Economics Letters 233 (2023) 111412

Contents lists available at ScienceDirect

Economics Letters

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

Urban consumption dissimilarities and subjective wellbeing: Evidence from mobile big data in China☆

Lanyue Hu, Wei Shi *

Institute for Economic and Social Research, Jinan University, China

ARTICLE INFO

JEL classification:
I31
D63
R2

Keywords:
Economic inequality
Subjective wellbeing
China

ABSTRACT

The inequality and spatial unevenness in the distribution of economic resources can both significantly impact individual wellbeing. Data on residents' income at a fine spatial scale are often not available in developing countries. In this study, we utilize a novel dataset derived from mobile big data with high spatial resolutions to estimate residents' income at the district level in China. By combining this measure of consumption dissimilarity with household survey data, our findings reveal that even after controlling for overall levels of inequality, the concentration of high-spending households within a district has a negative effect on residents' life satisfaction while the concentration of low-spending households has an opposite effect. These results contribute to the emerging research on understanding the spatial dimensions of economic inequalities.

1. Introduction

Cities typically attract individuals from diverse social and economic backgrounds, thereby fostering a rich tapestry of human experiences. Extensive scholarly research has explored the implications of diversity in group compositions or economic inequality on various individual and collective outcomes (e.g., Luttmer, 2005). Cities are fundamentally spatial entities, and the spatial distribution of individuals with diverse characteristics within a city is also significant. Social interactions are more likely to occur in close proximity. The benefits of increased social interactions between individuals from different social or economic backgrounds depend on the context; while such interactions may enhance mutual understanding, they can also accentuate disparities and escalate conflicts. Existing literature primarily focuses on residential segregation based on race or ethnicity (Cutler and Glaeser, 1997; Bharathi et al., 2023). Recent studies have expanded the scope of analysis to consumption segregation (Davis et al., 2019).

In developing countries, there is often a lack of fine-scale data on residents' income, posing challenges in examining the impact of spatial disparities among individuals with different income levels. In this study, we utilize a novel dataset derived from mobile big data capturing residents' consumption behavior at high spatial resolutions, enabling us to quantify variations in urban consumption across city districts in China. Subsequently, we integrate the dissimilarity measure of consumption with household survey data and ascertain that spatial concentrations of high-spending households within a district exert a negative impact on residents' life satisfaction in that particular area, whereas spatial concentrations of low-spending households yield an opposite effect. Moreover, the adverse consequences of spatial separations diminish as the group size expands. These findings remain consistent across various heterogeneity and robustness analyses.

Firstly, while economic inequality and racial segregation have garnered considerable attention, there has been a limited number of studies investigating the spatial divisions between social groups based on income levels and their implications, particularly in developing countries due to data constraints. In the Chinese context, Lei et al. (2018) find that consumption inequalities at county levels negatively affect residents' life satisfaction. Jiang et al. (2012) highlight the role of between group inequalities. We contribute to this line of research by adding the spatial dimension and showing that spatial separations of groups of different income levels can also affect life satisfaction. Secondly, we showcase the utilization of innovative data sources derived from mobile big data to address the dearth of income data at precise spatial scales obtained through conventional means. The integration of mobile phone-derived data is increasingly prevalent in economic research. Barwick et al. (2023) and Li et al. (2023). Due to space limitations, further discussions on the relevant literature can be found in the Online Appendix.

☆ This work received financial support from the Youth Project of Humanities and Social Sciences by the Ministry of Education of China (Grant Number: 18YJC790138) and the Fundamental Research Funds for the Central Universities, China (Grant Number: 23JNQMX31). We would like to thank the reviewer, Audra Bowlus (editor) and Jingliang Lu for constructive comments, and Yun Qiu for help with the remote sensing data.

