

# Northumbria Research Link

Citation: Lam, Jacqueline C. K., Han, Yang, Bai, Ruiqiao, Li, Victor O. K., Leong, Jeff and Maji, Kamal (2020) Household wealth proxies for socio-economic inequality policy studies in China. *Data & Policy*, 2. ISSN 2632-3249 (In Press)

Published by: Cambridge University Press

URL: <https://doi.org/10.1017/dap.2020.4> <<https://doi.org/10.1017/dap.2020.4>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/44179/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>


This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria**  
**University**  
NEWCASTLE

RESEARCH ARTICLE

# Household wealth proxies for socio-economic inequality policy studies in China

Jacqueline C. K. Lam<sup>1,2,3,4,\*</sup> , Yang Han<sup>1</sup>, Ruiqiao Bai<sup>1</sup>, Victor O. K. Li<sup>1,2</sup>, Jeff Leong<sup>5</sup> and Kamal J. Maji<sup>1</sup>

<sup>1</sup>Department of Electrical and Electronic Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong

<sup>2</sup>Energy Policy Research Group, Judge Business School, The University of Cambridge, Cambridge, United Kingdom

<sup>3</sup>Department of Computer Science and Technology, The University of Cambridge, Cambridge, United Kingdom

<sup>4</sup>CEEPR, MIT Energy Initiative, MIT, Cambridge, Massachusetts, USA

<sup>5</sup>Faculty of Economics, The University of Cambridge, Cambridge, United Kingdom

\*Corresponding author. Email: [jcklam@eee.hku.hk](mailto:jcklam@eee.hku.hk)

(Received 07 August 2019; revised 01 January 2020; accepted 09 January 2020)

**Keywords:** China; data privacy and security; fine-grained resolution; household wealth proxies; socio-economic status; environmental inequality

## Abstract

In China, one percent of the richest population holds more than one-third of the wealth, while the poorest 25% shares no more than two percent of the total. The country's rapid economic development has resulted in increasing socio-economic disparities, and a rapidly deteriorating environment. This puts the Chinese citizens, especially the most vulnerable and deprived socio-economic status (SES) groups, at high risks of environmental inequality (EI). In most SES-based EI studies conducted in China, household wealth has often been overlooked, though it potentially serves a good economic indicator to capture the socio-economic effect of environmental change in China. Nevertheless, existing SES databases in China are of low spatial resolution and are insufficient to support fine-grained EI studies at the intra-city level in China. The core research challenge is to develop a representative household wealth proxy in high-spatial resolution for China. This study highlights the research gaps and proposes a new household wealth proxy, which integrates both fine-grained data/features such as daytime satellite imagery and easily accessible wealth indicators such as house prices. We also capitalize on everyday economic activity data retrieved from personal mobile phones and online transaction/social platforms in the composition of our wealth proxy to achieve a higher accuracy in estimating household wealth at fine-grained resolution via machine learning. Finally, we summarize the challenges in improving both the quality and the availability of Chinese socio-economic datasets, while protecting personal privacy and information security during the data collection process for household wealth proxy development in China.

## Policy Significance Statement

Drawing from international literature covering household wealth proxies, we propose a five-criterion assessment framework to assess the performance of existing wealth proxies for high-spatial resolution intra-city environmental inequality (EI) study. The five criteria include spatial resolution, relevance with household wealth, applicability to urban China,<sup>1</sup> data availability, and temporal resolution or timeliness. Our study concludes that no existing household wealth proxies could meet all assessment criteria to capture the effects of SES on intra-city

<sup>1</sup> Urban China covers both cities and towns in China. The cities refer to areas of the municipal districts, other connected built-up areas not in the municipal districts, but under the management of the municipality residents' committees. The towns refer to areas not in the cities, but are under the management of the prefectural level government, and other connected built-up areas under the management of the municipality residents' committees (Chen et al., 2016).

© The Author(s) 2020. Published by Cambridge University Press in association with Data for Policy. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

EI in China, making it necessary to tailor-make a new household wealth proxy for China. Finally, we summarize challenges to sound data quality and availability, and to privacy and security, followed by recommendations of relevant data policy frameworks to ensure qualitative data collection with moral access control and ethical principles, while taking personal privacy and security dimension into consideration in the process of developing household wealth proxies, in order to facilitate sound evidence-based and fine-grained socio-economic-induced EI studies in China.

## Introduction

China's rapid economic development has been achieved at the expense of socio-economic inequality and environmental deterioration. Despite the rapid per capita gross domestic product (GDP) growth (8.6% per annum) recorded in between 1979 and 2014 (Chen *et al.*, 2016), the improvement in the living standards of common citizens is moderate and the gaps of wealth between the rich and the poor are widening (Cheng, 2013). In 2014, more than one-third of total household wealth is possessed by only one percent of the households in China, whereas the poorest one-fourth holds less than two percent of the total in China (Xie and Jin, 2015). Meanwhile, as a result of intensive economic development, citizens residing in some parts of China have been exposed to much higher risks of extreme weather and environmental pollution (Wang *et al.*, 2007; Wang *et al.*, 2012), jeopardizing their quality of life and health (Cheng, 2015; Gao and Li, 2013). According to a report published by the Organization for Economic Co-operation and Development (OECD, 2016), outdoor air pollution resulted in 662 premature deaths per million people in 2010 and the number is estimated to reach 1,563/million people per year by 2060 in China. Meanwhile, the loss of GDP due to extra health expenditure and labor productivity is estimated to increase by about 2.1% in 2060 (OECD, 2016).

Environmental pollution induced disparity, and the associated environmental inequality (EI) between the rich and the poor is evident in the cities with the fastest growing economies, such as Beijing and Shanghai in China (Kolenikov and Angeles, 2009; Taubenbock *et al.*, 2009; Xie and Jin, 2015). EI is defined as the disproportionate share of environmental burdens and adverse health outcomes among disadvantaged groups (e.g., ethnic minority communities or low-SES groups) (Davoudi and Brooks, 2012). The studies investigating EI in the developed and developing countries have revealed that people with low SES, low income, and education or nonprofessional occupation are more likely to experience a higher level of environmental catastrophe, in the form of air pollution, flood, drought, and extreme heat (Li *et al.*, 2018; Park *et al.*, 2018; Winsemius *et al.*, 2015), revealing the existence of climate or environmental inequality (Marlier and Atkinson, 2010; Sharath *et al.*, 2016). However, these studies have focused on the impacts of EI related to these three main SES indicators, without accounting the potential influencing factors, such as household wealth. Hence, these studies may not have adequately explained EI and the resulting SES disparities (Hajat *et al.*, 2015). A number of epidemiological studies have distinguished wealth from income and argued that the former serves a better indicator of economic wellbeing (Pastor-Satorras *et al.*, 2015). In parallel, much of the existing environmental justice literature has argued that household wealth may serve better in China than household income as an SES indicator (Aittomäki *et al.*, 2010; Benzeval and Judge, 2001; Fritzell *et al.*, 2004; Rahkonen *et al.*, 2000). A household's wealth condition may reflect better one's ability to purchase an apartment at one's desired location (Sun, 2015), eventually affecting the household members' air pollution exposure (Li *et al.*, 2018). To gain a better understanding of SES-based EI in China, it is important to investigate whether household wealth will influence household air pollution exposure in China.

EI study conducted at the intra-city level is important in capturing the crucial spatial characteristics on environmental exposure (Bowen, 2001; Jerrett *et al.*, 2005). Intra-city EI study identifies the most polluted areas in the city and hence offers insights to policy-makers for environmental management (Hallegatte and Rozenberg, 2017). Due to the *scaling effect*, EI studies conducted at the inter-city level in China might not be able to capture the effects of relevant EI studies conducted at the intra-city level. Two studies conducted in the United Kingdom (Huby *et al.*, 2009) and the United States (Krieger *et al.*, 2002) have

produced completely opposite results regarding the relationship between SES and EI at different spatial scales. When designing environmental policies that address the spatial effects of SES on EI, it is therefore necessary to couple SES data with environmental exposure data at the right spatial scale, especially at a finer spatial resolution, such that more specific, relevant results that facilitate area-specific environmental policy-making can be achieved (Fernández and Wu, 2016).

At the moment, neither fine-grained wealth data collected from existing household surveys, nor fine-grained proxies estimating household wealth distribution are readily available in China, rendering the EI studies at the intra-city level, such as at the block-level of a city in China, highly challenging. At the moment, proxies such as the nightlight satellite imaging are contextualized in the slow growth developing countries in Africa, making the extension of similar proxies to China inappropriate (Olivia et al., 2011). In contrast to Africa, there are sufficient nightlight facilities across both rich and poor urban areas in China, and the statistical correlation between nightlight and household wealth is weak (Mellander et al., 2015; Venerandi et al., 2015). Scholars also show that nightlight can hardly differentiate, especially in the case of a densely populated poor area, from a sparsely populated rich area (Graesser et al., 2012). Therefore, direct extension of similar socio-economic proxy models from studies in Africa to that of China is not recommended, given the very different economic environment and development patterns that China has exhibited as opposed to the relatively less-developed countries such as Africa (Mason, 2015). Hence, developing high resolution and precise household wealth proxies specifically relevant for urban China is urgently needed, as it provides the basis upon which meaningful exploration of the relationship between household SES and EI can be conducted (Henderson et al., 2012; Jean et al., 2016).

