Tax Me if You Can: Tax Evasion on Chinese Housing Market

Zhentong Lu* Sisi Zhang† Daxuan Zhao‡

October 15, 2022

Abstract

Tax evasion widely exists in Chinese housing market by misreporting the transaction price. This paper examines the impact of tax evasion behavior by develop a structural model with endogenous misreporting, mortgage borrowing decision of buyers and equilibrium house price. The model is calibrated and estimated by a unique data including both actual and fake registered price. The counterfactual exercise eliminates the tax evasion behavior and illustrates striking impacts of tax evasion: (1) Government loses about 14 billion yuan tax revenue from housing transaction in one year. (2) Strict taxation enforcement could decrease the house price slightly, but the effects are profound for cheap houses. (3) Wealthy cash buyers who do not need mortgage financing are benefited more from tax evasion.

*Bank of Canada, Email: zlu@bankofcanada.ca
†Institute of Economic and Social Research, Jinan University, Email: zhang_sisi@jnu.edu.cn
‡Department of Finance, School of Business, Renmin University of China, Email: zhaodaxuan@ruc.edu.cn

1 Introduction

Tax is the major source of public expenditure. Tax evasion is illegal, and it reduces the revenue of the government. Since wealthy and high-income people usually have an advantage to evade, tax evasion also rises income inequality and wealth inequality (Slemrod, 2007; Alstadsæter et al., 2019). Meanwhile tax is also widely used as a policy instrument to intervene in the economy, especially in asset market(Cutler, 1988; Thomasson, 2003; Hoynes et al., 2015). Determining the extent of tax evasion is not straightforward for obvious reasons. Tax evasion is both personally sensitive and potentially incriminating, self-reported information is vulnerable to substantial underreporting(Baumeister, 1982). Due to lack of enforcement, tax evasion is especially severe in many developing countries, which reduce the amount of resources that would have been allocated to public services and could further lead to negative consequences for economic growth.

In this paper we examine tax evasion behavior on China's housing market, using unique administrative records that contain both registered and actual transaction price of each property, so that we observe the amount of tax evasion directly. In China, multiple tax policies, including stamp duty tax, capital gain tax, and sales tax for non-ordinary houses are widely used to cool the overheated housing market. Given the extremely high housing price in large cities, individuals have strong incentive to underreport the transaction price to avoid the tax burden (also called "Yin-Yang contract"), despite the stringent monitoring policies in many cities. The huge tax evasion in China housing market is not only the loss of public revenue, but also affect the policy effectiveness of the cooling measures.

We develop a discrete choice model with endogenous tax evasion and mortgage financing, where the price is decided by the aggregate demand and seller's supply. Using transaction data from the largest real estate broker in Beijing in May 2015-April 2016, we calibrated and

estimated the model. The results allow us to conduct further counterfactual exercise and solve the endogenous house price, consumer surplus, seller profit, government tax revenue in the absence of tax evasion. The comparison between counterfactual and actual statuses highlights the effects of tax evasion in Beijing housing market.

First, tax evasion on housing market reduces government revenue significantly. Based on our estimation, Beijing government loses about 14 billion yuan tax revenue from housing resale market in one year. It is about 2.8% of the total government revenue of Beijing government in 2016[1].

Second, tax evasion behaviors mitigate the effects of housing price regulation. In particularly, our studies illustrate that the tax policy could effectively decrease the cheap house price since their buyers are more sensitive to the increasing cost and the cheap houses are more competitive between each other. However, the widely existing tax evasion in housing market destroys the effect of tax policy and weakens the cooling measures.

Third, the huge loss of government revenues are not evenly distributed to each buyers in the market. Our results show that wealthy cash buyer who do not need mortgage financing takes most of the benefits. Constrained budget buyers usually need to borrow mortgage to finance their transaction. Since the bank only take the registered price as the collateral value, they cannot evade from the tax payment.

Our studies highlight that the tax evasion in housing market does only not reduce the government revenue and invalid the policy target, but also increase inequality in the housing market. Wealthy buyers could easily evade from the tax payment and budget buyers suffer more.

Our study contributes to three streams of literature. First, we add new evidence to the tax evasion literature. Measuring tax evasion behavior is challenging due to lack of

[1]The total revenue of Beijing government is about 508.1 billion in 2016.

data. Existing studies either infer the extent of evasion (e.g., Fisman and Wei, 2004), or use stratified random audits data (e.g. Kleven et al., 2011), or comparing self-reported value with third-party reports (e.g., Carrillo et al., 2017). We obtain confidential data from the largest real estate brokerage company, which handles over 30% of all second-hand housing transactions in Beijing during our sample period May 2015 to April 2016. The data contain both true and reported price of each second-hand house transaction, together with rich information on housing characteristics, mortgage, and individual characteristics of buyer, seller, and agent.

Second, our study is related to the growing literature on the impact of housing market regulation (Quigley and Raphael, 2005; Sommer and Sullivan, 2018; Somerville et al., 2020). Due to the overheated housing market in recent years, government in China and several other countries in Asia imposes a series of cooling measures including home purchase restriction, lower limit on downpayment ratio, upper limit on sales price, and higher transaction taxes. Studies have documented that these measures only reduce housing price temporally.

Third, our study is closely related to the scarce literature on tax evasion on housing market. Montalvo et al. (2020) use Spanish housing actual transaction prices in 2005-2011 from a real estate intermediary and match to the price declared to tax authority to study the determinants of tax evasion. They find education of the buyer and local levels of corruption explain part of the observed heterogeneity in evasion. Ben-Shahar et al. (2020) identify underreporting by price significantly lower than the reported sale price at other time that the asset is traded, controlling for a series of factors. Using repeat sales data in Israel in 1998-2015, they conclude that about 8% of transactions are under-reported and with an average price report of 30% below the true price. The only published paper using data in China is Agarwal et al. (2020). They exploit a policy change in Feb 2013 that capital gains tax increase. They use bunching to identify tax avoidance (wait for holding period

exceeds 5 years), and use DID to identify tax evasion. They find that tax evasion (the difference between two prices) becomes 23.3% higher. They also find heterogeneous policy effect that cash buyers are 8.4% more likely to buy a house. We are among the first ones to empirically analyze housing market response to transaction tax policy change, and we are the only one that uses a structural model approach to answer this question and examine the welfare consequences of buyers and sellers. In addition, our identification does not reply on any policy change or threshold discontinuity.

The remaining paper is organized as follows: Section 2 briefly introduces the tax policy in Chinese housing market. Section 3 shows the data and some stylized facts of tax evasion in our sample. Section 4 presents the model. The estimation method and results are shown in Section 5. Section 6 Section 7 concludes.

2 Institutional Background

Chinese housing price has been experiencing dramatic growth since 1998 when the traditional housing allocation model was completely abolished, and housing properties were privatized (Glaeser et al., 2017; Fang et al., 2016). The skyrocketing housing price triggers big concerns about housing bubbles and housing affordability. As a result, Chinese government implemented various policies to curb the soaring house price and cool down the housing market, including imposing home purchase restrictions, increase minimum downpayment ratio, and increase housing transaction tax.

In China, government only tax households during housing transaction and the property tax is exempted. During our study sample period of May 1, 2015 to April 30, 2016 in Beijing, housing transaction-related tax includes three components. The first one is the deed tax. The tax rate is 1% for the first home under 90 square meters, 1.5% for the first home between

90 and 140 square meters, and 3% for the others. Second, the sale tax, 5.6% of the total price, is implemented to the non-ordinary house and the house with holding period less than 2 years[2]. The non-ordinary house is defined based on its attributes and price. Houses with floor-to-area ratio below 1.0 or the area is above 140 square meters is classified as a non-ordinary house. In addition, if both the total price and the unit price of the house are above a threshold value[3], it is also considered as a non-ordinary house. Third, a seller who owns more than one house in Beijing or holds the house for less than 5 years needs to pay 20% of the realized capital gain since the previous transaction. By the law, buyers pay for the deed tax and sellers pays for sale tax and capital gain tax. However, in practice, tax cost is internalized into the price and the buyer bears all the tax costs in the Chinese housing market.

