

Cai, Zhengyu; Yan, Yu

Working Paper

The Consumption Effects of Population Concentration

GLO Discussion Paper, No. 1631

Provided in Cooperation with:

Global Labor Organization (GLO)

Suggested Citation: Cai, Zhengyu; Yan, Yu (2025) : The Consumption Effects of Population Concentration, GLO Discussion Paper, No. 1631, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/320559>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Consumption Effects of Population Concentration

Zhengyu Cai (<https://orcid.org/0000-0002-3852-1253>)

Southwestern University of Finance and Economics,
School of Public Administration,
Research Institute of Labor Economics and Policy,
Chengdu, Sichuan 611130, China
and
Global Labor Organization (GLO),
Essen, Germany
Email: czy@swufe.edu.cn

Yu Yan

Sichuan University of Science & Engineering,
School of Economics,
Yibin, Sichuan 644005, China
E-mail: yanyuddd123@163.com

Abstract

This paper empirically examines the causal effects of population concentration on household consumption in China and explores the underlying mechanisms. After addressing endogeneity concerns from various sources, our results show that a 1% increase in urban population density leads to a 0.43 percentage point increase in household average propensity to consume and a 59.73 yuan increase in per capita consumption expenditure. Mechanism analysis reveals that, beyond the effects of urban income premiums, population density reduces consumption transaction costs and enhances social interactions, both of which encourage higher consumption. Additionally, population density raises non-housing consumption among some households by driving up housing prices.

Keywords: population density; household consumption; transaction costs; social interactions; housing prices

JEL Classification: R12, R23, D12

This article was originally written in Chinese and has been accepted for publication by *China Economic Quarterly* (《经济学》(季刊), ISSN: 2095-1086). The original title was “The Consumption Effect of Population Agglomeration (人口集聚的消费效应).” This article is an English translation of the published paper. Since the term “population agglomeration” is rarely used in English academic contexts and its meaning is similar to “population concentration” (measured by population density), we have replaced "population agglomeration" with "population concentration" throughout the translation.

The authors thank participants and discussants at the 4th China Urban Economists Forum, the 2022 Regional, Urban and Spatial Economics (RUSE) Workshop, the 21st Camphor Economics Seminar (Chengdu), the 2023 China Annual Conference of the Chinese Economists Society (CES), the 2024 Chengdu Annual Conference of the Young Economist Society (YES), and the seminar at the Institute for Economic and Social Research (IESR) at Jinan University, as well as the anonymous referees and editorial board of *China Economic Quarterly* for their helpful comments. The authors bear full responsibility for any remaining errors.

Table of Contents

1. Introduction	4
2. Theoretical Framework	6
2.1 Consumer	6
2.2 Firm	7
2.3 Consumption Effects of Population Concentration and Mechanisms	8
3. Empirical Framework and Data	10
3.1 Identification Strategy	10
3.2 Data and Variables	13
4. Empirical Results	15
5. Mechanism Tests	20
5.1 Consumption Matching	20
5.2 Social Interaction	22
5.3 Housing Prices	23
6. Policy Implications	27
Appendix I: Population Density and Transaction Costs	29
Appendix II: Data Sources	33
Appendix III: Variable Descriptions	35
Appendix IV: Detailed Results of Table 2	41
Appendix V: Robustness Checks	45
Appendix VI: Heterogeneity Tests	48
Appendix VII: Population Concentration and Non-tradable Service Density	52
Appendix VIII: Housing Expenditure Imputation	53
References	54

1. Introduction

Low household consumption has persistently constrained China's sustained economic growth (Ouyang et al., 2016). The existing literature has examined this issue through various lenses, including liquidity constraints (Wan et al., 2001), precautionary savings (Yang and Chen, 2009), income distribution (Gan et al., 2018), habit formation (Lei, 2009), cultural factors (Cheng and Zhang, 2011), and housing prices (Li, 2018). However, no research has approached this problem from the perspective of population concentration. The existing literature primarily documents correlations between population density and consumption (Glaeser et al., 2001; Rosenthal and Strange, 2004). Yet, there remains a lack of formal causal analysis examining how population concentration affects household consumption.

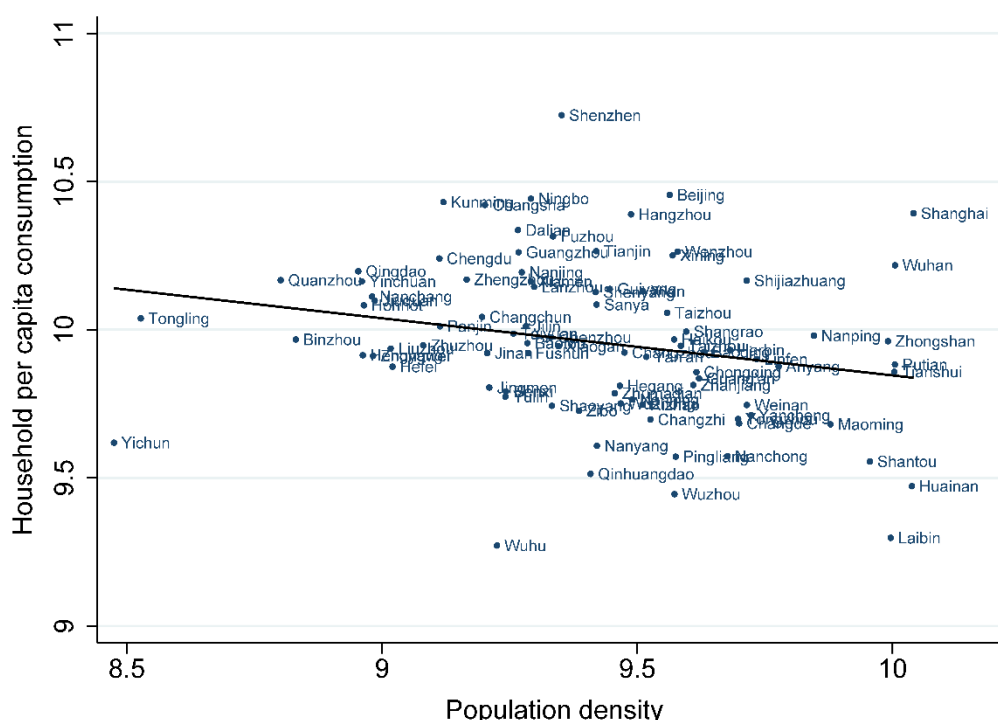
Population concentration generates both positive and negative externalities simultaneously. Existing literature demonstrates that urban population concentration produces positive externalities (Rosenthal and Strange, 2004), primarily manifested in productivity and income gains (Baum-Snow and Pavan, 2012), which subsequently stimulate consumption. Other studies document negative externalities of population concentration, such as crime (Mejía and Restrepo, 2016) and air pollution (Kang et al., 2019), which suppress consumption.

Assuming the aforementioned positive externalities dominate, if consumption also positively affects population concentration, this would lead to overestimation of the consumption effects of population concentration. Additionally, consumption preferences and residential location choices may be jointly determined by certain unobservable factors. If these unobservable confounding factors affect urban population density and household consumption in the same direction, the impact of population concentration on consumption would also be overestimated. For instance, high-skilled residents with preferences for consumption variety are more likely to choose to reside in large cities to fulfill their consumption needs (Lee, 2010).

Conversely, if certain confounding factors simultaneously affect population concentration and consumption in opposite directions, the relationship between them would be underestimated. Figure 1 illustrates a weak negative correlation between population

density and per capita household consumption expenditure across 89 cities in 2016.¹ Simple correlation analysis alone cannot establish whether this relationship reflects dominant negative externalities or endogeneity concerns. Our identification strategy addresses this challenge by accounting for different sources of endogeneity, enabling formal causal analysis of the consumption effects of population concentration and examination of the underlying mechanisms.

Figure 1: Population Density and Household Per Capita Consumption



Notes: The above figure presents a scatter plot of population density and household consumption for 89 prefecture-level and above cities in China in 2016. The horizontal axis represents the natural logarithm of urban population density (persons per square kilometer), while the vertical axis represents the natural logarithm of mean household per capita consumption (yuan). The definition and calculation of population density are detailed in the variable description section below. The calculation of urban average household per capita consumption employs household sampling weights. Data sources: China Household Finance Survey (CHFS) 2017, China City Statistical Yearbook, and LandScan database.

This paper employs panel microdata from the China Household Finance Survey (CHFS) spanning 2013 to 2017, utilizing two-way fixed effects models and instrumental variable

¹ In this paper, "cities" refers to municipal districts of prefecture-level and above cities.

methods to examine the causal relationship between population concentration and household consumption. Our findings reveal that, after controlling for different sources of endogeneity, population concentration exerts significant and robust positive effects on household consumption. A 1% increase in population density raises the average household propensity to consume by 0.430 percentage points and increases per capita household consumption expenditure by 0.355%. Mechanism analysis demonstrates that, given urban income premiums, population concentration reduces consumption transaction costs and enhances social interaction, providing both objective and subjective support for household consumption growth. Furthermore, evidence suggests that increased population density also raises non-housing consumption among households engaged in housing transactions or with housing purchase intentions by driving up housing prices.

This paper is the first study to directly quantify the causal effects of population density on household consumption using household microdata. The paper makes three primary contributions. First, by analyzing the causal relationship between population concentration and consumption and integrating existing consumption theory, we provide further empirical evidence for agglomeration economies. Second, this paper makes an important addition to the analysis of factors affecting household consumption. Urban population density constitutes a significant determinant of household consumption, with effects transmitted not only through income but also through reduced consumption transaction costs, increased social interaction, and elevated housing prices. Finally, this paper offers policy insights for boosting domestic consumption. Facilitating labor mobility and raising consumption across cities of varying sizes could promote more balanced regional development while sustaining economic growth.

2. Theoretical Framework

2.1 Consumer

The consumer's utility function is $U(C, H) = (C - \phi(d_j)\bar{C})^\alpha H^{1-\alpha}$, where $\alpha \in (0, 1)$. To capture the impact of social interactions on household consumption, we model consumption using an external habit formation specification, where C represents non-housing consumption, \bar{C} denotes the average consumption of other consumers (assumed

exogenous), and $\phi(d_j) \in (0,1)$ represents the information set regarding other consumers. A higher value of $\phi(d_j)$ indicates greater knowledge about other consumers' behavior, thereby amplifying the effect of relative consumption on utility. While many factors influence information transmission, this paper focuses primarily on the effects of population density. The existing literature demonstrates that geographic proximity has a positive impact on social learning and information diffusion (Bikhchandani et al., 1998), leading us to specify the information set as a function of population density (d_j) in region j . Additionally, H represents housing consumption in the utility function.

We assume that the representative consumer supplies one unit of labor to earn income W for consumption, subject to the budget constraint $W = C(1 + \tau(d_j)) + HP_H(d_j)$. Here, $P_H(d_j)$ denotes the housing price in region j , and $\tau(d_j) \geq 0$ represents generalized transaction costs, which include frictions in consumption matching as well as negative externalities generated by population concentration.²

Solving the consumer's utility maximization problem yields the consumption demand function:

$$C = \frac{\alpha}{1 + \tau(d_j)} W + (1 - \alpha)\phi(d_j)\bar{C}. \quad (1)$$

2.2 Firm

Following Glaeser and Maré (2001), we assume the firm's profit function is $A(d_j)K^\sigma L^{1-\sigma} - W_j L - RK$, where $\sigma \in (0,1)$. Here, K represents the capital stock with rental rate R , L denotes labor input, and W_j is the local average wage rate. $A(d_j)$ represents productivity determined by agglomeration economies (Rosenthal and Strange, 2004). From the first-order condition of the profit maximization problem, we obtain:

$$W_j = (1 - \sigma)\sigma^{\frac{\sigma}{1-\sigma}} R^{-\frac{\sigma}{1-\sigma}} A(d_j)^{\frac{1}{1-\sigma}}. \quad (2)$$

This equation represents an expression for the urban wage premium. Extensive literature

² For discussion of $\tau(d_j)$, see Appendix I.

documents the existence of urban wage premiums (Glaeser and Maré, 2001; Chen and Su, 2021), providing substantial evidence for agglomeration economies. Beyond agglomeration economies, worker wages are influenced by other factors, which can be expressed as $W = W(d_j) + \omega$, where ω represents wages and assets not determined by agglomeration economies. Combining Equation (2), we can rewrite Equation (1) as:

$$C = \frac{\alpha}{1 + \tau(d_j)} W(d_j) + \frac{\alpha}{1 + \tau(d_j)} \omega + (1 - \alpha) \phi(d_j) \bar{C}. \quad (3)$$

This equation demonstrates that, given the model parameters, population density affects consumption through its impact on wage income, generalized transaction costs, and social interaction.