* Corresponding author.
E-mail addresses: hulanyue_iesr21@163.com (L. Hu), wshi16@jnu.edu.cn (W. Shi).
URL: https://sites.google.com/view/wei-shi (W. Shi).

https://doi.org/10.1016/j.econlet.2023.111412
Received 21 June 2023; Received in revised form 29 September 2023; Accepted 14 October 2023
Available online 17 October 2023

2. Data and methods

The spatial unit of analysis in this paper is a city district in China, which refers to the county-level administrative division below the prefectural level. A district comprises multiple subdistricts. Suppose that the population is divided into two mutually exclusive groups with $a_{i,c}$ number of residents belonging to group A in subdistrict i of district c, and $b_{i,c}$ number of residents belonging to the other group A^C. To measure the relative concentration of group A in some subdistricts, we use the dissimilarity index, defined as:

$$D_c^A = \frac{1}{2}\sum_{i \text{ in district } c}\left|\frac{a_{i,c}}{\sum_{i \text{ in district } c} a_{i,c}} - \frac{b_{i,c}}{\sum_{i \text{ in district } c} b_{i,c}}\right|. \quad (1)$$

The dissimilarity index measures the unevenness in the spatial distribution of group A relative to the other group A^C.[1] The index takes values between 0 and 1, and the value can be interpreted as the share of group A residents that have to be relocated to other subdistricts such that the spatial distributions of residents from these two groups are the same in the district.

We use the share of group A residents in a district to measure the inequality level:

$$I_c^A = \frac{\sum_{i \text{ in district } c} a_{i,c}}{\sum_{i \text{ in district } c} a_{i,c} + \sum_{i \text{ in district } c} b_{i,c}}.$$

If the groups are classified according to social or economic status, D_c^A and I_c^A provide measures on different dimensions of inequality. I_c^A provides an aggregate measure of district level inequality,[2] while D_c^A measures the relative concentration or separation of the group from the rest of the population. A district can exhibit low levels of aggregate inequality while simultaneously displaying high levels of spatial dissimilarity between groups, and vice versa. Spatial dissimilarity between groups may shape one's perception of inequality levels, which is the focus of this paper.

Data on the consumption level of residents are obtained through a web interface from a big data analytics company.[3] Based on data regarding the spending behavior of mobile users in physical stores and their corresponding demographic information, the company employs a proprietary algorithm to accurately estimate residents' income levels at fine geographic scales, even down to the level of residential compounds or communities. As these income estimates are primarily derived from consumption patterns observed in stores, we consider individuals with higher incomes as those who exhibit greater spending tendencies. The company divides a city district into several subdistricts with sizes similar to the *jiedao* administrative division of China, and provide estimates on the number of residents and their income distributions in those subdistricts. The income distributions are reported as numbers of residents in a subdistrict with monthly income in RMB in the following categories: less than 3000, 3000–5000, 5000–8000, 8000–15,000, 15,000–20,000, 20,000–30,000, 30,000–50,000, and more than 50,000. We define the less than 3000 group as the low spending group (group L) and the more than 50 000 group as the high spending group (group H), and compute the district level dissimilarity indices according to Eq. (1). The data set covers 293 districts in 45 cities, which includes most provincial capitals of China. Section B in the online appendix provides an example of the web interface of the data.

To provide a validation on our measures of consumption dissimilarities, we compare our measure of spatial concentrations of high spending residents with spatial concentrations of households owning cars worth 500,000 RMB (71,100 USD) or more from the 2015 population census.[4] Fig. 1 plots the district level dissimilarity index of households owning cars worth 500,000 RMB or more with the dissimilarity index of high spending residents. To facilitate the visualization of the relationship, we have categorized the variable on the horizontal axis into 10 equally sized bins and reported the average value of the variable on the vertical axis for each bin. Spatial concentrations of consumption exhibit a significantly lower magnitude compared to those of assets. Nonetheless, there exists a distinct positive correlation between these two variables, thereby substantiating the validity of our consumption dissimilarity measure.

As a further validation of our measure of consumption dissimilarities, we show how they are influenced by the degree of geographic constraints. The constraints of geography can shape residential development (Saiz, 2010). Using data from the ASTER Global Digital Elevation Model Version 3 (GDEM v3), we calculate for each city district, the share of land area with slope less than 25%, which measures the availability of developable land. We also collect data on the share of built-up land area for each city district (Yang and Huang, 2023) and the distance from the geographic centroid of the district to the city center. We find that residential sorting along consumption levels is greater in city districts that are less geographically constrained or are less developed (Table C.1 in the appendix).

