NOT FOR CITATION
without permission of authors

Hidden Informality in Urban China:
A Mixed Method Approach to Shanghai's Market for Overcrowded Housing

by Julia G. Harten, Annette M. Kim, J. Cressica Brazier.

A Working Paper for the USC Lusk Center for Real Estate
January 2018

The abstract:
Currently urbanizing areas have refocused attention on densification and the struggle for space in the city. As people continue to migrate citywards for economic opportunity and housing markets are slow to catch up, informal markets emerge to cater to these unmet demands. However, the extent of contemporary informality in global cities may be concealed: informal housing may increasingly be hidden inside formal structures.

This paper presents evidence uncovering one such hidden housing informality. Drawing on remotely collected, large-n online advertisement data as well as findings from qualitative and quantitative field work, we document the market for bed spaces or 群租房 („group rentals") in Shanghai; a housing sub-market in which some formal commercial and residential units are illegally converted into extremely crowded dormitories. The ability to compare and contrast what is advertised online versus what is happening on the ground allows us to study (i) the targeted demographic of renters, (ii) the revealed preferences of the same, and (iii) the veracity of remotely collected online data.

We find an informal housing market catering to mostly young, educated migrants. Affordability is accrued by crowding on average 24 people into 3 bedroom apartments. The evidence presented is an illustration of the extreme housing shortage in today's urban centers. It is also evidences the importance of groundtruthing when using user generated content as we find considerable discrepancies between the online advertisement and the field work data.

1. Introduction:

> *"The example of China is like no other. It is possibly the only large country that has managed, so far, to urbanize rapidly without the creation of large slum areas or informal settlements." (UN-Habitat 2003: p. 126).*

By all accounts, China has urbanized at unprecedented scales and rates in human history (Zhang & Song 2003). Through this major transformation from a rural nation to an industrialized, urban one, China has accomplished rapid economic growth, lifting millions of people out of poverty (Ravallion et al. 2007, World Bank and State Council 2014). By some accounts, as in the UN quote above, China has been able to achieve this historic demographic and economic transformation without the usual accompanying dark side. Typically, when millions of people migrate to cities for economic opportunity, the under-capitalized migrants seek shelter in "slums" as has been the case in the western world at the turn of the 20th century and in many of today's global south cities. Formal real estate markets do not meet their housing demands and so informal markets emerge to supply residential spaces in what are initially undervalued areas of the city. These unregulated markets often provide low quality, unsafe, and unhealthy living environments.

Despite the seeming lack of "squatter settlements" in China's planned cities, there are now reports suggesting that informal housing does exist – but in surprising, less openly obvious locations (Lai & Ho 2001, Tanasescu et al. 2010, Huang & Yi 2015, Kim 2016). Migrants have been eking out space underground, on rooftops, in interstitial spaces, and behind closed doors. The extent of contemporary informality in global cities may be hidden.

This article focuses on one such housing sub-market, the phenomenon of 群租房 (qunzu fang), translated "group rental housing." The term loosely describes living arrangements in which ordinary units are turned into overcrowded dormitories filled with bunk beds by landlords that charge rent by individual bed. Such "group rentals" arrangements entail illegal crowding levels, with beds often also being placed even in the bathroom, kitchen, and storage areas in an effort to maximize rent extraction of urban space (Wang 2007, Li 2009).

Essentially, "group rental housing" is the market for a bed in an apartment full of beds rented to strangers. While the phenomenon is widely discussed in Chinese popular media, such as various news outlets, online magazines, cartoons and blogs (see above), academic literature on the topic is sparse. With the exception of one recent qualitative account (Shen 2015), English language academic sources barely acknowledge the existence of the phenomenon (e.g. Timberlake et al. 2014, Yang et al. 2015). They are related to the phenomenon of Chinese urban villages (城中村 chengzhongcun) which are the sizable informal housing markets that house rural-to-urban migrant workers generally on the periphery of cities (e.g., Zhang et al. 2003, Wang et al. 2010, Zhang 2011). However, the group rental housing market units are interspersed within formal developments and more often located in the central city. Scholars are still grappling with understanding the housing situation of migrants in China's extremely rapid urbanization (e.g. Li & Wu 2008, Stephens 2010, Huang & Tao 2014). This is an evolving and dynamic phenomenon in the midst of ongoing migrant housing demolition and redevelopment efforts (Wang et al. 2012, Wu et al. 2013), and the diminishing significance of employer provided migrant housing (Wu 2002, Liao & Wong 2015). Therefore, analysis of the little studied group rental housing sub-market is needed in order to better understand the pattern of migrant housing preferences.

Against this backdrop, group rental housing has emerged as a significant source of migrant housing in some major Chinese cities, such as Shanghai**.** In order to gain a sense of the scale of this market, we counted the number of daily internet ads in the group rental market in China's five largest cities. We searched for the term "床位出租" (chuangwei chu zu or "bed space") on the popular website ganji.com, China's largest classified ads website which happens to be organized by major cities. Table 1 below shows the tabulation of ads we found in one day in June 2017, in conjunction with a column showing the population size in the city's urban core. First we can see that the group rental housing market exists in different cities across China. The table also concurs with Li and Duda (2010) who argue that in bigger cities, which rely more on the service sector than on heavy industries, group rentals are the main alternative to employer provided housing. Shanghai is one of the biggest and wealthiest cities in China (Wu 2000) with an emergent private rental market, but very few urban villages for migrants (Wang et al. 212).

Table 1: Population data and number of daily online advertisements for cheap shared housing in top most searched Chinese cities

City	Number of Daily Advertisements	Urban Population in mill (2010 Census)
Shanghai	228,738	24.5
Beijing	216,388	21.5
Guangzhou	10,356	20.8
Tianjin	21,891	15.5
Shenzhen	22,545	12.4
Wuhan	53,328	10.7
Hangzhou	23,737	9.0
Nanjing	54,027	8.2

<u>Source:</u> *Table based on population data as compiled from the National Bureau of Statistic of China and evaluated by Chan (2007) and author collected online advertisements in June 2017 on gangji.com*.

In the midst of this phenomenon, the Chinese government policies have been trying to improve housing conditions. They have steadily decreased the numbers allowed for maximum housing occupancy in official regulatory standards for housing over the years (Zhang and Chen 2014). In the past, policies did not address the housing needs of hundreds of millions of migrants to urban centers who constituted a significant portion of the labor pool but who did not possess the "hukou" residency document that would entitle them to public services such as housing, education, and healthcare. However, more recently, policies have been acknowledging the migrant population and their housing needs and new programs have begun to increase the production of affordable public housing, including rental housing. However, these affordable housing programs target the college educated in key industries such as IT. Furthermore, the rent levels are still high relative to income levels and the supply is still small relative to the population numbers. Meanwhile, many of the informal housing markets that often house migrants (urban villages, underground housing, and group rentals) have been the target of government crackdowns and demolitions. In fact, starting in July 2017, the term for “group rental” was no longer allowed to be used on the online advertisement platforms.

China’s transition towards a market economy has profoundly impacted the socio-spatial make-up of the Chinese city. However, the continued importance of pre-reform institution of *hukou* has yielded sub-par housing outcomes for migrants, whose housing needs have until recently been ignored by official policymaking. Government funded surveys and Chinese scholars have started to document the housing conditions and rent levels of Chinese migrants by labor categories. The literature on migrant housing issues has primarily focused on the urban villages that have developed on the periphery of major cities (Liu, He, Wu, & Webster, 2010; Song, Zenou, & Ding, 2008; Wang, Wang, & Wu, 2009; Zheng, Long, Fan, & Gu, 2009). However, scholars examining government census figures have deduced that roughly one-half of migrants in cities like Beijing actually live in the urban center, approximately 4.3 million people (Liu, Wang, Cai, & He, 2013).

This paper aims to address the gap in knowledge about the revealed preferences of inner-city migrant housing consumption. Not many studies have examined the demonstrated tradeoffs migrants make in terms of price, location, and housing conditions. The patchwork of studies in this literature suggest that these choices are dependent on lifecycle issues of migrants, their education and economic prospects, and gender. In order to analyze central city migrant housing, we focus on group rental housing market in Shanghai.