2.3 Consumption Effects of Population Concentration and Mechanisms

Through comparative static analysis, we examine the consumption effects of population concentration and their transmission channels. Taking the derivative of Equation (3) with respect to population density d_j yields:

$$\frac{dC}{dd_j} = \frac{\alpha}{(1 + \tau_j)} \frac{dW_j}{dd_j} - \frac{\alpha(W_j + \omega)}{(1 + \tau_j)^2} \frac{d\tau_j}{dd_j} + (1 - \alpha) \bar{C} \frac{d\phi_j}{dd_j}. \quad (4)$$

Since $\tau(d_j)$ is nonlinear, if $d\tau_j/dd_j < 0$ (i.e., when positive externalities dominate), then $dC/dd_j > 0$. If $d\tau_j/dd_j > 0$, then $dC/dd_j > 0$ if and only if the consumption increase from wage premiums and social interaction exceeds the consumption decrease from increased transaction costs, i.e., when $\frac{\alpha}{(1 + \tau_j)} \frac{dW_j}{dd_j} + (1 - \alpha) \bar{C} \frac{d\phi_j}{dd_j} > \frac{\alpha(W_j + \omega)}{(1 + \tau_j)^2} \frac{d\tau_j}{dd_j}$; otherwise, $dC/dd_j < 0$. Therefore, determining the sign of the consumption effect of population concentration is an empirical question. However, we cannot yet attribute the negative correlation shown in Figure 1 to the dominance of negative externalities from population concentration, as endogeneity may exist in the data-generating process, which we will discuss in detail in Section 3. Moreover, since we cannot specify the exact functional form of $\tau(d_j)$, we employ a reduced-form approach in our empirical analysis. Based on the above discussion, our estimates should be interpreted as the net effect after offsetting positive and negative

externalities.

Equation (4) also reveals three channels through which population concentration affects household consumption. First, the marginal propensity to consume from local average wages is:

$$\frac{\partial C}{\partial W_j} = \frac{\alpha}{1 + \tau_j} \in (0,1).$$

Therefore, population concentration promotes consumption by increasing local average wages. Second,

$$\frac{\partial C}{\partial \tau_j} = -\frac{\alpha(W_j + \omega)}{(1 + \tau_j)^2} < 0.$$

The lower the transaction costs of consumption, the higher the consumption level. As previously discussed, the impact of population concentration on transaction costs is nonlinear. Higher population density improves the quality and opportunities for consumption matching, reducing $\tau(d_j)$ and thereby increasing consumption. Simultaneously, negative externalities such as congestion, crime, and pollution generated by increased population density raise $\tau(d_j)$, thereby suppressing consumption. Finally,

$$\frac{\partial C}{\partial \phi_j} = (1 - \alpha)\bar{C} > 0$$

Increased population density enhances social interaction, potentially leading to greater peer-influenced and conspicuous consumption.

Furthermore, if the utility function takes a constant elasticity of substitution form, $U(C, H) = (\alpha(C - \phi(d_j)\bar{C})^\rho + (1 - \alpha)H^\rho)^{1/\rho}$, population density may also affect consumption through its impact on housing prices:

$$\frac{\partial C}{\partial P_{Hj}} = \left(\frac{\rho}{1 - \rho}\right) \frac{\left[\frac{(1 + \tau_j)(1 - \alpha)}{P_{Hj}\alpha}\right]^{-\frac{1}{1-\rho}} [(1 + \tau_j)(C - \phi_j\bar{C}) + P_{Hj}H]}{\left\{(1 + \tau_j) \left[\frac{(1 + \tau_j)(1 - \alpha)}{P_{Hj}\alpha}\right]^{-\frac{1}{1-\rho}} + P_{Hj}\right\}^2}$$

If $\rho > 0$, meaning the elasticity of substitution between housing and non-housing

consumption $\delta = \frac{1}{1-\rho} > 1$, then $\frac{\partial C}{\partial P_{Hj}} > 0$. The housing demand effect of increased population density may drive up housing prices, and this effect is amplified because housing supply elasticity tends to be lower in areas with higher population density (Ahlfeldt and Pietrostefani, 2019). Rising housing prices generate substitution effects (Ahlfeldt and Pietrostefani, 2019) and wealth effects (Yan and Zhu, 2013) that reduce housing consumption while increasing non-housing consumption. However, if housing demand is inelastic, rising house prices may crowd out non-housing consumption under liquidity constraints (Fan and Liu, 2015; Li, 2018). Therefore, the theoretical impact of population density on consumption through housing prices remains ambiguous, which we will explore further in our subsequent empirical analysis.

3. Empirical Framework and Data

3.1 Identification Strategy

Based on the theoretical framework above, we estimate the following reduced-form of Equation (3):

$$consumption_{ijt} = \alpha \ln density_{jt} + \mathbf{X}_{ijt}\boldsymbol{\beta} + \varepsilon_{ijt}, \quad (5)$$

where the dependent variable $consumption_{ijt}$ represents the consumption measure for household i in city j at time t . We employ two approaches to measure household consumption: the natural logarithm of household per capita consumption expenditure and household average propensity to consume. Household per capita consumption expenditure measures the absolute quantity of consumption, while household average propensity to consume represents the household consumption rate, calculated as household consumption expenditure divided by household disposable income. The key explanatory variable $\ln density_{jt}$ is the natural logarithm of population density in city j at time t . α is the parameter of interest, representing the consumption effect of population concentration. \mathbf{X}_{ijt} is a set of control variables. ε_{ijt} is the error term. As shown in Figure 1, estimating Equation (5) using ordinary least squares (OLS) may yield a negative value for α , but we cannot conclude that negative externalities of population concentration dominate. This is because accurate estimation of α may be affected by the following channels.

First, unobservable confounding factors leading to sorting effects may bias OLS estimates in different directions. Assuming the true value of α is positive, if certain unobservable factors simultaneously affect population density and consumption in the same direction, OLS will overestimate α . For example, individuals with preferences for consumption variety are more likely to reside in high-density cities to satisfy their consumption needs. Conversely, if some unobservable factors have opposite effects on population density and consumption, OLS will underestimate α . For instance, individuals may accept lower real wages (and thus reduced consumption capacity) in order to work in large cities where their preferred career opportunities are concentrated. If this confounding effect dominates, then α could be negative. To reduce bias from omitted variables, we control for relatively exogenous observable variables in the model as much as possible. However, controlling for all exogenous confounding variables is impossible. This paper utilizes the panel nature of the data and employs two-way fixed effects models to address time-invariant household-level unobservable factors. The model specification is as follows:

$$consumption_{ijt} = \alpha \ln density_{jt} + X_{ijt}\beta + M_{jt}\gamma + \varphi_i + \pi_t + \mu_{ijt}, \quad (6)$$

where φ_i represents household fixed effects and π_t represents time fixed effects. Since the main explanatory variable in the equation is at the city level, we control for some observable city-level variables M_{jt} to examine the impact of such confounding factors. Additionally, time-varying unobservable confounding factors may also bias estimates. To explore the potential impact of such confounding factors, we employ two strategies. First, we augment Equation (6) by adding linear time trends for baseline values of household and city-level variables:

$$consumption_{ijt} = \alpha \ln density_{jt} + X_{ijt}\beta + M_{jt}\gamma + \varphi_i + \pi_t + \kappa_{i(j)0}t + \mu_{ijt}, \quad (7)$$

where $\kappa_{i(j)0}$ represents baseline values of household and city-level variables. Second, we use instrumental variables to reduce the impact of such endogeneity. This paper constructs a “shift-share” type instrumental variable (Bartik, 1991). We treat the annual population growth rate of other cities in the same province (excluding the city where the observed household resides) as an exogenous shock, and use the product of urban built-up area and the minimum distance to the Yangtze River, Yellow River, or coastline as weights. The product of these two

components serves as the instrumental variable (IV1). Using the product of urban built-up area and minimum distance to water bodies as weights is based on the following considerations: On one hand, extensive literature documents a strong correlation between population density and distance to coastlines (Beeson et al., 2001; Cai, 2019). Rivers and coastlines are typically areas of population concentration because they provide convenient access to water resources, transportation, and commercial activities. On the other hand, the minimum distance to water bodies combined with urban built-up area jointly characterize the constraints on urban development. Generally, land near rivers and coastlines is more restricted and may be affected by flood control measures, planning restrictions, or ecological protection, thereby influencing urban expansion. Urban built-up area refers to areas within a city's administrative boundaries that have been developed in clusters with basic municipal utilities and public facilities, and its size directly affects the city's population capacity. Combining these two factors provides a more complete measure of urban development constraints and thus better predicts population density.

Second, areas with high consumption levels may reflect superior consumption opportunities, potentially attracting migration inflows that increase population density. This potential reverse causality may overestimate α . The fixed effects model above cannot address this type of endogeneity. This paper uses instrumental variables to reduce the impact of such endogeneity. We use the linear time trend of urban population density in 1953 as an instrumental variable (IV2). Urban population data for 1953 comes from the first Census of the People's Republic of China in 1953. Using historical data as instrumental variables to mitigate bias from reverse causality is common in the literature (Lu et al., 2012). The literature typically uses historical values of urban population density as instruments for current values (Quintero and Roberts, 2023), with the advantage that the dependent variable is contemporaneous and cannot directly affect historical variables. Following Nunn and Qian (2014), we use the linear time trend of 1953 urban population density as an instrumental variable to satisfy model estimation conditions.

The two instrumental variables used in this paper target different sources of endogeneity. The "shift-share" instrumental variable (IV1) primarily addresses time-varying unobservable confounding factors; the historical data instrumental variable (IV2) primarily addresses

reverse causality. The validity of IV1 requires further discussion. A valid “shift-share” instrumental variable requires both the shock and weight components to be exogenous (Goldsmith-Pinkham et al., 2020). The shock component of IV1 is arguably exogenous; its weight component is the product of urban built-up area and minimum distance to water bodies, which may affect consumption through channels other than population density, thus raising questions about its exclusion restriction. This paper uses dual instrumental variables in regression analysis, which provides conditions for alleviating this concern: First, since the two instrumental variables have different sources of exogeneity, we can compare their independent estimation results to assess instrument validity. Second, using both instrumental variables simultaneously allows for overidentification tests to assess joint validity. Finally, we conduct strict exogeneity tests for IV1 following Grieser and Hadlock (2019).

3.2 Data and Variables

The main data sources for this paper include the China Household Finance Survey (CHFS) (Yin et al., 2015), LandScan, China City Statistical Yearbook, Amap POI data, Chinese General Social Survey (CGSS), CEIC database, and Juhui data.³ The dependent variables are household average propensity to consume and the natural logarithm of household per capita consumption expenditure. The consumption expenditure in our main analysis refers to non-housing consumption expenditure. The key explanatory variable is urban population density.⁴

The model also controls for household income and wealth variables, including household per capita income, household per capita financial assets, homeownership status, household per capita housing assets, and housing debt status; household characteristics, including self-employment status, youth dependency ratio, elderly dependency ratio, proportion of unemployed household members, pension insurance coverage for household heads and spouses, and medical insurance coverage for household heads and spouses; and household head characteristics, including risk preference, gender, age, years of education,

³ Detailed descriptions of data sources are provided in Appendix II.

⁴ Construction of key variables is detailed in Appendix III.

marital status, health status, and hukou type. Additionally, this paper attempts to control for city-level variables that may affect household consumption, including GDP, tertiary industry share of GDP, financial development level, digital inclusive finance, housing prices.⁵

Sample restrictions are as follows: We retain household samples that have not relocated since 2013; exclude samples with zero total household consumption, non-positive total income, or household heads younger than 20 or older than 80 years; exclude samples with missing or obviously erroneous key variables. This yields a balanced panel of 12,804 observations from 6,402 households across 83 cities in 29 provinces and centrally-administered municipalities, constituting our final dataset. Descriptive statistics for main variables are reported in Table 1. Unless otherwise specified, all empirical procedures use household survey weights to ensure population-representative estimates.