In the present study, we first highlight the significance of household wealth related fine-grained EI studies in China. Second, we survey the available Chinese household wealth database, and identify the current research gaps, including the lack of high spatial resolution and high economic representative household wealth data that could support EI studies in China. Third, we survey the available techniques for developing relevant wealth proxies across the world, and outline their strengths and weaknesses based on a five-criteria assessment framework. Fourth, we discuss the pros and cons of existing new household wealth proxies for measuring household wealth at the fine-grained spatial level, with reference to fine-grained SES-induced EI studies in China. Fifth, we summarize the challenges to data availability and quality with concerns about ethics and privacy, and recommend data policies to improve the quality and availability of fine-grained SES data by protecting data privacy and information security in relation to household wealth proxy development in China. Finally, we conclude our study and recommend future directions for developing new household wealth proxies to support fine-grained SES-induced EI studies in China.

## Methodology

Our research methodology is structured as follows. First, we surveyed the Chinese and international literature that cover SES-related EI studies, and highlighted the significance of using fine-grained household wealth as a proxy and the relevant datasets for improving EI research in China. Our selection of publications is guided by the following research questions:

1. On which datasets are the majority of SES-related EI studies based?
2. What are the critiques of such datasets?
3. Has household wealth been used as a proxy for EI studies?
4. Have household wealth datasets been available, and are they fine-grained enough for SES-induced EI studies?

Based on the state-of-the-art scientific literature guidelines, we used the following keywords to search for relevant literature, “environmental inequalities,” “socio-economic status,” “poverty estimation,” “poverty mapping,” “wealth proxy,” “fine-grained,” and “scaling effect”. These keywords were keyed into the databases including the Web of Science, Scopus, and the Google Scholar for selecting the relevant

review articles. The publications selected will provide a general idea regarding our household wealth proxy and database landscape.

Based on the publications we identified the main wealth proxies developed to further investigate their applications on SES-related EI studies. We further expanded our keyword search by keying in the names of individual proxies and their corresponding supporting technologies. We identified the lack of high-spatial resolution and high-economic representative household wealth data and selected 87 publications for further analysis. [Figure 1](#) summarizes how our key research gaps were identified.

## Research Gaps and Research Significance

### ***On wealth on SES-related EI study in China***

Traditional measures of SES are often defined by income, education and occupation (Adler and Newman, 2002). Some research has highlighted strong impacts of household SES on risks of environmental exposure (Evans and Kantrowitz, 2002), such as exposure to hazardous wastes and others, indoor and outdoor air pollution, water pollution, and ambient noise (Miao *et al.*, 2015; Mohai *et al.*, 2009; Pulido, 1996). Most EI studies have taken into account these three main components of SES, instead of household wealth, an important indicator of economic wellbeing (Morello-Frosch *et al.*, 2001), usually by default (Hajat *et al.*, 2015). However, household wealth can be an influential factor in environmental exposure and EI in the United States (Morello-Frosch *et al.*, 2001). Some scholars argue that the effect of household wealth on environmental exposure could have been easily accounted for by household income, which reflects a household's economic wellbeing and is positively correlated with one's environmental exposure. However, there is a significant difference in terms of the distribution of economic values represented by these two variables (Quadrini and Rios-Rull, 1997). In both the United States and China, household wealth is more unevenly distributed than household income, which may better capture the socio-economic disparity (Budria *et al.*, 2002). Some studies have reported that the relationship between household income and environmental exposure is not particularly strong (Quadrini and Rios-Rull, 1997; Taubenbock *et al.*, 2009). This renders household wealth a more relevant SES indicator for environmental exposure and EI studies in China.

Household wealth can reveal better the level of economic wellbeing than household income, given its higher stability and stronger influence on living standards over the longer term (Aittomäki *et al.*, 2010). In the United States, fluctuating income cannot fully capture the standards of living (Carle *et al.*, 2009). It reflects neither one's financial hardship experienced (Mendes De Leon *et al.*, 1994) nor one's affordability of goods and services (Krieger *et al.*, 1997; Tucker-Seeley *et al.*, 2013). In contrast, household wealth, which measures structural and chronic poverty, instead of stochastic and temporary poverty (Hallegatte and Rozenberg, 2017), is less volatile and more credible (Filmer and Pritchett, 2001; McKenzie, 2005). Compared to income, wealth is prone to be more indicative of household wellbeing in Australia (Headey and Wooden, 2004). This is especially the case in China. When compared to developed countries, households in China tend to take precautionary measures against any unexpected reductions in future income by accumulating household wealth, in order to ensure that they can enjoy relatively more stable living standards even when their household incomes fluctuate (Giles and Yoo, 2007).

If household wealth can better indicate one's economic wellbeing than household income in China, it is logical to deduce that household wealth can be a good indicator of one's capability to avoid household exposure to environmental hazards in China. Affluent families in China have more flexibility in choosing residence location (Huang and Batterman, 2000), as they can afford apartments of better environmental quality (Lucas *et al.*, 2004). Moreover, a number of studies have indicated that the long-term accumulation of material welfare, instead of income in the short term, tends to be a more significant cause of health inequality (Aittomäki *et al.*, 2010; Benzeval and Judge, 2001; Fritzell *et al.*, 2004; Rahkonen *et al.*, 2000). As household health condition tends to be correlated with exposure to environmental pollution, the poor health condition of low SES-population indirectly reveals the existence of EI among communities of different SES status (Aittomäki *et al.*, 2010).

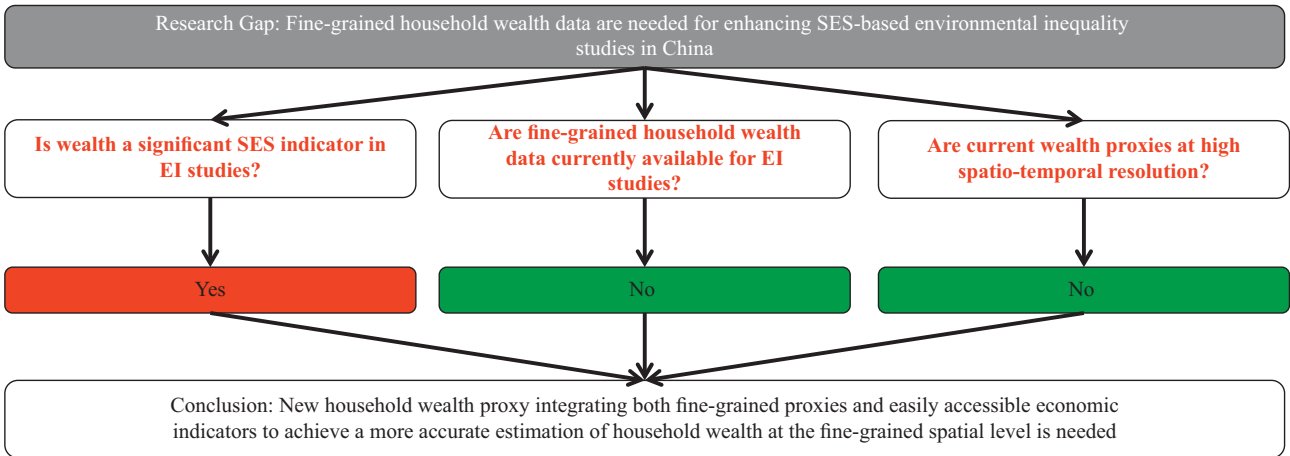


Figure 1. Methodology for identifying the research gaps.

### ***On selecting wealth as an SES indicator for intra-city EI studies in China***

Current SES-related EI studies in China are conducted at the inter-city or country level. However, EI, as reflected by differential environmental exposure, may be manifested at a finer spatial level. The scale of analysis is one of the most important methodological issues in EI studies (Bowen, 2001; Mennis, 2002). In general, homogeneous environmental exposure within the study unit is assumed. This assumption is usually made abruptly without reasonable justification (Huang and Batterman, 2000). Air pollution, for instance, could vary significantly at the intra-city level (Li et al., 2018). Hence, ignoring spatial heterogeneity could induce uncertainties and errors (Huang and Batterman, 2000). In addition, due to the spatial *scaling effect*, variations or even contradictions of statistical results could be identified when spatial data are aggregated at different spatial resolution units (Jelinski and Wu, 1996; Wu, 2007). In the past, such assumption may not be problematic, because the environmental monitoring data obtained may not be fine-grained due to restrictions in the number of monitoring stations established. In recent years, we have seen the environmental monitoring data evolving toward being more fine-grained with increasing monitoring coverage or interpolating techniques assisted by more advanced algorithm, but potentially relevant SES indicators are yet to achieve comparable resolution. In the United Kingdom, extensive high-quality national monitoring networks are required by the Environment Act 1995 to produce 1-km resolution emission inventory (Longhurst et al., 2009). In Hong Kong, Li et al.'s (2017) research has also obtained fine-grained air pollution data at 100 × 100 m spatial resolution, achieving 82% accuracy. EI studies are usually spatially dependent (Fernández and Wu, 2016; Huby et al., 2009; Krieger et al., 2002), and can be distorted by the scaling effect (Baden et al., 2007; Cutter et al., 1996). Research on the spatial pattern of EI in Chile has shown that the scale of study affects both distribution and correlation intensity between SES and environmental variables (Fernández and Wu, 2016). Another study conducted in Singapore has shown that EI appears to be stronger when examined at smaller spatial scales, whereas SES indicators are prone to be skewed in frequency distributions at finer scales (Tan and Samsudin, 2017). Hence, comparable SES data of similar fine-grained resolution as environmental variables are urgently needed.

