THREE ESSAYS ON HOUSING MARKETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ziye Zhang

August 2018

THREE ESSAYS ON HOUSING MARKETS

Ziye Zhang, Ph. D.

Cornell University 2018

This dissertation contains three chapters related to housing markets. First two chapters study the vertical variation in housing prices, and the third chapter models buy-sellers.

The first focus is the vertical dimension of housing markets. Location is the most important factor in determining housing prices. However, in multi-unit residential markets, two housing units in one building share the same location but may still have different transaction prices. The first chapter is to explore why. To identify the explanatory factors, the author collects over 200,000 transaction records from 2012-2016 in Beijing and provides rich and valid empirical evidence. Also, this chapter proposes a least nugget effects estimator (LNE) as an alternative to the conventional fixed effects estimator (FE). LNE is essentially a pairwise differences estimator (PD) whose comparison with FE in linear fixed effects models is studied in the second chapter.

The second focus of this dissertation is a type of households who are both home buyer and home seller at the same time (called "buy-sellers"). The traditional framework would model such households as two independent identities. However, this dichotomic tradition ignores the essential feature of the buy-sellers, that is, the interdependence between buying and selling decisions. As a buyer, a new purchase depends on a successful sale due to budget and policy restrictions; as a seller, a sale may rely on the success in bidding to avoid renting for a living. This interdependence results in a

connected bidding network. One’s success or failure in bidding or selling will influence others’. Chapter three, for the first time, studies this type of households and proposes an agent-based model (ABM) with buy-sellers. Then, the author uses this ABM to analyze the effects of housing purchase restrictions and home brokerage on market outcomes.

BIOGRAPHICAL SKETCH

The author was born in Beijing, China. He received his bachelor's degree in Public Policy and master's degree in Regional Economics from Peking University in 2011 and 2014. Later, he spent four years in Cornell University where he completed his Ph.D. in Regional Science. His research interests are urban housing markets, applied econometrics, and micro-simulation.

To my wife and my son

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Kieran Donaghy, for his guidance, instruction, patience, and support. I would also like to thank my other committee members, Dr. Ricardo Daziano and Dr. Nancy Brooks, who constantly provide insightful feedback and suggestions on my research.

I would like to thank my wife, Yi Li. She is the best friend in my life and always being there supporting me, listening to my endless "housing stories", and cheering me up through tough times. I would never have been able to finish my dissertation without her standing by my side.

I would also like to thank my parents and my parents-in-law. They encourage me to pursue my interest in research and support me in the daily life.

Finally, I would like to thank my friends, Yuanshuo, Karina, Abhi, and Fikiri for their valuable suggestions and feedbacks for my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

VERTICAL VARIATION IN HOUSING PRICES—A LEAST NUGGET EFFECT APPROACH

1.1 INTRODUCTION

Multi-unit residential markets (MRMs) have been developed around the world[1] and studied in many empirical hedonic analyses. One feature of these markets is the vertical variation in housing prices—that is, housing units in one building are transacted at different prices. Units in one building share the same location and neighborhood but may still differ in structural characteristics and transaction times. They could be on different floors, have different numbers of rooms, or face different directions. They could be transacted under different market conditions—prices may rapidly change in a housing boom or bust. Hence, the vertical variation in housing prices reflects people's preference for the structural attributes and temporal fluctuation of housing prices.

Most hedonic analyses of MRMs identified the structural and temporal effects using overall price variations on both vertical and horizontal dimensions (e.g., Gao & Asami, 2011; Jim & Chen, 2009; Jim & Chen, 2010; Wong, So, & Hung, 2002; Wong, Chau, Yau, & Cheung, 2011), largely because in this way they could estimate the effects of location factors (e.g., accessibility and neighborhood quality)—which are always of the most interest. However, location factors are inherently complex and subject to misspecification in variables (Can, 1992; Dubin, 1992). They may induce measurement errors and omitted

[1] Evidence can be found in studies worldwide, such as Singapore (Sing, 2001; Sun, Tu, & Yu, 2005), Hong Kong (Jim & Chen, 2009; Bao & Wan, 2004), Beijing (Ekblad & Werne, 1990), Tokyo (Kanemoto & Nakamura, 1986), Israel (Mitrany, 2005; Frenkel, 2007), US (Rosen & Walks, 2013), and large cities in Europe (Turkington, Van Kempen, & Wassenberg, 2004).

variable biases in parameter estimates (Kuminoff, Parmeter, & Pope, 2010). For some practices with primary interest in the structural and temporal factors—for example, designing floor plans of affordable housing projects or constructing housing price indices, the vertical dimension alone would provide better estimates than using overall variations because comparing housing units in one building naturally avoids the locational endogeneity issues. This paper uses the vertical information of the Beijing resale housing market from 2012 to 2016 to identify the explanatory variables for the vertical dimension.

Methodologically, to investigate housing price variation within one building, a conventional choice is to use the within model (fixed effects model) together with the fixed effects estimator (FE) [2], viewing the vertical variation as a particular type of within-group variation. In addition to applying FE, this paper explores another method using semivariograms. In spatial statistics, a semivariogram measures the variance of value differences between all pairs of observations with a certain distance apart (Cressie, 1993), which has been used in housing studies to model the spatial autocorrelation structure in error terms (Dubin, 1992; Basu & Thibodeau, 1998). When the distance between two observations in each pair tends to zero, the semivariogram is called the "nugget effect." This paper uses this term and denotes the vertical variation in housing prices as the "nugget effect of housing prices", because the vertical variation can be regarded as the variance of price differences between all pairs of housing transactions at the same location (i.e., with the certain distance apart tends to zero). To identify the explanatory variables for this

[2] It is also called within estimator or covariance estimator. In the context of clustered data, it is referred as cluster-specific fixed effects estimator (Cameron & Miller, 2015). It is known that FE and the least squared dummy variable estimator (LSDV) estimator are computationally equivalent (Wooldridge, 2010), thus this paper will not discuss LSDV separately.

nugget effect, this paper estimates a linear hedonic price model by minimizing the nugget effect of residuals, instead of minimizing the overall sum of squared residuals as in OLS. The estimator is called a least nugget effect estimator (LNE) and is designed to explain most of the vertical variation of housing prices—rather than the overall variation.