Table 1: Descriptive Statistics of Main Variables

Variable	Mean	Std. Dev.	Min	Max
Dependent Variables				
Ln(household per capita consumption)	9.354	0.863	3.790	13.296
Household average propensity to consume	1.055	0.757	0.269	3.929
Main Explanatory Variable				
Ln(population density)	9.573	0.340	8.408	10.480
Household Income and Wealth Variables				
Ln(household per capita income)	9.368	1.360	-1.894	14.852
Ln(household per capita financial assets)	7.639	2.571	0.000	15.406
Homeownership status	0.894	--	0	1
Ln(household per capita housing value)	7.421	5.484	0.000	15.742
Housing debt status	0.240	--	0	1
Household Characteristics				
Self-employment status	0.138	--	0	1
Youth dependency ratio	0.103	0.152	0	0.778
Elderly dependency ratio	0.158	0.300	0	1
Share of unemployed members	0.504	0.322	0	1
Number with pension insurance (respondent and spouse)	1.495	0.739	0	2
Number with medical insurance (respondent and spouse)	1.705	0.571	0	2
Household Head Characteristics				
Risk preference	4.108	1.180	1	5
Gender	0.768	--	0	1

⁵ Specific variable descriptions are provided in Appendix III.

Age	52.296	12.661	20	80
Years of education	9.506	3.786	0	22
Marital status	0.889	--	0	1
Health status	2.617	1.168	1	5
Hukou type	0.525	--	0	1
City Characteristics				
Ln(GDP)	16.762	1.514	13.693	19.179
Tertiary sector share of GDP	0.495	0.119	0.273	0.797
Financial development (loans/GDP ratio)	1.601	0.879	0.188	8.894
Ln(housing prices)	8.297	0.862	6.096	10.360
Ln(digital inclusive finance)	5.665	0.160	5.316	6.136

Note: Sample restrictions are described in the text. The sample includes 6,402 households observed over 2 years, yielding 12,804 observations. Sample households are located across 83 cities in 29 provinces and centrally-administered municipalities. All non-percentage variables are in natural logarithms. Household average propensity to consume has 10,883 observations after excluding outliers. Sample sizes reported represent the raw data sample; effective sample sizes used in empirical procedures may be smaller. Effective sample sizes for all empirical procedures are available from the authors upon request. All results use household sampling weights.

4. Empirical Results

Table 2 presents the regression results from pooled Ordinary Least Squares (OLS) for Equation (5) and the fixed effects models for Equations (6) and (7). Panel A uses household average propensity to consume as the dependent variable, while Panel B uses the natural logarithm of household per capita consumption.⁶ In Column (1), we control only for population density, which shows a positive but insignificant correlation with household average propensity to consume. In Panel B, consistent with the relationship shown in Figure 1, the elasticity between population density and household per capita consumption is negative but not significant. To control for potential confounding factors, Column (2) adds household asset variables, household characteristics, and household head characteristics. The results show that population concentration has a positive effect on household consumption, though it remains statistically insignificant. The OLS results preliminarily suggest that the relationship between population concentration and household consumption may be affected by omitted variables, with confounding factors that adversely affect both population density and consumption playing a dominant role, leading to an underestimation of the consumption effect of population concentration.

⁶ Detailed results for Table 2 are reported in Appendix IV.

To further control for unobservable household confounding factors, we employ a two-way fixed effects model in Column (3) to eliminate the influence of time-invariant household characteristics. Panel A results show that a 1% increase in population density leads to a 0.357 percentage point increase in household average propensity to consume, significant at the 1% level. In Panel B, a 1% increase in population density raises average household consumption expenditure by 0.271%, significant at the 5% level. Compared with Column (2), we find that pooled OLS estimates underestimate the consumption effect of population concentration. This suggests that time-invariant unobservable factors cause a systematic sorting pattern where high-consumption households are more likely to locate in low-density areas, thereby masking the true positive effect of population concentration on consumption. In Column (4), we further control for city characteristics. The results show that a 1% increase in population density increases household average propensity to consume by 0.430 percentage points (significant at the 1% level) and household per capita consumption by 0.355% (significant at the 1% level). Additionally, the coefficient for city average housing prices is not significant.

As discussed in the identification strategy, time-varying unobservable confounding factors may also bias estimates, so we estimate equation (7). Column (5) of Table 2 reports estimation results controlling for linear time trends of baseline values of household and city-level variables, which show no qualitative changes compared to Column (4). We also conduct further exploration using instrumental variable methods. Column (1) of Table 3 presents two-stage least squares (2SLS) estimation results using the product of urban built-up area and minimum distance to water bodies multiplied by the annual population growth rate of other cities in the same province as an instrumental variable (IV1). Results show that a 1% increase in population density leads to a 0.459 percentage point increase in household average propensity to consume and a 0.381% increase in household per capita consumption. The first-stage F-statistics are 978.044 and 1098.765, respectively, indicating minimal risk of weak instrument problems. The generalized Hausman test p-values are 0.199 and 0.142, respectively, suggesting we cannot reject the null hypothesis at conventional significance levels, meaning the 2SLS estimates in Column (1) of Table 3 are not statistically different from the OLS estimates in Column (4) of Table 2. Therefore, based on the above evidence, the impact of time-varying unobservable confounding factors on estimates may be limited.

Table 2: Impact of Population Concentration on Household Consumption

	OLS		FE		FE + Time Trends
	(1)	(2)	(3)	(4)	(5)
A. Dependent Variable: Household Average Propensity to Consume					
Ln(Population Density)	0.017 (0.061)	0.040 (0.040)	0.357*** (0.122)	0.430*** (0.134)	0.404*** (0.132)
Ln(Housing Price)				-0.056 (0.041)	-0.034 (0.040)
B. Dependent Variable: Ln(Household Per Capita Consumption)					
Ln(Population Density)	-0.101 (0.178)	0.004 (0.065)	0.271** (0.119)	0.355** (0.137)	0.343** (0.146)
Ln(Housing Price)				-0.037 (0.033)	-0.011 (0.045)
Household Controls	No	Yes	Yes	Yes	Yes
City Controls	No	No	No	Yes	Yes
Household Fixed Effects	No	No	Yes	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes	Yes
Time Trends	No	No	No	No	Yes

Notes: Household controls include: household per capita income, household per capita financial assets, homeownership status, household per capita housing value, housing debt status, self-employment in business, demographic structure (including youth dependency ratio, elderly dependency ratio, and proportion of unemployed household members), number of respondents and spouses with pension insurance, number of respondents and spouses with medical insurance, household head characteristic variables (including risk preference, gender, age, years of education, marital status, health status, and household registration type). City controls include: GDP, share of tertiary industry in GDP, financial development level, housing prices, and digital inclusive finance. All variables represented as discrete integers are entered into the regression as dummy variables. Time trends refer to linear time trends of baseline values of household and city-level variables. All results use household sampling weights. Robust standard errors clustered at the city level are in parentheses. Sample sizes for all columns in Panel A are 10,883; sample sizes for all columns in Panel B are 12,804. ** and *** indicate significance at the 5% and 1% levels, respectively.

As discussed earlier, besides omitted variable bias, endogeneity between population density and household consumption may also arise from reverse causality. Column (3) of Table 3 reports two-way fixed effects estimation results using the linear time trend of 1953 city population density as an instrumental variable (IV2). The 2SLS results show that a 1% increase in population concentration leads to a 0.431 percentage point increase in household average propensity to consume and a 0.356% increase in household per capita consumption expenditure. The generalized Hausman test p-values are 0.742 and 0.848, respectively, indicating that these 2SLS estimates are not statistically different from the OLS coefficients in Column (4) of Table 2. This may be because in the model of Column (4) in Table 2, the

dependent variable is at the household level, and its impact on the entire city's population density is minimal, so bias from reverse causality is negligible.

A potential concern with the above instrumental variable estimation is that the weight component of IV1, which is the product of urban built-up area and minimum distance to water bodies, may affect current household consumption by influencing the local institutional environment and economic development levels. Since the exclusion restriction of instrumental variables cannot be directly tested, we use indirect evidence to alleviate this concern. First, IV1 and IV2 have different sources of exogeneity. If their independent estimation results are similar, it suggests both instrumental variables may be valid. As mentioned earlier, the 2SLS estimation results in Columns (1) and (3) of Table 3 are not statistically different from the OLS results in Column (4) of Table 2. Therefore, both instrumental variables may be valid. Second, Column (4) of Table 3 reports 2SLS estimation results using both IV1 and IV2 as instrumental variables. The Hansen J statistic p-values are 0.139 and 0.194, respectively, indicating we cannot reject the null hypothesis at conventional significance levels, meaning all instrumental variables are jointly valid. The generalized Hausman test also shows this estimation result is not statistically different from the OLS estimates in Column (4) of Table 2. Finally, following Grieser and Hadlock (2019), we conduct a strict exogeneity test for IV1. If the instrumental variable is valid, then fixed effects and first-difference models should yield consistent estimates. Column (2) of Table 3 reports 2SLS estimation results for the first-difference model, with generalized Hausman test p-values of 0.128 and 0.207. Therefore, like the fixed effects model 2SLS results in Column (1), they are not statistically different from the OLS estimation results in Column (4) of Table 2, preliminarily indicating that IV1 has strict exogeneity. For further verification, Column (5) of Table 3 reports formal strict exogeneity test results. Adding the forward term of IV1 to the model, if its estimated coefficient is significantly different from zero, then IV1 does not have strict exogeneity. The estimated values of the forward IV1 term reported in Column (5) are not significant at conventional significance levels, indicating that we cannot reject that IV1 has strict exogeneity. Additionally, the generalized Hausman test also shows this estimation result is not statistically different from the OLS estimates in Column (4) of Table 2. Based on the above evidence, our instrumental variable estimation results are credible, but they are not

statistically different from OLS estimates, further indicating that potential endogeneity from time-varying unobservable confounding factors and reverse causality is limited. For estimation efficiency considerations, we will use the model in Column (4) of Table 2 as the benchmark model specification in subsequent analysis.

Table 3: Two-Stage Least Squares Estimation

	IV1		IV2		IV1 和 IV2
	FE-2SLS	FD-2SLS	FE-2SLS	FE-2SLS	FE-2SLS
	(1)	(2)	(3)	(4)	(5)
A. Dependent Variable: Household Average Propensity to Consume					
Ln(Population Density)	0.459*** (0.139)	0.522*** (0.154)	0.431*** (0.136)	0.432*** (0.136)	0.486*** (0.166)
IV1 Forward Term					0.030 (0.128)
First-Stage F-Statistic	978.044	962.383	10147.091	5252.027	742.363
Generalized Hausman Test	2.152	2.319	0.108	0.010	2.145
χ^2 Statistic [p-value]	[0.142]	[0.128]	[0.742]	[0.920]	[0.143]
Hansen J Statistic [p-value]	--	--	--	2.187 [0.139]	--
B. Dependent Variable: Ln(Household Per Capita Consumption)					
Ln(Population Density)	0.381*** (0.136)	0.416*** (0.142)	0.356*** (0.137)	0.356*** (0.137)	0.336* (0.176)
IV1 Forward Term					-0.048 (0.120)
First-Stage F-Statistic	1098.765	1100.254	9521.712	4995.011	817.563
Generalized Hausman Test	1.649	1.589	0.037	0.009	1.845
χ^2 Statistic [p-value]	[0.199]	[0.207]	[0.848]	[0.923]	[0.174]
Hansen J Statistic [p-value]	--	--	--	1.688 [0.194]	--

Notes: Except for Column (5) which additionally controls for the lead term of IV1, all model specifications are consistent with Column (4) of Table 2. Columns (1), (3), (4), and (5) report fixed effects two-stage least squares estimates; Column (2) reports first-difference two-stage least squares estimates. The instrumental variable IV1 used in Columns (1), (2), and (5) is the product of urban built-up area and minimum distance to water bodies, multiplied by the annual population growth rate of other cities in the same province during the corresponding sample years. The instrumental variable IV2 used in Column (3) is the linear time trend of 1953 city population density. Column (4) uses both IV1 and IV2 as instrumental variables. Sample sizes for Columns (1), (3) – (5) in Panel A are 10,883, Column (2) is 4,701; sample sizes for Columns (1), (3) – (5) in Panel B are 12,804, Column (2) is 6,402. All results use household sampling weights. Robust standard errors clustered at the city

level are in parentheses; p-values are in brackets. * and *** indicate significance at the 10% and 1% levels, respectively.

The above analysis provides evidence that population concentration has a significant positive impact on household consumption.⁷ Based on the benchmark model, a 1% increase in population concentration (calculated at the mean value, this equals 152 people per square kilometer) leads to a 0.430 percentage point increase in household average propensity to consume and a 0.355% increase in household per capita consumption expenditure (calculated at the mean value, this equals 59.73 yuan). Unobservable confounding factors causing sorting effects significantly underestimate this causal relationship, implying that high-consumption populations tend to be non-randomly distributed in low-density areas, while low-consumption populations tend to concentrate in high-density areas. This source of endogeneity has important implications for policy design.