The Chinese government adopt different policies for the first home of the household from other homes. The first home buyers usually can enjoy lower deed tax rates, more mortgage credit, and lower interest rates. Sellers are also exempted from the capital gain tax when they sell their first home. In addition, house purchasing rules in China treat married couples as a single unit. A married couple who own more than one house could rearrange their properties by getting a divorce[4]. One side of this couple would become either a first home buyer or a seller who owned one house. Given the huge benefits of a housing transaction,

[2]The sale tax is replaced by value-added tax since May 1st, 2016. The sample period in our research is before May 1st, 2016.

[3]The threshold value is different among different location. In Beijing, a house located within 5th ring road would be a non-ordinary house if its unit price is above 39,600 RMB and its total price is above 4,680,000 RMB. A house located between 5th ring road and 6th ring road would be a non-ordinary house if its unit price is above 31,680 RMB and its total price is above 3,744,000 RMB. A house located outside 6th ring road would be a non-ordinary house if its unit price is above 23,760 RMB and its total price is above 2,808,000 RMB.

[4]See some media reports: https://www.wsj.com/articles/china-cracks-down-on-home-buyers-with-fake-divorces-1479810605; https://www.nytimes.com/2016/10/17/business/international/china-home-price-bubble.html; https://www.bbc.com/news/blogs-trending-37257747.

this type of fake divorce is very common[5]. Thus, our paper assumes that buyers are the first home buyers and sellers only own one house in our sample.

Sales tax is a heavy burden for buyers in Beijing. In particular, the sale tax rate is very high, and the threshold of non-ordinary housing is quite low in the Chinese housing market. Based on the actual price records in our sample, *% of houses should be classified as non-ordinary houses. Thus, the minimum tax payment for these houses would be above 6% of the total value. To misreport the actual value could directly cut down the tax base, meanwhile, some houses whose price is around the threshold value of sale tax could be exempted from the sale tax. The buyers in the market have strong incentives to misreport the price of transacted houses.

Besides, the cost of misreporting is low in China. Few home buyers are punished by the authorities for tax evasion. First, the housing transaction-related tax income only weighs a small fraction of China's government revenue. According to the Statistical Yearbook of China National Bureau of Statistics, it contributes less than 10% of the government revenue in 2018 and most of them are directly from the developer. Second, the taxation regulation to the household level is relatively loose in China. The national-level personal information system in China has been under construction and improvement. Currently, the regulation cost to the household is still very high. Third, our research focuses on the housing transactions before the value-added tax reform. After May 1st, 2016, China government uses a value-added tax instead of a sale tax which only charge the increased house price since the previous transaction. However, the sale tax is calculated based on the total price and the misreported price could not affect the tax cost of the next transaction.

[5]China government also implemented some measures to stop the fake divorce in housing market. For example, buyers divorced within 12 months cannot enjoy the preferential interest rate for the first home after March 2017 in Beijing.

3 Data and Stylized Facts

3.1 Data

The data we use is resale transaction data in Beijing, May 2015-April 2016, from one of the largest real estate broker firm in China. It includes 81,645 valid transaction over 12 months. The market size of this real estate broker covers about half of the resale housing transactions in Beijing market. Our sample is a representative sample of the one-year full market in Beijing.

The data include transaction information, housing characteristics, demographics of buyer and seller, and mortgage. Transaction information include transaction date, actual transaction price, registered price, and number of days on market. Housing characteristics include location of the house, total area, number of bedrooms and living rooms, floor, and building year. There is limited information on buyer and seller, such as the birth place of buyer or seller, age, and gender.[6] The data also collect whether the transaction has mortgage and mortgage amount. Table 1 shows the summary statistics of the key variables in our research.

3.2 Stylized Facts

We first present some stylized facts in Beijing housing market and our research is inspired by these facts.

Tax evasion widely exists by misreporting the house price. As we mentioned before, buyers in Chinese housing market often report lower price to evade the tax payment. Table 1 shows that average ratio between registered and actual price is about 0.72, suggesting buyers could less report 30% of the actual price. The price gap between registered and actual

[6]When a house is sold or purchased by a couple, the data only collect information on one of the two individuals.

price is close to 1 million. It is huge number and cannot be neglected. Figure 1 also plots the density of the price gap and ratio between registered and actual price. Most of housing transactions in Beijing are underreported around 20% of actual price or 1 million yuan.

Tax evasion behavior is more severe for the expensive house transaction. Tax evasion behaviors are also heterogeneous across different types of transactions. Figure 2 plot the ratio between registered and actual price over the actual price of house. Buyers tend to report much lower house price for more expensive houses. For cheap houses in our sample, buyers only underreport 20% of the actual price. It is much lower than the average of our sample.

Buyers of low- and medium-priced houses rely more on mortgage financing. The misreporting could affect the mortgage borrowing of buyers since the bank only recognize the actual price. Thus, buyers who rely more on mortgage financing would be difficult to evade their tax. Figure 3 plots the actual mortgage ratio, defined by the total outstanding mortgage over actual price, relative to the actual price. We can find that the low- and medium-priced house transactions have higher mortgage debt ratio in our sample.

Most of houses in the market are low- and medium-priced. Figure 4 plots the distribution of actual price in our sample. It shows that most of house price in our sample is less than 5 million yuan. And the summarized statistics also shows that the median price in our sample is only 2.75 million yuan. It suggests that the low- and medium-priced houses are more competitive and they are more likely to be affected by the demand.

4 Model

To examine tax evasion behavior and its welfare consequences in Chinese housing market, we develop a Bertrand-Nash equilibrium model with endogenous house choice, financial decision,

and tax evasion behavior. Our model allows the buyers choose their house in the market and decide their finance plan of house purchasing. On the other side, the sellers also optimize their profit in response to demand. Housing price is endougenously decided given a market clearing condition.

Model Setting. A set of households indexed $i = 1, 2, 3, .., N$ are purchasing houses in a market which includes a set of houses indexed $j = 1, 2, 3, ...N$. Houses have heterogeneous attributes X_j and households vary by the attributes Z_i. Assume there is no search friction in the housing market and each household only purchases one house.[7]

Demand Side. We model the house choice of each household i as a discrete choice of a house j from a set of houses available in the market following Bayer et al. (2004, 2009). The utility function specification is based on the utility model developed by Berry et al. (1995). Household i chooses its house in the market to optimize the utility function

$$\max_j \{V_{ij} + \varepsilon_{ij}\}, \tag{1}$$

where is the indirect utility of household i who purchase house j, and ε_{ij} is an individual-specific preference shock [8]. We assume that ε_{ij} follows the extreme value distribution, in which case the probability of household i purchasing house j can be written as:

$$s_{ij} = \frac{\exp V_{ij}}{\sum_{k=1}^{N} \exp V_{ik}}. \tag{2}$$

The individual indirect utility V_{ij} could be decomposed into two parts. The first part is

[7]Search friction and the choice of number of houses are beyond the scope of this model. There are 17% households own more than one houses, according to the China Household Finance Survey in 2017. However, it is very rare that a household purchase multiple houses within one year.

[8]Each household is sufficiently small relative to the population such that households do not interact strategically with the specific shock ε_{ij}. It is uncorrelated with the price that arise in any equilibrium, and the other attributes of the houses and the households(Bayer et al., 2004) .

the utility of housing consumption from house j and the second part is the cost of purchasing house j. Thus,

$$V_{ij} = X_j\tilde{\beta}_i - \alpha C_{ij}, \tag{3}$$

where $\tilde{\beta}_i$ is the taste parameter for characteristics X_j of house j, C_{ij} is the cost of purchasing house j by household i, $\alpha > 0$ is paremeter to. For simplicity, we write the household i's taste parameter as

$$\tilde{\beta}_i = \beta + \sigma Z_i,$$

where Z_i is a vector of attributes of household i.

