

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

By JINGWEN TAN AND NEDA TRIFKOVIC

This paper 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 employed samples. The results of the linear probability model (LPM) show that for each 1% increase in the working hours of the migrant population, their settlement intention decreases by 0.0391%. This paper uses "labor union" and "chronic disease" as instruments for working hours to avoid potential endogeneity affecting the robustness. The results of the mechanistic analysis 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. In addition, decomposition using the KHB method found that social integration partially mediates the effect of this mechanism. Heterogeneity analysis reveals that the settlement intentions of female, low-educated, and married migrants are more likely to be influenced by labour hours. Intention to settle is not sensitive to labour hours for agricultural migrants, and 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 paper 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 average age of the sample ranged from 15-59 years. A stratified, multi-stage, large-scale PPS (Probability Proportionate to Size Sampling) sampling method was used to investigate in detail the developmental status, individual characteristics, social integration and employment of the migrants in China. The data relating to the urban control variables were obtained from the statistical yearbooks of the cities.

In this paper, the data were cleaned before empirical analysis. First, we excluded all samples under the age of 18 as well as age over 60. Secondly, the population studied in this paper is the migrants in cities with jobs, so we remove the sample without jobs. Finally, we matched

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

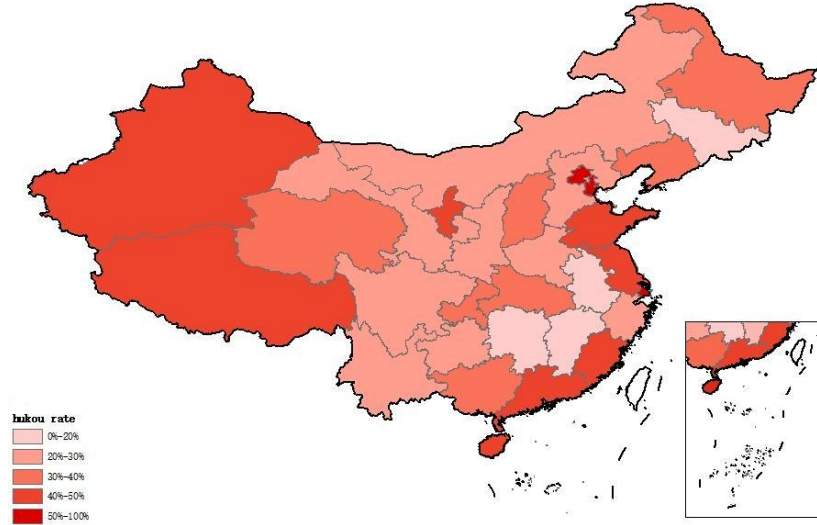


Figure 1. : Proportion of China migrants settlement intentions (provincial level).

the micro-questionnaire data of the sample with the data from the urban statistical yearbook and retained the successfully matched sample. The processed sample size with a total of 117,152.

B. Econometrics Model

The focus of this paper 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).

China's household registration system, also known as the "hukou system", is a system of population management based on family ties. Every Chinese citizen has a household register (also known as a "hukou"), which is used to record basic personal information about that citizen. There are two types of hukou for Chinese citizens: rural hukou and urban hukou. These two types of hukou have different entitlements and policies in terms of social welfare, education and healthcare. Moving the hukou requires a certain application and approval process, as well as a certain amount of time, effort and money, so being willing to move the hukou also indicates that the person is more willing and determined to settle in that city.

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 paper 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 paper can be expressed as

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

where $settle_{is}$ is the willingness of the migrants to settle, and this variable is a dummy variable. $workhour_{is}$ is the weekly labour hours of the migrants. X_{is} are other control variables. μ_i is the regional fixed effect, λ_s is the occupation fixed effect. This paper chooses to control for the regional fixed effects at the provincial level, while putting in control variables at the local municipality level for estimation.

Labour 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. This leads to the problem of endogeneity in forecasting that may generate two-way causality and affect the unbiasedness of the estimation results. Instrumental variables are widely used in the treatment of two-way causality problems.

In this paper, we choose "whether the migrant is a member of a union" and "whether the migrant has hypertension or diabetes" as instrumental variables for labour time. These two variables will directly affect the labour time of the migrants, which is in line with the correlation hypothesis of the instrumental variable; in addition, these two variables will not affect the willingness of the migrants to settle down, which is in line with the exogenous hypothesis of the instrumental variable. This paper conducts over-identification test on these two instrumental variables, which confirms the exogeneity. Based on the instrumental variables, this paper conducts a two-stage OLS regression analysis. The expression for the first stage is

$$(2) \quad workhour_{is} = \varphi_0 + \varphi_1 union_{is} + \varphi_2 chronic_{is} + \psi_{is}$$

where $workhour_{is}$ is the 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, and ψ_{is} is a random error term.