Comparing FE and LNE, the essential difference is that FE measures the variation using the deviation from the mean in each group; while LNE uses the pairwise differences. This paper shows that the LNE is essentially a pairwise differences estimator (PD), whose comparison with FE in linear fixed effects models is studied in Chapter 2 of this dissertation. Chapter 2 shows that PD (and hence LNE) is a generalized version of FE by weighting each observation by its group size. When the data is balanced, i.e., each group has the same number of observations, or, when the local estimate of each group is the same, FE and PD are equivalent. For MRM housing data, the observed housing transactions are grouped by buildings, and this group structure is always unbalanced—buildings have different numbers of observations due to their differences in height, transaction frequency, and sample size. As a result, applying FE and PD would generate different results. Chapter 2 also shows that both FE and PD are unbiased and consistent. However, using cluster-robust standard errors, FE is more efficient than PD when the number of groups is large. This condition would be satisfied in MRMs when the number of buildings is massive. The empirical results of this paper provide evidence to the theoretical discussions.

In what follows, Section 1.2 reviews related literature, Section 1.3 introduces the estimation approaches; Section 1.4 applies the model to housing data in Beijing, and Section 1.5 concludes with key results.

1.2 LITERATURE REVIEW

Hedonic price models regress housing transaction prices[3] on structural and spatial characteristics and on time variables (Butler, 1982; Freeman, 1981; Quigley, 1995; Tu, 1997). The regression coefficient of a variable is, ceteris paribus, the marginal change in the valuation of the housing unit (Rosen, 1974).

1.2.1 Estimation of Structural Effects in MRMs

For the structural charaterisitcs, literature on MRMs provides abundant empirical evidence on their estimates. Although the choice may differ by markets and regions, most studies involve several common structural variables, for example, floor areas, stories (floor level), number of rooms, views, orientations, and building age.

The floor area is the most common structural characteristic in hedonic housing studies (Sirmans, Macpherson, & Zietz, 2005). In the analyses of MRMs, the sign of the floor area depends on the choice of the price variable. Many studies use total selling prices and suggest a positive effect of the floor area (e.g., Chen, Clapp, & Tirtiroglu, 2011; Mason & Quigley, 1996; Xu, 2008; Yu, Han, & Chai, 2007); while some scholars find evidence of negative effects (Mok, Chan, & Cho, 1995) when using unit prices—the price per square foot (or meter), a common measure of the housing price in regions like Hong Kong and other cities in China.

The number of rooms is expected to correlate with the floor area. However, their effects on housing prices could be distinct. People may prefer housing units of a given floor area to be divided into more rooms (Borukhov, Ginsberg, & Werczberger, 1978). Positive

[3] Some scholars use listing prices or monthly rents (Kain & Quigley, 1970).

effect of the number of rooms is found in Tokyo (Shimizu, Takatsuji, Ono, & Nishimura, 2010) and Israel (Borukhov et al., 1978). Specifically, more bedrooms are associated with higher housing prices in Shenzhen, China (Xu, 2008) but with lower prices in Hong Kong (Jim & Chen, 2010). Some scholars include bedrooms as dummy variables to identify the most preferred number of bedrooms (e.g., Chen et al., 2011).

The floor level represents the vertical location of housing units and relates to the concept of vertical housing price gradients (Wong et al., 2011; Liu, Rosenthal, & Strange, 2018). In the literature, a higher floor level is correlated with a higher housing price (e.g., Yu et al., 2007), which is called a floor-level premium (see a review in Wong et al., 2011[4]). The higher the floor is, the fewer street conditions and the better view and air quality the housing unit takes (Wong et al., 2002). However, the vertical price gradient could be nonlinear. Wong et al. (2011) identify a concave vertical housing price gradient[5] and suggest a more flexible specification of the floor-levels. Mason & Quigley (1996) apply a Generalized Additive Model (GAM) and find a positive but nonlinear floor-level premium in Los Angeles. A ground-floor level could add a premium[6] in the buildings without lifts (Danton & Himbert, 2017) but affect housing prices adversely or insignificantly in other situations (Borukhov et al., 1978; Shimizu et al., 2010). Also, people may consider the floor level together with the building height and care about the relative position rather than the absolute altitude of a housing unit. Some scholars use relative floor or relative floor position zone (e.g., high/middle/low) instead of the absolute floor level (e.g., Borukhov et

[4] Other than those reviewed in Wong et al. (2011), evidence can also be found in works for Hong Kong (Bao & Wan, 2004; Jim & Chen, 2009; Jim & Chen, 2010) and Singapore (Yu et al., 2007).
[5] In rental market, Danton & Himbert (2017) find vertical rent gradient to be convex.
[6] This term is commonly used in commercial real estate markets (Liu, Rosenthal & Strange, 2016).

al., 1978). In contrast, Wong et al. (2011) find no evidence of relative level premium in Hong Kong.

The view is another critical factor in MRMs. The literature has investigated a variety of views. Positive correlations with housing prices have been found for the views of mountains, parks, gardens, and the sea (Bao & Wan, 2004; Wong et al., 2002; Wong et al., 2011; Yu et al., 2007); while negative ones for the views of graveyards, roads, and mass transit railway stations (Wong et al., 2002; Wong et al., 2011).