The cost of purchasing house j could be decomposed into three component as follows,

$$C_{ij} = p_{ij} + T_{ij} + F_{ij}. \tag{4}$$

They are purchased price p_{ij}, the tax cost T_{ij}, and the financial cost F_{ij}, respectively. First, the purchased price p_{ij} is the actual transaction price of house j for household i.

Second, the tax cost includes both the tax payment and other potential cost of tax evasion behavior. On one hand, tax evasion behavior could reduce the tax cost. Household could report a lower registered price to the government agency and pay a lower tax. Our model uses the ratio of registered price and actual price ρ_{ij} to measure the extent of tax evasion. $\rho_{ij} < 1$ implies a lower registered price is $\rho_{ij}p_{ij}$ and less tax payment. On the other hand, this tax evasion behavior also leads to some potential cost. For example, once the government agency detects the misreporting, the transaction would be declined or the buyer and seller even incur a financial punishment. To this end, a penalty term is included. We assume it increases with the amount of misreported housing value quadratically. Overall, the tax cost of household i is:

$$T_{ij} = \tau_j(\rho_{ij}p_j)\rho_{ij}p_j + \psi_i \left(1 - \rho_{ij}\right)^2 p_j^2,$$

where $\tau_j(\rho_{ij}p_j)$ is the tax rate for household i purchasing house j based on the registered price $\rho_{ij}p_j$, and ψ_i ($\psi_i > 0$) is a tax evasion coefficient describing the heterogeneous preferences on misreporting across households. For example, experienced households who have purchased houses multiple times could hide more value to evade the tax.

Third, household could borrow mortgage to purchase its house. Leveraged households need to pay the finance cost. First, the time value of money leads to the future payment is different from current dollar value. Second, leveraged households would pay the interest of the mortgage. Third, the borrowing increases the risk loading of households in the future. We define a risk cost term which related to the amount of the loan. If the downpayment ratio is η_{ij}, the finance cost is as following:

$$F_{ij} = \left(\gamma - 1 + \gamma r\right)\left(1 - \eta_{ij}\right)\rho_{ij}p_{ij} + \phi_i \left[\left(1 - \eta_{ij}\right)\rho_{ij}\right]^2 p_{ij}^2,$$

where γ is the discount factor ($0 < \gamma < 1$), r is the mortgage interest rate, and ϕ_i ($\phi_i > 0$) is a mortgage preference parameter describing the preference of household i on the outstanding loan. In general, wealthy households who own a large number of cash have higher ϕ_i, while constrained households who need mortgage to afford their houses have smaller ϕ_i .

Supply Side. We assume that each seller provides one house in the market. The seller of house j in the market would optimize its profit as:

$$\max S_j\left(p\right)\left(p_j - R_j\right),$$

where R_j is the reservation value of house j, and$S_j(p)$ is the demand function of house j. The probability of the house j to be sold out at price p is $S_j(p)$. The F.O.C. of seller's problem is

$$p_j + \frac{S_j(p)}{\partial S_j(p)/\partial p_j} = R_j. \tag{5}$$

Equilibrium. The objective function defined in equation 1 allows a household i with characteristics $\{Z_i, \psi_i, \phi_i\}$ to select a house j among a set of houses with with characteristics X_j. The aggregated demand on house j is

$$S_j = \frac{1}{N}\sum_{i=1}^{N} s_{ij}. \tag{6}$$

A Bertrand-Nash equilibrium is defined as a set of choice probability $\{s_{ij}\}$ and a vector of housing price $\{p_j\}$ such that

1. The set of choice probability $\{s_{ij}\}$ is defined by the solution of buyers' discrete choice problem according to equation 2.

2. The housing price $\{p_j\}$ is defined by the solution of sellers' problem according to equation 5.

3. The housing market clears. The demand of house is formed by explicit aggregation of choice probability according to equation 6.

5 Estimation

We calibrate and estimate the parameters in our model based on the housing transaction data in Beijing. The estimation is based on a discrete choice model with some endogenous variables closely related to that developed in Bayer et al. (2004) and Berry et al. (1995).

Calibration. We first calibrate several exogenous variables in our model based on our data and the other stylized facts in Beijing market. Our model shrinks the arrangement of mortgage into two periods. The discount factor and mortgage interest rate is calibrated based on the average duration of outstanding loans. The average loan term in China is substantially short. Meanwhile, the curtailment and prepayment are very common for Chinese households(see, e.g., Deng and Liu, 2009; Kuang et al., 2019; Ye et al., 2014). Thus, we assume that the average actual duration of the loans is 5 years. If the annual discount factor is 0.95 following some standing literature, $\gamma = 0.95^5 \approx 0.77$. Our sample shows that the average annual interest rate is 4.35%. The interest rate r is calibrated as the compounding rate over 5 years, 23.7%.

In addition, the individual misreporting preference ψ_i and mortgage borrowing preference ϕ_i is also calibrated based on our sample. To this end, we define a ratio of total mortgage loan as $d_{ij} \equiv (1 - \eta_{ij})\rho_{ij}$. Household choose ρ_{ij} and d_{ij} to maximize its utility. After substituting equation 4 into equation 3, the F.O.C. shows:

$$\tilde{\rho}_{ij} = \frac{2\psi_i p_j - \tau_{ij}}{2\psi_i p_j},$$

$$\tilde{d}_{ij} = \frac{1 - \gamma\,(1 + r)}{2\phi_i p_j}.$$

Our data has the ratio of registered price between actual price, $\tilde{\rho}_{ij}$, and the ratio of total mortgage loan, $\tilde{d}_{ij}$, for each household in our sample. ψ_i and ϕ_i could be directly calculated based on above equations. Specifically, if $\rho = 1$, $\psi \to \infty$ implies that the household extremely rejects the misreporting behavior, and if $d = 0$, $\phi \to \infty$ implies that the household extremely rejects the outstanding loans. Note that the ratio of outstanding mortgage is constrained by the requirement of downpayment ratio. $d \geq (1 - \underline{\eta})\rho$. The minimum downpayment requirement $\underline{\eta}$ is 0.3 in our sample. It implies that our calibration could underestimate the

preference of mortgage. [9]

Clustering. Our sample includes over 81,645 transaction over 12 months. Our estimates would allow over 81,645 households choose their houses among such a set of houses. To reduce the computation complexity, we cluster the households in our sample based on their attributes (Z_i, ψ_i, ϕ_i) into 200 groups by K-means clustering[10]. One representative household that is assigned by the mean of attributes and weighted by the number of households for each group, is used in our estimates.

Discrete Choice. The parameters of the individual utility function is estimated based on a discrete choice model as shown in equation 2. The quasi-log-likelihood function is defined as the sum of the log of the probability that each household makes its correct housing choice,

$$\widetilde{L} = \sum_j \sum_i I_{ij} \log\left(s_{ij}\right)$$

where $I_{ij}$is a dummy variable that equals to 1 if household i purchase house j in our sample, otherwise 0. To maximize the quasi-log-likelihood function, we can obtain the parameters of the individual utility function $\hat{\beta}$. The reservation value of each house $\hat{R}$ could be derived by equation 5.

Instrument and Control Function. The utility function in our model could be written as

$$U_{ij} = V_{ij}(p_{ij}, X_j, Z_i) + \varepsilon_{ij},$$

where X_j and Z_i are exogenous variables of house's and household's attributes to decided the value function, while the transaction price p_{ij} is endogenous variables which is correlated

[9]The lower preference of mortgage leads to less demand on mortgage loans. Thus, our estimates and simulations underestimates the effects of the mortgage. Our results are the lower bound of actual effects.