C. Variables and Descriptive Statistics

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

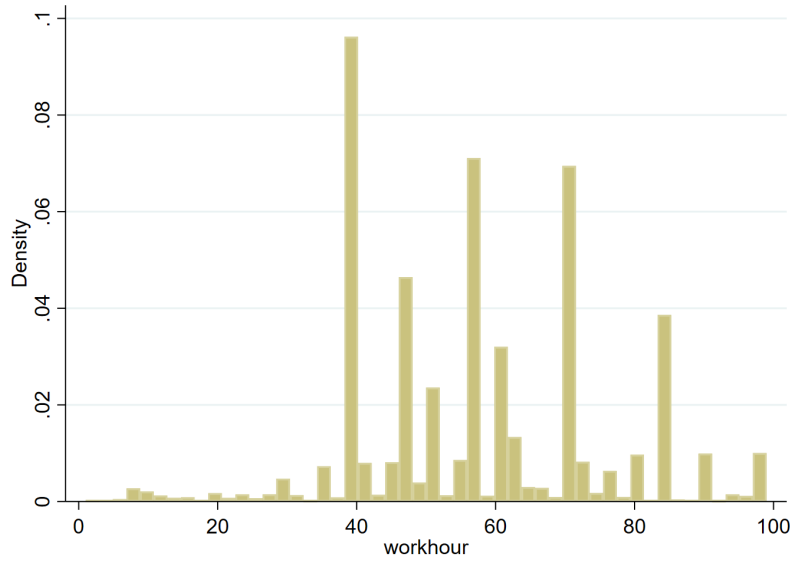


Figure 2. : Distribution of labour hours of the sample.

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 paper 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 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.

II. Empirical Analysis and Discussion

A. Baseline Regression Results

In the first part of the empirical study, the paper reports the results of estimating the effect of labour hours on migrant settlement intention. According to the regression results in Table 2, 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

Table 1—: Descriptive Statistics.

Variables	N	Mean	P50	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

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%.

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.

Table 3 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 endogeneity issues may underestimate 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 similar to 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 "being a member of a union" and "chronic disease" on migrants' intention to settle, and over-identification tests for these two instrumental variables

Table 2—: Linear Probability Model Regression Results.

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
workhour	-0.134*** (0.00403)	-0.0901*** (0.00395)	-0.0924*** (0.00383)	-0.0589*** (0.00381)	-0.0542*** (0.00387)	-0.0391*** (0.00383)
gender		-0.0160*** (0.00284)		-0.0134*** (0.00277)	-0.0115*** (0.00287)	-0.0107*** (0.00282)
age		0.00101*** (0.000180)		0.000628*** (0.000177)	0.000643*** (0.000177)	0.000708*** (0.000175)
edu		0.0133*** (0.000521)		0.0111*** (0.000511)	0.0105*** (0.000528)	0.0119*** (0.000526)
hukou		0.147*** (0.00422)		0.140*** (0.00410)	0.139*** (0.00411)	0.129*** (0.00409)
marriage		-0.00183 (0.00417)		0.00961** (0.00411)	0.0108*** (0.00412)	0.0130*** (0.00411)
income		0.0754*** (0.00258)		0.0398*** (0.00256)	0.0419*** (0.00258)	0.0267*** (0.00262)
pgdp			0.185*** (0.00392)	0.172*** (0.00392)	0.170*** (0.00395)	0.0881*** (0.00599)
structure			0.00585*** (0.000105)	0.00527*** (0.000106)	0.00502*** (0.000107)	0.00253*** (0.000162)
green			0.00170*** (0.000216)	0.000802*** (0.000215)	0.000933*** (0.000214)	-0.00200*** (0.000363)
pbed			0.000165 (0.000103)	0.0000342 (0.000102)	0.0000174 (0.000102)	0.000624*** (0.000156)
pteacher			0.000683*** (0.0000880)	0.000869*** (0.0000876)	0.000883*** (0.0000880)	0.000546*** (0.000141)
constant	0.924*** (0.0162)	-0.0970*** (0.0279)	-1.738*** (0.0456)	-2.166*** (0.0474)	-2.126*** (0.0491)	-0.742*** (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 3—: IV Regression Results.