5. Mechanism Tests

In the theoretical framework presented in Section II, we discussed potential channels through which population concentration affects household consumption: generating urban wage premiums, reducing consumption transaction costs, increasing social interaction, and driving up housing prices. Given that extensive literature documents urban income premiums (Glaeser and Maré, 2001; Chen and Su, 2021), we focus our analysis on testing the latter three mechanisms, subject to data availability. On the one hand, the consumption effects of population concentration require cities to provide objective supply-side conditions that reduce transaction costs. On the other hand, population concentration increases social interaction, which may subjectively enhance residents' consumption propensity. Finally, we examine whether population density affects non-housing consumption by driving up housing prices.

5.1 Consumption Matching

Since transaction costs cannot be directly measured, we examine this mechanism from the perspective of consumption matching quality. Appendix I demonstrates that higher

⁷ Robustness check results and discussions are provided in Appendix V, with main results remaining robust. Due to space limitations, heterogeneity analysis is presented in Appendix VI.

population density leads to a smaller expected distance to product characteristics, resulting in higher consumption matching quality. To estimate the expected distance to product characteristics, we assume that the distribution of product characteristics y demanded by consumers remains constant. The greater the variety of available product characteristics, the smaller the expected distance to desired product characteristics. Entropy measures the expected information needed to purchase products with characteristic x : $H(x) = -\sum_x p(x) \ln p(x)$, where $p(x)$ is the probability of purchasing products with characteristic x . Higher entropy $H(x)$ indicates greater diversity of unique product characteristics available in the market, a higher likelihood of satisfying consumers' specific needs, and higher consumption matching quality.

We use the variety of locally non-tradable services (Li et al., 2019) to represent different product characteristics for calculating the entropy of local non-tradable service markets. Existing research finds a positive correlation between population concentration and non-tradable services (Glaeser et al., 2001; Li et al., 2019), and people are willing to accept lower real income to live in large cities in order to consume these non-tradable services (Lee, 2010). Specifically, we use the number of service types within five categories of points of interest (POI) in each city: dining services, shopping services, education and cultural services, daily life services, and sports and leisure services. These serve as product characteristics to calculate the entropy of non-tradable service markets in each city.⁸ Entropy density, obtained by dividing entropy by urban built-up area, measures consumption matching quality per unit area within cities.

Table 4 reports estimation results for the impact of population concentration on entropy density. Column (1) presents results where the dependent variable is the total entropy density of non-tradable services in each city. Columns (2)-(6) show results for entropy density of dining services, shopping services, education and cultural services, daily life services, and sports and leisure services, respectively. Results indicate that population concentration significantly enhances entropy density for both aggregate and individual service categories. This demonstrates that population concentration affects consumption by improving

⁸ Non-tradable service data comes from Amap POI data for 2014 and 2016.

consumption matching quality.⁹

Table 4: Impact of Population Concentration on Consumption Matching Quality

	Dependent Variable					
	Non-tradable Service Matching Quality	Dining Service Matching Quality	Shopping Service Matching Quality	Education & Cultural Service Matching Quality	Daily Life Service Matching Quality	Sports & Leisure Service Matching Quality
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Population Density)	0.047*** (0.010)	0.020*** (0.004)	0.035*** (0.008)	0.029*** (0.007)	0.034*** (0.009)	0.023*** (0.006)
Sample Size	166	166	166	166	166	166

Notes: The dependent variables are entropy density for various non-tradable services. All regressions are unweighted. Control variables include: GDP, tertiary industry share of GDP, financial development level, housing prices, digital inclusive finance, city dummies, and year dummies. Standard errors clustered at the city level are reported in parentheses. *** indicates significance at the 1% level.

5.2 Social Interaction

In relatively dense urban areas, social interaction is more easily achieved (Glaeser and Gottlieb, 2006), enabling individuals to more readily obtain consumption information from others, which affects their consumption preferences (George and Waldfogel, 2003) and consumption behavior (Kuhn et al., 2011).

In Column (1) of Table 5, we first examine the correlation between population concentration and social interaction using data from the China General Social Survey (CGSS) 2013. Results show a significant positive correlation between population concentration and social interaction. To further explore whether population concentration affects household consumption through social interaction, we replace the dependent variables in our baseline model with total household gift expenditure and per capita gift expenditure using CHFS data. Gift expenditure is defined as the actual monetary value of gifts and cash given to non-household members during traditional holidays and ceremonial events. On one hand, relevant literature (Guo and Liang, 2014) indicates that household gift expenditure can serve as a proxy for household social interaction levels. On the other hand, gift expenditure in the form of presents can be viewed as consumption expenditure arising from social interaction,

⁹ Appendix VII presents results for the impact of population concentration on non-tradable service density.

providing direct evidence for this channel. Gift expenditure in cash form increases recipients' budgets, thereby increasing consumption expenditure. Columns (2) and (3) of Table 5 report the impact of population concentration on household gift expenditure. Results show that higher urban population concentration increases residents' social spending. This reflects both deeper social engagement and represents a direct form of consumption.

Table 5: Impact of Population Concentration on Residents' Social Interaction

	Dependent Variable		
	Social Frequency	Ln(Total Gift Expenditure)	Ln(Per Capita Gift Expenditure)
	(1)	(2)	(3)
Ln(Population Density)	0.259** (0.119)	1.970** (0.929)	1.713** (0.815)
Control Variables	Yes	Yes	Yes
Household Fixed Effects	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes
Sample Size	3,027	12,804	12,804

Notes: Column (1) dependent variable is social frequency, using data from the China General Social Survey (CGSS). Respondents were asked to rate “How frequently do you engage in social and recreational activities with neighbors” and “How frequently do you engage in social and recreational activities with other friends” on a scale from 1-7, ranging from “never” to “almost daily.” The social frequency measure is the average of these two responses, with higher values indicating more frequent social activities. Columns (2) and (3) dependent variables are the natural logarithms of total household gift expenditure and per capita gift expenditure, respectively, using data from the China Household Finance Survey (CHFS). Model specifications for Columns (2) and (3) are identical to Table 2 Column (4) except for the dependent variables. Control variables in Column (1) include: household income, gender, age, years of education, marital status, health status, household registration type, self-rated socioeconomic status, GDP, tertiary industry share of GDP, financial development level, housing prices, and digital inclusive finance. All discrete integer variables are included as dummy variables. All results use sampling weights. Standard errors clustered at the city level are reported in parentheses. ** indicates significance at the 5% level.

5.3 Housing Prices

First, we estimate the relationship between population density and housing prices. Columns (1) and (2) of Table 6 show that both new and existing housing prices exhibit positive elasticity with respect to population density. This initially supports the possibility that population density affects household non-housing consumption through housing prices. Groups that are more sensitive to housing prices, including households that engaged in housing transactions or have housing purchase plans during the sample period, are more likely to have their non-housing consumption affected by population density through housing prices. Results in Columns (3) and (4) of Table 6 show that for households insensitive to

housing prices, a 1% increase in population density increases household per capita consumption expenditure by 0.384%, while population density increases have no significant effect on their per capita housing expenditure.¹⁰ For households sensitive to housing prices, increases in population density both raise per capita non-housing consumption expenditure and significantly reduce per capita housing expenditure. This suggests that higher population density may create substitution effects between non-housing consumption and housing expenditure by driving up housing prices. We therefore test this mechanism further.

Table 6: Impact of Population Concentration on Housing Prices and Consumption

	Dependent Variable			
	Ln(New Housing Prices)	Ln(Existing Housing Prices)	Ln(Household Per Capita Consumption)	Ln(Household Per Capita Housing Expenditure)
	(1)	(2)	(3)	(4)
Ln(Population Density)	0.351* (0.178)	0.323* (0.182)	0.384*** (0.135)	-0.127 (0.352)
Price Sensitive × Ln(Population Density)			-0.002 (0.003)	-0.035*** (0.008)
Housing Debt × Ln(Population Density)			-0.157** (0.076)	-0.230 (0.187)
Price Sensitive × Housing Debt × Ln(Population Density)			0.003 (0.006)	0.002 (0.013)
Household Controls	No	No	Yes	Yes
City Controls	Yes	Yes	Yes	Yes
Household Fixed Effects	No	No	Yes	Yes
City Fixed Effects	Yes	Yes	No	No
Year Fixed Effects	Yes	Yes	Yes	Yes
Sample Size	166	97	12,804	12,355

Notes: Control variables in Columns (1) and (2) include: GDP, tertiary industry share of GDP, financial development level, digital inclusive finance, city dummies, and year dummies. Price sensitive is a dummy variable that equals 1 if households engaged in housing transactions or have home purchase plans during the sample period, and 0 otherwise. Housing debt is a dummy variable that equals 1 if households have housing debt during the sample period, and 0 otherwise. Columns (3) and (4) follow the same model specification as Table 2 Column (4), except for the addition of interaction terms between price sensitivity and housing debt with the natural logarithm of population density. See text for detailed variable and sample

¹⁰ For rental households, housing expenditure is actual rent; for owners, housing expenditure is estimated using the estimated city average rent-to-price ratio multiplied by housing value. See Appendix VIII for specific estimation methods.

descriptions. Columns (1) and (2) use city population as weights, while the remaining columns use household sampling weights. Standard errors clustered at the city level are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

New and existing housing prices may have different effects on household consumption decisions, so we distinguish between new and existing housing prices in subsequent tests. Panel A of Table 7 presents results for new housing prices. Column (1) shows that for price-sensitive households, after controlling for new housing prices, the impact of population density remains positive but is no longer significant. The coefficient for new housing prices on household average non-housing consumption propensity is positive but also lacks statistical significance. Therefore, we cannot definitively conclude that population density affects non-housing consumption of these residents through housing prices. These results may be due to non-random entry into the price-sensitive sample, affecting our estimation of population concentration's consumption effects. We therefore use the two-step Heckman method to correct for potential sample bias. Column (2) of Panel A reports results using new housing prices as the exclusion restriction. The exclusion restriction here assumes that new housing prices only affect household average non-housing consumption propensity through influencing housing transactions or purchase intentions. Results show that after correcting for sample bias, the estimated consumption effect of population concentration is numerically larger than in Column (1) and achieves statistical significance. This means that if we randomly sample households sensitive to housing prices, a 1% increase in population density would increase their household average non-housing consumption propensity by 1.693 percentage points. The significant inverse Mills ratio coefficient also confirms the existence of sample selection bias. Therefore, combining these results, we can conclude that population concentration affects non-housing consumption of price-sensitive households by driving up housing prices.

To verify whether this mechanism also holds for households insensitive to housing prices, Columns (3) and (4) of Panel A report placebo test results. Column (3) shows that after controlling for new housing prices, the estimated consumption effect of population concentration in the price-insensitive sample does not differ significantly from our main results. The coefficient for new housing prices is also insignificant. Column (4) uses the

Heckman method to correct for the non-randomness of households being insensitive to housing prices. Results show no significant sample selection bias problem, and the random sample results remain robust compared to main results. Therefore, we can further confirm that the mechanism of population concentration affecting household consumption through housing prices operates in the price-sensitive sample.

Panel B of Table 7 reports corresponding results replacing new housing prices in Panel A with existing housing prices. Column (5) shows that in the price-sensitive sample, after controlling for existing housing prices, the consumption effect of population concentration is significant and numerically large. Although the coefficient for existing housing prices is positive, it is not significant. Column (6) shows that after correcting for significant sample selection bias, the consumption effect of population concentration increases numerically. These results are similar to those in Panel A, indicating that in the price-sensitive sample, population concentration affects household consumption by driving up housing prices. Results in Columns (7) and (8) are similar to corresponding results in Panel A, showing insufficient evidence for this mechanism in the price-insensitive sample.

Table 7: Housing Price Mechanism of Population Concentration's Consumption Effects

Dependent Variable: Household Average Consumption Propensity				
A. New Housing Prices				
	Price Sensitive	Price Sensitive	Price Insensitive	Price Insensitive
	(1)	(2)	(3)	(4)
	FE	Heckman	FE	Heckman
Ln(Population Density)	0.446 (0.747)	1.693* (0.881) [0.796]	0.359* (0.191)	0.422* (0.247) [0.249]
Ln(New Housing Prices)	0.740 (0.616)		-0.043 (0.180)	
Inverse Mills Ratio		23.118** (10.363) [9.354]		-1.246 (2.485) [2.374]
Sample Size	2,390	2,390	8,493	8,493
B. Existing Housing Prices				
	Price Sensitive	Price Sensitive	Price Insensitive	Price Insensitive
	(5)	(6)	(7)	(8)
	FE	Heckman	FE	Heckman
Ln(Population Density)	4.541*** (0.940)	4.681*** (0.904)	0.640** (0.310)	0.652** (0.310)

		[0.733]		[0.305]
Ln(Existing Housing Prices)	1.739 (1.062)		0.161 (0.254)	
Inverse Mills Ratio		18.553** (6.784)		-0.817 (1.307)
		[5.024]		[1.262]
Sample Size	1,817	1,465	6,054	5,242

Notes: The samples in Columns (1), (2), (5), and (6) are price-sensitive samples, namely households that have engaged in housing transactions or have housing purchase intentions during the sample period; the samples in Columns (3), (4), (7), and (8) are the remaining price-insensitive samples. The model specifications for odd-numbered columns are consistent with Column (4) of Table 2 except for housing prices; even-numbered columns use the two-stage Heckman method for estimation, with model specifications consistent with Column (4) of Table 2 except for the exclusion restriction variables in the second stage. Panel A uses new housing prices as the exclusion restriction variable in Heckman estimation, while Panel B uses existing housing prices. All results use household sampling weights. Numbers in parentheses are robust standard errors adjusted for city-level clustering; numbers in brackets are Wild Bootstrap standard errors. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively; significance levels are based on p-values calculated using clustered standard errors.