[10]We also use the other number of groups to repeat our estimates. That does not significantly affect the final results.

to X_j and Z_i. Thus, ε_{ij} is not independent and could be decomposed into two component as follows:

$$U_{ij} = V_{ij}(p_{ij}, X_j, Z_i) + \varepsilon^1_{ij} + \varepsilon^2_{ij},$$

where ε^2_{ij} is independent and follows the extreme value distribution, and ε^1_{ij} is correlated with p_{ij}. We adopt a control function developed by Petrin and Train (2010) to capture ε^1_{ij} in our estimation. In the control function, a novel instrument "the listing duration of a house" is used since the price of a house could be lowered by seller if it has been listed for a long time. This instrument captures the exogenous features of the housing market, but have no effects on the utility of household staying in the house.

We first regress the price p_{ij} on the house's attributes X_j, buyer's attributes Z_i, their interacted terms, and the instrument to obtain the residual of the regression. The results are reported in Table 2. In Column (1), we only include the instrumental variable *Stay_ market* and the house's attributes to test the correlation between the listing duration and the house price. The coefficient on *Stay_ market* is positive and significant, suggesting the instrument variable is related to the house price. We further included the buyer's attributes and the interacted term between house's and buyer's attributes in Column (2) and (3), respectively. The coefficients on *Stay_ market* are still positive and significant. Moreover, the magnitude is also consistent across different columns. The residuals from the regressions in Column (3) are used to control the price-correlated error term ε^1_{ij} in our estimation.

Estimation Results. The estimated parameters are shown in Table 3. We estimate the coefficients on the house's attributes and the interacted terms between house's and buyer's attributes. The coefficient α on the cost function is also estimated by the discrete choice model. Most of estimated parameters in the table are significant.

To better illustrate our estimation, we also calculate the willingness-to-pay (WTP) to

each house's attribute across different buyer's attributes and report them in Table 4. The first row of Table 4 shows that the average WTP in our sample for 100 square meter house is about 3.121 million yuan. And the average buyers in Beijing market prefer to houses in higher building, with more bedrooms and built in earlier age. And the remaining rows further show that the WTP varies across the buyers. For example, male buyers could pay more for a house in higher floor, while female buyers are more likely to choose a house in lower floor in our sample.

6 Counterfactual Exercise

Using the estimated parameters of our model, we conduct a counterfactual exercise to find the status of Beijing housing market without tax evasion behavior. The comparison between actual and counterfactual statuses allows us to explore the the effects of tax evasion on Beijing housing market.

Counterfactual Exercise Procedure. In the counterfactual exercise, $\rho \equiv 1$ implies tax evasion does not exist in our model. And then, we set the initial price p_j^0 as the actual transaction price in our sample. Given p_j^0 , household i solves its problem and find a optimal V_{ij} . Using the equation 2 and 6, the aggregate demand of house j at price p_j^0 is derived. We update the house price based on equation 5 and then repeat this process. Finally, the counterfactual house price could be solved with the following contract mapping,

$$p_j^{t+1} = R_j + \frac{S_j\left(p_j^t\right)}{\partial S_j\left(p^t\right)/\partial p_j^t}.$$

Counterfactual Price. Our counterfactual exercise provide a counterfactual price for each house. We calculate the change between counterfactual and actual price to illustrates

the effects of tax evasion on house price. Averagely the counterfactual house price decrease by about 0.5% or 15.2 thousand yuan for each house. The magnitude is not very significant in average, but it varies across different houses.

We plot the change between counterfactual and actual price relative to the actual price in Figure 5 to show the heterogeneity among different houses. First, the counterfactual price decreases more significantly for cheaper houses since the increasing cost of houses reduce a large proportion of cheaper house demand. However, the increasing housing cost due to the diminishing tax evasion has few impacts on the expensive houses. The extra tax cost plays little role in the value function if the house price is very high.

Second, this figure illustrates that houses in our sample could be clustered into four types. Each type has different impact from tax evasion. The first type of houses (Type 1) has few or small price gap between actual and registered price in our sample. Thus, the counterfactual price would be very close to actual price since their price are not misreported. Type 2 houses are purchased with mortgage and tax evasion behavior in the actual market. The suspension on tax evasion behavior increase the cost of buyers on one hand, and increase the mortgage borrowing capacity on the other hand. Overall, the counterfactual price decreases a little bit. On the contrary of Type 2, Type 3 houses are purchased without mortgage and the tax evasion cannot be neglected. Higher tax payment would decrease the price in our counterfactual exercise. The purchasers of Type 4 houses could avoid the additional 5.6% sale tax if they can misreport the transaction price. However, our counterfactual exercise does not allow the misreporting. The house price would pricing into the additional tax cost.

To highlight the effects of additional sale tax, we also plot the change between counterfactual and actual price relative to the total area of house in Figure 6. Based on the regulation, the house would recognized as a non-ordinary house if its total area is larger than 140 square meter for sure. Thus, the purchasers of houses larger than 140 square meter

would have less motivations to manipulate the registered price. As figure 6 shows, the price of houses whose area is a little bit larger than 140 square meter decrease by less than 1% in the counterfactual exercise. However, the counterfactual price decreases significantly if their sale tax could be evaded by misreporting.

Who Loses? Besides the change of house price, we also check the change of welfare in our sample. Since we have cluster the buyers in our sample into 200 clusters, the consumer surplus for each of the representative buyers could be written as

$$CS_i = \sum_j V_{ij}(p_j, X_j, Z_i)s_{ij}.$$

Our model uses ψ and ϕ to describe the preference of buyers on tax evasion and mortgage. We first plot the change of consumer surplus between counterfactual and actual status relative to the tax evasion preference. The results are shown in Figure 7. First, the average consumer surplus decreases if the tax evasion is not allowed. Second, the change in consumer surplus for buyers who are more likely to evade their tax (smaller ψ) are more profound. The counterfactual assumption constrains their tax evasion behavior and heavily affects their consumer surplus.

Figure 8 plots the change of consumer surplus over the mortgage preference parameter ϕ. The figure presents a very clear negative relation between consumer surplus change and the mortgage preference parameter. It suggests that buyers who relies on mortgage financing are less likely to be affected by the tax evasion behavior. In other word, cash buyers who do not borrow would take more advantage from the tax evasion behavior. This results clearly highlight that the tax evasion enhance the benefit of cash buyers in housing market, and most of them are wealthy people.

The aggregate welfare changes in the whole economy are also calculated in Table 5.

Comparing the counterfactual and actual status, we find that the aggregate consumer surplus in our economy increase about 1% due to tax evasion. In terms of sellers' profit defined by $p_j - R_j$, it decrease by about 1.5% if the tax evasion is not allowed. More importantly, the government lose huge tax revenue, about 7.1 billion yuan, in our sample. Given our sample only includes half of housing transactions in Beijing housing market, about 14 billion yuan tax revenue is evaded in one year in Beijing housing market. Finally, the social welfare including buyers, sellers and government could be improved even the number is very small if the taxation is enforced.

7 Conclusion

We study the impact of tax evasion behavior on Chinese resale housing market by developing a discrete choice model with endogenous buyers' tax evasion and mortgage decision, and in which the housing price is also decided by the aggregate demand of buyers. Using the resale transaction data on both reported price and actual price in Beijing market, we estimate the parameters of our model and then conduct a counterfactual exercise without tax evasion behavior. The comparison between actual and counterfactual statuses highlights the effects of tax evasion on housing market.

Our paper find that the tax evasion behavior slightly increase the price of house and the welfare of buyers. But the benefit is unevenly among different buyers. The price of cheap houses are more sensitive to the tax evasion behavior in the market since their buyers value the increasing tax cost more than wealthy people. Moreover, the consumer surplus of cash buyers in the housing market increase a lot due to tax evasion. Budgeted buyers take less benefit from the tax evasion since they rely on mortgage financing.