Specifically, our research questions are:

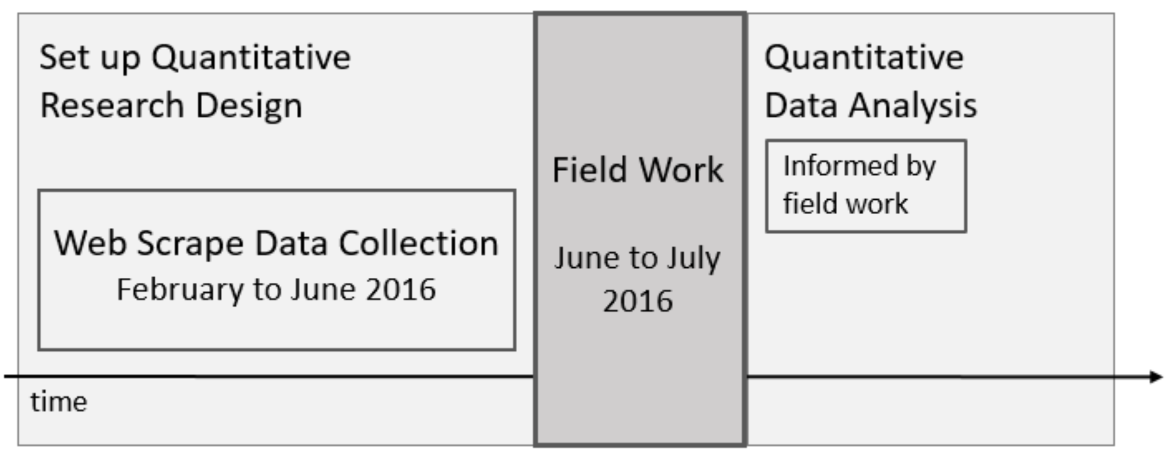
1) What kind of demographic uses group rental housing/is this housing being marketed to? Past studies found differences by education level, income level, age, etc.
2) What is the niche this sub-market is providing in terms of rent, space, amenities, location value?
    a. In particular, unique to group rental housing and our data source is that the room’s crowding levels are often advertised. So, we can find how the market prices extreme crowding levels at a micro-level, the price for a bed dependent on the number of people sharing the room or apartment.
3) Does gender play a significant role in this housing market? We can ask this because the group rental housing market ads specify if the rooms are female only, male only, or if the apartment contains a mix of female and male rooms. This could potentially provide the rare opportunity to see if there is rent discrimination based on gender.
4) What is the discrepancy between scraped online market data and field survey data about market conditions and living conditions? This issue is described below.

2. Data and Research Design

As others have noted before us, studying informality poses additional difficulties as data collection is complicated due to illegal status of informal activities and vulnerability of subjects involved. For the study of the market of group rentals in Shanghai, we employed a mixed method approach and draw on several different sources for data collection. A web-crawl algorithm set up to collect online advertisements for group rentals yielded a large set of web-scraped data points. Additionally, survey data collected in the field allows me to compare and contrast advertised arrangements with their real life counterparts. The following discusses the data collection processes**.**

We were fortunate to happen to scrape online market data before the government shut down the use of the term for "group rental" (subsequently, alternative terms are being used). While using web-scraped data is being used more often (Kim, 2016), these need to be approached critically and with groundtruth verification (Acolin and Kim, forthcoming). Indeed, when we went to Shanghai to verify the online ads, we found that the real estate brokers invariably asked to meet us on the street and took us to a completely different housing situations, in terms of location, price, crowding levels, and amenities. It appeared that the internet ads were being used to locate potential renters. So, we quickly assembled a team to gather original market data through field survey. Therefore, this paper also allows us to measure the discrepancy between scraped market data and field market data addressing research methods issues.

Figure 1: Embedded Mixed Method Research Design



Online Advertisement Data

Using the digital traces of online markets for research data is particularly appropriate in urban China. Mobile internet penetration in urban China is pervasive and widely affordable (CNNIC, 2017). Turning to the marketplace itself through the collection of information from housing advertisements is revelatory for analyzing market operations. Particularly for informal housing markets, this open and extensive source of data can be the only possibility (Kim 2004, 2007, 2016). While advertisements do not reveal the ultimate transaction price for housing sales, this should not be debilitating unless the negotiated price is correlated with the error term. In the case of this study's focus on rental housing, the differential between advertised price and rented price should be even more negligible given the housing shortage and the low bargaining power of migrants.

We compared several online advertising websites popular in China to check whether there were differences between them in what kinds of group rentals were advertised and found no significant differences with the same ads populating different sites. We selected the website 58.com for our web scraping because advertisers are required to list whether the apartment is open to one sex only (female or male tenants only), or would allow co-ed living situations. As the literature increasingly discusses the different migration experiences based on gender (e.g. Meng & Zhang 2001, Gaetano & Jacka 2004), we sought to study gender's potential relevance for this housing submarket.

The pilot run in January and February 2016 yielded a test sample of scraped online advertisement data. Descriptive statistics and hedonic regression results based on this initial test sample suggested that there is indeed a (informal) market for beds in group rentals and that the market rationally prices crowding. Initial results suggested that advertisements featuring beds in more crowded rooms also show cheaper rents. Consequently, we proceeded with setting up the web crawl algorithm to consistently collect date from the online advertisements for the following 5 months.

Largely following the methodology developed in Kim (2016), we set up a Python web scrape script to regularly collect data in roughly 10 day intervals. Between February and June 2016,[1] the web crawl algorithm would extract and store the entire set of online advertisements into a database of variables with an initial collection of 33,084 ads. We developed conservative criteria for deleting potentially duplicate advertisements as well as ads that were missing key variables. Because of our research question to price housing density levels, the largest number of ads we deleted (19,040 ads) did not specify how many people were sharing a room. In the end, we retained 3450 unique advertisements (roughly 10%). Another 300 advertisements were excluded because retrieved XY–coordinates mapped too far outside Shanghai city boundaries.

Fieldwork Market Data

In addition to mining this remotely collected data, we invested time in fieldwork. This is imperative since given the lack of scholarly familiarity with the group rental market, we may be missing key variables or misinterpreting the ads (Pickles, 1995). Fieldwork is also generally important because there may be contextual differences with housing markets in Shanghai compared to other cities where the same terms in the ad may mean different things (Kim 2007). Recently, an important issue has been raised by data scientists about the lack of critical thinking

[1] During the scrape period 58.com had a subsection of listings in the 合租 (hezu, "co-living") category, called "床位出租" (chuangwei chuzu, "bed space for rent"), which made it easy to isolate the group rental advertisements. In July 2016, the category was removed and the search term became illegal. Searching 床位出租 now redirects to: http://sh.58.com/sou/?badword=床位出租.

when researchers apply their own professional, race/class, nationality assumptions in interpreting remotely detected data, essentially embedding bias into algorithms (Noble 2018; Sandvig et al 2014; Tucker 2017).

Initially intended to field verify accuracy of data collected online, we planned a field trip to Shanghai in June and July of 2016. In order to test how the data collected online compares to the actual situation in Shanghai, we started responding to randomly selected online advertisements and booking appointments to see the apartments. Rather than meet at the apartment, the real estate brokers would ask to meet us in the streets or at transit stops and then take us to rental. Clearly, this was an informal market whose landlords and brokers were operating surreptitiously. Those initial site visits also made obvious that actual group rentals were different from the picture painted by the online advertisements, including the key variables of location and rent. As it become clear that the actual market situation was different than the remotely collected data suggested, we embarked on collecting our own market data through surveying.

Because of Shanghai's polycentric urbanization pattern, we first investigated where to choose the location of sample observations. Whereas the city government announces official economic centers, we also visited activity centers that were identified by a study based on geocoded mobile phone signaling data (Ding et al. 2016) from Shanghai's biggest cell phone carrier Shanghai Mobile for five consecutive working days in 2011 and covering approximately 12.4 million users. This information combined with our mapping of "hot spots" from our initial pilot scraping of data informed focusing our field survey group rental housing market data collection on 11 locations.

To carry out the survey we recruited a research team of six, all of which are familiar with the study area and context as well as fluent in both English and Chinese. The team consisted of five Chinese graduate students (two males, three females) and one American (female) doctoral student researcher. During the time from June 18th – July 1st 2016 the research groups visited roughly 190 group rental apartments. After one training day, the surveyors were split up in teams of two such that one person could be the primary interviewer and the second person could take detailed notes and pictures about the living conditions.