Current SES-related EI studies in China conclude that the wealthier the city, the higher the air pollution exposure. This may contradict the results obtained at the intra-city level, as revealed by a recent study conducted in Hong Kong (Li et al., 2018). This fact is consistent with the deduction of the *scale effect*, and provides the hints on why inter-city level is not adequate for EI studies. More specifically, research shows that the 30 most economically powerful cities in China are exposed to higher levels of PM<sub>2.5</sub> pollution, based on the census data at prefecture city level (Wan and Su, 2017). However, research in Hong Kong has demonstrated that, when examining environmental exposure at the Constituency Area (CA)-level (with Hong Kong divided into 412 CAs) across the same city, the lower SES communities are the victims of higher air pollution (Li et al., 2018). Evidences of EI shown at the intra-city, including EI that are associated with differential exposure to air pollution and water pollution, differential access to waste disposal, urban greening, and natural resources (Davoudi and Brooks, 2012), have been identified in cities such as the United Kingdom and the United States (Bowen, 2001). Without accounting for the scale-dependency of EI, policymakers might risk making too general a conclusion that people with higher SES are exposed to worse environmental quality across all geographic scales, which would affect quality environmental decision-making.

Environmental policy addressing EI will become more effective if one can specify which specific geographical regions and which specific SES groups will require specific adjustments or attention, instead of providing general guidelines or directions. EI is the result of the complex interaction between individuals in the environment (Davoudi and Brooks, 2012). It is important to investigate the fairness of distributing environmental goods not only at the national or the provincial level, but also at the intra-city level (Schlosberg, 2004). Intra-city EI research offers new insights on how EI is at work at the micro-level, and creates opportunities for more bottom-up environmental policy-making (Hallegatte and Rozenberg, 2017). For instance, even when climate change-related EI has been widely identified at the global level, uncovering the experience of micro-level EI could still offer significant insights to advance local climate policy-making (Fisher, 2015). Research in East Africa highlights that choosing a more fine-grained spatial



scale model for EI research could improve management strategies (MacKenzie et al., 2014). Environmental planners have also advocated the benefit of conducting more fine-grained analysis at the block or the community level rather than the city or the regional level to advance local environmental planning (Tan and Samsudin, 2017). In short, the mismatch between geographical scales of EI research and policy-making would lead to erroneous misinterpretation of EI, and weaken the effectiveness of environmental policy-making (Fernández and Wu, 2016).

To the best of our understanding, except for Li et al. (2018), who focus on Hong Kong, no SES-related EI study conducted in China has so far examined the relationship between wealth and EI at the fine-grained intra-city level. The relationship between SES and environmental risks has been investigated at the national or the provincial level in China (Pratt et al., 2015); such results may not be sufficient to provide the needed evidence for location-based EI alleviation. Even though fine-grained wealth-related study is crucial for the Chinese Government to understand the potential crisis induced by EI and to develop effective precautionary measures, it has yet to be conducted in China (Pratt et al., 2015).

### ***On high spatio-temporal resolution and representative wealth database for intra-city China***

The major challenge of conducting fine-grained SES-based EI analysis comes from the low spatial resolution of current wealth databases. A list of databases that provides Chinese household wealth data is summarized in Table 1, and we will discuss below why these databases are not suitable for our use.

Chinese censuses should in theory obtain household information covering the entire population in China. However, micro-data are not commonly available to the public. Only a few research institutions with specific permission given by the government can have access to the data, representing only 0.1–1% of the sampled data. Furthermore, the time interval between two population censuses is large—it is only taken roughly every 10 years. Hence, the outdated Chinese censuses have failed to provide a true representation of the country's annual socio-economic development.

Other household surveys besides the national census have been conducted. One of the challenges of such household surveys is their very small sample size; the China Family Panel Studies (CFPS) and Chinese Household Income Project Survey (CHIP) in Table 1 for instance, just sample annually around 16,000 and 18,000 households, respectively. Another challenge being that the disclosure of geographic information in most such household surveys is limited to the provincial level only. As observed from Table 1, only a few surveys, such as the Chinese General Social Survey (CGSS) have broken this down into 125 counties. In reality, China consists of more than 2,800 counties and 660 cities. The data collected from household surveys thus has a low spatial resolution and not representative enough of the entire population. This makes it difficult to determine the wealth distribution and household differences within the intra-city regions, and presents one of the biggest obstacles to conduct fine-grained socio-economic analysis across China. Furthermore, it usually takes one to 2 years for the data to become available to the public after the corresponding surveys have been completed. The lack of timely data is especially inconvenient for researchers to investigate the current intra-city EI reality in China and provide meaningful recommendations for policymakers.

To summarize, the two critical defects for utilizing the current household database in fine-grained EI analysis are:

1. Low coverage and low spatio-temporal resolution—not all blocks or communities are covered in the research; in addition, some data are often infrequent and not timely.
2. Statistical unreliability—the sampling size within each geographical unit is too small, thus undermining the power of statistical tests and the precision of estimations.

In order to overcome these challenges, it is essential to investigate new fine-grained proxies gathered from other data resources to support high spatio-temporal EI studies in China.

**Table 1.** Database of household wealth in China

| Database   | Design  | Year  | Sample frame and size   | Area   |
|--|---|---|---|--|
| Census   | Panel   | 1953, 1964, 1982, 1990, 2000, 2010                                    | Some research institutions could get 0.1–1% sampling data from the database.  | China and all its provinces  |
| China National/<br>Provincial<br>Statistical<br>Yearbooks        | Panel   | 1981–   | –   | China and all its provinces  |
| China Household<br>Finance Survey<br>(CHFS)                      | A cross-section has been completed  | 2011, 2013, 2015  | Multistage semi-probability sample. 8,438 households in 2011, 28,000 in 2013 and 40,000 in 2015.                            | 25–29 provinces in China   |
| China Labor Force<br>Dynamic Survey<br>(CLDS)                    | Panel (rotating panel)  | 2012, 2014, 2016  | Family members aged 15–64. Multistage cluster, stratified, PPS sampling. Over 900 urban households.                         | 29 provinces in China  |
| China Health and<br>Nutrition Survey<br>(CHNS)                   | Panel (rotating panel) with revised modules of questionnaires each wave                     | 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009                        | Multistage semi-probability sample. Around 4,400 urban and rural households.  | 10 provinces in China  |
| Urban Household<br>Survey (UHS)                                  | Panel (rotating panel)  | 1986–2009   | Stratified sampling and equidistant sampling or two-phase sampling. 3,500–4,000 households annually.                        | Data from all provinces have been collected, while data for six provinces have been accessed |
| China Family Panel<br>Studies (CFPS)                             | Panel, with a new wave of interviews each year  | Since 2010, once every 2 years  | National probability sampling (excluding ethnic minority autonomous districts in western region). Around 16,000 households. | 25 provinces in China  |
| Chinese General<br>Social Survey<br>(CGSS)                       | Repeated cross-sections, with new modules of questionnaires each wave; panel from 2010 wave | 2003, 2004, 2005, 2006, 2008, 2010                                    | National probability sampling of people aged 18 and over. 6,000–12,000 households.  | 125 counties in China  |
| Chinese Household<br>Income Project<br>Survey (CHIP)             | Repeated cross-sections, with revised modules of questionnaires each wave                   | 1988, 1995, 2002, 2007, 2013  | Systematic sampling. 18,948 households.   | 234 counties in 126 cities in 15 provinces in China  |
| China Health and<br>Retirement<br>Longitudinal Study<br>(CHARLS) | Panel with a new wave of interviews every 2 years   | Since 2011, once every 2 years (pilot survey and resurvey 2008, 2012) | Multistage clustering, PPS sampling. Around 10,200 households.  | 150 counties from 30 provinces in China  |

## Use of Wealth Proxies for High Spatio-Temporal Resolution

As discussed in Section “Research Gaps and Research Significance,” despite the current research gaps in fine-grained wealth database for intra-city China, such database is yet to be established. Hence, proxies that support high spatio-temporal resolution SES-based EI studies in China would be needed. Table 2 summarizes some of the proxies, which have been proposed. We shall next investigate each of these proxies in details below.

### ***Night-time satellite imagery data***

Past studies show that nightlight emission is a legitimate measure of wealth distribution at a spatial level (correlation: 0.73–0.87) (Weidmann and Schutte, 2017). Therefore, rather than relying solely on the nonrepresentative, time-consuming and often inconsistent household survey data as in traditional research, there is a recent trend in developing socioeconomic models that use satellite-based data as proxy measurements and to calibrate such models based on survey data (Elvidge et al., 2009). Key benefits of utilizing satellite imagery data is shown in Figure 2, which include its relatively low acquisition cost, public availability, data consistency, and most importantly, the high spatial and temporal accuracy it provides to improve our understanding in socio-economic activities of humanity (Elvidge et al., 2009).

However, there are a few limitations on this methodology (Fernández and Wu, 2016), which has been reported in Figure 2. If we simply employ nightlight satellite data alone, we might face the difficulty of distinguishing between densely populated poor areas and sparsely populated wealthy areas. Nightlights only display little variation in extremely poor regions such as Africa since it is simply uniformly dark at night. Hence, this motivates us to use daytime satellite imageries such as street maps, land features and water resources as a complement of the nightlight data.