References

Agarwal, S., K. Li, Y. Qin, J. Wu, and J. Yan (2020). Tax evasion, capital gains taxes, and the housing market. *Journal of Public Economics 188*, 104222.

Alstadsæter, A., N. Johannesen, and G. Zucman (2019). Tax evasion and inequality. *American Economic Review 109*(6), 2073–2103.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological bulletin 91*(1), 3.

Bayer, P., N. Keohane, and C. Timmins (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management 58*(1), 1–14.

Bayer, P., R. McMillan, and K. Rueben (2004). An equilibrium model of sorting in an urban housing market. Technical report, National Bureau of Economic Research.

Ben-Shahar, D., R. Golan, and E. Sulganik (2020). Tax evasion in the housing market: Identification and exploration. *Journal of Real Estate Research 42*(3), 315–340.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Carrillo, P., D. Pomeranz, and M. Singhal (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics 9*(2), 144–64.

Cutler, D. M. (1988). Tax reform and the stock market: An asset price approach. *The American Economic Review 78*(5), 1107–1117.

Deng, Y. and P. Liu (2009). Mortgage prepayment and default behavior with embedded forward contract risks in china's housing market. *The Journal of Real Estate Finance and Economics 38*(3), 214–240.

Fang, H., Q. Gu, W. Xiong, and L.-A. Zhou (2016). Demystifying the chinese housing boom. *NBER macroeconomics annual 30*(1), 105–166.

Fisman, R. and S.-J. Wei (2004). Tax rates and tax evasion: evidence from "missing imports" in china. *Journal of political Economy 112*(2), 471–496.

Glaeser, E., W. Huang, Y. Ma, and A. Shleifer (2017). A real estate boom with chinese characteristics. *Journal of Economic Perspectives 31*(1), 93–116.

Hoynes, H., D. Miller, and D. Simon (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy 7*(1), 172–211.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Kuang, W., C. Liu, Q. Wu, and H. Zeng (2019). How do interest rate changes affect mortgage curtailments? evidence from china. *Real Estate Economics*.

Montalvo, J. G., A. Piolatto, and J. Raya (2020). Transaction-tax evasion in the housing market. *Regional Science and Urban Economics 81*, 103526.

Petrin, A. and K. Train (2010). A control function approach to endogeneity in consumer choice models. *Journal of marketing research 47*(1), 3–13.

Quigley, J. M. and S. Raphael (2005). Regulation and the high cost of housing in california. *American Economic Review 95*(2), 323–328.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic perspectives 21*(1), 25–48.

Somerville, T., L. Wang, and Y. Yang (2020). Using purchase restrictions to cool housing markets: A within-market analysis. *Journal of Urban Economics 115*, 103189.

Sommer, K. and P. Sullivan (2018). Implications of us tax policy for house prices, rents, and homeownership. *American Economic Review 108*(2), 241–74.

Thomasson, M. A. (2003). The importance of group coverage: How tax policy shaped us health insurance. *American Economic Review 93*(4), 1373–1384.

Ye, G., G. Deng, and Z. Li (2014). Mortgage rate and choice of mortgage length: A quasi-experimental evidence from chinese transaction-level data. *Journal of Housing Economics 25*, 96–103.

Table 1: Summart Statistics

Variable	Mean	Std. Dev.	25th	Median	75th
Price of House (million yuan)					
Price_actual	3.18	1.66	2.02	2.75	3.85
Price_register	2.20	1.12	1.5	1.88	2.73
Price_register/Price_actual	0.72	0.18	0.57	0.72	0.89
Price_gap	0.98	0.94	0.27	0.706	1.37
House Attribute					
Area (100 square meter)	0.84	0.35	0.59	0.756	0.99
Floor (in 10)	0.72	0.57	0.3	0.5	1
Floor_total (in 10)	1.32	0.77	0.6	1.1	1.9
Bedroom	2.04	0.74	2	2	2
Age_build (in 10)	1.61	0.86	1	1.4	2.2
Stay_market	71.67	95.01	12	39	92
Buyer Attribute					
Age (in 10)	3.57	0.97	3	3.3	3.8
Local	0.28	0.45	0	0	1
Gender (Male = 1)	0.52	0.50	0	1	1

Table 2: First Stage Regression of Control Function

	(1)	(2)	(3)
Stay_market	-0.0009***	-0.0009***	-0.0009***
	(0.0000)	(0.0000)	(0.0000)
House Attribute	Yes	Yes	Yes
Buyer Attribute		Yes	Yes
Interacted Term			Yes
Observations	81,645	81,645	81,645
R-squared	0.4235	0.4250	0.4276

Notes: This table shows the first stage regression of our control function. The dependent variable is the actual price of each house, and the key independent variable is $Stay\_market$, the number of listing days of each house .Standard errors are reported in the parentheses. Significance at 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

Table 3: Estimates of Value Function

House Attribute	Buyer Attribute	Coefficient		Std. Dev.	T-value
Area	1	-2.138	***	0.471	-4.54
	Age	2.214	***	0.136	16.31
	Local	-5.669	***	0.370	-15.33
	Gender	-2.437	***	0.454	-5.37
Floor	1	0.354	*	0.283	1.25
	Age	-0.124	**	0.073	-1.71
	Local	-0.065		0.183	-0.35
	Gender	0.124		0.267	0.46
Floor_total	1	-2.465	***	0.134	-18.37
	Age	0.800	***	0.038	20.93
	Local	-1.106	***	0.101	-10.91
	Gender	0.568	***	0.122	4.66
Bedroom	1	-2.247	***	0.258	-8.71
	Age	0.440	***	0.069	6.41
	Local	-0.712	***	0.180	-3.96
	Gender	1.961	***	0.263	7.46
Age_build	1	0.034		0.163	0.21
	Age	0.203	***	0.042	4.85
	Local	-0.767	***	0.111	-6.93
	Gender	-0.371	***	0.156	-2.38
Control function		1			
Cost Function		-0.925	***	0.002	-462

Notes:This table shows the estimated paramters in the value function. Standard errors and T statistics are reported in different column. Significance at 1%, 5%, and 10% levels are indicated by ***, **, and *, respectively.

Table 4: The willingness-to-pay of houses' attributes

	Area (in 100 sqm)	Floor (in 10)	Floor_total (in 10)	Bedroom	Age_build (in 10)
Mean	3.121	-0.047	0.400	0.149	0.375
Young	1.757	0.030	-0.092	-0.122	0.250
Old	3.671	-0.078	0.599	0.258	0.425
Local	-1.260	-0.097	-0.454	-0.402	-0.218
Non-Local	4.867	-0.027	0.741	0.368	0.611
Male	1.854	0.018	0.696	1.168	0.182
Female	4.488	-0.117	0.082	-0.951	0.583