We then merge the district level dissimilarity indices with the 2019 wave of the China Household Finance Survey (CHFS) (Gan et al., 2014)[5] using district information from the China Household Employment Survey. Subjective wellbeing is measured by the responses to the survey question "*Overall, do you feel happy?*". The responses take integer values from 0 to 4.[6] We also construct a binary variable that is 1 if the respondent feels happy[7] and 0 otherwise. We only keep cities that have at least two districts in the merged data set, such that within city between district variations in consumption dissimilarities and inequalities are used in the empirical analysis. The merged data set includes 9278 households in 25 cities and 77 city districts.

We use the following empirical model to examine the effect of consumption dissimilarities on residents' subjective wellbeing:

$$\text{happiness}_{ic} = D_c^{\mathrm{H}}\alpha_1 + D_c^{\mathrm{L}}\alpha_2 + I_c^{\mathrm{H}}\alpha_3 + I_c^{L}\alpha_4 + X_{ic}\beta + \eta_{\text{city}(c)} + \epsilon_{ic}. \quad (2)$$

The key variables of interest are D_c^{H} and D_c^{L} which respectively measures the spatial separations of high spending and low spending households from the rest of the population within a city district. The aggregate inequality levels are controlled by the shares of high spending and low spending households of the district (I_c^{H} and I_c^{L}). X_{ic} includes individual controls that correlate with life satisfaction. City fixed effects are included to control for city level unobserved confounding factors. ϵ_{ic} is an error term. Summary statistics of the variables are reported in Table 1.

3. Results

We estimate the linear probability model of happiness (Eq. (2)) and report estimates of the effects of consumption dissimilarities on residents' subjective wellbeing in Table 2.[8] In columns (1)–(4), the dependent variable is binary with 1 indicating being happy with life and 0 otherwise. In columns (5)–(8), the dependent variable takes

[1] The dissimilarity index is often used to measure residential segregation between races (e.g. Cutler and Glaeser, 1997). Davis et al. (2019) use this index to measure dissimilarities in consumption between races in New York City.

[2] Other measures of aggregate levels of inequality include the Gini coefficient and the Theil index.

[3] Shuwei Guancha, www.swguancha.com.

[4] The census questionnaire does not cover income or the values of other assets.

[5] The CHFS is organized and managed by China Household Financial Survey and Research Center, Southwestern University of Finance and Economics. The 2019 wave is the most recent wave of the CHFS that is publicly released.

[6] 4: very happy; 3: happy; 2: neutral; 1: unhappy; 0: very unhappy.

[7] The response to the survey question is "very happy" or "happy".

[8] The results from logit and ordered logit regressions (Table C.3 in the appendix) are qualitatively similar.