### ***Daytime satellite imagery data***

As discussed above, nightlight may serve a good intermediate proxy measurement of wealth, but we also need other proxies to further facilitate cases when nightlight alone cannot predict wealth at high accuracy. Past studies showed that daytime visible features such as roofing materials and urban morphologies have demonstrated a significant linear relationship with local expenditures at the block level (Fernández and Wu, 2016); attributes of cars and other environmental variables such as rainfall and topography also have appeared to be legitimate measures for predicting local wealth at high-spatial resolution (Cao et al., 2016; Dargay et al., 2007; Watmough et al., 2016).

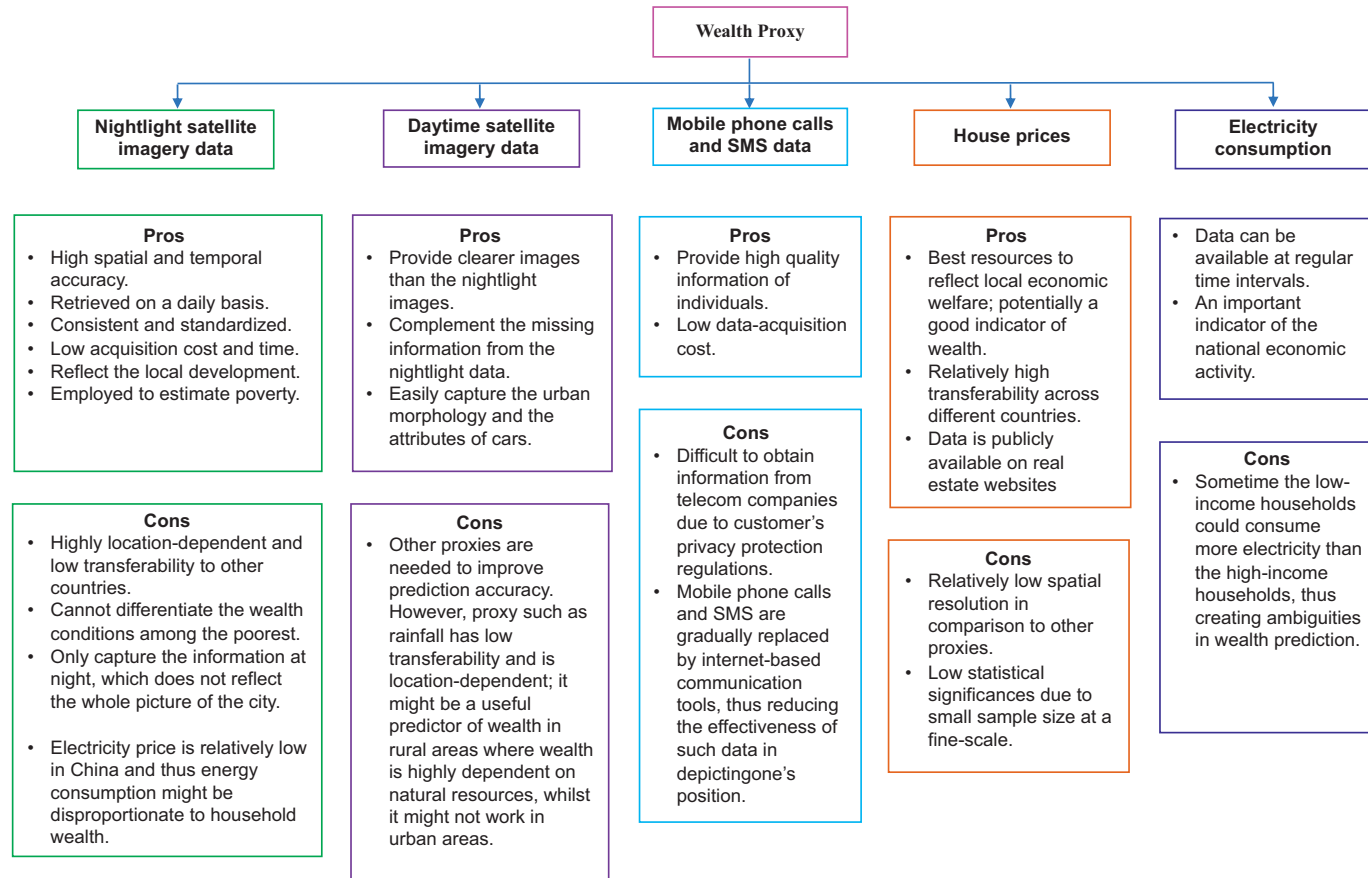
### ***Mobile phone calls and SMS data***

Another popular proxy for estimating one’s socio-economic status is the use of personal mobile phone usage history. The rationale for using mobile phone metadata as a proxy is not just due to its ability to capture one’s social network and characteristics of communication events and travels, but also its capability to reflect other welfare information like consuming behaviors and histories (Candia et al., 2008). There is enough evidence to justify the significant relationship between mobile phone usage and many socioeconomic factors such as demographic statistics and education levels (Frias-Martinez and Virseda, 2012). In particular, previous research showed that an individual’s wealth can be predicted at high accuracy from features of personal social network retrieved from his/her mobile phone calls and SMS metadata (Luo et al., 2017). Therefore, mobile phone metadata provides an alternative to collecting localized and timely information and again, serves as a complement of the traditional methods such as household survey data and national censuses, often at a relatively lower data acquisition cost.

Nevertheless, mobile phone usage can only provide localized information since the communication behaviors could vary dramatically across countries due to cultural differences, which hampers the transferability of this proxy data (Fernández and Wu, 2016). Moreover, mobile phone data are usually owned by private companies and might not be available to researchers due to the customers’ privacy protection regulations.

**Table 2.** Summary of five wealth proxies

| Proxy                             | Source data  | Applicable area   | Core methodology  | Reference   |
|-----------------------------------|--|---|---|---|
| Nightlight satellite imagery data | Nighttime satellite imagery  | Poor areas  | Let $X$ be the total observable light in a region and $Y$ be the total income. The assumed relationship is given by $X = Y^\beta$ , where $\beta$ is the elasticity of lights with respect to income. The methodology is a response to poor countries, which do not only lack funding for national censuses or surveys, but are also in short of illumination/electricity facilities in under-developed regions across the world.   | Henderson et al. (2012)   |
| Daytime satellite imagery data    | Satellite imagery on urban morphology/attributes of cars/environmental variables | Urban/rural areas   | Features of the daytime satellite image, such as buildings, streets, open spaces, water, road, and so forth, can be identified using convolutional neural network or object-oriented image recognition technologies. Based on these features, the average wealth in urban or rural areas has been estimated in previous research through artificial classification or automatic estimation by deep learning technology.   | Ekbohm and Bojő (1999), Gebru et al. (2017), Jean et al. (2016), Olivia et al. (2011), and Taubenbock et al. (2009)   |
| Mobile phone calls and SMS data   | Telecom operator records   | Areas with mobile phone as the most frequent contact approach | Personal social network and personal wealth are highly correlated, and supported by differences in personal access to job opportunities, consumer behaviors, human mobility, and so forth among population possessing different amounts of wealth. This can be done as follows: Obtain the personal social network structure, communication pattern and geographical location through mobile phone records, which involves the total volume, the intensity and timing of phone calls, the geospatial mobility pattern marked by phone calls, etc. Collect individuals' wealth information through mobile phone survey. Establish a model between regional wealth and mobile phone data. | Blumenstock et al. (2015), Deville et al. (2016), Luo et al. (2017), Salah et al. (2011), Šćepanović et al. (2015), Seidel et al. (2000), and Toole et al. (2015) |
| House prices                      | Online real estate websites records  | Areas with online house prices                                | This approach can be implemented to determine the value of consumers in e-commerce. The average price of houses near the mailing address recorded in e-commerce are gathered from the websites of real estates companies, and the average housing price is regarded as a reflection of consumers' consumption ability.  | Li et al. (2015), Xie and Jin (2015), Zhu and Chen (2017)   |
| Electricity consumption           | Electricity distributor records  | Areas with high electricity penetration                       | Establish the regression model based on historical data between electricity consumption and household wealth, and make predictions based on the model.  | Chen et al. (2007), and Zhang and Cao (2012)  |



**Figure 2.** Pros and cons of five existing wealth proxies.

### **House prices**

In contrast to European countries where financial assets and pension funds are the main components of an individual's wealth, scholars have shown that housing assets account for over 70% of household wealth in China, and that the wealth-to-income ratio is approximately 9.2:1 (Xie and Jin, 2015). These figures are still growing and thus it is natural to consider using house prices as a predictor of household wealth. On the other hand, house prices can also provide useful information about the land resources and reflect the local economic development. It is also convenient to gather house prices from online real estate company websites. A study has shown that the household income is positively correlated to house price (Sun, 2015), and thus house price can be used as a potential predictor of the household's economic status and consuming ability.

### **Electricity consumption**

In addition, electricity expenditure derived from the records of electricity distributors could be a potential proxy for household wealth, although it has not been used in the literature yet. One advantage is that the data can be gathered at regular time intervals. A study shows that electricity usage is an important indicator of national economic activity in Asian countries (Chen *et al.*, 2007). Some Chinese scholars have conducted a survey in China, and have claimed that there is a positive correlation between household income and electricity consumption (Zhang and Cao, 2012).