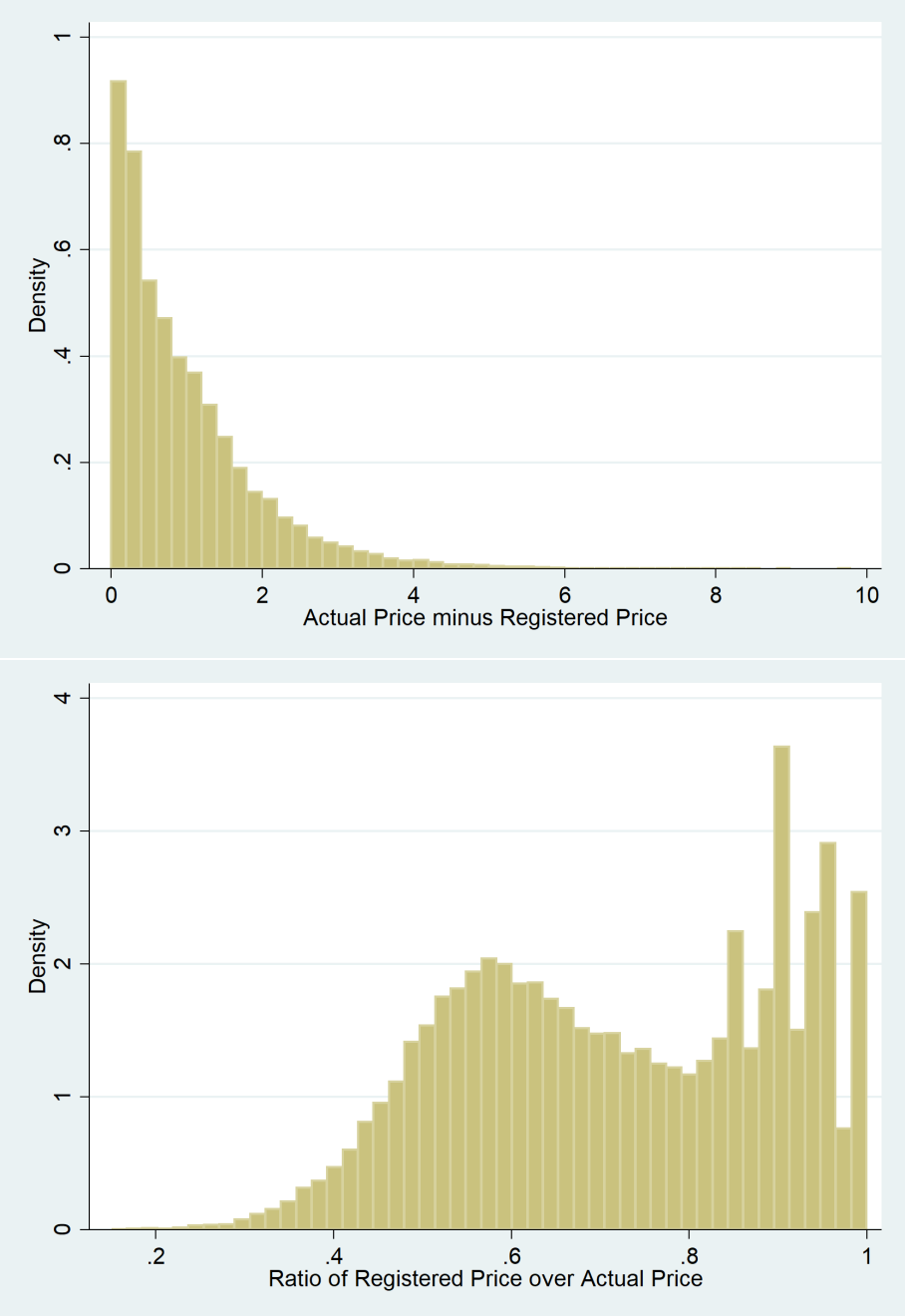
Notes:This figure shows the willingness-to-pay of different house's attributes for different buyers. The number in the table is in million yuan.

Table 5: Change of Welfare

	Actual	Counterfactual	Change
Consumer Surplus	417,840	413,727	-0.98%
Seller Profit	84,948	83,706	-1.46%
Tax Revenue	6,244	13,320	113.32%
Total	509,032	510,754	0.34%

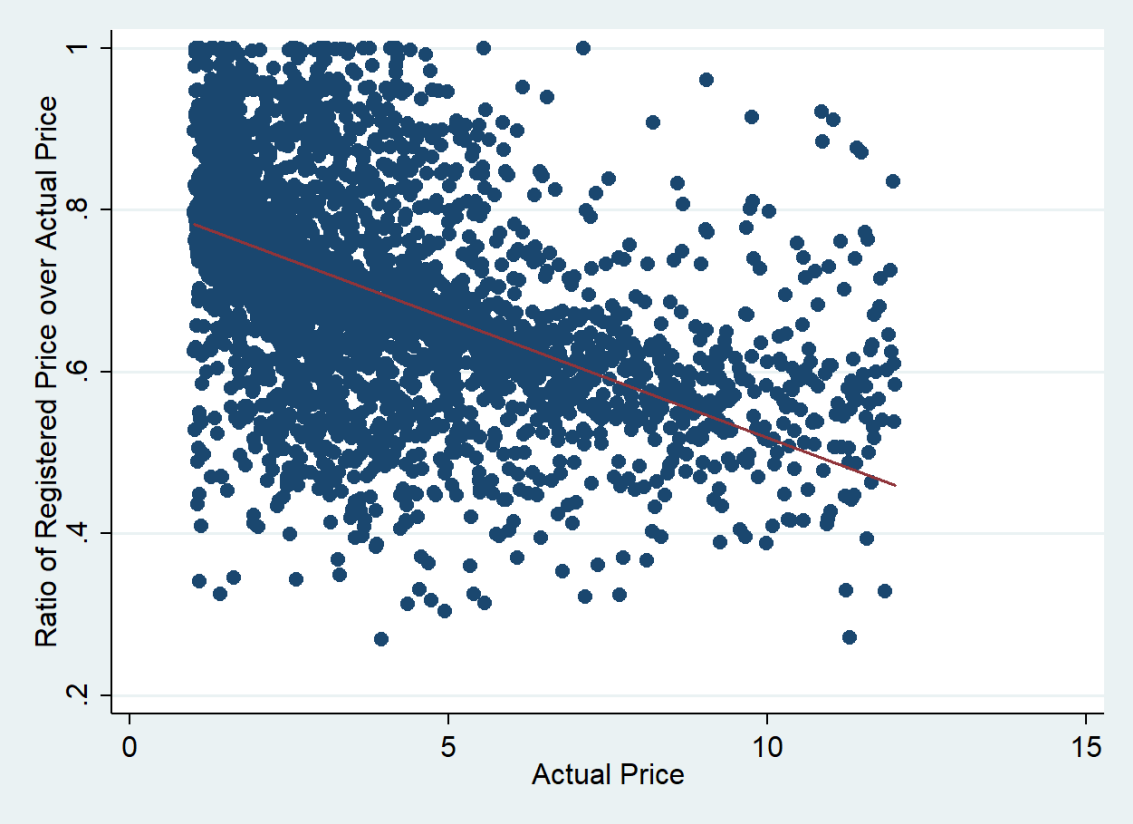
Notes: This table shows the consumer surplus, seller profit and government tax revenue of our sample in the actual senario and the counterfactual exercise.

Figure 1: Tax evasion in the sample



Notes: The figure above shows the density distribution of price gap between actual price and registered price, and the figure below illustrates the density distribution of the ratio between registered price and actual price.