Each morning, the entire team would meet at one of the locations as identified by the clusters and call online advisements that would fall within the respective loosely defined radius. It turned out to be necessary to do so the same day, as advertisements would frequently disappear or become unavailable within hours. Importantly, location and gender restrictions needed to be verified over the phone and actual site visits were distributed across the small groups accordingly. At the end of each surveying day, the entire team would meet to debrief and consolidate findings (Survey questions listed in Appendix 1). One thing we discovered is that sometimes the same broker would respond to several different advertisements and so we made sure that we were not duplicating observations. It appeared that there was a super structure of either landlords owning multiple group rentals and/or a localized brokerage business but this question is beyond the scope of our study.

In total, 132 online advertisements led to roughly 190 interviews and site visits. The number of actual site visits exceeds the number of online advertisements used because frequently brokers/ landlords would show additional vacancies in other apartments upon request. In this way, responding to one advertisement could lead to up to six distinct site visits with the same or related brokers/ landlords. The survey resulted in 177 observations that met all our data requirements. Because we recorded both online advertisements and corresponding actual site visit survey information, this data collection process allows us to match one-to-one actual group rentals versus advertised group rentals.

Three Comparative Data Sets
In summary, the mixed methods data collection process yielded three distinct data sets:

(1) Field Survey data collected during field work ("real data"), 177 apartments
(2) Matched online advertisements that were the basis for the field work, 133 apartments
(3) Scraped online advertisement data, 3144 apartments

Essentially, this group rental market is a market for a bed rather than an apartment. In order to be able to investigate the relationship between crowding and price in this housing submarket, for each data set we created additional observations to uniquely pair room level crowding with a reported rent price level. For field survey data, assistants had recorded the bed rental price which varied according to how crowded the room was (2-12 people per room) as well as whether the bed was on the top bunk or bottom bunk (consistently more expensive). For the online advertisement data, on the other hand, rent and crowding levels were not always explicitly identified as pairs in the advertisement text. In these cases, we threw out any observations that did not mention crowding levels. For those that mentioned a range of rents and a range of crowding levels, we went with a conservative matching process informed by our field work and matched the lowest rent price with the highest level of crowding and the highest rent with the lowest level of crowding, deleting any values in between. Therefore, the number of our observations are:

(1) Field survey data, [n=747]
(2) Matched online data [n=194]
(3) Scraped data [n=4209]

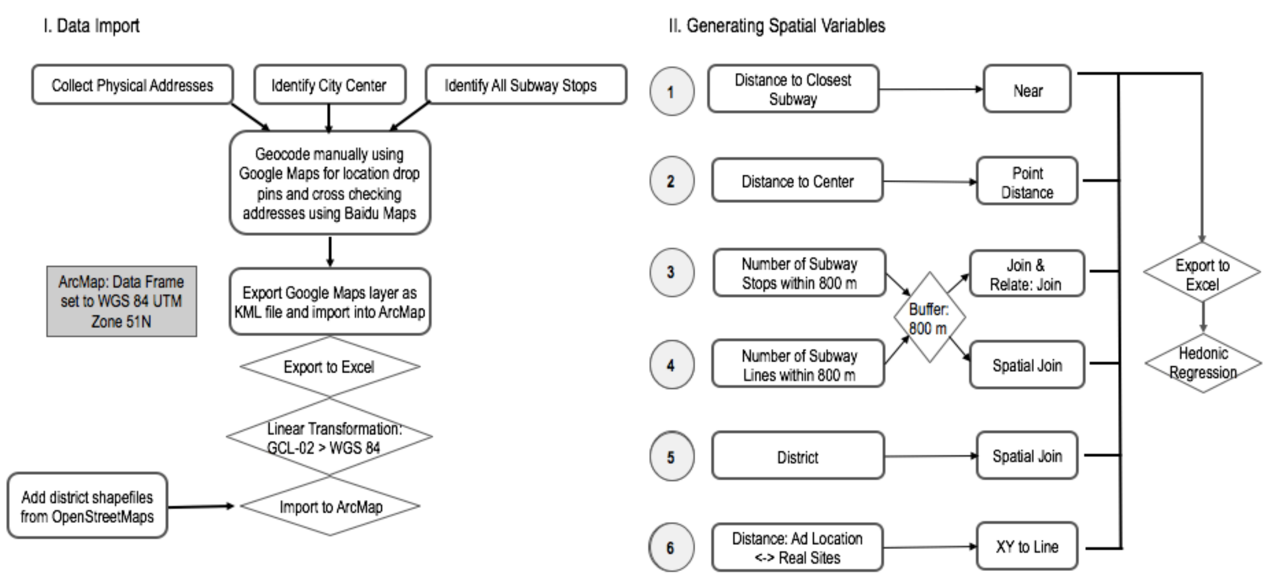
Methodology
In our analysis we apply a hedonic model to the market for bed spaces in Shanghai's group rentals. In order to be able to take into account both physical and locational characteristics, we generate spatial variables using ArcMap 10.4.
(more on use of hedonics in China)

The web scrape algorithm directly extracted latitude and longitude from each online advertisement, such that each advertisement observation was readily geocoded. Interestingly, however, extracted XY-coordinates are based on a geographic coordinate system that is not available in ArcMap. The geodetic datum formulated by the Chinese State Bureau of Surveying and Mapping (GCJ-02) is based on WGS-84 and uses an obfuscating algorithm leading to erroneous coordinate shifts of 100 to 700 meters. The geographic information science community has been able to figure out a linear transformation that – if applied to XY-coordinates – corrects for the offset in Shanghai.[2] We apply this linear transformation for extracted coordinates before importing them into ArcGIS labeled as WGS-84.

Site visit observations and corresponding online advertisements were geocoded manually using mobile phone markers in google maps, cross-referenced with baidu maps; the Chinese counterpart. The same linear transformation of extracted XY-coordinates was necessary before processing the data in ArcGIS.

Once imported, for each observation we calculate the distance to the center[3], distance to closest subway stop[4], the number of subway stops and lines accessible within a 800 meters radius, and the associated district. Additionally, we calculated the distance between the real site visits and the matched online advertisements used as basis for the field work. The detailed workflow in ArcGIS is depicted schematically in Figure 2 below.

Figure 2: Using ArcMap to Create Spatial Variables



Source: Author's own illustration.

[2] https://gis.stackexchange.com/questions/141542/what-causes-the-gps-offset-shift-in-china
[3] After testing multiple possible urban centers, we concurred with the location used in previous hedonic models of Shanghai: the municipal government building at People's Square (Wang & Huang 2007, Chen & Hao 2008, 2010).
[4] We found transit layers provided by OpenStreetMap to be outdated. Hence, we manually created a layer file with all the subway stops of Shanghai's transit network.

3. <u>Results</u>

Mapping

In a first step, we mapped geocoded online advertisement observations and real site visit data points for visual inspection (see figure 3).

Because we had selected online advertisements to respond to for site visits based on the clusters as produced by the initial sample set, the spatial distributions of the survey data observations and the web scraped data are similar. As can be observed from Figure 3, the majority of the observations are tightly distributed in a ring around the historic core. This distribution is consistent with Shanghai's urban spatial structure: most of the historic core is either commercial, old-style public rental housing and some luxury developments. Spatial analysis of the different housing types and spatial segregation based on housing tenure as conducted by Li and Wu (2008) using census data supports our findings. It is important to notice that the observations are highly clustered. In mapping the large-n web scrape data, it turns out that the online advertisements are concentrated in 730 unique locations. This is in line with our field work: brokers/ landlords would frequently show several different apartments on different levels of the same high-rises, or in different buildings of the same gated community.