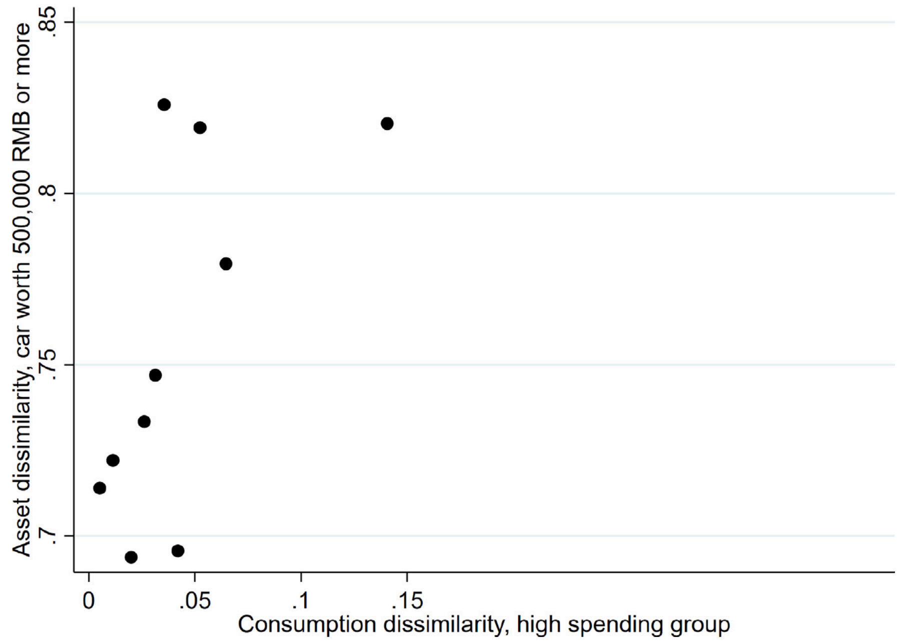

Fig. 1. Comparisons with population census. *Notes*: The variable on the horizontal axis is classified into 10 equally sized bins and the average value of the variable on the vertical axis for each bin is reported.

Table 1
Summary statistics.

	N	Mean	SD	Min	Median	Max
Dependent variables						
Happiness, 0 or 1	9278	0.642	0.479	0.000	1.000	1.000
Happiness, 0–4 scale	9278	2.786	0.827	0.000	3.000	4.000
City district level variables						
Dissimilarity, low spending	77	0.061	0.038	0.000	0.056	0.189
Dissimilarity, high spending	77	0.045	0.061	0.000	0.034	0.451
Share, low spending	77	0.047	0.020	0.020	0.042	0.110
Share, high spending	77	0.016	0.008	0.000	0.016	0.030
Share of land with slope <25%	77	0.955	0.068	0.695	0.985	1.000
Share of built-up land area	77	0.491	0.307	0.015	0.433	0.988
Distance to city center[a]	77	1.385	2.162	0.005	0.500	14.309
Individual level variables						
Income[b]	9278	10.998	1.996	0.000	11.333	16.311
Consumption[b]	9278	11.231	0.804	8.226	11.208	15.142
Asset[b]	9278	13.669	1.745	0.000	13.941	18.596
Debt[b]	9278	2.851	5.086	0.000	0.000	17.521
Gender (1 if female)	9278	0.557	0.497	0.000	1.000	1.000
Age	9278	55.253	16.108	16.000	57.000	96.000
Married	9278	0.794	0.404	0.000	1.000	1.000
Good health	9278	0.436	0.496	0.000	0.000	1.000
Years of education	9278	10.891	3.784	0.000	12.000	21.000

Notes:
[a] In 10 km.
[b] In logarithms.

integer values between 0 and 4, with 0 indicating very unhappy and 4 very happy. We include as controls factors that are shown to be associated with life satisfaction in the literature, such as household income, expenditure, asset, debt, respondent's gender, age, marital status, health condition and education level.

We find that controlling for household income, the level of household consumption and its inequality in the city district the individual resides in do not affect subjective wellbeing. On the other hand, spatial concentrations or separations of high spending or low spending households from the rest affect individuals' subjective wellbeing. From column (3) of Table 2, a 0.01 unit increase in the dissimilarity score of low spending households, which means that an additional 1% of the low spending households need to move to other subdistricts within the district in order to make the spatial distributions of low spending households equal to those of the others, increases the likelihood of happiness by 0.49%. Spatial concentrations of high spending households have the opposite effect: a 0.01 unit increase in the dissimilarity score of high spending households decreases the likelihood of happiness by 3.19%. Similar conclusions hold in models with the scale of happiness ranging from 0 to 4. We interact the dissimilarity and the inequality measures to examine if the effect of consumption dissimilarities depends on the levels of inequality. Based on the coefficient estimates, we plot the average marginal effects in Fig. 2. We find that the negative effect of spatial concentrations of high or low spending households intensifies as its population share decreases. The result is consistent with the findings in Bharathi et al. (2023) that spatial segregation of minority groups harms development.