Figure 2: The distribution of tax evasion over actual price

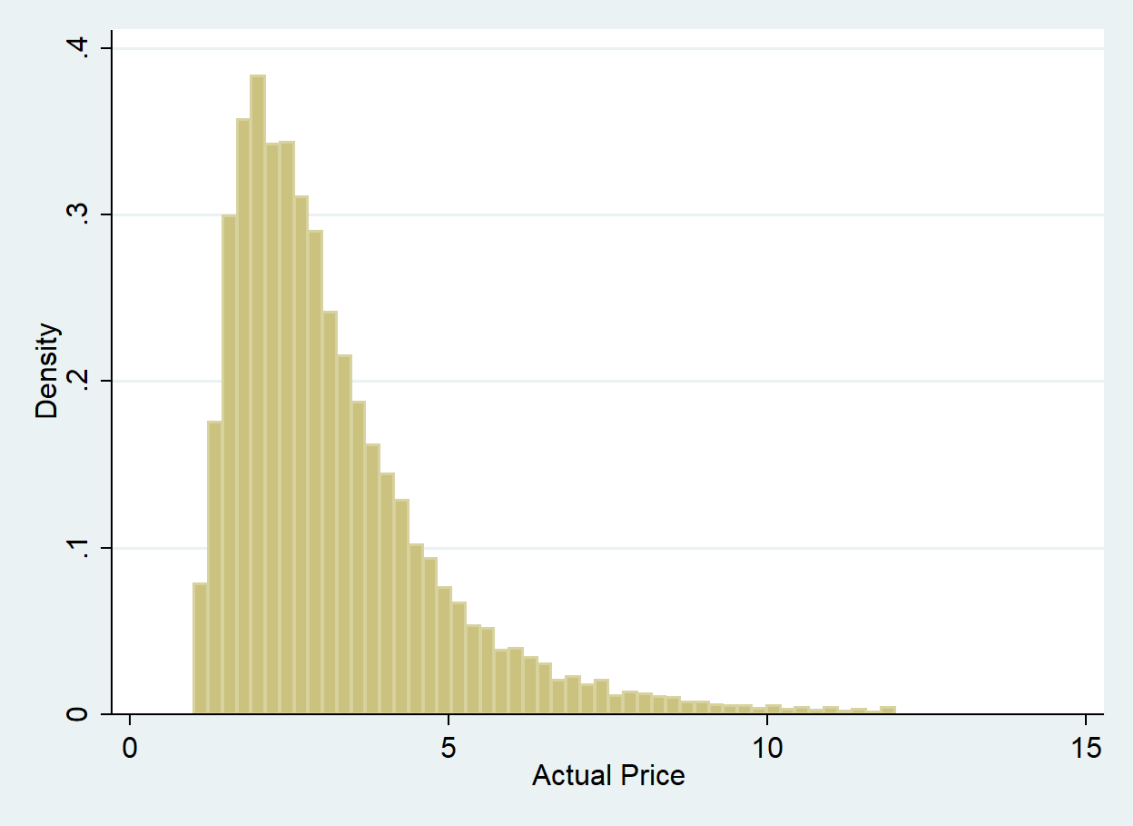


Notes: The horizontal axis is the actual transaction price (million yuan), and the vertical axis is the ratio of registered price over actual price. The solid line illustrated the fitted values.

Figure 3: The distribution of mortgage over actual price

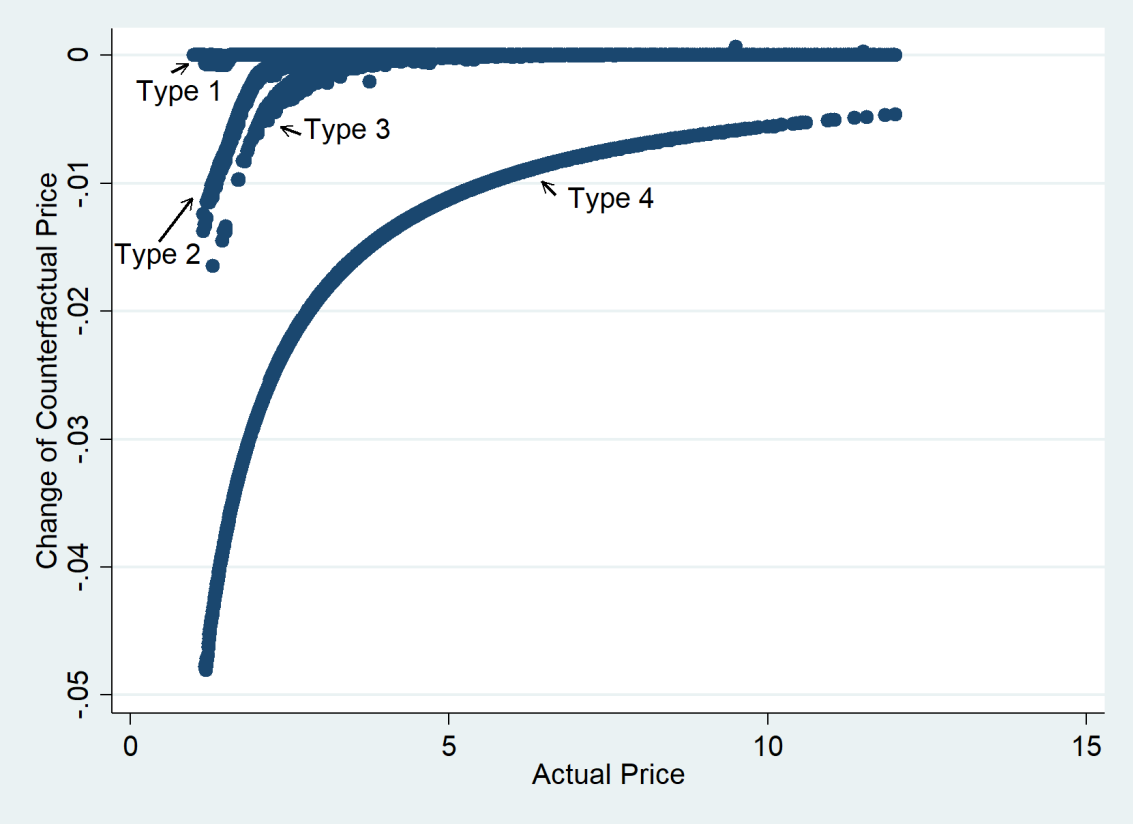
Notes: The horizontal axis is the actual transaction price (million yuan) , and the vertical axis is the ratio of total mortgage over actual price. The solid line illustrated the fitted values.

Figure 4: The distribution of actual price



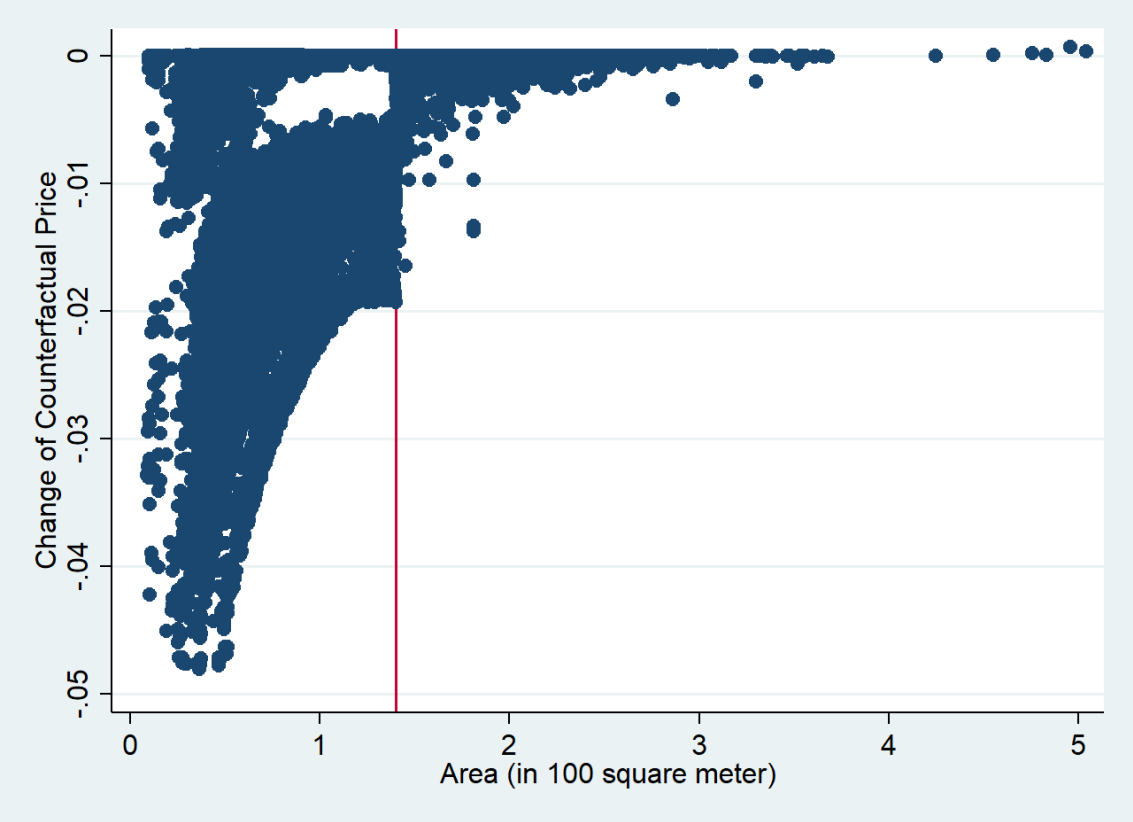
Notes: This figure shows the shows the density distribution of the actual transaction price (million yuan).