Facing south might be crucial in many regions. For example, a general belief of Chinese people is that housing units facing south have the best "fung-shui" (Wong et al., 2002). Generally, for cities in the Northern Hemisphere, windows facing south bring in light and warmth, which are influential on the housing choice and price.

In general, many structural factors are significantly affecting housing prices. However, expect for views and south facing windows, most attributes have controversial effects and their expected signs depend on model specifications and empirical contexts. Additionally, most estimates are obtained by using both vertical and horizontal variations and thus might be biased due to the locational unobserved effects. Robust evidence from more areas are needed for this strand of literature.

1.2.2 Temporal Factors in MRMs

Many hedonic analyses on MRMs choose cross-sectional data in one year (e.g., Jim & Chen, 2009, 2010; Kanemoto & Nakamura, 1986) to keep a balance between a sufficient sample size and minimal temporal disturbance. A few use an extended period to build price indices with time dummies (Mason & Quigley, 1996). Constructing price indices has been a primary motivation for hedonic modelling (Sheppard, 1999). The simplest practice is to estimate the regression coefficients of time dummy variables in a hedonic housing price model. In contrast to the repeated sales method (Bailey, Muth, & Nourse, 1963), which uses dwelling units that have been transacted more than once, hedonic methods are supported for the larger sample size but criticized for potential measurement errors and omitted variable issues (Meese & Wallace, 1997). In this sense, using vertical information is a balance between sample size and endogeneity issue. One may view housing transactions in one building as the building's repeated sales.

1.2.3 Study and Practice in Beijing

Following the horizontal tradition, most hedonic housing price studies in Beijing focus on the effects of location factors, such as distance to CBD (Dong, Ding, & Zhao, 2009), ring road location (Ma and Li, 2003; Gong, 2009), subway proximity (Dong et al., 2009; Li, Yang, Qin, & Chonabayashi, 2016), and school quality (Zheng, Hu, & Wang, 2016). Most studies use complex-level data with average prices and exclude the unit-level structural attributes from analysis. One exception is the study of the effect of housing sizes on prices by Gao and Asami (2011). Using 279 apartments for sale in the period of one week before and after May 1st, 2006 in Beijing, Gao and Asami (2011) classify the continuous housing size variable into five categories, estimate the hedonic price for each category, and find that housing properties with size 50-60m^2 and 80-189m^2 have higher prices. Thus, they conclude that housing properties within those sizes are valued at a high level than other sizes. Their study is inspiring, but the small sample size undermine the validity of their results. This paper uses a much larger data set to provide new evidence to this discussion.

The price indices for housing markets in Beijing are reported each year in the statistical yearbook. These price indices are constructed in a uniformed way in China[7]. The Survey Office of the National Bureau of Statistics in Beijing collects a sample of transacted housing units from large local housing agencies. Simple average prices are calculated by dividing the sum of the total prices by the sum of transacted floor area in the sample. These average prices are then used to calculated price indices. The quality of the indices is controlled by requiring local agencies to report a comparable and representative sample. However, it is unclear about the sample size and the comparability among units. This study uses information from the local agency as well but demonstrates a different feasible way to obtain temporal variation in housing prices.

1.2.4 Fixed Effects Models in MRMs Studies

In the literature, although no studies explicitly aim at explaining the vertical variations, some of them essetially do this by applying fixed effects approach using location or building dummies (e.g., Chen et al., 2011, Mason & Quigley, 1996; Yu et al., 2007). Individual building identifiers are ideal for the vertical investigation. In the literature, a dataset of downtown Los Angeles from 1980.01 to 1991.12 contains four high-rise buildings' identifiers and has been used by Quigley (1995) with 463 repeated sales and by Mason & Quigley (1996) with 843 all sales. More commonly, identifiers refer to larger spatial units such as blocks, projects, and estates, containing one or more adjacent high-rise structures. For example, Yu et al. (2007) use a dataset of 841 sales in 10 residential projects in Singapore during 1995.01 and 2003.07; Chen et al. (2011) use 10,252 sales in

[7] http://www.stats.gov.cn/tjgz/tzgb/201102/t20110216_57581.html

16 projects in Shenzhen, China from 2004.08 to 2006.01. Many studies have documented a significant increase in estimation performance after including building or location dummy variables (e.g., Quigley, 1995; Xu, 2008).

1.3 METHODOLOGY

Following the hedonic housing price model (Can, 1992; Rosen, 1974; Goodman, 1978) and assuming a linear form, for each i at location l,

$$y_{li} = \boldsymbol{x}_{li}'\boldsymbol{\beta} + \boldsymbol{z}_l'\boldsymbol{\gamma} + \alpha_l + \varepsilon_{li}, \qquad \text{for } l = 1, \dots, M;\ i = 1, \dots, N_l \tag{1.1}$$

where y_{li} is the price of housing unit i at location l; $\boldsymbol{x}_{li}$ is a $k \times 1$ column vector of observed housing characteristics which vary across locations and units at one location; $\boldsymbol{z}_l$ is a $g \times 1$ column vector of observed housing characteristics which vary across locations but invariant across units at each location; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $k \times 1$ and $g \times 1$ column vectors of coefficients associated with $\boldsymbol{x}$ and $\boldsymbol{z}$; α_l includes the unobserved location-specific characteristics; ε_{li} is the independent error term; M is the number of locations and N_l is the number of observations at location l.