We conduct a series of tests to assess the robustness of the findings and effect heterogeneities among subpopulations. Firstly, our results hold without controlling for individual characteristics (Table C.2 in the appendix). Secondly, we assess whether the effects of consumption dissimilarities differ substantially between subgroups of people. For the binary outcome of happiness,[9] Fig. 3 displays the coefficients and associated 95% confidence intervals of consumption dissimilarities in different subsamples. Overall, the effects are similar across subsamples. The negative effects of the spatial concentration of high spending households are stronger for residents with more years of education and weaker for residents who are younger. On the other hand, the positive effects of the spatial concentration of low spending households are weaker for residents with good health or more assets. Thirdly, we check whether there are unobserved regional confounders that affect both district consumption dissimilarities and residents' life satisfaction through a placebo test. Residents of a city district are randomly matched to a different district and we estimate the effects of consumption dissimilarities and inequalities of those placebo districts on residents' life satisfaction. The process is repeated 1000 times. As shown in Figures C.3 and C.4 in the online appendix, the robustness of our findings is confirmed because the estimated coefficients from most of these 1000 placebo samples deviate from the estimates in Table 2 and center around zero.

[9] Figure C.2 in the online appendix displays the estimates for the outcome of happiness in the scale of 0 to 4.

Table 2
Consumption dissimilarities and subjective wellbeing.

	(1) Happiness 0 or 1	(2) Happiness 0 or 1	(3) Happiness 0 or 1	(4) Happiness 0 or 1	(5) Happiness 0–4 scale	(6) Happiness 0–4 scale	(7) Happiness 0–4 scale	(8) Happiness 0–4 scale
Dissimilarity, low spending	0.404*		0.487**	−1.135	0.793**		1.011***	−2.185*
	(0.230)		(0.216)	(0.764)	(0.396)		(0.361)	(1.267)
Dissimilarity, high spending	−0.202*		−0.319**	−1.106***	−0.356		−0.655**	−2.561***
	(0.115)		(0.138)	(0.423)	(0.262)		(0.279)	(0.768)
Share, low spending		−0.542	0.321	0.586		0.0421	1.819	2.439
		(1.324)	(1.436)	(1.373)		(2.076)	(2.182)	(2.067)
Share, high spending		−4.724	−3.129	0.568		−6.980	−3.694	1.194
		(3.029)	(3.186)	(3.301)		(4.932)	(5.066)	(5.309)
Dissimilarity, low spending × Share, low spending				41.20***				72.35***
				(15.71)				(26.49)
Dissimilarity, high spending × Share, high spending				43.31				174.4**
				(52.43)				(88.50)
Income[a]	0.0128***	0.0128***	0.0126***	0.0127***	0.0253***	0.0253***	0.0250***	0.0251***
	(0.00273)	(0.00274)	(0.00274)	(0.00273)	(0.00480)	(0.00479)	(0.00480)	(0.00479)
Consumption[a]	0.000477	0.00299	0.00304	0.00314	−0.0111	−0.00598	−0.00588	−0.00606
	(0.00750)	(0.00754)	(0.00754)	(0.00757)	(0.0129)	(0.0130)	(0.0129)	(0.0130)
Asset[a]	0.0148***	0.0158***	0.0154***	0.0154***	0.0334***	0.0354***	0.0345***	0.0345***
	(0.00377)	(0.00375)	(0.00373)	(0.00372)	(0.00675)	(0.00672)	(0.00668)	(0.00667)
Debt[a]	−0.00311***	−0.00327***	−0.00327***	−0.00320***	−0.00785***	−0.00818***	−0.00819***	−0.00806***
	(0.00113)	(0.00113)	(0.00113)	(0.00113)	(0.00185)	(0.00186)	(0.00186)	(0.00185)
Gender (1 if female)	0.0117	0.0139	0.0140	0.0137	0.0392**	0.0438**	0.0440**	0.0440**
	(0.0102)	(0.0102)	(0.0101)	(0.0101)	(0.0186)	(0.0185)	(0.0184)	(0.0184)
Age	0.00443***	0.00462***	0.00465***	0.00461***	0.00765***	0.00803***	0.00809***	0.00806***
	(0.000420)	(0.000424)	(0.000424)	(0.000419)	(0.000738)	(0.000749)	(0.000749)	(0.000741)
Married	0.0443***	0.0401***	0.0404***	0.0406***	0.0755***	0.0669***	0.0676***	0.0680***
	(0.0131)	(0.0131)	(0.0131)	(0.0130)	(0.0222)	(0.0220)	(0.0221)	(0.0219)
Good health	0.184***	0.184***	0.183***	0.183***	0.317***	0.316***	0.316***	0.315***
	(0.0104)	(0.0104)	(0.0104)	(0.0104)	(0.0183)	(0.0183)	(0.0183)	(0.0182)
Years of education	−0.00425**	−0.00320*	−0.00311*	−0.00319*	−0.0143***	−0.0121***	−0.0119***	−0.0118***
	(0.00172)	(0.00172)	(0.00172)	(0.00171)	(0.00306)	(0.00306)	(0.00306)	(0.00305)
Constant	−0.0349	0.0220	−0.0557	−0.130	1.677***	1.698***	1.538***	1.412***
	(0.0884)	(0.144)	(0.146)	(0.148)	(0.158)	(0.239)	(0.241)	(0.244)
No. of respondents	9278	9278	9278	9278	9278	9278	9278	9278
No. of cities	25	25	25	25	25	25	25	25
No. of districts	77	77	77	77	77	77	77	77
R^2	0.067	0.068	0.069	0.070	0.076	0.077	0.078	0.079
City fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors, clustered at community levels, are reported in parentheses.
[a] In logarithms.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