Figure 3: Mapped Real Fieldwork and Web-Scraped Data

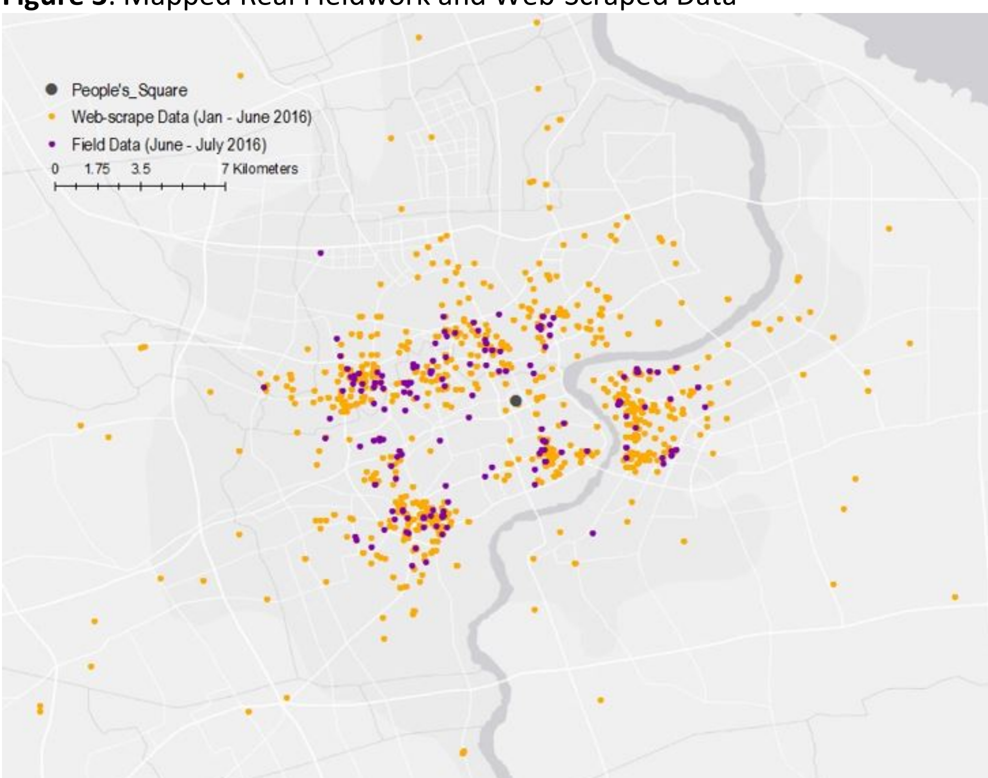
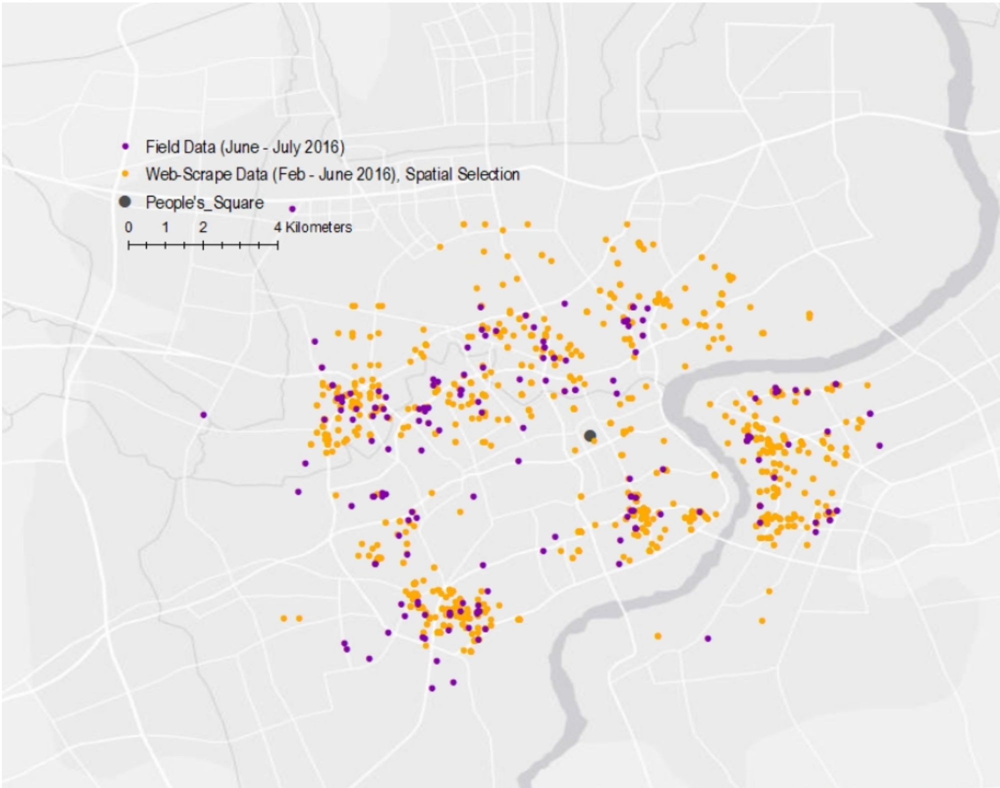


Figure 4: Mapped Real Fieldwork and Web-Scraped Data (Spatial Selection, n=3147)



As alluded to above, due to Shanghai's size we may likely to encounter spatial non-stationarity. Therefore, we still split the large-n web-scraped "fake data" by distance to CBD and look at differences in descriptive statistics and hedonic regression results using OLS estimation at various intervals. As it turns out, splitting the sample at 6.5 kilometers makes a difference: postings advertising group rentals at a distance farther than 6.5 km from the city center are considerably smaller, more crowded and advertise fewer bedrooms and bathrooms. Hence, we confine our analysis to the spatially restricted sample, which results in the loss of 1062 (out of 4209) observations. The geographically restricted data set is depicted in figure 4. In a subsequent step we also test for spatial autocorrelation using Moran's I and work with the whole data set employing GWR estimation to allow for coefficients to vary spatially.

In the following we present descriptive statistics, as well as hedonic regression results estimated using OLS and GWR drawing on our three distinct data sets. The discussion is structured to answer the research questions laid out in the introduction of this paper.

Research Question 1: What Is the Targeted Population?

In addition to quantitative data inputs for our hedonic price model, we found that online advertisements frequently feature rich descriptive text. In particular, the ads often state the kind of renter that they are seeking. In our scraping algorithm we made sure to capture and store the entire descriptive text for each advertisement, allowing us to search for and count the terms we found most commonly used. Table 2 presents a frequency table of the most the terms used most often. Additionally, surveyors collecting the market data would ask brokers and/or landlords about the kind of person that typically rents from them and whether there were any restrictions.

Table 2: Desired Tenant Traits Descriptive Statistics (Online Advertisements, n=3144)

Desired Tenant Traits	Chinese Term	Frequency[5]
Looking for a job	求组, 找工作	2041
Currently employed/ in training	上班, 工作，员工，培训	2405
White collar worker	白领	607
Student/ Graduate	学生, 学习, 毕业	2193
"Young"	年轻人, 青年人	781
Age restriction (<30-35)	岁	564
ID	身份证	878
"Newcomer"	刚来上海	175
"Obedient"	服从	360
Unwanted traits	酗酒, 赌博, 打闹,上夜班	987

As Table 2 illustrates, online advertisements are targeted at students and recent graduates (70%) that are either looking for a job (65%) or are already employed (76%). Advertisers are "picky": roughly one third mentions traits that are unwanted such as drunkenness, gambling, nosiness and night worker, white collar workers are preferred in roughly 20% of the ads, and more than 10% even explicitly mention "being obedient".

We found that these expressed "desirable" tenant traits were actually in line with our fieldwork experience. Many of the tenants were indeed recent graduates, mostly coming from smaller cities and lesser known university and at the beginning of their careers. Interestingly, brokers/ landlords expressed a much stronger preference for young renters in our interviews than the online advertisements would suggest. We found that real apartments have an age restriction in more than half of the cases (57%), while advertisements more often use language that implies they are seeking a younger clientele ("young", "student", "recent graduate") rather than explicitly mention any age restriction (6%). Additionally, several brokers/ landlords voiced a preference for male renters because they were easier to manage. These preferences also bear out in the descriptive statistics for our large-n web-scraped data, where more male beds are advertised in this market as is discussed in more detail below.

Beyond allowing us to understand the target population from the advertisers' point of view, we also interpret the advertisers' ability to be rather specific about desired and undesired tenants as evidence for a sellers' market. If market conditions are such that the good in demand is scarce,

---

[5] We created dummy variable if an ad text mentions any one of the terms specified. Hence, n here is equivalent to the number of online advertisements that mention one of the Chinese terms in the respective category.

sellers are usually able to get better prices and/or better conditions of sale. In our case the ability of brokers/ landlords to be very selective with regard to potential renters – despite this being an informal market – is suggestive of the extreme demand overhang for space in Shanghai.

Research Question 2: What Are the Revealed Preferences?