1.3.1 Least Nugget Effect Estimator (LNE)

A variogram shows the variance of value differences between all pairs of observations with $s_i - s_j$ distant apart,

$$2\gamma(s_i - s_j) = var\left(Z(s_i) - Z(s_j)\right), \quad for\ all\ (s_i, s_j) \in D \tag{1.2}$$

where $Z(s)$ is a regionalized variable, a function of locations. γ is called a semivariogram (Cressie, 1993). As $(s_i - s_j) \to 0$, $\gamma(s_i - s_j)$ is expected to be zero. If one observes $\gamma(s_i - s_j) \to c_0 > 0$ as $(s_i - s_j) \to 0$, then c_0 is called the "nugget effect" (Matheron, 1963). Cressie(1993) argues that a non-zero nugget effect includes the variance of a white-noise microscale process and the variance of a measurement error. This original definition of the nugget effect restricts the regional variables to take only one value at each location.

Inspired by this concept, if multiple values at each point are allowed in the space, the vertical variation in housing prices can be measured as the semivariogram at a distance zero. In the following, I call this "nugget effect of housing prices". For all pairs of observations at the same location,

$$NE = \gamma(0) = \frac{1}{2}var\left(x_i - x_j\right),$$

for $\{(i,j)|i \neq j; x_i \text{ and } x_j \text{ are variables at the same location}\}$. Assuming a constant mean at each location, the nugget effect is calculated as $NE = \frac{1}{2}E\left(x_i - x_j\right)^2$. The empirical nugget effect can be estimated as

$$\widehat{NE} = \hat{\gamma}(0) = \frac{1}{2|D|}\sum_D (x_i - x_j)^2, \tag{1.3}$$

where $D \equiv \{(i,j)|i \neq j; x_i \text{ and } x_j \text{ at the same location}\}$, and $|D|$ is the number of pairs.

For the original model in (1.1), $y_{li} = \boldsymbol{x}_{li}\boldsymbol{\beta} + \boldsymbol{z}_l\boldsymbol{\gamma} + \alpha_l + \varepsilon_{li}$, I apply the nugget effect to the loss function of estimation: instead of minimizing the overall sum of squared residuals, this paper minimizes the nugget effect of residuals (e_{li}). This leads to a least nugget effect estimator (LNE):

$$\widehat{\boldsymbol{\beta}}_{LNE} \equiv \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \widehat{NE}(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2|D|}\sum_{l=1}^{M}\sum_{D_l}(e_{li} - e_{lj})^2. \tag{1.4}$$

where $e_{li} = y_{li} - \hat{y}_{li} = y_{li} - \boldsymbol{x}_{li}'\widehat{\boldsymbol{\beta}}_{LNE} - \alpha_l$; $D_l \equiv \{(i,j)|i \neq j; i,j \text{ at the location } l\}$.

1.3.2 Equivalency between LNE and Pairwise Differences Estimators (PD)

As shown in Chapter 2, a pairwise differences estimator (PD) is derived by pair-wisely differencing observations in each group to eliminate the location fixed effects α_l.

The original model (1.1) becomes

$$y_{li} - y_{lj} = \left(\boldsymbol{x}_{li} - \boldsymbol{x}_{lj}\right)'\boldsymbol{\beta} + (\varepsilon_{li} - \varepsilon_{lj}). \tag{1.5}$$

Then, PD is defined as the OLS estimator of (1.5)

$$\widehat{\boldsymbol{\beta}}_{PD} \equiv \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_M \sum_{D_l} (e_{li} - e_{lj})^2. \tag{1.6}$$

From (1.4) and (1.6), it is easy to see the equivalency between LNE and PD

$$\widehat{\boldsymbol{\beta}}_{LNE} \equiv \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2|D|} \sum_{l=1}^{M} \sum_{D_l} \left(e_{li} - e_{lj}\right)^2 = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_M \sum_{D_l} (e_{li} - e_{lj})^2 \equiv \widehat{\boldsymbol{\beta}}_{PD}.$$

1.3.3 PD Generalizes Fixed Effects Estimator (FE)

LNE and PD are not easy to compute directly, and their relationships with the conventional fixed effects estimators (FE) are not apparent. Using the algebraic relationship between the sum of the squared differences and the sum of the squared deviations from the means (see Appendix B for details),

$$\sum_M \sum_{D_l} \left(e_{li} - e_{lj}\right)^2 = \sum_M N_l \sum_{N_l} (e_{li} - \overline{e_l})^2 \ , \tag{1.7}$$

where $\overline{e_l} = \frac{1}{N_l} \sum_{N_l} e_{li} = \overline{y_l} - \overline{\boldsymbol{x}_l}' \widehat{\boldsymbol{\beta}}_{LNE} - \alpha_l$ [8] with $\overline{y_l} = \frac{1}{N_l} \sum_{N_l} y_{li}$ and $\overline{\boldsymbol{x}_l} = \frac{1}{N_l} \sum_{N_l} \boldsymbol{x}_{li}$.