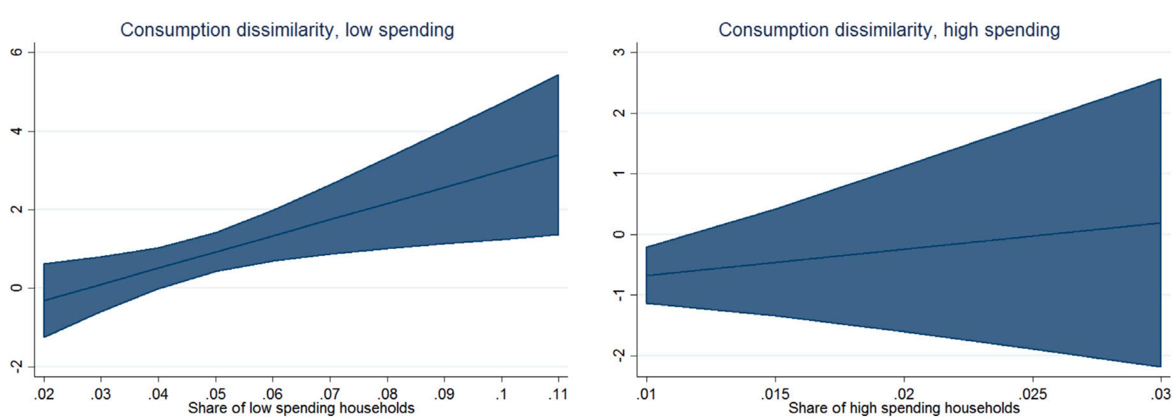

Fig. 2. Average marginal effects of consumption dissimilarity on inequality levels. *Notes*: The figure is based on the coefficient estimates of Column 4 of Table 2. The shaded areas indicate 95% confidence intervals.

Furthermore, noting that life satisfaction is subjective and perceptions of what constitutes happiness can vary among individuals, we consider how consumption dissimilarities affect changes in life satisfaction between the 2017 and 2019 waves of the CHFS. If one's happiness scale does not vary between 2017 and 2019, the time differencing will eliminate those time invariant heterogeneities. The estimates are reported in Table C.4 in the appendix. Lastly, we use the measures of district geographic constraints as instrumental variables for consumption dissimilarities and estimate 2SLS regressions (Table C.5 in the appendix). The results are qualitatively similar with our main findings.