6. Policy Implications

China's economy is at a critical juncture of structural transformation. Boosting domestic consumption is essential for sustainable economic growth. Our analysis of household microdata from 2013 – 2017 establishes a positive causal link between population concentration and household consumption, with significant implications for regional policy design.

China's regional policies involve a fundamental efficiency-equity trade-off. Economically, concentrating resources and population in productive regions would maximize overall efficiency. Yet concerns about equity have driven policies that favor less productive areas to reduce regional disparities (Lu et al., 2015). This equity focus may sacrifice economic efficiency. Meanwhile, China's hukou (household registration) system restricts population mobility and hampers regional development (Lu et al., 2012; Zhang et al., 2020). Studies suggest that removing migration barriers and allowing market forces to guide population flows would create larger, denser cities concentrated in major metropolitan areas (Zhong et al., 2020). While this would boost economic efficiency, it might worsen regional imbalances.

Our findings shed new light on the efficiency-equity debate. We confirm that population

density generally boosts economic efficiency and household consumption. However, this positive relationship becomes statistically insignificant in the largest metropolitan areas, where population density no longer exerts a meaningful effect on household consumption behavior. These findings indicate that population concentration beyond optimal thresholds generates diseconomies of scale, calling into question the assumption that unrestricted urban expansion maximizes efficiency.

Our results point toward a potential solution: redirecting population from megacities to smaller urban centers could reduce inefficiencies in large cities while allowing smaller cities to reap agglomeration benefits. This approach might achieve regional balance without sacrificing overall efficiency. While our analysis is limited to consumption effects, the evidence of diminishing returns in this domain raises questions about whether the efficiency-equity trade-off in regional policy may be less severe than commonly assumed.

Before these findings can inform policy decisions, several important questions require further research. First, effective population policies require understanding the factors that drive household sorting, particularly why high-consumption households gravitate toward low-density areas while low-consumption households concentrate in high-density areas. Second, policies designed to stimulate urban consumption should also account for distributional concerns, as boosting aggregate consumption may not address underlying inequality within metropolitan areas. Finally, further research is needed to examine whether the consumption effects of population concentration represent level effects or growth effects for migrant residents.

Appendix I: Population Density and Transaction Costs

The quality and opportunities of consumption matching are influenced by population density. The quality of consumption matching refers to the gap between consumers' idiosyncratic preferences for product characteristics and the product characteristics available in the market. Higher population density leads to better consumption matching quality, reducing the discrepancy between consumers' preferences for product characteristics and the characteristics of products available in the market. This can be explained using the unit circle model commonly employed in urban labor market studies (Duranton and Puga, 2004; Helsley and Strange, 1990).

Assume there are n consumers, each corresponding to a unique preference for product characteristics. These heterogeneous product characteristics are uniformly distributed on a unit circle with coordinates y , where n represents the population density in this unit circle space. In the same space, assume there are m heterogeneous firms. Each firm provides a product with unique characteristics based on consumer demand, with these product characteristics located at coordinates x on the unit circle. The quality of consumption matching is measured by the distance between the product characteristics provided by firms and those desired by consumers, $|x - y|$; the smaller the distance, the higher the matching quality. Different consumers can consume products with the same characteristics. Let $Y(x)$ denote the coordinate set of all consumers who consume products with characteristic x . According to the proof by Helsley and Strange (1990), the expected total characteristic distance for product characteristic x is $E \sum_{Y(x)} |x - y| = n/4m^2$. Since each firm provides products with consistent market coverage θ , we have $E \sum_{Y(x)} |x - y| = \theta^2/4n$. Clearly, the first derivative of this expression with respect to n is negative; therefore, the greater the population density, the smaller the expected total characteristic distance, and the higher the expected consumption matching quality. From the perspective of consumption matching quality, $\tau(d_j)$ is a decreasing function of d_j .

We provide a numerical example of product characteristic matching. Assume each consumer corresponds to a unique product characteristic preference, and these characteristic

preferences are uniformly distributed on a unit circle. Figure I1 below shows a case with four consumers. These four consumers, or four types of heterogeneous product characteristics, are uniformly distributed clockwise on a unit circle. For analytical simplicity, assume consumer 1's coordinate is 0. The coordinate set for consumers is $\{0, 1/4, 1/2, 3/4\}$. Firms enter the market and provide products with unique characteristics. Assuming each firm's market coverage can serve two consumers, there are 2 firms in this example. Firm 1 provides products with characteristic coordinate $1/8$, and Firm 2 provides products with characteristic coordinate $5/8$. Each consumer faces a characteristic distance of $1/8$, and the total characteristic distance for each product is $1/4$.

Now assume the number of consumers increases to 6, with other conditions unchanged. As shown in Figure I2, the coordinate set for consumers becomes $\{0, 1/6, 1/3, 1/2, 2/3, 5/6\}$, with three firms providing products with characteristics $\{1/12, 5/12, 3/4\}$ respectively. Each consumer now faces a characteristic distance of $1/12$, and the total characteristic distance for each product becomes $1/6$. Figure I3 shows the case where the number of consumers increases to 12, with 6 firms. Each consumer faces a characteristic distance of $1/24$, and the total characteristic distance for each product is $1/12$. This demonstrates that as consumer density increases, characteristic distance decreases, consumption matching quality improves, consumers are willing to pay more for products that better match their preferences, and therefore consumption expenditure increases.

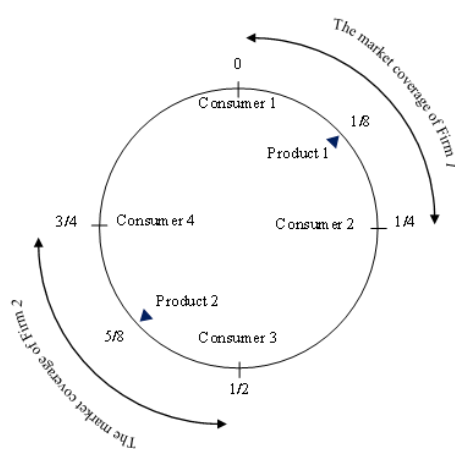


Figure I1: Consumer density of 4

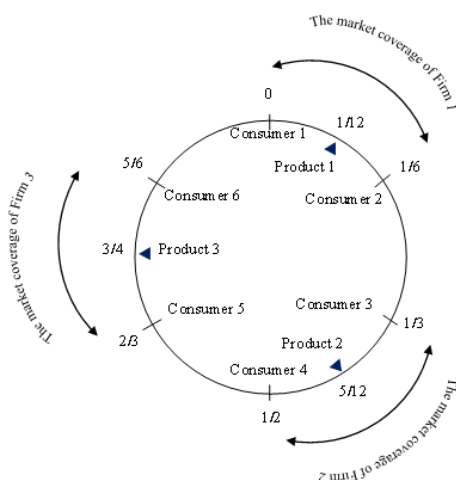


Figure I2: Consumer density of 6

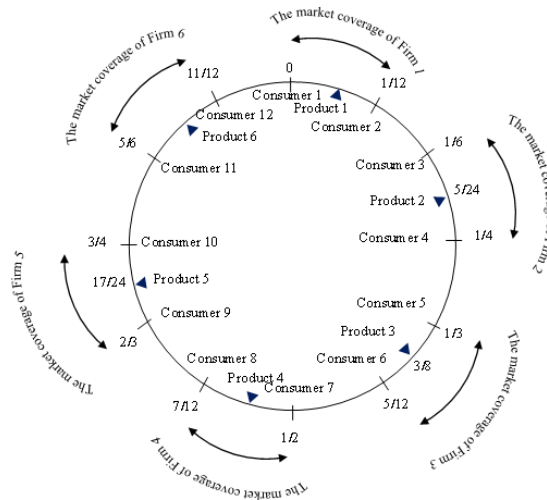


Figure I3: Consumer density of 12

On the other hand, population density may affect consumption matching opportunities. If we view successful transactions as successful matching between buyers and sellers, then the more buyers and sellers in a region, the greater the opportunities for consumption transactions to occur. Duranton and Puga (2004) provide a comprehensive review of search and matching models that explain this mechanism. Although these models are built on job-search scenarios, Duranton and Puga (2004) also note that these models essentially model the matching between buyers and sellers.

The first strand of literature studies non-coordinated random matching. This approach yields an aggregate matching function with decreasing returns to scale, but as the number of trading parties increases, the degree of decreasing returns to scale diminishes, approaching constant returns to scale in the limit. The second strand of literature focuses on how search frictions are endogenously generated when market participants choose their locations to better target their search objectives; this model also yields an aggregate matching function with constant returns to scale. In the third type of model, Coles (1994) and Coles and Smith (1998) obtain an aggregate matching function with increasing returns to scale, indicating that consumption behavior tends to concentrate in areas with high participant density. Therefore, while population density improves consumption matching opportunities, this may occur with constant returns to scale. Nevertheless, due to improved consumption matching quality, increased population density will still increase total household consumption expenditure.

The negative externalities generated by population concentration may increase transaction costs and suppress consumption. Increased population density leads to traffic congestion, resulting in longer commuting times. Longer commute times may crowd out leisure time, indirectly reducing individuals' consumption time and consumption demand. The literature shows that densely populated cities have higher crime rates (Glaeser and Sacerdote, 1999). Increased crime rates distort household consumption decisions, significantly reducing household consumption of conspicuous goods (Mejía and Restrepo, 2016). Higher population density in cities may lead to more severe overall pollution (Gaigné et al., 2012). Worsening pollution makes people reluctant to go out to purchase goods and services, thereby affecting consumption behavior. Evidence shows that air pollution significantly reduces offline consumption, and this negative impact is unlikely to be offset by online shopping (Kang et al., 2019).

In summary, $\tau(d_j)$ may be nonlinear, but similar to the nonlinearity of agglomeration economies (Glaeser and Gottlieb, 2008), its specific functional form is difficult to determine.

Appendix II: Data Sources

The primary data used in this study are drawn from three waves of the China Household Finance Survey (CHFS) conducted by the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics from 2013 to 2017. The CHFS collects comprehensive information on household demographics, wealth, income and expenditure, insurance coverage, and employment status, providing detailed records of household income and consumption patterns that offer reliable data support for studying household consumption behavior (Yin et al., 2015).

The data for our dependent variables come from CHFS household consumption and income records. Since CHFS collects consumption and income data for the year preceding the survey year, the 2015 and 2017 survey data actually reflect consumption and income for 2014 and 2016, respectively. For analytical convenience, we use the 2015 and 2017 survey data (representing actual 2014 and 2016 outcomes) to construct our dependent variables, while using data from the same tracked households in 2013 and 2015 to construct control variables. The calculation of our main explanatory variable, namely urban population density, also uses 2013 and 2015 data. Consequently, all control variables in our regression models are introduced in lagged form.

Urban population data is sourced from the LandScan population database. The LandScan Global Population Distribution Database utilizes high-quality census data and employs spatial data analysis, imagery analysis, and multivariate dasymetric modeling techniques, combined with movement patterns and collective mobility behaviors, to estimate the spatial distribution of ambient population. This database has been widely applied in population research (Li and Liu, 2018). Compared to administrative statistical data, LandScan population data is high-resolution gridded data, offering the advantage of enabling researchers to extract population data at various geographic scales using geographic information system software to meet diverse research needs. Given that systematic differences in consumption behavior may exist between urban and non-urban areas, our study defines “cities” as municipal districts of prefecture-level and above cities, for which administrative statistical data on resident population is difficult to obtain. We therefore

leverage the aforementioned characteristics of LandScan population data to obtain resident population data for municipal districts. A potential limitation of LandScan global population data is that its model parameter calibration may not fully align with China's specific circumstances, potentially introducing non-random bias in population estimates that could lead to attenuation bias in our study's estimates. To assess the impact of this potential bias on our empirical results, we also use prefecture-level population density calculated from administrative statistical data in our robustness checks. Prefecture-level resident population data are obtained from the *China Regional Economic Statistical Yearbook* and the CEIC China Economic Database. Urban area data come from the *China City Statistical Yearbook*.