What did we learn about the target population's revealed preferences from the data? We first consider the market data collected through the survey of actual group rentals.

Table 3 offers descriptive statistics about the most important variables on location, apartment specific amenities, and the gender situation. While most of these variables are straightforward in their interpretation, some need further elaboration.

Table 3: Rental Characteristics Descriptive Statistics (Real Data, n= 747)

	Variable	Mean	Std.	Min	Max
Location	**Distance to CBD (km)**	4.41	1.68	1.27	9.12
	Distance to closest subway stop (km)	0.43	0.25	0	1.86
	Subway stops within 800m radius	1.67	0.97	0	5
	Subway lines within 800m radius	2.27	1.29	0	5
Apartment	**Rent**	850.16	314.18	450	1650
	Total number of beds	24.22	11.01	3	62
	Crowding (people per room)	6.6	2.4	1	12
	Number of bedrooms	3.24	1.11	1	7
	Number of bathrooms	1.61	1.06	0	3
	Cooking allowed	0.32	0.46	0	1
Gender Apt	**Female Apt**	0.21	0.41	0	1
	Male Apt	0.31	0.46	0	1
	Co-ed Apt	0.48	0.5	0	1
Gender Room	**Female Room**	0.32	0.47	0	1
	Male Room	0.47	0.5	0	1
	Not Identified	0.21	0.4	0	1

Crowding. The survey allowed us to collect two sets of crowding data: room level crowding (i.e. how many people share a room), and apartment level crowding (i.e. how many people live in the apartment in total). For all data sets we have information on how many beds are in one room, but only for the real market data we are able to observe the total number of beds per apartment as well. Note that numbers are even, because beds are bunk beds most of the time.

Gender. For all advertisements and site visits we recorded the gender arrangement on the apartment level, i.e. whether living there was restricted to one gender only, or whether co-ed living was permitted. Additionally, for our real market data, we were able to recover the gender arrangement down to the room level for the majority of the observations (73%).

Table 4: Observations by Gender

Gender (apartment)		
Female	160	21.42
Male	229	30.66
Co-ed	358	47.93
Gender (room)		
Female	238	31.90
• Female (all female apt)	160	
• Female (co-ed apt)	78	
Male	351	46.92
• Male (all male apt)	229	
• Male (co-ed apt)	122	
Not identified	158	21.15
Total	747	

Cooking. Although most apartments (79.43%) are in residential buildings and thus should have a kitchen, whether cooking was allowed depended on other factors. Frequently kitchens would be converted to additional "bedrooms" or functioned as the room for the house manager. Sometimes cooking on portable electrical kitchen devices (e.g. rice cooker) would be allowed even if kitchen was not available. We were able to ask through the survey in all real site visits whether cooking was allowed. However, online advertisements rarely (in 6% of the observations for the large-n online data) mention the ability to cook.

As can be seen from the table 3, group rentals are on average about 4.4 kilometers away from the city center (see also figures 3 and 4 above), and generally with good access to public transit: average distance to closet subway is less than 500 meters, on average more than 1 subway stop is within walking distance and on average more than 2 lines can be accessed within the same walkable radius. An average rent for a bed is 850 yuan (approx. 129 dollars), most commonly found in a room with 6 beds, in an apartment with 24 beds in total. Average apartments have three bedrooms and one or two bathrooms. Cooking is only allowed in about 30% of the cases. Most apartments are co-ed, but most rooms – overall – are rented to male renters.

Table 5: Price Per Bed Space Depending on Crowding (Real Data, n=747)

Number of People in a Room	Frequency	Mean Rent in Yuan
1	4	1250
2	35	1007
4	153	1078
6	249	931
8	188	668
10	86	625
12	32	589

As can be seen from Table 5, the market seems to rationally price crowding: beds in rooms with more people are less expensive. Note the slight irregularity for rooms with only two people in them. Most often, these would be kitchen or storage spaces with extremely small square meter counts, thus explaining the slightly lower mean rents in comparison to four-people bedrooms.

We conducted hedonic price model regression estimates to identify the marginal willingness to pay for different housing characteristics in this informal sub-market for bed spaces. The regression results can be found in table 6. In all model specification the logarithm of price per bed per month is the dependent variables. Overall, the model performs – as reflected by relatively high R squared. Coefficient estimates are in line with what traditional housing market theory suggests: the farther away from the city center, the cheaper bed; the better the access to transit, the costlier the bed. The fact that the hedonic model seems to work well in explaining the observed price patterns suggest that rents for bed spaces are governed by market mechanisms.

Note that the coefficient estimate for distance to center is highly significant and stable across specifications, especially after the introduction of the crowding variables. Interestingly, the coefficient estimate on the variable measuring distance to the closest subway stop shows the expected sign, but is statistically insignificant across specification. This might be due to the network density and general great accessibility to transit in the inner city of Shanghai.

Crowding, measured both at the room level (people per room) and the apartment level (total number of beds per apartment) is a significant price determinant throughout. Note that once crowding variables are introduced to the hedonic price model, adjusted R-square, the general measure for goodness of fit, jumps to 46%. Since beds are bunks, we should multiple coefficient estimates by 2 in order to get the impact of adding another bunk bed on the price. According to our regression estimates, adding another bunk bed to a room yields a rent price decrease of 3-9%. In addition, adding another bunk bed to the apartment, irrespective of the room in question, decreases the rent by 2-3%. Interestingly, the number of bathrooms is highly relevant for the bed rent. Coefficient estimates are highly significant and economically large: having an additional bathroom increases the rent by 12-24%.

In our fieldwork we noticed that there was often a price difference between the rent for the upper versus the lower bunk: in almost half of the cases, lower bunks were more expensive (43.5%).[6] Hence, we include it in the regression model and find it to be a significant price structuring variables (models 5-7, table 6). Another variable, the inclusion of which was informed (and made possible) by our fieldwork, is the dummy variables indicating whether cooking is allowed (cooking =1) or not (cooking =0). Note statistically and economically significant coefficient estimate in model 7: being allowed to cook raises the price per bed by almost 40%. We suspect that this variable might be a proxy for the general apartment quality and possibly

[6] None of the online advertisements mention such distinctions. Note that there's a long tradition the pricing for bunk in sleeping wagons in trains in China. The lower bunk is priced higher, because in tight spaces having the lower bunk means having somewhere to sit comfortably as well.

lower densities. Hence we conduct hedonic analysis of subsets of the data; divided according to ability to cook.

Table 6: Hedonic OLS Regression Results (Real Data, n=747)