However, the credibility of using this proxy for household wealth is still questionable, as the Tianjin study excludes the records of both low-income households with high electricity consumption and high-income households with low electricity consumption, without providing sufficient details on the exclusion criteria.

Each proxy has its own pros and cons. In [Figure 2](#), we summarize the advantages and disadvantages of the proxies used in the previous studies.

## **Assessment of the performance of existing proxies for estimating wealth at the intra-city level in China**

In this section, we provide a systematic comparison of the performance of household wealth proxies (see [Table 3](#)). Five criteria are used to assess the household wealth proxies' performance:

1. Spatial resolution, which refers to the size of the spatial dimension that a proxy covers.
2. Direct relevance with household wealth proxy, which refers to how directly a proxy is related to household wealth.
3. Applicability to urban China, which refers to how suitable a proxy can be applied to the context of urban China.
4. Data availability, which refers to how easily accessible the data are to the public.
5. Temporal resolution, which refers to how frequently such data is updated.

The criteria are listed in rank order in terms of importance in proxy performance. The higher the performance, the higher the reliability of the existing proxy model to estimate wealth. In [Table 3](#), H indicates a high performance in any of the aspects (1–5) described above and L indicates a low performance. Detailed descriptions of performance are included as Notes to [Table 3](#).

In terms of spatial resolution, house prices posted on real estate websites may not perform well since the number of houses on sale in a small sub-district may be relatively limited, whereas the sample size may limit the proxy's statistical representativeness, while other proxies derived from satellite images have higher spatial resolution (ranging from  $1 \times 1$  m to  $10 \times 10$  m) (Jin and Davis, 2005; Karantzas and Argialas, 2004). In terms of direct relevance with household wealth, house prices are considered more directly relevant for two reasons: first, a residential property takes up the greatest share of assets in most families, while wealthy households are more capable of purchasing expensive houses; second, house prices is a reflection of the scarcity of land resources, and serves an indicator of local economic development.

**Table 3.** Assessment of the performance of household wealth proxies based on five criteria

|                                   | Spatial resolution | Direct Relevance with household wealth | Applicability to urban China | Data availability | Temporal resolution |
|-----------------------------------|--------------------|--|------------------------------|-------------------|---------------------|
| Nightlight satellite imagery data | H                  | L                                      | L                            | H                 | H                   |
| Daytime satellite imagery data    | H                  | L                                      | H                            | H                 | H                   |
| Mobile phone calls and SMS data   | H                  | L                                      | H                            | L                 | H                   |
| House prices                      | L                  | H                                      | H                            | H                 | H                   |
| Electricity expenditure           | H                  | L                                      | L                            | L                 | H                   |

Spatial resolution—H means high spatial resolution corresponding to square grids less than 100 m per side, L means low spatial resolution corresponding to square grids greater than 100 m per side.

Direct relevance with household wealth—H means directly related to household wealth, L means indirectly related to household wealth.

Applicability to urban China—H means highly applicability to urban areas in China, L means low applicability to urban areas in China.

Data availability: H means data readily accessible by the public, L means data not readily accessible by the public.

Temporal resolution—H means data updated at least yearly, L means data not updated at least yearly.

As compared to house prices, other proxies have a less direct relationship with household wealth. In terms of the proxy's applicability to urban China, using nightlight image as a proxy may have low performance, since China is relatively well developed, cities have been provided with very good illumination facilities at night, even in some poorer districts, standard illumination systems are available, the applicability of nightlight image as a proxy to household wealth in urban China appears to be low. However, daytime satellite image has been utilized in estimating wealth in both rural and urban areas (Taubenbock et al., 2009), and thus has a high adaptability. Mobile phone data (calls and SMS) may not be too applicable in China in the future since calls and SMS will likely be phased out in China due to the increasing popularity of internet-based communications via WeChat and Weibo.

In terms of data availability, compared to data from telecom operators or electricity suppliers, house prices or satellite images are more easily accessible online to avoid privacy disclosure issues. Finally, in terms of temporal resolution, it is expected that proxies based on satellite images will be updated daily. Mobile phone call and SMS data and electricity consumption data are usually aggregated over a period of time, for instance, updated monthly. Online house prices are updated hourly or daily. As such, none of the current proxies can achieve high performance in these five dimensions, it will thus be good to look for a new household wealth proxy that can be catered for fine-grained SES-based EI studies in China.

### New Wealth Proxy Development

As shown above, none of the existing proxies support our need for a fine-grained and highly accurate wealth proxy for intra-city EI studies in China. In order to achieve a high performance meeting all five criteria, we propose three principal guidelines for future wealth proxy development in China:

1. Combining existing proxies to create a hybrid proxy.
2. Identifying new proxies to account for social and technological trends.
3. Deploying unsupervised or semi-supervised machine learning methodologies.

A hybrid proxy attempts to take advantage of the merits of existing wealth proxies, while avoiding their disadvantages. For China, the features extracted from daytime satellite images, mobile phone data (calls and SMS), and house prices are recognized as applicable to intra-city studies (see Table 3). Among these three types of data, only the house prices are of direct relevance to household wealth, hence it shall be kept in the new model. However, the spatial resolution of house price data is low, and should be compensated by fine-grained daytime satellite images or mobile phone data. Comparing the daytime satellite images

and mobile phone data, the former has higher data availability. Hence, a combination of daytime satellite images and house prices is suggested.

More specifically, if only the daytime satellite images are utilized, it might be hard to differentiate the average household wealth between two communities with similar building structures but different locations, for instance, one in the city center and another in the suburbs. However, with the hybrid proxy, such issue can be fixed by incorporating the local house prices. Moreover, the accuracy of using the house prices to estimate the household wealth could be improved by the satellite data. Apart from improving the spatial and temporal resolution, the estimation accuracy can also be improved by observing the housing structures captured by the satellite images.

For the wellbeing of our societies, high spatial resolution gridded data sources are urgently needed, since they play crucial roles in the examination of statistical relationships/correlations for robust socio-economic analysis and technological developments. However, such data sources are unavailable in countries such as China, which creates unnecessary barriers to conduct evidence-based socio-economic research and sound decision-makings in environmental policy evaluations. Previous research has indicated that mobile phone calls and SMS data can predict the wealth of the older population, but are less powerful in predicting that of the younger population (Luo *et al.*, 2017). This might be related to the emerging online communication platforms such as Weibo and WeChat, which have witnessed an increasing popularity among the younger generations. According to a report (Lee, 2019), the number of daily active WeChat users in China had reached over one million in 2018, which was equivalent to around 72% of the total Chinese population (World Bank, 2019). Hence, mobile phone data could be a valuable economic activity proxy if the access of online social media network data is permissible/available. Besides, geo-coded data and satellite imagery provided by Google or Gaode map in China can also serve as useful inputs for household wealth proxy constructions. Moreover, various wealth proxy data can be collected via opportunistic sensing or passive crowdsourcing. In some pilot studies, personal mobility data extracted from Uniform Resource Locator (URL) requests (Zhao *et al.*, 2019) and geo-coded text data extracted from online forum platforms (Saeidi *et al.*, 2017) are found to be correlated with the economic conditions/characteristics at the individual and the neighbourhood level; this could be left as a future research area in our studies of wealth proxy development in China.

Given the lack of relevant ground truths for relevant wealth proxies in high spatial resolution in China (see Methodology), it would be important to develop a more representative household wealth proxy based on unsupervised or semi-supervised machine learning methods from big data.

For instance, an integrated wealth proxy that reflects a household's spending across all aspects of life can be developed based on proxies that represent expenditures on accommodation, transportation, daily necessities, clothing, accessories, education, and so forth. Zhu *et al.* (2018) conduct a related study and extract individual home locations, mobility histories via smart card records, local housing prices of the areas travelled by individuals via real estate websites, consumer price data via business review websites, and levels of individual consumption, to develop a composite wealth index based on statistical clustering and feature selection methods such as principal component analysis. The study does not use survey data as the primary input, and is yet to comprehensively account for different types of expenditures of individuals. Furthermore, it has yet made good use of the advanced unsupervised or the semisupervised machine learning techniques, which may be more relevant for multimodal data input and more powerful for discovering any nonlinear relationship between wealth and wealth proxies.

To address these research gaps, a proxy model that follows the three principal guidelines might provide important insights into how the household wealth differs among Chinese citizens residing across different parts of a city, which may subsequently benefit intra-city EI studies.

### Implications on Data Policies for China

Our study points to the need for new data policies, as well as high quality data with high availability that meet the principles of privacy protection and ethics in China, for developing high resolution proxy for



fine-grained EI analysis and for facilitating relevant policy decision-makings in China. Our policy recommendations and the potential implications include the followings:

### **Data quality and availability**

Given that survey data serve the basis for household wealth analysis, poor survey data quality will inevitably carry negative consequences and misleading statistics (Zhang, 2018). It is therefore extremely important to ensure the quality of survey data for reliable causal inference analysis and sound decision-makings. In order to improve the availability and quality of household wealth survey data in China, data policies should champion openness, representativeness, security and accountability (Gil-Garcia et al., 2009), and reduce bias and errors (Lutes, 2015).

