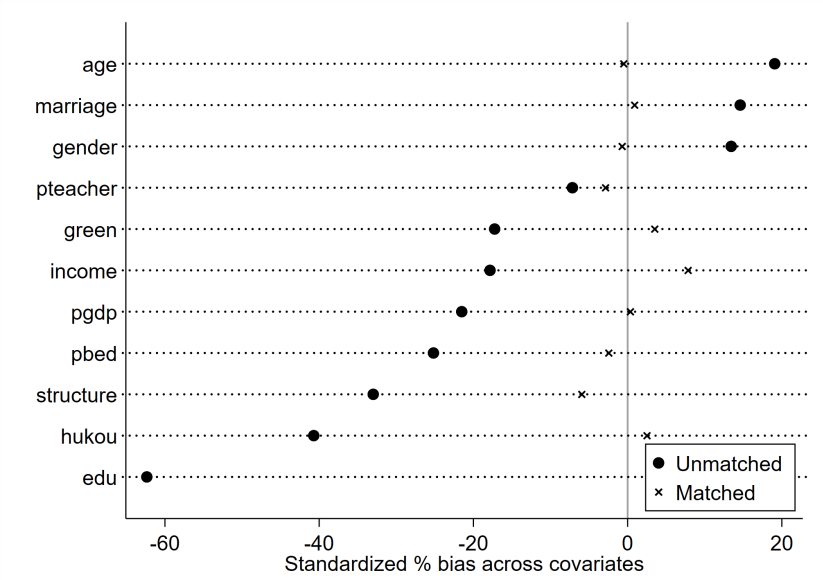

Figure A1. : PSM Balance Test.

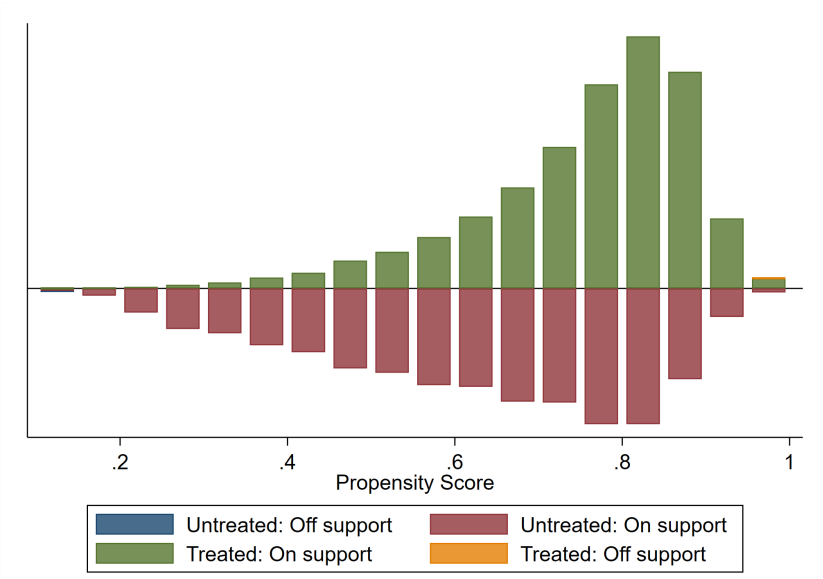

Figure A2. : PSM Common Value Interval.