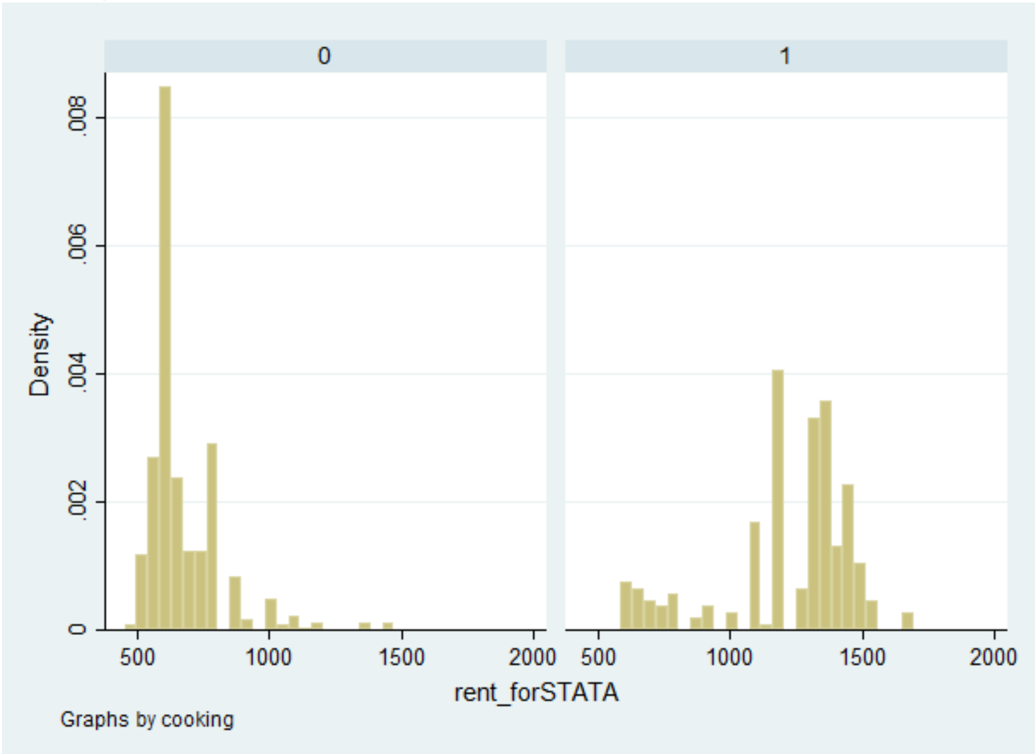
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Distance to CBD (km)	-0.037*** (-5.07)	-0.026*** (-4.11)	-0.023*** (-4.06)	-0.018*** (-3.64)	-0.018*** (-3.69)	-0.018*** (-3.78)	-0.018*** (-5.28)
Distance to Closest Subway (km)	-0.0776 (-1.27)	-0.0628 (-1.22)	-0.0114 (-0.25)	-0.0377 (-0.93)	-0.0406 (-1.03)	-0.0393 (-0.99)	-0.0547 (-1.91)
Subway Lines (within 800 m)	0.0291* (2.51)	0.0314** (3.22)	0.0199* (2.26)	0.030*** (3.91)	0.028*** (3.71)	0.028*** (3.74)	0.007 (1.27)
Crowding (People per Room)		-0.075*** (-17.43)	-0.047*** (-9.24)	-0.027*** (-6.54)	-0.027*** (-6.84)	-0.024*** (-6.79)	-0.015*** (-4.87)
Total Number of Beds (People per Apt)			-0.013*** (-13.15)	-0.019*** (-19.46)	-0.018*** (-19.62)	-0.018*** (-19.63)	-0.011*** (-13.83)
Number of Bathrooms				0.244*** (15.11)	0.230*** (14.69)	0.231*** (14.73)	0.123*** (10.11)
Dummy Lower Bunk					0.127*** (7.43)	0.126*** (7.37)	0.106*** (8.50)
Dummy Female Room						-0.0032 (-0.19)	
Dummy Cooking							0.397*** (25.43)
N	743	743	743	743	743	742	743
Adj. R-sq	0.058	0.332	0.458	0.586	0.614	0.615	0.795
F	16.25	93.17	126.5	175.9	169.8	148.7	360.0

Table 7: Does the Cooking Variable Indicate a Submarket? Descriptive Statistics: No Cooking (n=509) vs. Cooking (n=238)

	Variable	Mean		Std.		Min		Max	
		NO	YES	NO	YES	NO	YES	NO	YES
Location	Distance to CBD (km)	4.5	4.2	1.7	1.6	1.23	1.27	9.12	6.61
	Distance to closest subway stop (km)	0.44	0.39	0.26	0.19	0	0.12	1.86	1.02
	Subway stops within 800m radius	1.61	1.79	0.98	0.93	0	0	4	5
	Subway lines within 800m radius	2.14	2.55	1.29	1.25	0	0	5	5
Apartment	Number of bedrooms	3.29	3.14	1.14	.83	1	1	7	6
	Number of bathrooms	1.55	1.75	0.56	0.5	0	1	3	3
	Rent	**677**	**1221**	149	246	450	600	1450	1650
	Total number of beds	**27.86**	**16.42**	10.49	7.48	3	3	62	45
	Crowding	**7.32**	**5.06**	2.33	1.72	1	1	12	12
Gender	Female Rooms	**0.29**	**0.39**	0.45	0.49	0	0	1	1
	Male Rooms	**0.54**	**0.32**	0.59	0.47	0	0	1	1
	Unidentified Rooms	0.17	0.3	0.38	0.46	0	0	1	1

Table 7 above presents summary statistics for the submarkets of apartments that allow cooking versus those that do not. Note that while locational variables appear similar, there exist considerable differences with regard to pricing and crowding. The price for apartment that allow cooking is almost double that for apartments that do not; at the same time overall apartment level crowding is almost half for apartment that allow cooking in comparisons for those that do not and rooms are generally also less crowded (an average of 5 versus 7 people per room). The histogram of rent price frequency by cooking dummy variable clearly shows the clustering of rent around 600 and 1200 yuan respectively (see figure 5).

Figure 5: Histogram of prices based on whether cooking is allowed (cooking=1) vs. not allowed (cooking=0)



In analyzing regression estimates by subsets we find that locational variables have similar impacts on price across subgroups. However, apartment and room level crowding and as well as the number of bathrooms are valued less in apartments that do not allow cooking.

We conducted a spatial autocorrelation measure and find evidence for spatial clustering of values. Moran's I index and z-scores are positive and the p-value is below 0.05. We hence conclude that the spatial distribution of rents in this dataset of bed space rents is more spatially clustered than would be expected if underlying spatial processes were random. Looking at the attribute table and examining the coefficient estimate values, we find no significant differences in comparison to the simple OLS regression. Hence, we conclude that the presence of spatial autocorrelation is of minor importance for the estimation of our hedonic price model.

Research Question 3: Does Gender Play a Role?

We tested for gender rent price discrimination by employing a variety of different gender dummy variables in the hedonic regression models. Model 6 in table 5 shows the coefficient estimate for one such dummy variable. Like all other gender specifications (regression outputs not included), the coefficient estimate is insignificant, leading us to conclude that women do not pay more for the same arrangement. Splitting up the raw data by gender, however, does reveal that women pay for a bed on average. The descriptive statistics by gender displayed in table 8 help understand this discrepancy. While women do not pay more for the same group rentals, preferences seem to be gender: beds in female rooms are on average approximately 70 yuan more expensive then their male counterparts, but they also tend to be in less crowded rooms and apartments, having slightly better access to the city center and public transit (see table 7).

Table 8: Descriptive Statistics Based on Gender of Room (Real Data: Female = 248, Male = 351)

Variable	**Obs.**	**Mean**	**Std.**	**Min**	**Max**
Female Room					
Rent	238	867.563	326.4318	450	1650
Total number of beds	238	23.17647	10.99148	3	50
Crowding (people per room)	238	6.172269	2.27065	1	12
Distance to CBD (km)	237	4.463833	1.588817	1.272768	8.878531
Distance to subway (km)	237	.490671	.2622701	.1197837	1.102472
Male Room					
Rent	351	796.8091	279.4353	500	1500
Total number of beds	351	25.48718	10.1119	7	50
Crowding (people per room)	351	7.05698	2.428909	1	12
Distance to CBD (km)	348	4.406375	1.754805	1.287669	9.124877
Distance to subway (km)	348	.4278157	.241184	0	1.862639

Research Question 4: How Do Real Market Data Compare to Online Advertisement Data?

Table 9: Descriptive Statistics – Comparison Across Data Sets

	Variable	Real (n=747)	Matched (n=194)	Web Scrape (n=3147)
Location	Distance to CBD (km)	4.41	4.40	4.34
	Distance to closest subway stop (km)	0.43	0.36	0.38
	Number of subway stops within 800m radius	1.67	1.82	1.65
Apartment	Rent	850.16	592.78	688.03
	Crowding (people per room)	6.60	5.46	6.56
	Total number of beds (Apt)	24.22	-	-
	Size (Sqm)	-	142.46	145.41
	Number of bedrooms	3.24	2.89	2.95
	Number of bathrooms	1.61	1.90	1.98
	Cooking	0.32	0.06	0.03
	Female Only Apartments	0.21	0.10	0.02

Since we are able to draw on both remotely collected online advertisement data and real market data collected in the field, we discuss the discrepancies in findings between these different data sources in this section. Table 9 above offers an overview of how descriptive statistics vary over our three distinct data sets.

Overall, the online advertisements suggest slightly better access both to the city center and to public transit. Prices are significantly lower online highlighting the importance of considering actual transaction data in comparison to advertised rent levels. Interestingly, we find that room crowding levels in the large-n data are actually similar to what is found on the ground. Both field and scraped data suggest that group rentals are most likely to be found in 3 bedroom apartments with 1 or 2 bathrooms, but online ads tend to list fewer bedrooms and more bathrooms. Fewer rooms and apartments are offered for women both online and in reality. Co-ed apartments usually consist of one room for female tenants, with the rest of the rooms for male tenants. The fact that 20% of the observations of real market data were beds in female apartments might be due to the fact that we had more female surveyors, which possibly skewed the sample to be biased in the above illustrated direction. However, qualitative interview data suggests that gender imbalance this is not an expression of supply and demand, but rather a preference on the landlord/ house manager side.