Although it is stated by laws that any governing statistics in China with census data obtained at or above the county level should be made publicly available in a timely manner (National Bureau of Statistics of China, 2012) it is still unclear how the regulations should be implemented; for instance, when and what type of information should be disclosed, at the fine-grained resolution (National Bureau of Statistics of China, 2002). At the moment, though existing household wealth datasets in China collected via the census method have the highest accuracy and can represent a sizeable population across a broad geographical scale, there is by no publicly available means for researchers to access these census data at the fine-grained resolution. In addition to the government-led census initiatives in China, a number of household wealth surveys have also been conducted by nonofficial institutions, yet the sampling methods in these surveys vary significantly which induces inconsistency and biases in the data (see Table 1).

Establishing online statistical data platforms to provide socio-economic information at the fine-grained resolution can provide the needed data for evidence-based socio-economic analysis and decision-making in China. Such open data can be customized and downloadable for future research studies. To improve data quality, nonofficial statistics can be an important complement to the official census data; nevertheless, it is important to set statistical standards/guidelines on data sampling and collection. An official guideline detailing the sampling methodology can enhance the representativeness of the sampled data and reduce the potential biases and improve accuracy in evidence-based decision-making. Moreover, public consultation and stakeholder engagement exercises should play a role in shaping the design and content of the data collected, which has been a practice adopted by the government in the United Kingdom to ensure that the data collected is usable to the public (Cabinet Office, 2018). In the future, China's online statistical data platforms may also host specific stakeholder meetings to ensure that the data released can meet the demands of their users.

### **Data privacy and security incorporating moral and ethical principles**

The release and use of open data for wealth proxy in developing economies may raise concerns about data privacy and ethics. In the conflict-prone regions, exposing the individuals' identities may produce grave consequences on individuals. Given the need and research significance of open household wealth data in China, data policies should promote the development of moral- and ethics-driven data system for sound evidence-based decision-makings in China.

As compared to personal privacy protection, the policy of governing moral and ethical access of data by the general public remains relatively under-developed in China. Most western countries have established privacy protection units to protect their citizens' private information (Zhang, 2018). China has also set up a new department, the Public Security Bureau, subsumed under the Ministry of Industry and Information Technology, to protect its citizens' privacy. However, privacy protection has yet become a dominant function of the department.

Establishing a specific government agency to regulate data privacy and security can facilitate implementation of relevant laws, and protect the citizens' privacy (Zhang, 2018). In addition, a national guideline outlining the moral and ethical principles for safeguarding personal data privacy and security can be developed by the government. The guideline may cover the following ethical and moral

principles: the principle of respect; the principle of informed consent; the principle of equilibrium; and the principle of social rectification (Yu, 2004). Ensuring the moral and the legal use of data can pave the way for social stability and harmony (Lü, 2005). Furthermore, the government, academia and industry can work together to facilitate the development and implementation of data security and privacy protection technologies across the public and the private sectors, to enable privacy-heightened evidence-based decision-makings. To protect survey data security, access control mechanisms that segregate the data depending on the level of data sensitivity can be introduced among institutions of data science (Conrad *et al.*, 2012). A decentralized data sharing system can be established to enable sensitive data exchange between government agencies and individuals/companies, based on blockchain distributed ledgers and smart contract technologies (Engin, 2018; Zyskind and Nathan, 2015). For wealth data proxy development that involves individual data such as mobile phone records and electricity expenditure, data privacy can be enhanced via data minimization, anonymization, and encryption (Li *et al.*, 2019). Such measures can be applied throughout the lifecycle of wealth data proxies, including data collection, transmission, storage and analysis. When processing confidential data and releasing results of personal information associated with personal health, financial or daily activities (e.g., shopping history and smart card records), a number of privacy and security techniques can be adopted to improve computational security and to regulate information disclosure (Crowcroft and Gascón, 2018). Measures can be undertaken to protect personal privacy and security, via differential privacy, homomorphic encryption, secured multi-party computation, enclave and edge cloud (for instance, the databox technology developed by Prof. Jon Crowcroft and his team of the University of Cambridge) (The Royal Society, 2019).

## Conclusion

Our study has demonstrated that household wealth may serve a more relevant predictor of differential exposure and related EI at the intra-city level than income in China. With the growing availability of fine-grained environmental data, more fine-grained household wealth-related EI studies are urgently needed. We put forward the following conclusions and recommendations to improve household wealth socio-economic EI study in fine-grained resolution in China:

1. Fine-grained household wealth-related EI studies that carry both research and policy significance, and provide better understandings of intra-city household wealth distribution are needed. Spatial resolution can greatly influence the statistical results of EI studies, and policy interventions based on intra-city analysis may reduce city-level EI in China. However, only few significant studies have been conducted at the intra-city level, thus restricting our understanding of the intra-city level air pollution induced EI.
2. The small sample size extracted from household economic surveys lack representativeness for any fine-grained socio-economic studies in China. Moreover, existing freely accessible household wealth databases do not provide the locations of Chinese households, due to the privacy protection laws in China. Therefore, current household wealth databases could hardly support fine-grained socio-economic studies in China. There is a need to develop a new representative household wealth proxy for fine-grained socio-economic and EI study in China.
3. The fine-grained wealth proxies developed from the nightlight imagery, the daytime satellite imagery, the mobile phone calls and the SMS data, the house prices and the electricity expenditure, can hardly meet our five proposed criteria, namely, spatial resolution, temporal resolution, direct relevance with household wealth, applicability to urban China, and data availability.
4. A new household wealth proxy is recommended to facilitate the fine-grained intra-city EI study in China. A hybrid household wealth proxy incorporating the daytime satellite images, the house prices, and the mobile phone data retrievable via internet-based communication tools may support

fine-grained socio-economic-related EI study, assist the formulation of proper urban policies to close the EI gap, and to improve the quality of life of citizens in China.

5. Legal, policy and technology measures for improving data quality and availability, and for improving morals and ethics, for securing privacy and security of data collection and access, are fundamentally important to ensure that in future, good quality SES data in fine-grained resolution, such as the wealth proxy data, can be fully developed and openly shared in an ethical and moral manner to facilitate sound evidence- and location-based decision-makings in China.

**Acknowledgments.** We gratefully acknowledge the insights shared by Stephanie Sun, on using the house prices as a proxy for household wealth in China, the research assistance of Shanshan Wang, and the editorial assistance of Clinton Ohlers on the Section, “Implications on Data Policies for China”.

**Funding Statement.** This research is supported in part by the Theme-based Research Scheme, Research Grants Council, under Grant No. T41-709/17-N. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests.** The authors declare no competing interests exists.

**Data Availability Statement.** Data sharing is not applicable to this article as no datasets were generated or analysed during this study.

**Author contributions.** Conceptualization, V.O.K.L., J.C.K.L.; Methodology, J.C.K.L.; Formal Analysis, J.C.K.L. and R.B.; Data Curation, R.B.; Writing—Original Draft, J.C.K.L., Y.H., R.B., J.L., Writing—Review and Editing, J.C.K.L., Y.H., J.L., and K.J.M.; Supervision, J.C.K.L. and V.O.K.L.; Funding Acquisition, J.C.K.L. and V.O.K.L.

## References

- Adler NE and Newman K** (2002) Socioeconomic disparities in health: pathways and policies. *Health Affairs* 21(2), 60–76.
- Aittomäki A, Martikainen P, Laaksonen M, Lahelma E and Rahkonen O** (2010) The associations of household wealth and income with self-rated health—A study on economic advantage in middle-aged Finnish men and women. *Social Science & Medicine* 71(5), 1018–1026.
- Baden BM, Noonan DS and Turaga RMR** (2007) Scales of justice: is there a geographic bias in environmental equity analysis? *Journal of Environmental Planning and Management* 50(2), 163–185.
- Benzeval M and Judge K** (2001) Income and health: the time dimension. *Social Science & Medicine* 52(9), 1371–1390.
- Blumenstock J, Cadamuro G and On R** (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264), 1073–1076.
- Bowen WM** (2001) *Environmental Justice Through Research-Based Decision-Making*. New York: Routledge.
- Budria S, Díaz-Giménez J, Quadri V and Rios-Rull J-V** (2002) Updated facts on the U.S. distributions of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review* 26, 2–35.
- Cabinet Office (2018) Help Shape Our Future: The 2021 Census of Population and Housing in England and Wales. Available at the UK Government <https://www.gov.uk/government/publications/the-2021-census-of-population-and-housing-in-england-and-wales>. 8 Dec 2019.
- Candia J, González MC, Wang P, Schoenharl T, Madey G and Barabási A-L** (2008) Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* 41(22), 224015.
- Cao L, Wang C and Li J** (2016) Vehicle detection from highway satellite images via transfer learning. *Information Sciences* 366, 177–187.
- Carle AC, Bauman KJ and Short K** (2009) Assessing the measurement and structure of material hardship in the United States. *Social Indicators Research* 92(1), 35.
- Chen S-T, Kuo H-I and Chen C-C** (2007) The relationship between GDP and electricity consumption in 10 Asian countries. *Energy Policy* 35(4), 2611–2621.
- Chen Z, Yang H and Feng Y** (2016) Shehui ziben, jiating shouru yu chengzhen jumin xingfu gan: Jiyu zhongguo da yangben weiguan shuju de shizheng yanjiu. [Social capital, household income and urban residents’ happiness: an empirical study based on the micro-data of a large sample in China]. *Xexue Juece* 12, 24–44.
- Cheng C** (2013) Zhongguo pinfu chayi toushi. [A perspective on the differences between rich and poor in China]. *Jiage yu Shichang* 9, 41–43.
- Cheng D** (2015) *Jingji fazhan yu huanjing wuran xianghu guanxi de kongjian tongji fenxi*. (Master’s Thesis). Hunan University, China.
- Conrad E, Misener S and Feldman J, & Riggins, K** (2012) *CISSP Study Guide* (2nd ed.). Waltham, MA: Syngress.
- Crowcroft J and Gascón A** (2018) Analytics without tears or is there a way for data to be anonymized and yet still useful? *IEEE Internet Computing* 22(3), 58–64.