	(1) TSLS	(2) LIML	(3) GMM	(4) IGMM
workhour	-0.601*** (0.0879)	-0.602*** (0.0881)	-0.601*** (0.0879)	-0.601*** (0.0879)
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	0.0372	0.0372	0.0372	0.0372
Sargan statistic			0.643	
P-value			0.423	

are also conducted in this paper. 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 paper, 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 paper, 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 paper, 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.

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 4, 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,

Table 4—: Robustness Tests

	(1)	(2)	(3)	(4)	(5)
	OLS_clustercity	OLS_nolog	OLS_longstay	OLS_quadratic	Logit
workhour	-0.0391*** (0.00606)	-0.00105*** (0.0000843)	-0.0416*** (0.00355)	0.152*** (0.0240)	-0.182*** (0.0179)
workhour^2				-0.0271*** (0.00335)	
Control Vars	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES	YES
N	117152	117152	117152	117152	117152
R^2	0.1334	0.1338	0.1410	0.1339	0.1037

but the sign and confidence level did not change (Model 3). After adding the quadratic term of working hours (Model 4), the model identifies an "inverted U-shaped" relationship and calculates an inflexion point of 2.80, which proves that labour hours are still negatively correlated with the intention to settle. Model (5) uses a logit model based on likelihood estimation, and the direction and significance of the coefficients remain unchanged from the baseline regression. The above results indicate that the findings of this paper are robust.

C. Mechanism Analysis

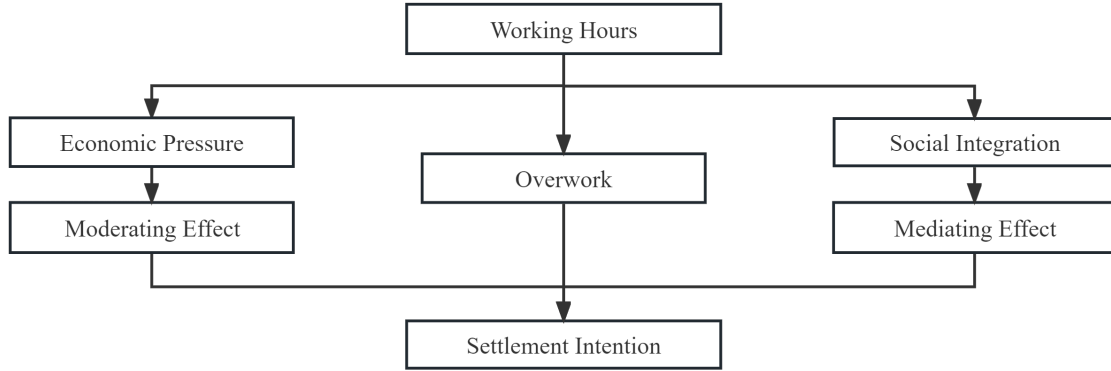


Figure 3. : Mechanism analysis

The mechanism of labour hours on migrants' willingness to settle is affected by a variety of factors, which are analysed in this paper from three perspectives: excessive labour, economic situation and social integration. Firstly, the migrants' excessive labour, this paper 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. Second, the economic situation of migrants also has an impact on the regression mechanism in this paper. This paper 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 paper 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 paper 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 paper 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 paper 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 5 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.

Table 5—: Overwork and Settlement Intention

	PSM_1:1	PSM_1:2	PSM_radius	PSM_kernel
workhour44	-0.0519 (0.0049)	-0.0519 (0.0048)	-0.0519 (0.0048)	-0.0519 (0.0048)
workhour48	-0.0501 (0.0048)	-0.0525 (0.0044)	-0.0535 (0.0048)	-0.0487 (0.0048)
workhour56	-0.0464 (0.0042)	-0.0464 (0.0042)	-0.0469 (0.0042)	-0.0464 (0.0039)
workhour70	-0.0478 (0.0042)	-0.0476 (0.0041)	-0.0479 (0.0042)	-0.0478 (0.0042)