Put (1.7) into (1.4) and (1.6), LNE and PD can be rewritten as

$$\widehat{\boldsymbol{\beta}}_{LNE} = \widehat{\boldsymbol{\beta}}_{PD} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_M N_l \sum_{N_l} (e_{li} - \overline{e_l})^2. \tag{1.8}$$

The first order condition yields

$$\widehat{\boldsymbol{\beta}}_{LNE} = \widehat{\boldsymbol{\beta}}_{PD} = \left[\sum_M N_l \sum_{N_l} \widetilde{\boldsymbol{x}}_{li} \widetilde{\boldsymbol{x}}_{li}{}'\right]^{-1} \sum_M N_l \sum_{N_l} \widetilde{\boldsymbol{x}}_{li} \widetilde{y}_{li},$$

[8] $\overline{e_l} = \frac{1}{N_l} \sum_{N_l} e_{li} = \frac{1}{N_l} \sum_{N_l} (y_{li} - \boldsymbol{x}_{li}' \widehat{\boldsymbol{\beta}}_{LLV} - \alpha_l) = \frac{1}{N_l} \sum_{N_l} y_{li} - \left(\frac{1}{N_l} \sum_{N_l} \boldsymbol{x}_{li}'\right) \widehat{\boldsymbol{\beta}}_{LLV} - \alpha_l = \overline{y_l} - \overline{\boldsymbol{x}_l}' \widehat{\boldsymbol{\beta}}_{LLV} - \alpha_l$

$$= [\sum_M N_l\, \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \sum_M N_l\, \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l \ ,$$

where $\widetilde{\boldsymbol{x}}_{li} = \boldsymbol{x}_{li} - \overline{\boldsymbol{x}}_l$; $\tilde{y}_{li} = y_{li} - \overline{y}_l$; $\widetilde{\boldsymbol{X}}_l$ is a $N_l \times k$ matrix; and $\widetilde{\boldsymbol{y}}_l$ is a $N_l \times 1$ column vector.

Recall that the fixed effects estimator (FE) is $\widehat{\boldsymbol{\beta}}_{FE} = [\sum_M \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \sum_M \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l$. Then it is easy to see that PD generalizes FE by weighting observations in each group by its group size. In another word, PD is simply the FE for $\sqrt{N_l} y_{li} = \sqrt{N_l} \boldsymbol{x}_{li}{}' \boldsymbol{\beta} + \sqrt{N_l} \alpha_l + \sqrt{N_l} \varepsilon_{li}$. The idea is that PD gives less weight to locations with less observations; while FE gives each location the same weight. If every location has the same number of observations (i.e., balanced data with equal group size N/M), then $\widehat{\boldsymbol{\beta}}_{PD}$ can be simplified into $\widehat{\boldsymbol{\beta}}_{FE}$

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{PD} &= [\sum_M N_l\, \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \sum_M N_l\, \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l \\ &= [\sum_M (\tfrac{N}{M}) \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \sum_M (\tfrac{N}{M}) \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l \\ &= [\sum_M \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \sum_M \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l \\ &= \widehat{\boldsymbol{\beta}}_{FE} \ . \end{aligned}$$

As shown in Chapter 2, the balancedness is one of the two equivalence conditions for PD and FE. The second one is the equal local estimates of each group $[\widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{X}}_l]^{-1} \widetilde{\boldsymbol{X}}_l' \widetilde{\boldsymbol{y}}_l$.

1.3.4 Application in This Paper

The least nugget effect estimator (LNE) has an intuitive spatial statistical meaning for measuring the vertical variation in housing prices. Comparing its performance to the conventional fixed effects estimator (FE) provides instructions for relevant empirical works. Since LNE is equivalent to PD, and since PD and FE have been theoretically compared in Chapter 2, the performance of LNE in comparison to FE could be expected.

First, since the housing transaction dataset in this paper is highly unbalanced, the results from LNE and FE are expected to be different.

Second, since both estimators are unbiased and consistent, their estimates should be similar to each other when the sample size is large. The dataset in this paper contains over 200,000 transactions. Thus the estimates of LNE and FE are expected to be close to each other.

Third, this paper is using clustered data and cluster-robust standard errors, in which cases FE would be more efficient if the number of clusters is large; while LNE would be more efficient only when the number of clusters is small (6-10) and the within-cluster regressor correlations are very high (shown in Chapter 2). The housing data in this paper has more than 4000 clusters (locations/buildings), which should be large enough to show FE's efficiency.

1.4 EMPIRICAL APPLICATION

1.4.1 Study Area

Beijing is the capital of China and one of the most populated cities in the world[9]. In 2016, Beijing has a permanent population[10] of 21,729,000 in 16,807 square kilometers. Abundant multi-story buildings and massive transaction volumes in the past decade make Beijing an ideal choice to investigate vertical housing markets.

The housing type in Beijing has experienced a transition from traditional courtyard

[9] According to the United Nation's *The World's Cities in 2016: Data Booklet*, Beijing rank 6th among 31 megacities (i.e., cities with 10 million inhabitants or more).
[10] According to *Beijing Statistical Yearbook 2017*, the permanent population refers to persons living for more than half a year in Beijing. http://www.bjstats.gov.cn/nj/main/2017-tjnj/zk/indexeh.htm.

houses (one-story) to mid-rise buildings (3-6 story buildings), and to high-rise buildings (7 or more story buildings) [11]. See Ekblad & Werne (1990) for a review of the history. At the end of 1997, the multi-story flats take 80% of all housing types in the housing stock (Wang, 2001). The abolishment of the subsidized housing allocation system in 1998 transformed the housing market in Beijing from a centrally planned system to a free-market one (Fu, David, & Zhou, 2000; Jim & Chen, 2007). The subsequent emergence of the commercial housing sector led to a large amount of high-rise building construction. In 2015, Beijing had 85.2% of the residents living in condominiums of multi-story buildings[12].

The housing market boom in Beijing in the past decade has drawn attentions from the world. A growing population together with increasing demand and speculation sent housing prices skyrocketing (see Figure 1.1 for resale and new housing price indices and population index). This research takes the resale housing market of Beijing as objective since compared with the new housing market, the resale one has become the dominant market in Beijing since 2008 with larger transaction volume, more even spatial distribution, and more accessible data.

[11] Some scholars argue the mid-rise buildings are 7–20 stories and the high-rise buildings are more than 20 stories (Jim & Chen, 2007).