Additionally, digital inclusive finance data is sourced from the Digital Finance Research Center at Peking University. Related data for mechanism testing comes from Amap POI data, the Chinese General Social Survey (CGSS), and JuHui Data. Consumer price index data is sourced from provincial statistical yearbooks, city statistical yearbooks, and municipal national economic and social development statistical bulletins. Specific data sources for individual variables are detailed in Appendix III.

Appendix III: Variable Descriptions

The main dependent variables include household per capita consumption expenditure, measured as the natural logarithm of total household consumption divided by household size, and household average propensity to consume, calculated as total household consumption divided by household disposable income. All nominal variables are converted to real values using provincial consumer price indices with 2010 as the base year. Total household consumption includes expenditures on food, clothing, housing (including property management fees, heating costs, and housing renovation and repair expenses), household goods and services, durable goods, education and entertainment, transportation and communication, healthcare, and other expenses.

The key explanatory variable, urban population density, is measured as the ratio of population size in the municipal districts of cities to the built-up area of these municipal districts. The LandScan dataset displays the resident population in each grid cell, with a grid resolution of 30 arcseconds. The total population across all grid cells within the municipal districts constitutes the urban population size. Built-up areas represent the land actually requisitioned by cities for non-agricultural production and living purposes, serving as the location for urban population and economic activities. Following Zhong et al. (2020), we use built-up area to calculate urban population density.

Other control variables in the model include: household per capita income, household per capita financial assets, housing ownership, household per capita housing assets, housing debt status, self-employment status, demographic structure (including youth dependency ratio, elderly dependency ratio, and proportion of unemployed family members), number of respondents and spouses with pension insurance, number of respondents and spouses with medical insurance, and household head characteristics (including risk preference, gender, age, years of education, marital status, health status, and household registration type). Additionally, this study controls for city-level variables that may affect household consumption, including gross domestic product (GDP), share of tertiary industry in GDP, financial development level, digital inclusive finance, and housing prices.

Detailed variable descriptions are presented in Table III1.

Table III1: Variable Descriptions

Variable Name	Variable Definition	Data Source
Dependent Variables		
Household Per Capita Consumption (yuan)	Total household consumption (including expenditures on food, clothing, housing [property management fees, heating costs, and housing renovation and repair expenses], household goods and services, durable goods, education and entertainment, transportation and communication, healthcare, and other expenses, same definition applies below) divided by household size	CHFS
Household Average Propensity to Consume	Total household consumption divided by household disposable income	CHFS
Main Explanatory Variable		
Population Density (persons/km ²)	Urban municipal district population size (total population across grid cells within municipal districts) divided by built-up area of municipal districts	LandScan Dataset, Urban Statistical Yearbook
Family Asset Variables		
Household Per Capita Income (yuan)	Household disposable income divided by household size	CHFS
Household Per Capita Financial Assets (yuan)	Total household financial assets (including demand deposits, time deposits, stocks, funds, wealth management products, bonds, derivatives, non-RMB assets, gold, other financial assets, cash, and loans made) divided by household size	CHFS
Homeownership	Binary variable, 0 = no housing, 1 = owns housing	CHFS
Household Per Capita Housing	Total household housing value divided by household size	CHFS

Variable Name	Variable Definition	Data Source
Value (yuan)		
Housing Debt Status	Binary variable, 0 = no housing debt, 1 = has housing debt	CHFS
Family Characteristic Variables		
Self-Employment Status	Binary variable, 0 = not self-employed, 1 = self-employed	CHFS
Youth Dependency Ratio	Number of household members aged 14 and below divided by household size	CHFS
Elderly Dependency Ratio	Number of household members aged 65 and above divided by household size	CHFS
Share of Unemployed Members	Number of unemployed household members divided by household size	CHFS
Number with Pension Insurance	Number of respondents and spouses with pension insurance	CHFS
Number with Medical Insurance	Number of respondents and spouses with medical insurance	CHFS
Household Head Characteristic Variables		
Risk Preference	<p>Risk preference measured through the following question: “If you had a sum of money, which investment project would you choose?”</p> <p>Options: 1. High-risk, high-return projects; 2. Slightly high-risk, slightly high-return projects; 3. Average risk, average return projects; 4. Slightly low-risk, slightly low-return projects; 5. Unwilling to take any risks. We assign values 5, 4, 3, 2, 1 to options 1-5 respectively, with higher values indicating greater risk preference</p>	CHFS
Household Head Gender	Binary variable, 0 = female, 1 = male	CHFS

Variable Name	Variable Definition	Data Source
Household Head Age (years)	Survey year minus birth year	CHFS
Household Head Years of Education	Converted based on education level categories: 0 = no schooling, 6 = primary school, 9 = junior high school, 12 = senior high school/technical secondary school/vocational high school, 15 = college/higher vocational education, 16 = undergraduate degree, 19 = master's degree, 22 = doctoral degree	CHFS
Household Head Marital Status	Binary variable, 0 = unmarried, 1 = married	CHFS
Household Head Health Status	Scale from 1-5: 1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good	CHFS
Household Head Hukou Type	Binary variable, 0 = agricultural household registration, 1 = non-agricultural household registration	CHFS
City Characteristic Variables		
GDP (yuan)	City GDP	Urban Statistical Yearbook
Share of Tertiary Industry in GDP	Tertiary industry GDP divided by city GDP	Urban Statistical Yearbook
Financial Development Level	Year-end balance of loans from financial institutions divided by GDP	Urban Statistical Yearbook
Housing Prices (yuan/m ²)	Household housing market value divided by housing floor area, then weighted at the municipal district level using household survey weights	CHFS
Digital Inclusive Finance	For detailed compilation process, see Guo et al. (2020)	Peking University Digital Finance Research Center
Other Variables		
Non-tradable	Entropy of non-tradable services (for specific definition of information	Amap POI Data, Urban

Variable Name	Variable Definition	Data Source
Service Matching Quality	entropy, see description in main text, same applies below) divided by built-up area of municipal districts	Statistical Yearbook
Dining Service Matching Quality	Entropy of food services divided by built-up area of municipal districts	Amap POI Data, Urban Statistical Yearbook
Shopping Service Matching Quality	Entropy of shopping services divided by built-up area of municipal districts	Amap POI Data, Urban Statistical Yearbook
Education and Cultural Service Matching Quality	Entropy of educational and cultural services divided by built-up area of municipal districts	Amap POI Data, Urban Statistical Yearbook
Daily Life Service Matching Quality	Entropy of life services divided by built-up area of municipal districts	Amap POI Data, Urban Statistical Yearbook
Sports and Leisure Service Matching Quality	Entropy of sports and recreation services divided by built-up area of municipal districts	Amap POI Data, Urban Statistical Yearbook
Social Interaction Frequency	CGSS asks respondents about “frequency of social and recreational activities with neighbors” and “frequency of social and recreational activities with other friends.” Based on response options ranging from “never” to “almost daily,” values 1-7 are assigned respectively. The average of these two questions represents respondent’s social interaction frequency, with higher values indicating more frequent social activities	CGSS
Total Gift Expenditure (yuan)	Actual monetary value of gifts and cash given to non-family members during traditional holidays and ceremonial occasions (weddings, funerals, celebrations)	CHFS
Per Capita Gift Expenditure (yuan)	Actual monetary value of gifts and cash given to non-family members during traditional holidays and ceremonial occasions, divided by household size	CHFS

Variable Name	Variable Definition	Data Source
New Housing Prices (yuan/m ²)	Prefecture-level city total residential sales revenue divided by total residential sales area	CEIC China Economic Database
Existing Housing Prices (yuan/m ²)	District existing housing prices weighted at the municipal district level by district population share (district population size divided by municipal district population size)	LandScan Dataset, Juhui Data
Instrumental Variable IV1	Product of city built-up area and nearest distance to the Yangtze River, Yellow River, or coastline, multiplied by annual population growth rate of other cities in the household's province excluding the city itself	Urban Statistical Yearbook, Chinese Academy of Sciences Resource and Environmental Science Data Center, LandScan Dataset
Instrumental Variable IV2	Linear time trend of 1953 city population density	First National Population Census Data 1953, Urban Statistical Yearbook

Appendix IV: Detailed Results of Table 2

Table IV1: Detailed Results of Table 2

	A. Dependent Variable: Average Propensity to Consume				
	OLS		FE		FE + Time Trends
	(1)	(2)	(3)	(4)	(5)
Ln(Population density)	0.017 (0.061)	0.040 (0.040)	0.357*** (0.122)	0.430*** (0.134)	0.404*** (0.132)
Ln(Per capita household income)		-0.531*** (0.017)	-0.688*** (0.026)	-0.689*** (0.026)	-0.696*** (0.028)
Ln(Per capita household financial assets)		0.037*** (0.006)	0.007 (0.006)	0.008 (0.006)	0.006 (0.008)
Ln(Per capita household housing value)		0.009*** (0.003)	0.000 (0.002)	0.001 (0.002)	-0.003 (0.003)
Youth dependency ratio		-0.415*** (0.101)	-0.391** (0.175)	-0.403** (0.177)	-0.353* (0.186)
Elderly dependency ratio		-0.215*** (0.062)	0.128 (0.140)	0.123 (0.139)	0.155 (0.132)
Share of unemployed members		0.241*** (0.051)	0.064 (0.071)	0.068 (0.071)	0.046 (0.079)
Ln(City GDP)				0.101 (0.156)	0.098 (0.150)
Share of tertiary industry				-0.040 (0.496)	-0.143 (0.553)
Level of financial development				0.064*** (0.022)	0.069** (0.034)
Ln(Housing price)				-0.056 (0.041)	-0.034 (0.040)
Ln(Digital inclusive finance level)				-0.216 (0.278)	0.187 (0.529)
Time trend of Ln(Per capita household income)					-0.001 (0.012)
Time trend of Ln(Per capita household financial assets)					-0.002 (0.004)
Time trend of Ln(Per capita household housing value)					-0.002 (0.002)
Time trend of youth dependency ratio					0.048 (0.059)
Time trend of elderly dependency ratio					0.061* (0.033)
Time trend of share of unemployed members					-0.023 (0.028)

Time trend of homeownership					-0.030 (0.028)
Time trend of housing debt					-0.050* (0.026)
Time trend of self-employment					-0.032 (0.027)
Time trend of pension coverage = 1					0.023 (0.039)
Time trend of pension coverage = 2					0.011 (0.029)
Time trend of health insurance coverage = 1					-0.086** (0.040)
Time trend of health insurance coverage = 2					-0.023 (0.033)
Time trend of Ln(City GDP)					0.001 (0.013)
Time trend of share of tertiary industry					-0.017 (0.125)
Time trend of level of financial development					0.005 (0.015)
Time trend of Ln(Housing price)					0.017 (0.017)
Constant	0.892 (0.580)	5.344*** (0.428)	4.135*** (1.232)	3.342 (4.034)	-99.237 (390.401)
Number of observations	10,883	10,883	10,883	10,883	10,883

B. Dependent Variable: Ln(Per Capita Household Consumption)

	OLS		FE		FE + Time Trends
	(1)	(2)	(3)	(4)	(5)
Ln(Population density)	-0.101 (0.178)	0.004 (0.065)	0.271** (0.119)	0.355** (0.137)	0.343** (0.146)
Ln(Per capita household income)		0.184*** (0.012)	0.110*** (0.011)	0.110*** (0.011)	0.116*** (0.016)
Ln(Per capita household financial assets)		0.060*** (0.005)	-0.004 (0.006)	-0.004 (0.006)	-0.009 (0.007)
Ln(Per capita household housing value)		0.016*** (0.003)	0.000 (0.002)	0.000 (0.002)	-0.007** (0.003)
Youth dependency ratio		-0.740*** (0.081)	-0.397*** (0.129)	-0.409*** (0.129)	-0.302** (0.141)
Elderly dependency ratio		-0.315*** (0.082)	0.266** (0.106)	0.257** (0.106)	0.304*** (0.107)
Share of unemployed members		0.273*** (0.050)	-0.017 (0.051)	-0.011 (0.051)	-0.017 (0.056)
Ln(City GDP)				0.169	0.171