Figure 5: The change of counterfactual price



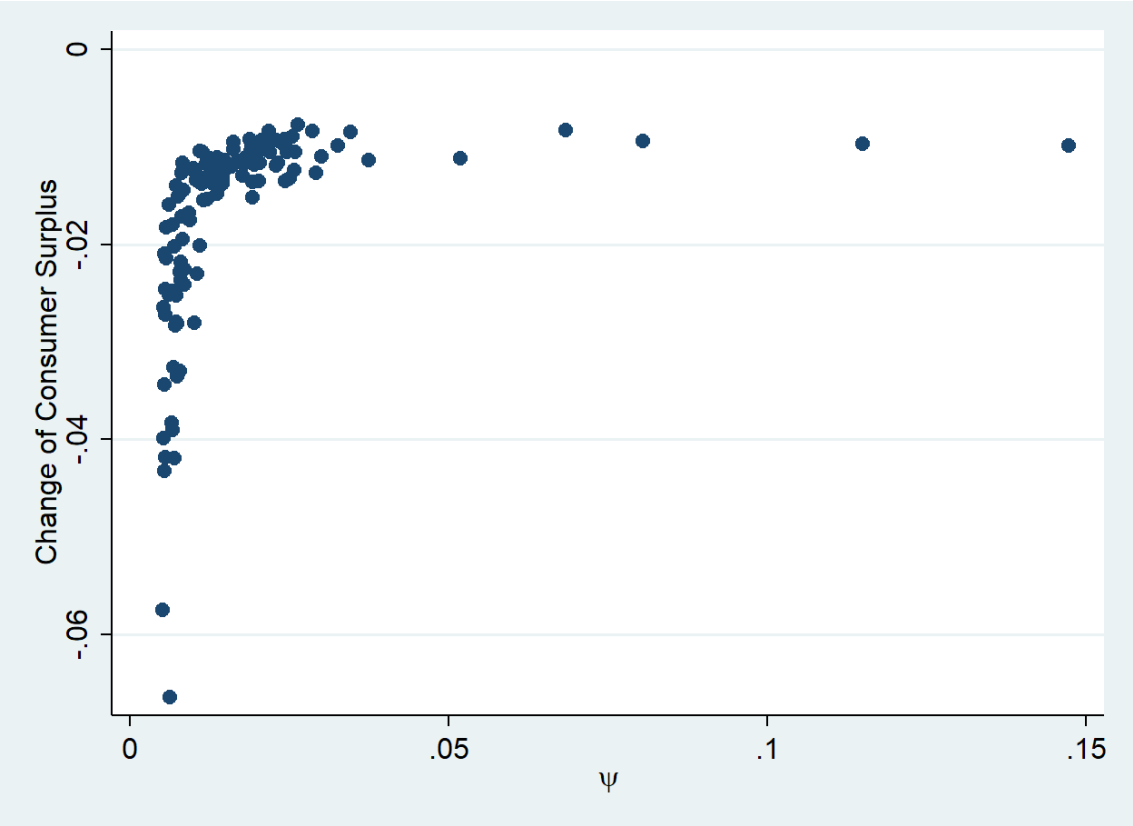
Notes: The horizontal axis is the actual transaction price (million yuan), and the vertical axis is the change between counterfactual price and actual price.

Figure 6: The change of counterfactual price over area



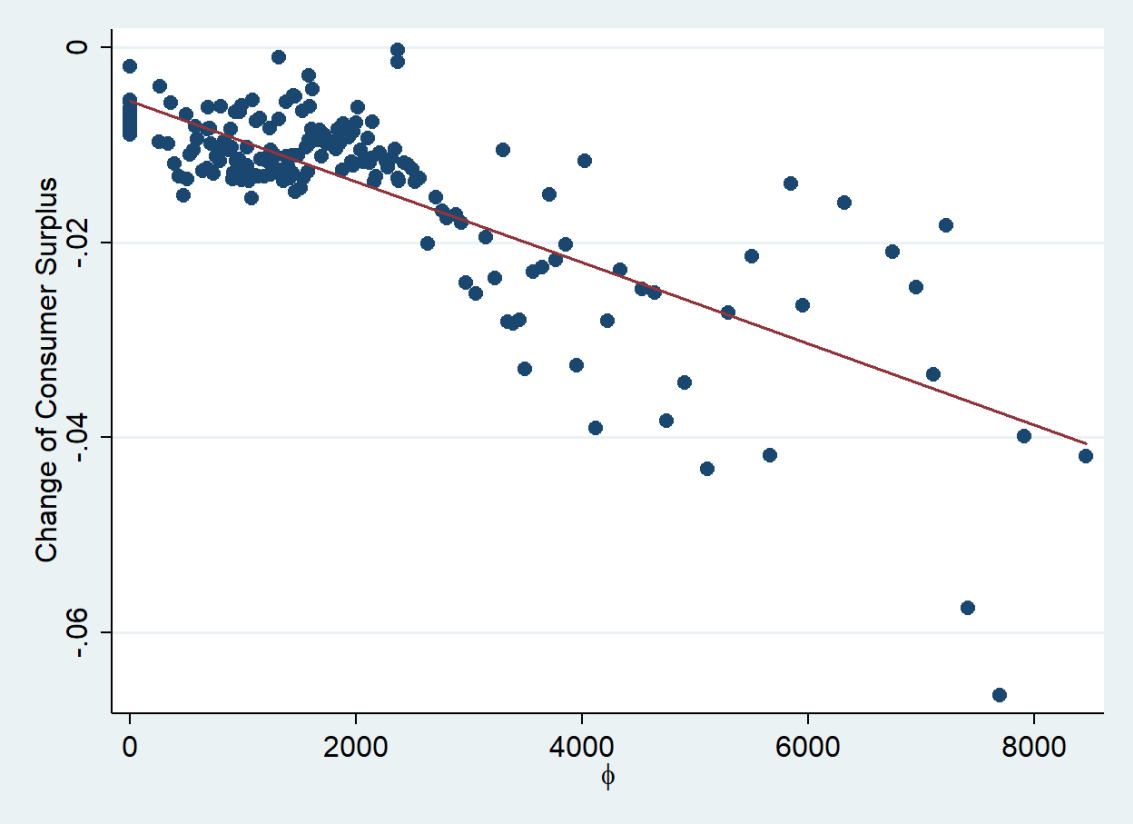
Notes: The horizontal axis is the total area (100 square meter) , and the vertical axis is the change between counterfactual price and actual price. The solid line indicates the 140 square meter cutoff.

Figure 7: The change of counterfactual price over tax evasion coefficient



Notes: The horizontal axis is the tax evasion coefficient ψ in our model, and the vertical axis is the change between counterfactual price and actual price.

Figure 8: The change of counterfactual price over mortgage preference coefficient



Notes: The horizontal axis is the mortgage preference coefficient ϕ in our model, and the vertical axis is the change between counterfactual price and actual price.