- Cutter SL, Holm D and Clark L** (1996) The role of geographic scale in monitoring environmental justice. *Risk Analysis* 16(4), 517–526.
- Dargay J, Gately D and Sommer M** (2007) Vehicle ownership and income growth, worldwide: 1960–2030. *Energy Journal* 28(4), 143–170.
- Davoudi S and Brooks E** (2012) *Environmental Justice and the City: Full Report*. Available at Global Urban Research Unit, Newcastle University: <http://www.ncl.ac.uk/media/wwwnclacuk/globalurbanresearchunit/files/electronicworkingpapers/environmental-justice-full-report.pdf>. 24 Dec 2019.
- Deville P, Song C, Eagle N, Blondel VD, Barabási A-L and Wang D** (2016) Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences* 113(26), 7047–7052.
- Ekbom A and Bojö J** (1999) Poverty and the Environment: Evidence of Links and Integration Into Country Assistance Strategy Process (Discussion paper no. 4). Available at Environment Group, Africa Region, The World Bank: <https://web.worldbank.org/archive/website00672/WEB/PDF/POVERT-3.PDF>. 24 Dec 2019.
- Elvidge CD, Sutton PC, Ghosh T, Tuttle BT, Baugh KE, Bhaduri B and Bright E** (2009) A global poverty map derived from satellite data. *Computers & Geosciences* 35(8), 1652–1660.
- Engin Z** (2018) Digital Ethics: Data, Algorithms, Interactions. Available at <https://doi.org/10.5281/zenodo.1689674>. 24 Dec 2019.
- Evans GW and Kantrowitz E** (2002) Socioeconomic status and health: the potential role of environmental risk exposure. *Annual Review of Public Health* 23(1), 303–331.
- Fernández IC and Wu J** (2016) Assessing environmental inequalities in the city of Santiago (Chile) with a hierarchical multiscale approach. *Applied Geography* 74, 160–169.
- Filmer D and Pritchett LH** (2001) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography* 38(1), 115–132.
- Fisher S** (2015) The emerging geographies of climate justice. *The Geographical Journal* 181(1), 73–82.
- Frias-Martinez V and Virseda J** (2012) *On the Relationship Between Socio-Economic Factors and Cell Phone Usage*. Paper presented at the Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, Atlanta, Georgia, USA.
- Fritzell J, Neremo M and Lundberg O** (2004) The impact of income: assessing the relationship between income and health in Sweden. *Scandinavian Journal of Public Health* 32(1), 6–16.
- Gao J and Li L** (2013) Qihou bianhua dui renqun jiankang yingxiang de yanjiu jinzhan. [Research progress on the effects of climate change on the public health]. *Shanghai Yixue* 4, 39–48.
- Gebru T, Krause J, Wang Y, Chen D, Deng J and Li F-F** (2017) *Fine-Grained Car Detection for Visual Census Estimation*. Paper presented at the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA.
- Gil-Garcia JR, Chun S and Janssen M** (2009) Government information sharing and integration: combining the social and the technical. *Information Polity* 14(1, 2), 1–10.
- Giles J and Yoo K** (2007) Precautionary behavior, migrant networks, and household consumption decisions: an empirical analysis using household panel data from rural China. *The Review of Economics and Statistics*, 89(3), 534–551.
- Graesser J, Cheriyyadat A, Vatsavai RR, Chandola V, Long J and Bright E** (2012) Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(4), 1164–1176.
- Hajat A, Hsia C and O'Neill MS** (2015) Socioeconomic disparities and air pollution exposure: a global review. *Current Environmental Health Reports* 2(4), 440–450.
- Hallegatte S and Rozenberg J** (2017) Climate change through a poverty lens. *Nature Climate Change* 7(4), 250–256.
- Headey B and Wooden M** (2004) The effects of wealth and income on subjective well-being and ill-being. *Economic Record* 80, S24–S33.
- Henderson JV, Storeygard A and Weil DN** (2012) Measuring economic growth from outer space. *American Economic Review* 102(2), 994–1028.
- Huang Y-L and Batterman S** (2000) Residence location as a measure of environmental exposure: a review of air pollution epidemiology studies. *Journal of Exposure Science and Environmental Epidemiology* 10(1), 66.
- Huby M, Cinderby S, White P and de Bruin A** (2009) Measuring inequality in rural England: the effects of changing spatial resolution. *Environment and Planning A* 41(12), 3023–3037.
- Jean N, Burke M, Xie M, Davis WM, Lobell DB and Ermon S** (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301), 790–794.
- Jelinski DE and Wu J** (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology* 11(3), 129–140.
- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, Morrison J and Giovis C** (2005) A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Science and Environmental Epidemiology* 15(2), 185.
- Jin X and Davis CH** (2005) Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing* 2005(14), 745309.
- Karantzalos K and Argialas D** (2004) Towards automatic olive tree extraction from satellite imagery. In O Altan (Ed.), *Proceedings of International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences: XXth ISPRS*

- Congress, Commission III (pp. 1173-1177). Istanbul, Turkey. This publication can be found here: <https://www.isprs.org/proceedings/XXXV/congress/comm3/comm3.aspx>.
- Kolenikov S and Angeles G** (2009) Socioeconomic status measurement with discrete proxy variables: is principal component analysis a reliable answer? *Review of Income and Wealth* 55(1), 128–165.
- Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian S and Carson R** (2002) Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the Public Health Disparities Geocoding Project. *American Journal of Epidemiology* 156(5), 471–482.
- Krieger N, Williams DR and Moss NE** (1997) Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annual Review of Public Health* 18(1), 341–378.
- Lee C** (2019) Daily Active Users for WeChat Exceeds 1 Billion. Available at ZDNet website: <https://www.zdnet.com/article/daily-active-user-of-messaging-app-wechat-exceeds-1-billion/>. 7 Dec 2019.
- Li H, Yu L and He W** (2019) The impact of GDPR on global technology development. *Journal of Global Information Technology Management* 22(1), 1–6.
- Li R, Xiong H and Zhao H** (2015) *More Than Address: Pre-Identify Your Income with the Open Data*. Paper presented at the 2015 International Conference on Cloud Computing and Big Data (CCBD), Shanghai, China.
- Li VOK, Han Y, Lam JCK, Zhu Y and Bacon-Shone J** (2018) Air pollution and environmental injustice: Are the socially deprived exposed to more PM2.5 pollution in Hong Kong? *Environmental Science & Policy* 80, 53–61.
- Li VOK, Lam JCK, Chen Y and Gu J** (2017) *Deep Learning Model to Estimate Air Pollution Using M-Bp to Fill in Missing Proxy Urban Data*. Paper presented at the 2017 IEEE Global Communications Conference, Singapore.
- Longhurst J, Irwin J, Chatterton T, Hayes ET, Leksmono N and Symons J** (2009) The development of effects-based air quality management regimes. *Atmospheric Environment* 43(1), 64–78.
- Lü Y-H** (2005) Privacy and data privacy issues in contemporary China. *Ethics and Information Technology* 7(1), 7–15.
- Lucas K, Walker G, Eames M, Fay H and Poustie M** (2004) Environment and Social Justice: Rapid Research and Evidence Review. Available at Sustainable Development Research Network: <http://www.psi.org.uk/sdrm/SDRN%20environment%20and%20social%20justice%20report%20-%20revised.pdf>. 24 Dec 2019.
- Luo S, Morone F, Sarraute C, Travizano M and Makse HA** (2017) Inferring personal economic status from social network location. *Nature Communications* 8, 15227.
- Lutes T** (2015) Data-Driven Government: Challenges and a Path Forward. Available at [https://www.ibm.com/downloads/cas/BJE8Z48J?mhq=data%20driven%20government&mhsrc=ibmsearch\\_a](https://www.ibm.com/downloads/cas/BJE8Z48J?mhq=data%20driven%20government&mhsrc=ibmsearch_a). 24 Dec 2019.
- MacKenzie CA, Baird TD and Hartner J** (2014) Use of single large or several small policies as strategies to manage people–park interactions. *Conservation Biology* 28(6), 1645–1656.
- Marlier E. and Atkinson AB** (2010) Indicators of poverty and social exclusion in a global context. *Journal of Policy Analysis and Management* 29(2), 285–304.
- Mason R** (2015) Patterns and Consequences of Economic Engagement Across Sub-Saharan Africa: A Comparative Analysis of Chinese, British and Turkish Policies (LSE Working Paper CIS/2015/01). Available at Centre for International Studies, London School of Economics: <http://www.lse.ac.uk/international-relations/assets/documents/cis/working-papers/cis-working-paper-2015-01-mason.pdf>. 24 Dec 2019.
- McKenzie DJ** (2005) Measuring inequality with asset indicators. *Journal of Population Economics* 18(2), 229–260.
- Mellander C, Lobo J, Stolarick K and Matheson Z** (2015) Night-time light data: a good proxy measure for economic activity? *PloS One* 10(10), e0139779.
- Mendes De Leon CF, Rapp SS and Kasl SV** (1994) Financial strain and symptoms of depression in a community sample of elderly men and women: a longitudinal study. *Journal of Aging and Health* 6(4), 448–468.
- Mennis J** (2002) Using geographic information systems to create and analyze statistical surfaces of population and risk for environmental justice analysis. *Social Science Quarterly* 83(1), 281–297.
- Miao Q, Chen D, Buzzelli M and Aronson KJ** (2015) Environmental equity research: review with focus on outdoor air pollution research methods and analytic tools. *Archives of Environmental & Occupational Health* 70(1), 47–55.
- Mohai P, Pellow D and Roberts JT** (2009) Environmental justice. *Annual Review of Environment and Resources* 34, 405–430.
- Morello-Frosch R, Pastor M and Sadd J** (2001) Environmental justice and Southern California’s “riskycape” the distribution of air toxics exposures and health risks among diverse communities. *Urban Affairs Review* 36(4), 551–578.
- National Bureau of Statistics of China (2002) Detailed Rules for Implementation of Statistics Law of People’s Republic of China. Available at [http://www.stats.gov.cn/english/LF/SR/200204/t20020402\\_27179.html](http://www.stats.gov.cn/english/LF/SR/200204/t20020402_27179.html). 10 Dec 2019.
- National Bureau of Statistics of China (2012) Statistics Law of the People’s Republic of China. Available at [http://www.stats.gov.cn/english/LF/SL/201209/t20120921\\_27177.html](http://www.stats.gov.cn/english/LF/SL/201209/t20120921_27177.html). 10 Dec 2019.
- OECD (2016) Economic Consequences of Outdoor Air Pollution. Available at <https://www.oecd.org/environment/indicators-modelling-outlooks/Policy-Highlights-Economic-consequences-of-outdoor-air-pollution-web.pdf>. 24 Dec 2019.
- Olivia S, Gibson J, Rozelle S, Huang J and Deng X** (2011) Mapping poverty in rural China: how much does the environment matter? *Environment and Development Economics* 16(2), 129–153.
- Park J, Bangalore M, Hallegatte S and Sandhoefner E** (2018) Households and heat stress: estimating the distributional consequences of climate change. *Environment and Development Economics* 23(3), 349–368.
- Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A** (2015) Epidemic processes in complex networks. *Reviews of Modern Physics* 87(3), 925.