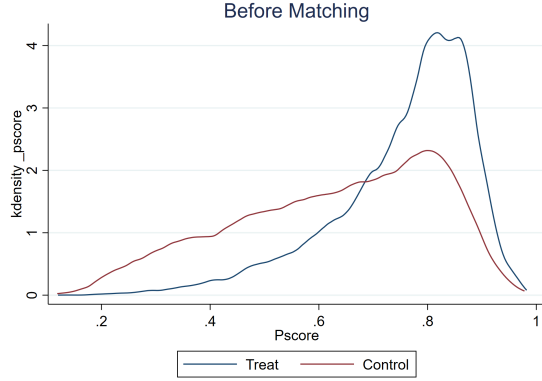


Figure 4. : Pre-matching kernel density

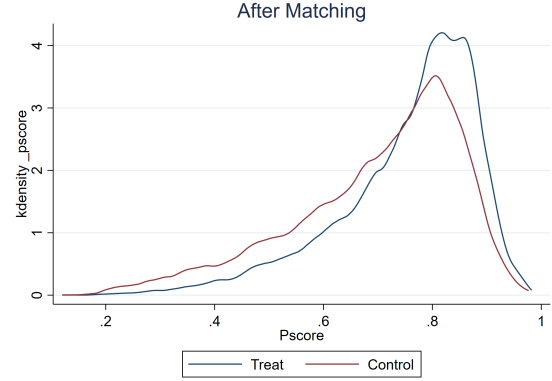


Figure 5. : Post-matching kernel density

ECONOMIC SITUATION AND SETTLEMENT INTENTION

Table 6 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 (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 paper conducted a regression using social integration indicators as the dependent variable and labour hours as the independent variable. In this paper, 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 7. According to Table 8, 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 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. It can be seen that the mediating effect of Perceptual Integration is -0.0024, Community Integration is -0.0087, and Perceptual and Community Integration is -0.009.

Table 6—: Economic Situation Interaction Term Regression Results

	(1) OLS	(2) OLS	(3) OLS
workhour	-0.0508*** (0.00714)	-0.0036 (0.0222)	-0.0528*** (0.0104)
house	-0.00484 (0.00478)		
workhour*house	0.00225* (0.00120)		
income		0.0173*** (0.0112)	
workhour*income		-0.0044** (0.0027)	
ratio			0.2276*** (0.0977)
workhour*ratio			-0.0456** (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 7—: 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 8—: Social Integration and Settlement Intention

Panel A. Dependent Variable: Perceptual Integration Indicator			
	(1)	(2)	(3)
	Indicator 1	Indicator 2	Indicator 3
workhour	-0.00606*** (0.00145)	-0.0144*** (0.00178)	-0.00292 (0.00401)
Control vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
<i>N</i>	117152	117152	117152
<i>R</i> ²	0.004	0.015	0.044
Panel B. Dependent Variable: Community Integration Indicator			
	(4)	(5)	(6)
	Indicator 1	Indicator 2	Indicator 3
workhour	-0.0173*** (0.00221)	-0.0165*** (0.00226)	-0.0487*** (0.00341)
Control vars	YES	YES	YES
Province FE	YES	YES	YES
Occupation FE	YES	YES	YES
<i>N</i>	117152	117152	117152
<i>R</i> ²	0.028	0.030	0.089

Table 9—: KHB Analysis Results

	Perceptual Integration	Community Integration	Perceptual+Community
Reduced	-0.0390837*** (0.0037668)	-0.0390837*** (0.0037239)	-0.0390837*** (0.0037182)
Full	-0.0366433*** (0.003768)	-0.0303187*** (0.0037274)	-0.0300261*** (0.003722)
Diff	-0.0024404*** (0.0003852)	-0.008765*** (0.0006988)	-0.0090576*** (0.0007311)

D. Heterogeneity Analysis

The heterogeneity analysis in this paper is divided into two parts, group regressions according to the sample's personal characteristics and work industry, respectively. In this paper, regressions are first conducted according to gender (male, female), education (with higher education diploma, without higher education diploma), and marital status (unmarried, married). The results in Table 9 find that the settlement intentions of migrants who are female, do not have a higher education diploma, and are married are more sensitive to labour hours. Labourers without higher education are mainly engaged in physically demanding low-end industries, where overtime work is physically demanding and the wages are low. Married migrants tend to be more resistant to overtime work because they have to take care of their families and have children.

Table 10—: Individual Heterogeneity Analysis Results

	(1) female	(2) male	(3) with diploma	(4) without diploma	(5) married	(6) unmarried
workhour	-0.0449*** (0.00564)	-0.0338*** (0.00511)	-0.0390*** (0.0102)	-0.0398*** (0.00408)	-0.0408*** (0.00932)	-0.0374*** (0.00413)
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	50654	66498	22148	95004	22751	94401
R^2	0.146	0.125	0.181	0.096	0.105	0.146

Table 11—: Sector Heterogeneity Analysis Results

	(1) agriculture	(2) industrial	(3) services
workhour	0.0133 (0.0254)	-0.0358*** (0.00767)	-0.0373*** (0.00442)
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