Additional descriptive differences between our field survey data and the scraped data include the following: Age restrictions are much more prevalent in reality (57%) than the small (7%) or big (18%) fake data suggest. The ability to cook is almost never advertised (small 6%, big 2%), but much more prevalent –and indeed important for the rent– in reality (31%). Online, the vast majority of beds are advertised as being in co-ed apartments (small 68%, big 86%), while in reality less than 40% of all apartments visited where co-ed (38%).

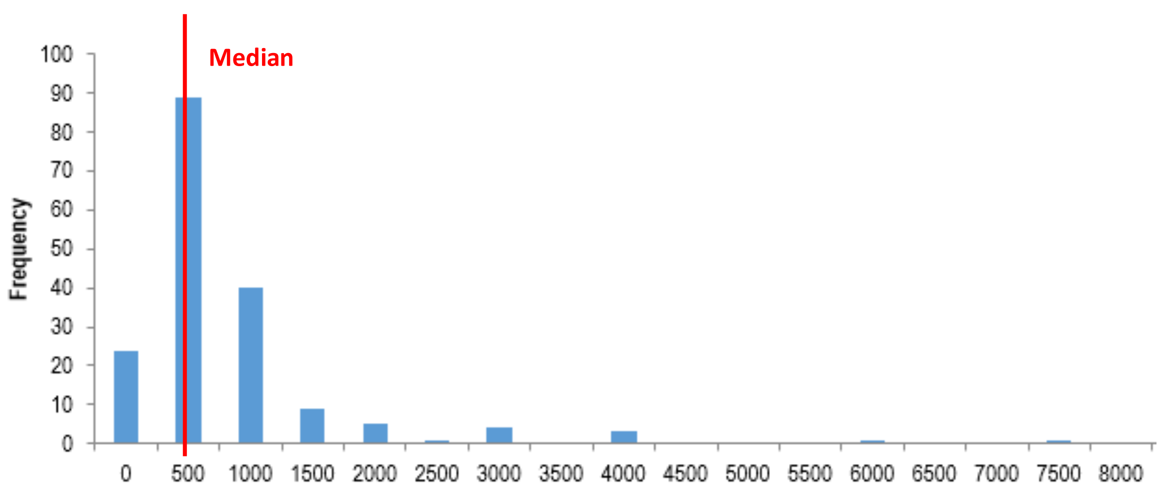
Table 10: Price Per Bed Space Depending on Crowding – Comparison Across Data Sets

	Real Data		Large-n Online Data	
Number of People in a Room	**Frequency**	**Mean Rent in Yuan**	**Frequency**	**Mean Rent in Yuan**
1	4	1250	0	0
2	35	1007	228	1398
4	153	1078	354	1051
6	249	931	1134	725
8	188	668	1161	444
10	86	625	269	511
12	32	589	-	-

The pricing of crowding in the market differs significantly across data sets. Although the online advertisement data also suggests a rational pricing of crowding, we see much more beds in 2 and 4 people rooms, comparing proportions to the real market situation. Additionally, the price drop at six people per room upwards is much starker online and crowding beyond 10 people per room is never mentioned.

Comparing the real site visit with the matched online data yields additional insight. For only 6% of the cases, everything (price and location) were as advertised. In addition to the price, locations would be different from what was advertised. However, most the real apartments where in walking distance (800 meter, ½ a mile) of the advertised location. As figure 6 below illustrates, the mean distance between advertised and actual location was 593 meter; due to a highly skewed distribution, the median was even less (332 meters).

Figure 6: Histogram of Distances – Advertised Locations vs. Actual Sites



Since price and location differed – but perhaps not by important amounts – advertisement data could still be important, unless error is systematic and indicative of a missing variable bias. We still conducted hedonic model regression estimations to test how the large-n online data set fares in terms of accurately reflecting the price mechanisms in this submarket as uncovered by our hedonic regression analysis using the real market data. Table 11 below shows the regression outputs. Note that three important price structuring variables are not available: total number of beds per apartment, the lower bunk dummy, and the cooking dummy. Nevertheless, overall goodness of fit (adjusted R-squared) is still reasonably high.

First, note that in comparison to the real data hedonic regression estimation, the access to the center has a much higher price impact.[7] Additionally, access to transportation seems to have no impact on price: coefficient estimates are insignificant and show a counter intuitive sign.

Table 11: Hedonic OLS Regression Results (Larg-n Online Advertisement Data, n= 3147)

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Distance to CBD (km)	-0.164*** (-24.64)	-0.115*** (-22.20)	-0.106*** (-21.17)	-0.107*** (-21.17)	-0.102*** (-20.25)	-0.107*** (-21.16)
Distance to Closest Subway (km)	0.0178 (0.43)	0.116*** (3.64)	0.0151 (0.49)	0.0157 (0.50)	0.0235 (0.76)	0.0155 (0.50)
Subway Lines (within 800 meters)	-0.00800 (-1.04)	-0.00571 (-0.97)	-0.00784 (-1.39)	-0.00785 (-1.39)	-0.00982 (-1.76)	-0.00795 (-1.41)
Crowding (People per Room)		-0.148*** (-47.47)	-0.145*** (-48.34)	-0.146*** (-48.10)	-0.140*** (-45.45)	-0.146*** (-48.08)
Number of Bathrooms			-0.251*** (-16.59)	-0.246*** (-15.04)	-0.312*** (-18.10)	-0.245*** (-14.98)
Number of Bedrooms				-0.00890 (-0.88)		-0.00910 (-0.90)
Apartment Size (Sqm)					0.0013*** (6.30)	
Dummy All Female Apartment						0.0137 (0.33)
N	3146	3146	3135	3135	3118	3135
Adj. R-sq	0.173	0.518	0.559	0.559	0.571	0.559
F	220.0	846.8	794.8	662.4	693.4	567.6

[7] Future research should explore potential correlation with missing variables. There's a high chance coefficient estimates are higher because of missing variable bias.

Interestingly, the crowding variable (people per room) is much more important for pricing: adding an additional bunk bed now decreases price by up to 30%. Note that perhaps this is reflecting the fact that the other important crowding variables (total number of beds per apartment) is missing from the specifications below. Another curiosity are the coefficient estimates on the number of bathrooms. According to the data, having an additional bathroom yields a significant price decrease. One explanation could be that more bathrooms are viewed as a signal for more overall apartment level crowding in this market. Further research is needed to interpret these coefficient estimates conclusively.

As with the real market data, we also calculated Moran's I index to test for presence of autocorrelation in the data. Again, we have to reject the null hypothesis of random spatial distribution of rent levels. Similar to the analysis using the real market data, we again find the regression estimates from using the GWR tool instead of a simple OLS method to be only slightly different and not important in terms of overall conclusions.

4. Conclusion

Rentals are an important housing source, especially for new migrants to urban centers. A 2013 government survey reports that in Shanghai 78% of migrants live in self-rented accommodation (Shanghai Municipal Statistics Bureau, 2013). However, until recently, data availability and the informal character of it have prevented this market activity from being studied rigorously.

For this paper we collected, cleaned, analyzed, and mapped more than 33,000 online postings advertising bed spaces for rent in Shanghai. In addition, we collected real market data through shopping. This way we are able to investigate the target demographic and revealed preferences of renters in this housing submarket. Additionally, because we can draw on both remotely collected, large-n online advertisement and hand collected survey data, we are able to assess the veracity of the volunteered geographic information.

Employing hedonic modelling, qualitative text analysis, mapping, and spatial analysis, we find an active informal housing market that caters to young, high-skilled, and recent migrants. Group rentals occur in high-amenity locations in Shanghai's inner city, which are made accessible by crowding on average 24 people into three-bedroom apartments. We find evidence in support of rational pricing of crowding (i.e., beds in less crowded rooms and apartments are more expensive), but not in support of gender discrimination. Comparing our remotely collected online advertisement to the real market data, we find that beds are more expensive than advertised in more than 80% of the cases and that location is advertised incorrectly more than 85% of the time. Comparing hedonic regressions on both data sets, we find that hedonic modelling works reasonably well on the advertisement data, but that coefficient estimates are distorted in ways only apparent through comparison with the real market data.