				(0.175)	(0.173)
Share of tertiary industry				0.329	0.076
				(0.502)	(0.578)
Level of financial development				0.053***	0.095***
				(0.013)	(0.026)
Ln(Housing price)				-0.037	-0.011
				(0.033)	(0.045)
Ln(Digital inclusive finance level)				-0.338	-0.104
				(0.277)	(0.466)
Time trend of Ln(Per capita household income)					0.009
					(0.007)
Time trend of Ln(Per capita household financial assets)					-0.005
					(0.004)
Time trend of Ln(Per capita household housing value)					-0.005***
					(0.002)
Time trend of youth dependency ratio					0.117**
					(0.050)
Time trend of elderly dependency ratio					0.055*
					(0.029)
Time trend of share of unemployed members					-0.015
					(0.023)
Time trend of homeownership					-0.057*
					(0.030)
Time trend of housing debt					-0.029
					(0.021)
Time trend of self-employment					-0.037*
					(0.019)
Time trend of pension coverage = 1					0.055**
					(0.027)
Time trend of pension coverage = 2					-0.013
					(0.028)
Time trend of health insurance coverage = 1					-0.136***
					(0.042)
Time trend of health insurance coverage = 2					-0.033
					(0.038)
Time trend of Ln(City GDP)					0.007
					(0.010)
Time trend of share of tertiary industry					-0.134
					(0.130)
Time trend of level of financial development					0.025**
					(0.012)
Time trend of Ln(Housing price)					0.005
					(0.017)
Constant	10.324***	7.004***	5.769***	4.113	-79.683

	(1.664)	(0.628)	(1.195)	(4.245)	(272.323)
Number of observations	12,804	12,804	12,804	12,804	12,804

Notes: This table presents the detailed results of Table 2 in the main text. For brevity, only results for continuous variables are reported. Variables represented as discrete integers are included in the model as dummy variables, with some results unreported, including: homeownership, housing debt, self-employment, risk preference, gender, age, years of education, marital status, health status, household registration type, city fixed effects, and year fixed effects. Time trends refer to the linear time trends of the baseline values of household- and city-level variables. All results use household sampling weights. Numbers in parentheses are robust standard errors clustered at the city level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix V: Robustness Checks

Table V1 reports the results of a series of robustness checks.¹¹ Since the CHFS data does not survey the purchase years of durable goods, it is impossible to amortize usage expenses (Li, 2018). In Column (1) of Table V1, we subtract durable goods expenditure from household consumption expenditure and then calculate the average propensity to consume, using this as the dependent variable to test the robustness of our main results after avoiding the aforementioned amortization issue. The results are quite robust.

This paper focuses on residents' real consumption expenditure. To eliminate the effects of price changes, as mentioned earlier, we convert nominal variables to real variables using provincial consumer price indices. Price levels often differ among cities within the same province, so we collected city-level consumer price indices for the sample period.¹² Column (2) of Table V1 reports the results after deflating nominal variables using city-level consumer price indices. The results are highly robust compared to the baseline estimates.

To examine the robustness of results using administrative data, in Column (3) of Table V1, we replace the LandScan-based population density of municipal districts with population density calculated from prefecture-level administrative data on resident population.¹³ The results remain robust compared to the baseline estimates.

Considering that urban population in China is mainly concentrated in central urban areas, particularly in certain monocentric megacities, our population density calculation method may underestimate the degree of population concentration in these cities. This may cause measurement error problems, leading to attenuation bias in the estimates. The instrumental variable results in the main text show that the consumption effects estimated using population density changes induced by exogenous shocks do not differ significantly from the OLS estimates. This alleviates, to some extent, concerns about significant bias that

¹¹ For ease of exposition, this paper only discusses the results in Table V1 using household average propensity to consume as the dependent variable. Table V2 presents the results for household per capita consumption expenditure.

¹² Due to missing data for some cities, we still use provincial-level consumer price indices for the baseline analysis.

¹³ We also follow the approach of Zhong et al. (2020), using employment density in built-up areas of municipal districts (total employment in secondary and tertiary industries/built-up area) to measure population density. The results still confirm that population concentration has a significantly positive effect on household per capita consumption.

measurement error might cause. For rigor, we exclude from the sample certain distinctive cities where population density may be underestimated, including Beijing, Shanghai, Guangzhou, Shenzhen, Chongqing, and Tianjin. Additionally, as China's major metropolitan areas, these cities are subject to special policy treatments that could bias our estimates. Column (4) of Table V1 shows that the estimation results remain robust, although they are somewhat larger in magnitude; we cannot determine whether they differ significantly from the main estimation results.

To further reduce measurement error, Column (5) of Table V1 excludes samples where respondents do not live in the same district/county as the main household members. Considering that household heads who are too elderly may not be the primary decision-makers for household consumption activities, Column (6) of Table V1 excludes samples where the household head is older than 70. Households with excessively large population sizes may have characteristics, preferences, and consumption behaviors that differ from typical households; Column (7) of Table V1 excludes samples with household size greater than 8. Due to data quality concerns, Column (8) of Table V1 includes only households where the household head or spouse was interviewed. The estimation results are robust across all specifications.

Table V1: Robustness Checks (Household Average Propensity to Consume)

	(1)	(2)	(3)	(4)
	Excluding durable goods expenditure	City price index	Administrative data	Excluding distinctive cities
Ln(Population density)	0.418*** (0.126)	0.443*** (0.135)	0.353** (0.150)	0.459*** (0.114)
Sample size	10,884	9,971	10,883	7,736
	(5)	(6)	(7)	(8)
	Excluding non-local residents	Excluding household heads aged >70	Excluding household size >8	Excluding non- head/spouse interviews
Ln(Population density)	0.475*** (0.147)	0.402*** (0.151)	0.478*** (0.133)	0.434*** (0.145)
Sample size	9,579	9,800	10,824	9,975

Notes: The dependent variable is the original household average propensity to consume for all columns except Column (1), which uses the household average propensity to consume calculated after subtracting durable goods expenditure from total household expenditure. All results in the table are obtained using the model specification from Column (4) of Table 2 in the main text. For specific variable and sample settings, please refer to the main text. Numbers in parentheses are robust standard errors clustered at the city level. ** and *** indicate significance at the 5% and 1% levels, respectively.

Table V2: Robustness Checks (Natural Logarithm of Household Per Capita Consumption Expenditure)

	(1)	(2)	(3)	(4)
	Excluding durable goods expenditure	City price index	Administrative data	Excluding distinctive cities
Ln(Population density)	0.375*** (0.134)	0.377*** (0.133)	0.319** (0.138)	0.386*** (0.130)
Sample size	12,804	11,706	12,804	9,196
	(5)	(6)	(7)	(8)
	Excluding non-local residents	Excluding household heads aged >70	Excluding household size >8	Excluding non- head/spouse interviews
Ln(Population density)	0.402*** (0.128)	0.312** (0.152)	0.386*** (0.135)	0.347** (0.141)
Sample size	11,146	11,519	12,726	11,726

Notes: The dependent variable is the natural logarithm of original household per capita consumption expenditure for all columns except Column (1), which uses the natural logarithm of household per capita consumption expenditure calculated after subtracting durable goods expenditure from total household expenditure. All results in the table are obtained using the model specification from Column (4) of Table 2 in the main text. For specific variable and sample settings, please refer to the main text. Numbers in parentheses are robust standard errors clustered at the city level. ** and *** indicate significance at the 5% and 1% levels, respectively.

Appendix VI: Heterogeneity Tests

Table VII reports the results of heterogeneity tests.¹⁴ First, we examine whether population concentration affects how households allocate their consumption spending. We analyze the shares of essential, development-oriented, and discretionary consumption in total household expenditure.¹⁵ The estimation results in Columns (1) – (3) of Panel A in Table VII show that population concentration reduces the shares of essential and development-oriented consumption, though these effects are not statistically significant. However, a 1% increase in population density significantly raises the share of discretionary consumption by 0.072 percentage points at the 5% significance level. As urban population density increases, residents shift their spending toward discretionary goods and services such as cultural entertainment and tourism, reflecting improved living standards. This suggests that population concentration not only boosts overall household consumption but also drives a shift toward higher-quality consumption patterns. From a policy standpoint, these findings highlight the need for supply-side reforms that strengthen service sector development to meet evolving consumer preferences.

Next, we further examine the impact of population concentration on consumption expenditure across different income groups. Panel B of Table VII reports the regression results.¹⁶ The results show that population concentration has heterogeneous effects on the average propensity to consume across households with different income levels. For low-income households, the consumption-promoting effect of population concentration is significant at the 5% level. The impact of population concentration on the average propensity to consume of middle-income households is positive but not significant. For high-income households, population concentration has a significant promoting effect on the average

¹⁴ For convenience of exposition, this paper only explains the results in Table VII using the average propensity to consume as the dependent variable. The results for household per capita consumption expenditure are shown in Table VI2.

¹⁵ Essential consumption covers food, clothing, household necessities, and housing; development-oriented consumption includes transportation, communication, education, and healthcare; discretionary consumption encompasses household services, durable goods, entertainment, travel, and luxury items.

¹⁶ Using CHFS 2017 income data as the standard, this paper classifies households with the lowest 30% of annual income as low-income households; households with the highest 30% of income as high-income households; the remainder as middle-income households.

propensity to consume. A 1% increase in population concentration raises the average propensity to consume of low-income and high-income households by 0.957 and 0.578 percentage points, respectively. Table VI2 shows similar results using the natural logarithm of household per capita consumption expenditure as the dependent variable. These findings reveal that population concentration primarily boosts consumption among middle- and high-income households. This pattern aligns with our earlier finding that population concentration promotes discretionary consumption. While population concentration does increase the average propensity to consume among low-income households, middle- and high-income households show a stronger tendency toward consumption upgrading when measured by actual expenditure levels. When formulating policies to promote consumption through population concentration, policymakers should consider several key factors. First, they must account for differences among target population groups and understand how sorting effects may influence policy effectiveness. Specifically, understanding why high-consumption groups tend to locate in low-density areas while low-consumption groups concentrate in high-density areas is crucial for effective policy design and warrants further research. Second, policymakers should recognize that such policies may exacerbate consumption inequality. Future research should therefore examine strategies to promote consumption among low-income households in high-density areas.

Finally, we consider the heterogeneity of consumption effects of population concentration across different city sizes. Panel C of Table VII reports the results.¹⁷ We find that the consumption effect of population concentration is mainly manifested in large city samples. Population concentration has a positive but statistically insignificant impact on household consumption propensity in small and medium-sized cities. In large cities, a 1% increase in population density significantly increases the average propensity to consume by 0.335 percentage points at the 5% significance level. For megacities or super-large cities, the estimate is positive but not significant. This suggests that population concentration's effect on consumption follows an inverted U-shaped pattern, initially rising with city size before

¹⁷ Using the 2015 population size of urban districts as the standard, this paper defines cities with less than 1 million people as small and medium cities, cities with 1 to 5 million people as large cities, and cities with more than 5 million people as megacities or super-large cities.

declining in the largest cities. Policymakers should therefore tailor their approaches based on city size. In megacities and super-large cities, policies should focus on mitigating the negative externalities of excessive population density. Large cities should leverage the strong positive consumption effects of population concentration to foster economic growth. For small and medium-sized cities, the priority should be promoting urban growth to reach optimal population thresholds. Moreover, these city size differences suggest that strategic population migration policies could potentially increase overall consumption. For instance, encouraging migration from megacities to smaller urban areas could serve dual purposes: reducing overcrowding pressures and consumption-dampening negative externalities in megacities while simultaneously boosting consumption effects through increased population concentration in smaller cities. This could yield net consumption gains across the urban system.

However, empirical evidence remains limited on whether migrants from large cities actually change their consumption behavior when they relocate to smaller cities. A key unresolved question for future research is whether population concentration affects consumption through level effects or growth effects for migrant populations.