[12] According to a survey conducted by Beijing Municipal Bureau of Statistics and Survey Office of the National Bureau of Statistics in Beijing. The survey interviews 2,300 18-70 years old residents by phone in 2015. Some results are reported in this link http://tjj.beijing.gov.cn/zxfb/201601/t20160129_335975.html.

Figure 1.1 Resale Housing Price and Population Indices (Year 2007 = 100) [13]

1.4.2 Data Source

Data are obtained from the official website of Lianjia—the largest real estate brokerage firm in China and in Beijing. During 2012-1016, Lianjia made 51%-63% of the total number of transactions in the resale housing market of Beijing (Figure 1.2), more than three times larger than the second largest firm 5i5j. The website (lianjia.com) provides detailed listing and historical transaction information. The earliest transaction dates to June 2002. Each record contains a unique ID corresponding to a signed contract, which makes it trackable and verifiable by the buyer and the seller. This paper uses records from January 2012 through December 2016 to ensure sufficient sample size in each year. In addition, this single-listing data helps to control for brokerage effects and other incentive-related effects existing in multiple-listing systems (Ong & Koh, 2000).

[13] Data source: From *Beijing Statistical Yearbooks* and the website of Beijing Municipal Bureau of Statistics and Survey Office of the National Bureau of Statistics (http://tjj.beijing.gov.cn/tjsj/yjdsj/fj_5661/2018/).

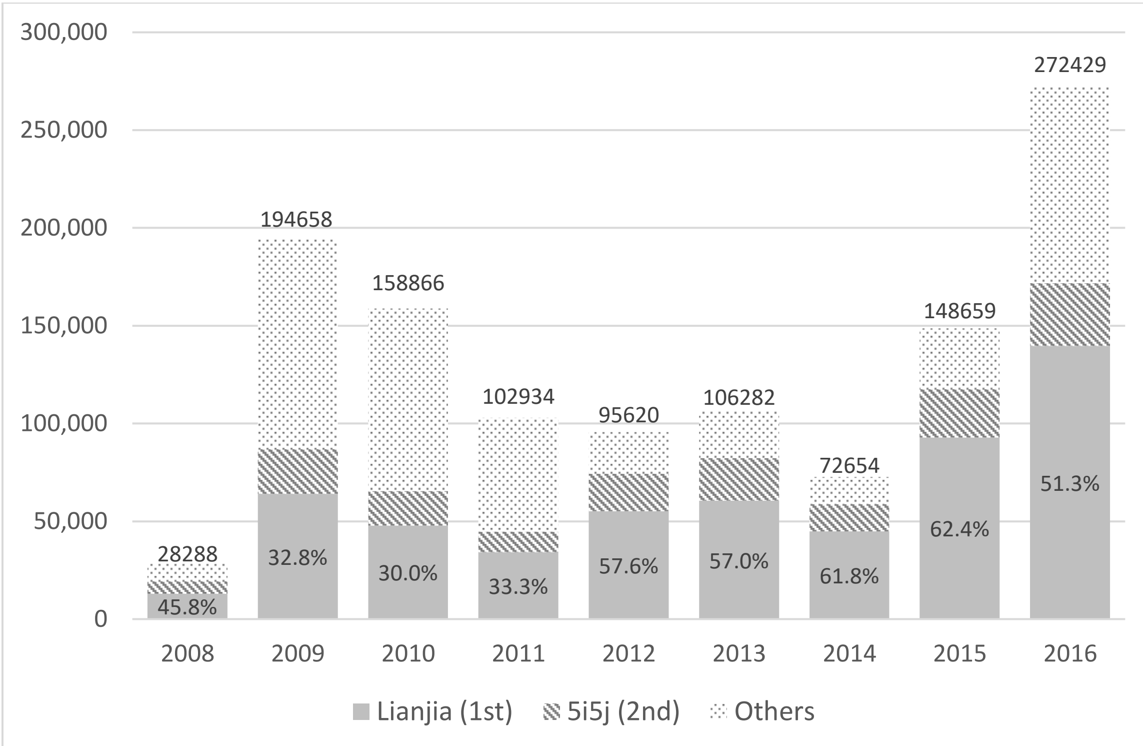


Figure 1.2 Resale Transaction Volume (Units) and Market Share (%) of Lianjia[14]

The biggest challenge of this data is the lack of accurate building identifiers. The most detailed identifier corresponds to an area containing one or multiple buildings, which could be either a complex or an estate project with clear borders (called "Xiaoqu" in Chinese), or a tract of land sharing the same street address (called "Jiedao" in Chinese). Although the data set contains building-related variables (e.g., building age, elevator accessibility) which could be potentially used to identify different buildings in each area, we cannot guarantee those variables are sufficient for all areas. To be rigorous, this paper studies the variation in housing prices not only on the vertical dimension (i.e., within one building) but also on a small scale of horizontal dimension (i.e., between buildings in one area). All available building-related variables are included to control for between-building

[14] Data source: Historical data from Beijing Municipal Commission of Housing and Urban-Rural Development http://www.bjjs.gov.cn/bjjs/fwgl/fdcjy/fwjy/index.shtml

variations.

The original data set contains 386,790 transaction records from 2012-2016. After cleaning the data (such as missing values and outliers; see Appendix A for details), this paper keeps 262,533 observations for analysis. All observations in the data set are resale commodity housing units, which were built for sale at market prices (Fu et al., 2000). Figure 1.3 shows that our sample represents half of the total transactions in Beijing and the cleaned sample represents about one third.