- Pratt G, Vadali M, Kvale D and Ellickson K** (2015) Traffic, air pollution, minority and socio-economic status: addressing inequities in exposure and risk. *International Journal of Environmental Research and Public Health* 12(5), 5355–5372.
- Polido L** (1996) A critical review of the methodology of environmental racism research. *Antipode* 28(2), 142–159.
- Quadrini V and Rios-Rull J-V** (1997) Dimensions of inequality: facts on the US distribution of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review* 21(2), 3–21.
- Rahkonen O, Arber S, Lahelma E, Martikainen P and Silventoinen K** (2000) Understanding income inequalities in health among men and women in Britain and Finland. *International Journal of Health Services* 30(1), 27–47.
- Saeidi M, Venerandi A, Capra L and Riedel S** (2017) Community Question Answering Platforms vs. Twitter for Predicting Characteristics of Urban Neighbourhoods. *arXiv preprint arXiv: 04653*.
- Salah AA, Lepri B, Pianesi F and Pentland AS** (2011) *Human Behavior Understanding for Inducing Behavioral Change: Application Perspectives*. Paper presented at the International Workshop on Human Behavior Understanding, Amsterdam, The Netherlands.
- Šćepanović S, Mishkovski I, Hui P, Nurminen JK and Ylä-Jääski A** (2015) Mobile phone call data as a regional socio-economic proxy indicator. *PLoS One* 10(4), e0124160.
- Schlosberg D** (2004) Reconciling environmental justice: global movements and political theories. *Environmental Politics* 13(3), 517–540.
- Seidel M-DL, Polzer JT and Stewart KJ** (2000) Friends in high places: the effects of social networks on discrimination in salary negotiations. *Administrative Science Quarterly* 45(1), 1–24.
- Sharath R, Nirupam K, Sowmya B and Srinivasa K** (2016) *Data Analytics to Predict the Income and Economic Hierarchy on Census Data*. Paper presented at the 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, Karnataka, India.
- Sun C** (2015) *Estimating household income with proxy data*. Unpublished manuscript, The University of Hong Kong, Hong Kong.
- Tan PY and Samsudin R** (2017) Effects of spatial scale on assessment of spatial equity of urban park provision. *Landscape and Urban Planning* 158, 139–154.
- Taubenbock H, Wurm M, Setiadi N, Gebert N, Roth A, Strunz G, Birkmann, J, Dech S** (2009) *Integrating Remote Sensing and Social Science*. Paper presented at the 2009 Joint Urban Remote Sensing Event, Shanghai, China.
- The Royal Society (2019) Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis. Available at <https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf>. 8 Dec 2019.
- Toole JL, Lin Y-R, Muehlegger E, Shoag D, González MC and Lazer D** (2015) Tracking employment shocks using mobile phone data. *Journal of The Royal Society Interface* 12(107), 20150185.
- Tucker-Seeley RD, Harley AE, Stoddard AM and Sorensen GG** (2013) Financial hardship and self-rated health among low-income housing residents. *Health Education & Behavior* 40(4), 442–448.
- Venerandi A, Quattrone G, Capra L, Quercia D and Saez-Trumper D** (2015) *Measuring Urban Deprivation From User Generated Content*. Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, Canada.
- Wan C and Su S** (2017) China's social deprivation: measurement, spatiotemporal pattern and urban applications. *Habitat International* 62, 22–42.
- Wang X, Ling H, Huang J, Fan Z, Wang Y and Yong Y** (2012) Quanqiu qihou bianhua yu hekou chengshi qihou cuiruo xing shengtai quhua yanjiu – yi Shanghai weilì. [Global climate change and ecological regions planning based on the climatic fragility of estuarine cities – a case study of Shanghai]. *Shanghai Chengshi Guihua* 6, 1–6.
- Wang X, Qi Y, Wang Z, Guo H and Yu T** (2007) Zaocheng Beijing PM10 zhong wuran de erlei dianxing tianqi xingshi. [Two typical weather conditions causing heavy PM10 pollution in Beijing]. *Qihou yu Huanjing Yanjiu* 12(1), 81–86.
- Watmough GR, Atkinson PM, Saikia A and Hutton CW** (2016) Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. *World Development* 78, 188–203.
- Weidmann NB and Schutte S** (2017) Using night light emissions for the prediction of local wealth. *Journal of Peace Research* 54(2), 125–140.
- Winsemius H, Jongman B, Veldkamp T, Hallegatte S, Bangalore M and Ward P** (2015) *Disaster Risk, Climate Change, and Poverty: Assessing the Global Exposure of Poor People to Floods and Droughts*, 7480, Policy Research Working Paper. Washington, DC: World Bank Group.
- World Bank (2019) Population, Total – China. Available at <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=CN>. 7 Dec 2019.
- Wu, J.** (2007) Scale and scaling: a cross-disciplinary perspective. In *Key Topics in Landscape Ecology*. Cambridge University Press, pp. 115–142.
- Xie Y and Jin Y** (2015) Household wealth in China. *Chinese Sociological Review* 47(3), 203–229.
- Yu Y** (2004) Wangluo yinsi quan baohu lunli chutan. [On the ethics of protection of right to privacy in networks]. *Guangxi Social Sciences* 7, 52–53.
- Zhang C and Cao M** (2012) Jiating shouru, richang xiaofei zhichu yu yongdian liang de guanlian guanxi. [The relationship between household income, daily consumption expenditure and electricity usage]. *Nengyuan jishu jingji* 24(7), 44–48.
- Zhang D** (2018) *Big Data Security and Privacy Protection*. Paper presented at the 8th International Conference on Management and Computer Science (ICMCS 2018), Shenyang, China.

- Zhao T, Huang H, Yao X and Fu X** (2019) Predicting individual socioeconomic status from mobile phone data: a semi-supervised hypergraph-based factor graph approach. *International Journal of Data Science and Analytics* 1–12. <https://doi.org/10.1007/s41060-019-00195-z>
- Zhu D and Chen X** (2017) Fangchan jiage, jiating caifu zai fenpei yu huobi zhengce youxiao xing – jiyu dongtai suiiji yiban junheng moxing de fenxi. [Housing prices, redistribution of household wealth and effectiveness of monetary policy – an analysis based on dynamic stochastic general equilibrium model]. *Nanfang Jinrong* 5, 18–36.
- Zhu Y, Chen F, Li M and Wang Z** (2018) Inferring the economic attributes of urban rail transit passengers based on individual mobility using multisource data. *Sustainability* 10(11), 4178.
- Zyskind G and Nathan O** (2015) *Decentralizing Privacy: Using Blockchain to Protect Personal Data*. Paper presented at the 2015 IEEE Security and Privacy Workshops, San Jose, California, USA.