This paper also conducted regressions by group according to the industry of migrants, and Table 9 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,

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 paper 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 employed samples. The results of the linear probability model (LPM) show that for each 1% increase in the working hours of the migrant population, their settlement intention decreases by 0.0391%. This paper uses "labour union" and "chronic disease" as instruments for working hours to avoid potential endogeneity affecting the robustness. The results of the mechanistic analysis 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. In addition, decomposition using the KHB method found that social integration partially mediates the effect of this mechanism. Heterogeneity analysis reveals that the settlement intentions of female, low-educated, and married migrants are more likely to be influenced by labour hours. Intention to settle is not sensitive to labour hours for agricultural migrants, and is sensitive to labour hours for industrial and service migrants.

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—: 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

Table A2—: IV Test Results.

Underidentification test (Anderson canon. corr. LM statistic):	253.718
Chi-sq(2) P value =	0
Weak identification test (Cragg-Donald Wald F statistic):	127.053
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
Source: Stock-Yogo (2005). Reproduced by permission.	
Sargan statistic (overidentification test of all instruments):	0.643
Chi-sq(1) P value =	0.4227

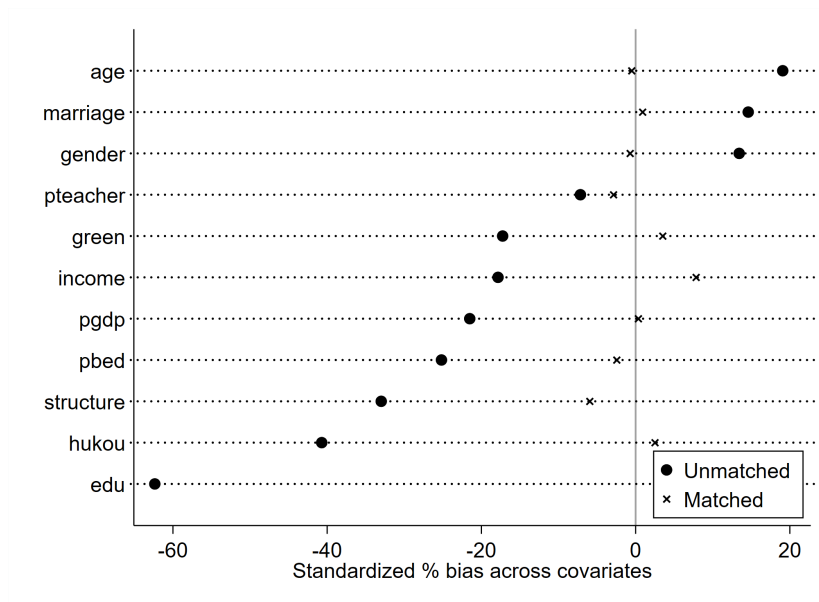


Figure A1. : PSM Balance Test.

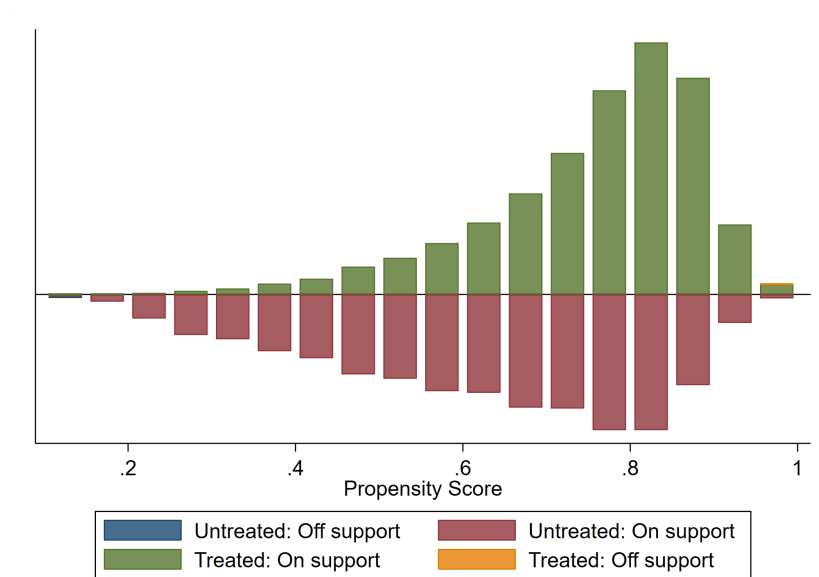


Figure A2. : PSM Common Value Interval.