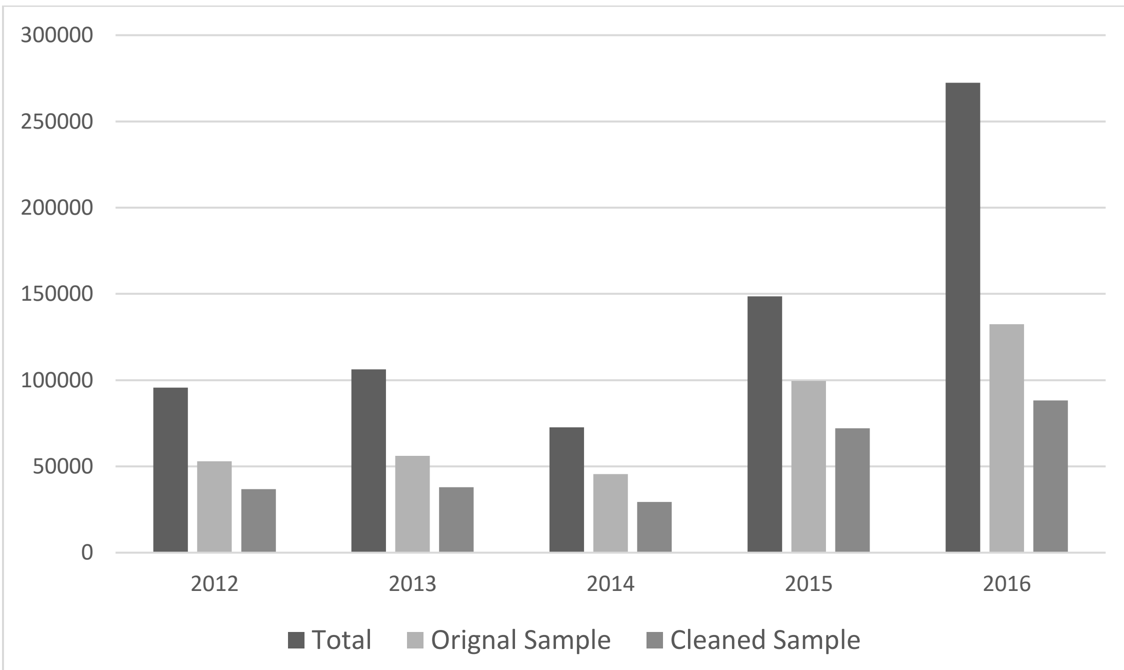


Figure 1.3 Population and Sample Sizes

1.4.3 Variable

The descriptive statistics of all variables are shown in Table 1.1. The dependent variable is unit selling price (per square meter), which is the standard measure of housing prices in China[15]. All prices are in real term in 2009 RMB level. One feature of the Beijing housing

[15] See examples in Hong Kong (Bao & Wan, 2004).

prices is the concentric pattern. See Figure 1.4-1.8 for the sample distribution across locations in Beijing from 2012 to 2016.

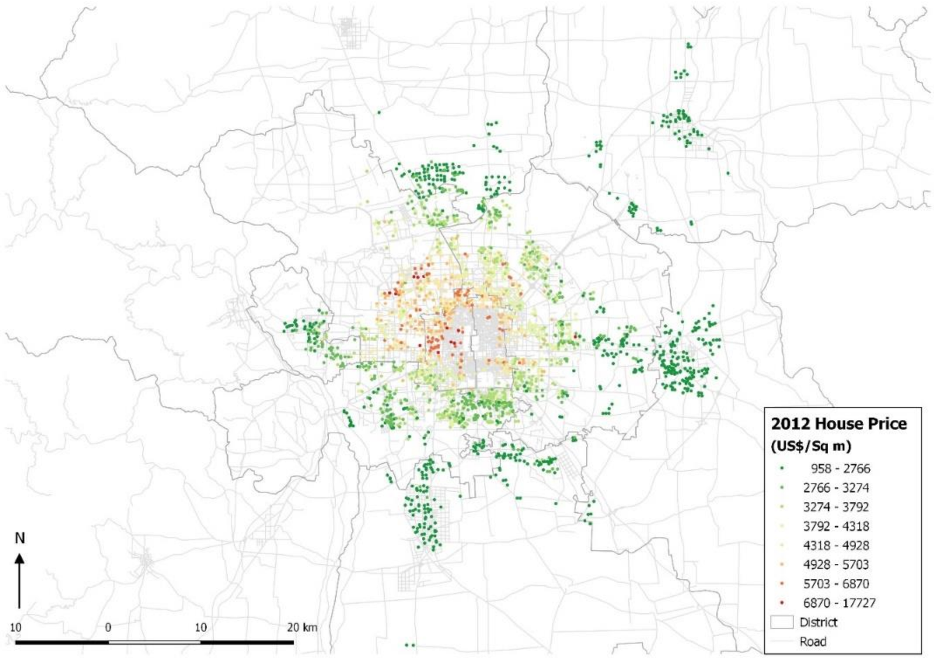


Figure 1.4 Beijing Resale Housing Price (2012 Sample)