Table VII: Heterogeneity Tests (Average Propensity to Consume)

A. Consumption Structure	Dependent Variables		
	Share of Essential Consumption	Share of Development-Oriented Consumption	Share of Discretionary Consumption
	(1)	(2)	(3)
Ln(Population Density)	-0.030 (0.042)	-0.090 (0.082)	0.072** (0.032)
Sample Size	12,804	12,804	12,804
Dependent Variable: Average Propensity to Consume			
B. Household Income	Low-Income Households	Middle-Income Households	High-Income Households
	(1)	(2)	(3)
	(1)	(2)	(3)
Ln(Population Density)	0.957** (0.426)	0.218 (0.189)	0.578*** (0.168)
Sample Size	1,859	4,921	4,103
Dependent Variable: Average Propensity to Consume			
C. City Size	Small and Medium Cities	Large Cities	Megacities or Super-Large Cities
	(1)	(2)	(3)
	(1)	(2)	(3)
Ln(Population Density)	0.117	0.335**	0.226

	(0.236)	(0.131)	(0.230)
Sample Size	1,657	4,517	4,709

Notes: The dependent variables in Panel A are the shares of household essential, development-oriented, and discretionary consumption in total household consumption. The dependent variables in Panels B and C are the average propensity to consume. Each column in Panel A uses all samples. Panel B uses samples of low-, middle-, and high-income households, respectively. Panel C uses samples of households in small and medium cities, large cities, and megacities or super-large cities, respectively. Except for the different dependent variables or samples mentioned above, the model specifications are consistent with Column (4) in Table 2 of the main text. All results use household sampling weights. Robust standard errors clustered at the city level are shown in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

Table VI2: Heterogeneity Tests (Natural Logarithm of Household Per Capita Consumption Expenditure)

A. Household Income	Dependent Variable: Ln(Household Per Capita Consumption)		
	Low-Income Households	Middle-Income Households	High-Income Households
	(1)	(2)	(3)
Ln(Population Density)	-0.053 (0.333)	0.397** (0.183)	0.479*** (0.152)
Sample Size	2528	5394	4882
B. City Size	Dependent Variable: Ln(Household Per Capita Consumption)		
	Small and Medium Cities	Large Cities	Megacities or Super-Large Cities
	(1)	(2)	(3)
Ln(Population Density)	-0.451 (0.431)	0.360** (0.136)	0.351** (0.125)
Sample Size	2006	5394	5404

Notes: The dependent variables in Panels A and B are the natural logarithm of household per capita consumption. Panel A uses samples of low-, middle-, and high-income households, respectively. Panel B uses samples of households in small and medium cities, large cities, and megacities or super-large cities, respectively. Except for the different dependent variables or samples mentioned above, the model specifications are consistent with Column (4) in Table 2 of the main text. All results use household sampling weights. Robust standard errors clustered at the city level are shown in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

Appendix VII: Population Concentration and Non-tradable Service Density

As described in Appendix I, the expected product characteristic distance $E \sum_{Y(x)} |x - y| = \theta^2 / 4n$ decreases as population density increases based on the assumption that θ remains constant. Under this assumption, it follows that $\frac{dm}{m} / \frac{dn}{n} = 1$, indicating that the elasticity of non-tradable service density with respect to population density is unit elastic, with no scale effects present. The results reported in Table VII1 validate this assumption. When using non-tradable service density as a proxy for consumption matching opportunities, this result is also consistent with the theoretical framework, namely that the aggregate matching function exhibits constant returns to scale in most of the literature. Based on the results, population density enhances consumption matching opportunities. Although it does not exhibit increasing returns to scale effects, overall, increases in population density reduce consumption transaction costs to some extent, thereby promoting household consumption.

Table VII1: Impact of Population Concentration on Non-tradable Service Density

	Dependent Variable					
	Ln(Non-tradable Service Density)	Ln(Dining Service Density)	Ln(Shopping Service Density)	Ln(Education and Cultural Service Density)	Ln(Daily Life Service Density)	Ln(Sports & Leisure Service Density)
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Population Density)	1.010** (0.431)	1.244* (0.661)	1.141** (0.563)	1.057*** (0.196)	0.752** (0.330)	1.219*** (0.206)
Unit Elasticity Test P-value	0.982	0.713	0.803	0.773	0.455	0.290
Sample Size	166	166	166	166	166	166

Notes: The density of various non-tradable services is measured as the ratio of the total number of points of interest of each category within the urban districts to the built-up area of the urban districts. Control variables include: GDP scale, tertiary industry share of GDP, financial development level, housing prices, digital inclusive finance, city fixed effects, and year fixed effects. Robust standard errors clustered at the city level are reported in parentheses. The null hypothesis for the unit elasticity test is that the coefficient estimate of Ln(population density) equals 1. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix VIII: Housing Expenditure Imputation

The CHFS2017 survey asked selected households two questions: “If this house were to be rented out currently, how much rent could it generate per month?” and “What is the current market value of this house?” Using the respondents’ answers, we calculated the rent-to-price ratio for individual housing units by dividing the monthly rental income by the market value. Since some households own multiple properties (CHFS2017 collected information on up to six properties per household), we calculated rent-to-price ratios for each housing unit owned by multi-property households. Based on this data, we computed weighted average rent-to-price ratios at the city level using survey weights. For cities with missing data, we used the mean value from other cities.

Given that urban housing market characteristics remain relatively stable in the short term, we assumed that the average city-level rent-to-price ratios remained constant from 2013 to 2017. We then multiplied the city-level rent-to-price ratio by the market value of each household’s owner-occupied housing to approximate the housing expenditure for owner-occupying households.

References

- Ahlfeldt, G. M., and E. Pietrostefani, “The Economic Effects of Density: A Synthesis”, *Journal of Urban Economics*, 2019, 111, 93-107.
- Bartik, T. J., *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: WE Upjohn Institute for Employment Research, 1991.
- Baum-Snow, N., and R. Pavan, “Understanding the City Size Wage Gap”, *Review of Economic Studies*, 2012, 79(1), 88-127.
- Beeson, P. E., D. N. DeJong, and W. Troesken, “Population Growth in U.S. Counties, 1840-1990”, *Regional Science and Urban Economics*, 2001, 31(6), 669-700.
- Bikhchandani, S., D. Hirshleifer, and I. Welch, “Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades”, *Journal of Economic Perspectives*, 1998, 12(3), 151-170.
- Cai, Z., “Hours Worked of the Self-employed and Agglomeration”, *Growth and Change*, 2019, 50(1), 352-380.
- Chen, F., and S. Su, “Urban Size Wage Premium: Sources and Economic Mechanisms”, *Management World*, 2021, (1), 19-32+2+15-16. [in Chinese]
- Cheng, L., and Y. Zhang, “Did Early Famine Experience Impact People’s Saving Behavior? A New Explanation for High Saving Rate in China”, *Economic Research Journal*, 2011, (8), 119-132. [in Chinese]
- Coles, M.G., “Understanding the Matching Function: The Role of Newspapers and Job Agencies.” *CEPR Discussion Papers*, 1994.
- Coles, M.G., and E. Smith, “Marketplaces and Matching.” *International Economic Review*, 1998, 39(1), 239-273.
- Duranton, G., and D. Puga, “Chapter 48 - Micro-Foundations of Urban Agglomeration Economies.” In J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Volume 4. Amsterdam: Elsevier, 2004.
- Fan, Z., and J. Liu, “Saving for House Purchase: On the Income Distribution Effect of Property Tax Reform”, *Management World*, 2015, (5), 18-27+187. [in Chinese]
- Gaigné, C., S. Riou, and J.-F. Thisse, “Are Compact Cities Environmentally Friendly?” *Journal of Urban Economics*, 2012, 72(2), 123-136.

- Gan, L., N. Zhao, and Y. Sun, "Income Inequality, Liquidity Constraints and China's Household Saving Rate", *Economic Research Journal*, 2018, (12), 34-50. [in Chinese]
- George, L., and J. Waldfogel, "Who Affects Whom in Daily Newspaper Markets?", *Journal of Political Economy*, 2003, 111(4), 765-784.
- Glaeser, E. L., and J. D. Gottlieb, "Urban Resurgence and the Consumer City", *Urban Studies*, 2006, 43(8), 1275-1299.
- Glaeser, E.L., and J.D. Gottlieb, "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity*, 2008, 2008(2), 155-239.
- Glaeser, E. L., J. Kolko, and A. Saiz, "Consumer City", *Journal of Economic Geography*, 2001, 1(1), 27-50.
- Glaeser, E. L., and D. C. Maré, "Cities and Skills", *Journal of Labor Economics*, 2001, 19(2), 316-342.
- Glaeser, E.L., and B. Sacerdote, "Why is There More Crime in Cities?" *Journal of Political Economy*, 1999, 107(S6), S225-S258.
- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift, "Bartik Instruments: What, When, Why, and How", *American Economic Review*, 2020, 110(8), 2586-2624.
- Grieser, W. D., and C. J. Hadlock, "Panel-Data Estimation in Finance: Testable Assumptions and Parameter (In) Consistency", *Journal of Financial and Quantitative Analysis*, 2019, 54(1), 1-29.
- Guo, F., J. Wang, F. Wang, T. Kong, X. Zhang, and Z. Cheng, "Measuring China's Digital Financial Inclusion: Index Development and Spatial Characteristics." *China Economic Quarterly*, 2020, (4), 1401-1418. [in Chinese]
- Guo, S., and P. Liang, "Social Interaction, Information Channels and Household Stock Market Participation: Empirical Evidence from China Household Finance Survey 2011", *Economic Research Journal*, 2014, (S1), 116-131. [in Chinese]
- Helsley, R.W., and W.C. Strange, "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics*, 1990, 20(2), 189-212.
- Kang, H., H. Suh, and J. Yu, "Does Air Pollution Affect Consumption Behavior? Evidence from Korean Retail Sales", *Asian Economic Journal*, 2019, 33(3), 235-251.
- Kuhn, P., P. Kooreman, A. Soetevent, and A. Kapteyn, "The Effects of Lottery Prizes on

- Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery”, *American Economic Review*, 2011, 101(5), 2226-2247.
- Lee, S., “Ability Sorting and Consumer City”, *Journal of Urban Economics*, 2010, 68(1), 20-33.
- Lei, Q., “Wealth Accumulation, Habit, Preference Change, Uncertainty and Household Consumption Decision”, *China Economic Quarterly*, 2009, (3), 1029-1046. [in Chinese]
- Li, B., D. Guo, and S. Liu, “City Size, Population Structure and Non-tradable Goods Diversity: Big Data Analysis Based on Dianping.com”, *Economic Research Journal*, 2019, (1), 150-164. [in Chinese]
- Li, J., “Does the ‘House Slave Effect’ Lead to Sluggish Household Consumption?”, *China Economic Quarterly*, 2018, (1), 405-430. [in Chinese]
- Li, Y., and X. Liu, “How did Urban Polycentricity and Dispersion Affect Economic Productivity? A Case Study of 306 Chinese Cities.” *Landscape and Urban Planning*, 2018, 173, 51-59.
- Lu, M., H. Gao, and H. Sato, “City Size and Inclusive Employment”, *Social Sciences in China*, 2012, (10), 47-66+206. [in Chinese]
- Lu, M., H. Zhang, and W. Liang, “How Land Supply Biased toward Central and Western China Pushed up Wages in Eastern China”, *Social Sciences in China*, 2015, (5), 59-83+204-205. [in Chinese]
- Mejía, D., and P. Restrepo, “Crime and Conspicuous Consumption”, *Journal of Public Economics*, 2016, 135, 1-14.
- Nunn, N., and N. Qian, “US Food Aid and Civil Conflict”, *American Economic Review*, 2014, 104(6), 1630-1666.
- Ouyang, Y., Y. Fu, and S. Wang, “Scale Effect of Household Consumption and Its Evolution Mechanism”, *Economic Research Journal*, 2016, (2), 56-68. [in Chinese]
- Quintero, L. E., and M. Roberts, “Cities and Productivity: Evidence from 16 Latin American and Caribbean Countries”, *Journal of Urban Economics*, 2023, 136, 103573.
- Rosenthal, S. S., and W. C. Strange, “Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies”, in J. V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Elsevier, 2004.

- Wan, G., Y. Zhang, and J. Niu, “Liquidity Constraints, Uncertainty and Chinese Household Consumption”, *Economic Research Journal*, 2001, (11), 35-44+94. [in Chinese]
- Yan, S., and G. Zhu, “‘House Slave Effect’ or ‘Wealth Effect’? A Theoretical Analysis of the Impact of Housing Price Rise on National Consumption”, *Management World*, 2013, (3), 34-47. [in Chinese]
- Yang, R., and B. Chen, “Higher Education Reform, Precautionary Saving and Household Consumption Behavior”, *Economic Research Journal*, 2009, (8), 113-124. [in Chinese]
- Yin, Z., Y. Wu, and L. Gan, “Financial Accessibility, Financial Market Participation and Household Asset Choice”, *Economic Research Journal*, 2015, (3), 87-99. [in Chinese]
- Zhang, J., J. Huang, J. Wang, and M. Huang, “Urban Settlement Threshold and Labor Return Flow”, *Economic Research Journal*, 2020, (7), 175-190. [in Chinese]
- Zhong, Y., M. Lu, and X. Xi, “Agglomeration and Service Industry Development: From the Perspective of Population Spatial Distribution”, *Management World*, 2020, (11), 35-49. [in Chinese]