The findings of this paper highlight the importance of groundtruthing when working with remotely collected data. We hope that the growing literature seeking to exploit spatialized big data for the study of urban phenomena takes note of the value of complementary field work.

5. References

Ding, L., Niu, X.Y., Song, X.D. (2016). Measurement of Shanghai Center City Employment Center System - Based on Mobile Phone Signaling Data. 上海中心城就业中心体系测度 ——基于手机信令数据的研究 (Shànghǎi zhōngxīn chéng jiùyè zhōngxīn tǐxì cèduó ——jīyú shǒujī xìn lìng shùjù de yánjiū). ACTA Geographica Sinica (地 理 学 报, dìlǐ xuébào), 71(3).

Chan, K. W. (2007). Misconceptions and complexities in the study of China's cities: Definitions, statistics, and implications. Eurasian Geography and Economics, 48(4), 383-412.

Chen, J., & Hao, Q. (2008). The impacts of distance to CBD on housing prices in Shanghai: a hedonic analysis. Journal of Chinese Economic and Business Studies, 6(3), 291-302.

Chen, J., & Hao, Q. (2010). Submarket, heterogeneity and hedonic prediction accuracy of real estate prices: Evidence from Shanghai. International Real Estate Review, 13(2), 190-217.

China Internet Network Information Center, CNNIC (2017). Statistical Report on Internet Development in China. https://cnnic.com.cn/IDR/ReportDownloads/

Gaetano, A.M., & Jacka, T. (2004). On the move: Women and rural-to-urban migration in contemporary China. Columbia University Press

Huang, Y., & Yi, C. (2015). Invisible migrant enclaves in Chinese cities: Underground living in Beijing, China. Urban Studies, 52(15), 2948-2973.

Huang, Y. & Tao, R. (2015). Housing migrants in Chinese cities: Current status and policy design. Environment and Planning C: Place and Politics, 33(3).

Kim, A. M. (2004). A market without the 'right'property rights. Economics of Transition, 12(2), 275-305.

Kim, A. M. (2007). North versus South: the impact of social norms in the market pricing of private property rights in Vietnam. World Development, 35(12), 2079-2095.

Kim, A. M. (2016). The extreme primacy of location: Beijing's underground rental housing market. Cities, 52, 148-158.

Lai, L. W., & Ho, D. C. (2001). Unauthorised structures in a high-rise high-density environment-The case of Hong Kong. Property Management, 19(2), 112-123.

Li, Y (2009), "The economic and social analyses of group renting (qunzu): a case study of Shanghai", Shanghai Real Estate (Shanghai Fangdi) Vol 6, pages 16–23.

Li, B., & Duda, M. (2010). Employers as landlords for rural-to-urban migrants in Chinese cities. *Environment and Urbanization*, *22*(1), 13-31.

Li, Z., & Wu, F. (2008). Tenure-based residential segregation in post-reform Chinese cities: a case study of Shanghai. Transactions of the Institute of British Geographers, 33(3), 404-419.

Liao, B., & Wong, D. W. (2015). Changing urban residential patterns of Chinese migrants: Shanghai, 2000–2010. Urban Geography, 36(1), 109-126.

Meng, X., & Zhang, J. (2001). The two-tier labor market in urban China: occupational segregation and wage differentials between urban residents and rural migrants in Shanghai. Journal of comparative Economics, 29(3), 485-504.

Ravallion, M., Chen, S., & Sangraula, P. (2007). New evidence on the urbanization of global poverty. Population and Development Review, 33(4), 667-701.

Redfearn, C. L. (2009). How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. Regional Science and Urban Economics, 39(3), 297-306.

Shen, Y. (2015). Why does the government fail to improve the living conditions of migrant workers in Shanghai? Reflections on the policies and the implementations of public rental housing under neoliberalism. Asia & the Pacific Policy Studies, 2(1), 58-74.

Stephens, M. (2010). Locating Chinese urban housing policy in an international context. Urban Studies. 47(14), 2965-2982.

Tanasescu, A., Wing-tak, E. C., & Smart, A. (2010). Tops and bottoms: State tolerance of illegal housing in Hong Kong and Calgary. Habitat International, 34(4), 478-484.

Timberlake, M., Wei, Y. D., Ma, X., & Hao, J. (2014). Global cities with Chinese characteristics. Cities, 41, 162-170.

UN-HABITAT (2003). The Challenge of Slums: Global Report on Human Settlements 2003. London: Earthscan. https://www.un.org/ruleoflaw/blog/document/the-challenge-of-slums-global-report-on-human- settlements-2003/

Wang, Y (2007), "Calling for cheap housing – group renting: banning may not be as good as facilitating" (huhuan lianzu fang – qunzu xianxiang: qudi heru shudao), Shanghai Economics (Shanghai jingji) Vol 11, pages 6–11

Wang, D. and Huang, W. (2007). Hedonic Study on the Impact of External Environment on Housing Price: A Case Study of Shanghai. (外部环境对住宅价格影响的 Hedonic 法研究 —以上

海市为例 Wàibù huánjìng duì zhùzhái jiàgé yǐngxiǎng de Hedonic fǎ yánjiū —yǐ shànghǎi shì wéi lì). City Planning Review, 31(9): 34-46.

Wang, Y. P., Wang, Y., & Wu, J. (2010). Housing migrant workers in rapidly urbanizing regions: a study of the Chinese model in Shenzhen. Housing Studies, 25, 83-100.

Wang, M., Lin, X., & Ning, Y. (2012). Migrant populations, temporary residence, and urban village renovation: A survey of migrant settlements in Shanghai. City Planning Review, 36(7), 73–80 (in Chinese).

World Bank and Development Research Center of the State Council, the People's Republic of China (2014). Urban China: Toward Efficient, Inclusive, and Sustainable Urbanization. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/18865

Wu, F. (2000). The global and local dimensions of place-making: remaking Shanghai as a world city. Urban Studies, 37(8), 1359-1377.

Wu, W. (2002). Migrant housing in urban China choices and constraints. Urban Affairs Review, 38(1), 90-119.

Wu, F., Zhang, F., & Webster, C. (2013). Informality and the development and demolition of urban villages in the Chinese peri-urban area. Urban Studies, 50(10), 1919-1934.

Yang, S., Wang, M. Y., & Wang, C. (2015). Socio-spatial restructuring in Shanghai: Sorting out where you live by affordability and social status. Cities, 47, 23-34.

Ye, F. (2014). Group-Renting in Big Cities Has Hidden Risks, the Government Official Claims That It Will Continue (大城市群租乱象丛生藏隐忧 政府人士称短期难消失 Dachengshi Qunzu Luanxiang Chongsheng Cang Yinyou, Zhengfu Renshi Cheng Duanqi Nan Xiaoshi). Xinhuanet.http://news.xinhuanet.com/politics/2014-07/28/c_1111822184.htm

Zhang, K.H., & Song, S. (2003). Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. China Economic Review, 14(4), 386-400.

Zhang, L. (2011). The political economy of informal settlements in post-socialist China: the case of chengzhongcun (s). Geoforum, 42(4), 473-483.

Zhang, L., Zhao, Simon X. B., & Tian, J. P. (2003). Self-help in housing and Chengzhongcun in China's urbanization. International Journal of Urban and Regional Research, 27(4), 912–937.

Appendix 1:
Elements of Field Survey Instrument to Collect Group Rental Market Housing Data

- Date & cell phone number of contact person(s)
- Is this the same address as in the advertisement? If not, what is the actual address?
- Total number of beds in unit, Total number of beds per room.
- Total number of bedrooms, toilets, bathing facilities
- Is there a kitchen? Is cooking allowed?
- Gender of apartment, Gender by room.
- Building type (residential, commercial, mixed-use, lilong[8])
- Minimum duration of stay
- Age restrictions
- ID required?
- Any information about the identity/ function of the person showing the room
  - Is this the same person as on the phone?
  - Is this the owner, landlord, sub lesser or house manager?
- Any information about current tenants?
  - Job, commute, province, education
- Is a contract needed?

In addition to the survey data, more qualitative accounts and visual impressions were collected and documented during end-of-day debriefing session. Moreover, the surveyors took pictures to document the conditions on site.

[8] Lilong (里弄) or Longtang (弄堂) are a type of vernacular Shanghainese architecture, akin to Beijing’s Hutongs (胡同).