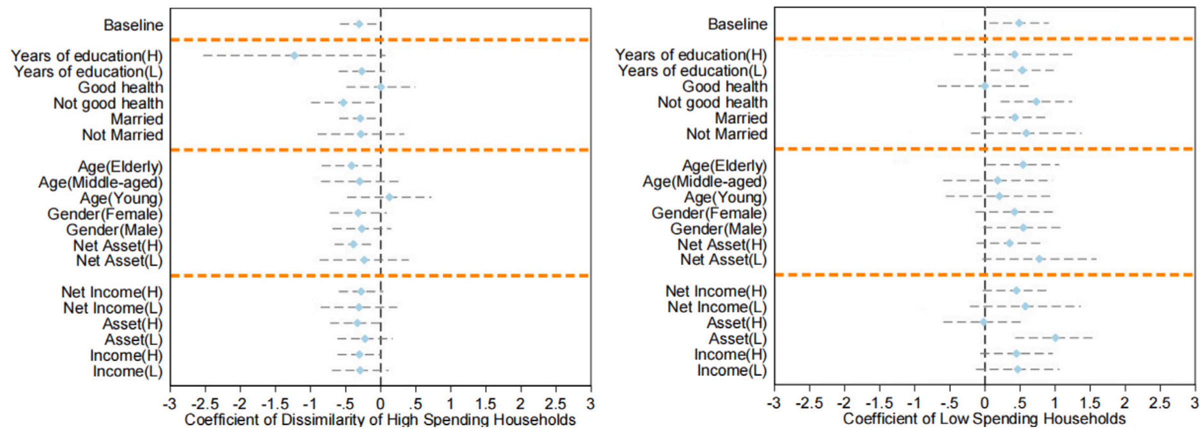

Fig. 3. Heterogeneity analysis by subgroups, dependent variable: happiness 0 or 1. *Notes*: Each line represents a separate regression using a subsample and reports the coefficient estimate and its 95% confidence interval of the variable indicated in the figure. The model specification follows Column 3 of Table 2 in each separate regression. "H" ("L") indicates the subsample where the value of the variable is above (below) its median.

4. Conclusion

One's perception of the unequal distribution of economic resources is influenced by the outcomes observed among nearby individuals, which may deviate from general levels of inequality. Research on this issue, particularly in developing countries, often faces challenges due to limited data availability at fine spatial scales. In this study, we utilize a novel data source derived from mobile big data and develop metrics to capture the spatial disparities between high or low spending households and others at the city district level in China. Our findings reveal that after controlling for city-level inequalities, the spatial separations of high spending households within a city district exert a negative impact on residents' life satisfaction, whereas the spatial separations of low spending households have the opposite effects. The results have implications for governmental policies aimed at mitigating the adverse effects of economic inequalities on individuals' well-being. Future research should further validate the utility of mobile big data and explore the underlying mechanisms that contribute to this intriguing phenomenon.

Data availability

The CHFS is organized and managed by China Household Financial Survey and Research Center, Southwestern University of Finance and Economics.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2023.111412.

References

Barwick, P.J., Liu, Y., Patacchini, E., Wu, Q., 2023. Information, mobile communication, and referral effects. Amer. Econ. Rev. 113 (5), 1170–1207.
Bharathi, N., Malghan, D., Rahman, A., 2023. Ethnic diversity and economic development with spatial segregation. Econom. Lett. 222, 110951.
Cutler, D.M., Glaeser, E.L., 1997. Are Ghettos Good or Bad? Q. J. Econ. 112 (3), 827–872.
Davis, D.R., Dingel, J.I., Monras, J., Morales, E., 2019. How segregated is urban consumption? J. Polit. Econ. 127 (4), 1684–1738.
Gan, L., Yin, Z., Jia, N., Xu, S., Ma, S., Zheng, L., 2014. Data you need to know about China: Research report of China household finance survey•2012. Springer Berlin Heidelberg, Berlin, Heidelberg.
Jiang, S., Lu, M., Sato, H., 2012. Identity, inequality, and happiness: Evidence from urban China. World Dev. 40 (6), 1190–1200.
Lei, X., Shen, Y., Smith, J.P., Zhou, G., 2018. Life satisfaction in China and consumption and income inequalities. Rev. Econ. Househ. 16 (1), 75–95.
Li, T., Barwick, P.J., Deng, Y., Huang, X., Li, S., 2023. The COVID-19 pandemic and unemployment: Evidence from mobile phone data from China. J. Urban Econ. 135, 103543.
Luttmer, E.F.P., 2005. Neighbors as negatives: Relative earnings and well-being. Q. J. Econ. 120 (3), 963–1002.
Saiz, A., 2010. The geographic determinants of housing supply *. Q. J. Econ. 125 (3), 1253–1296.
Yang, J., Huang, X., 2023. The 30 m annual land cover datasets and its dynamics in China from 1985 to 2022.