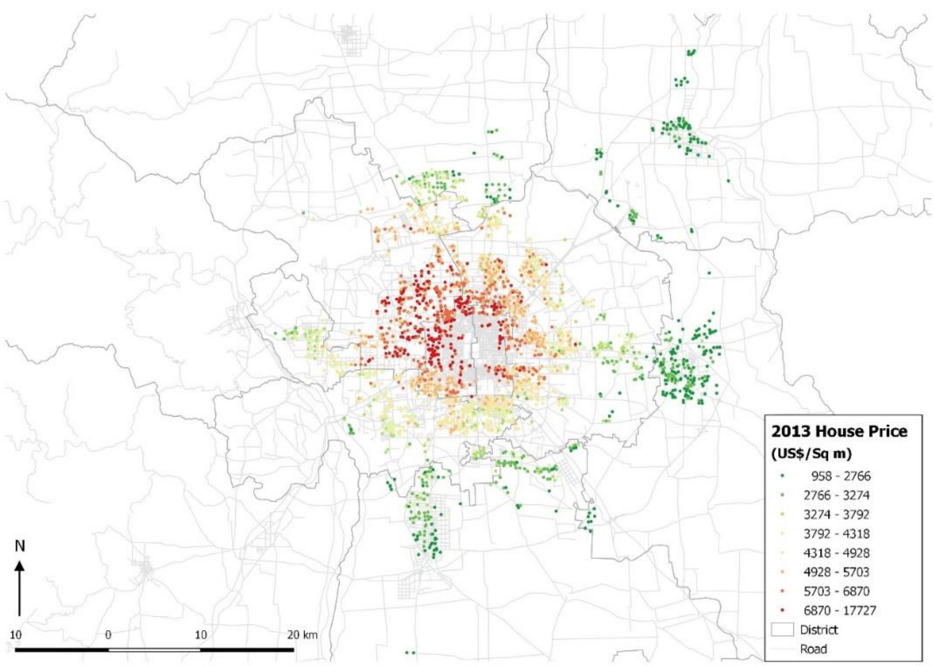


Figure 1.5 Beijing Resale Housing Price (2013 Sample)

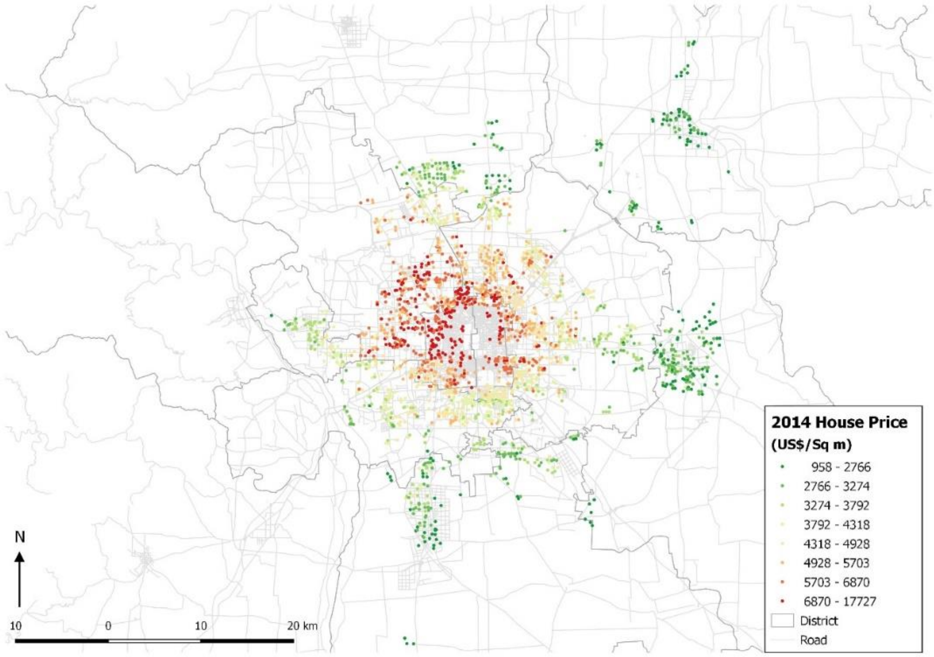


Figure 1.6 Beijing Resale Housing Price (2014 Sample)

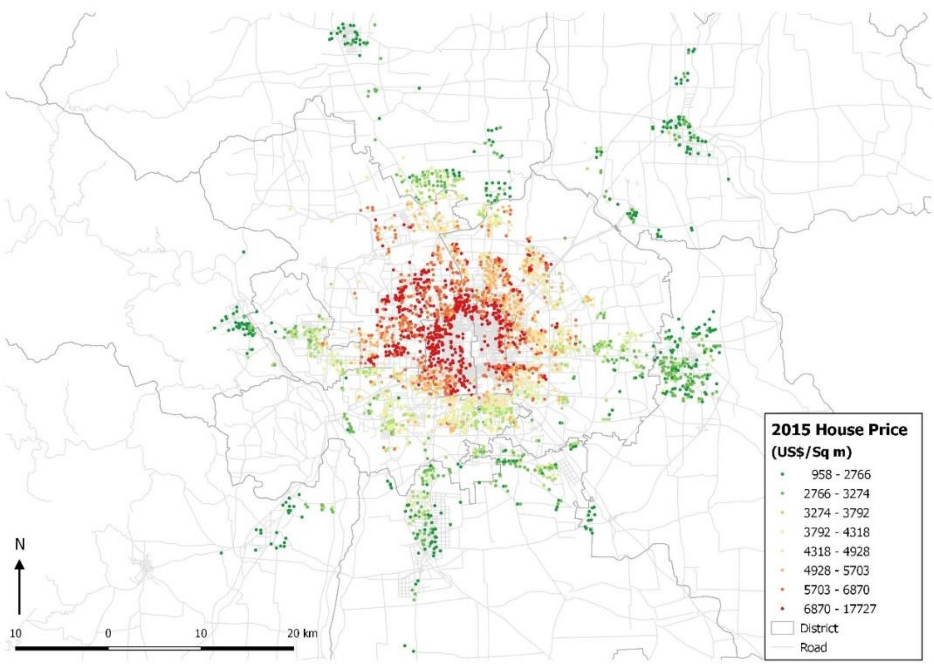


Figure 1.7 Beijing Resale Housing Price (2015 Sample)

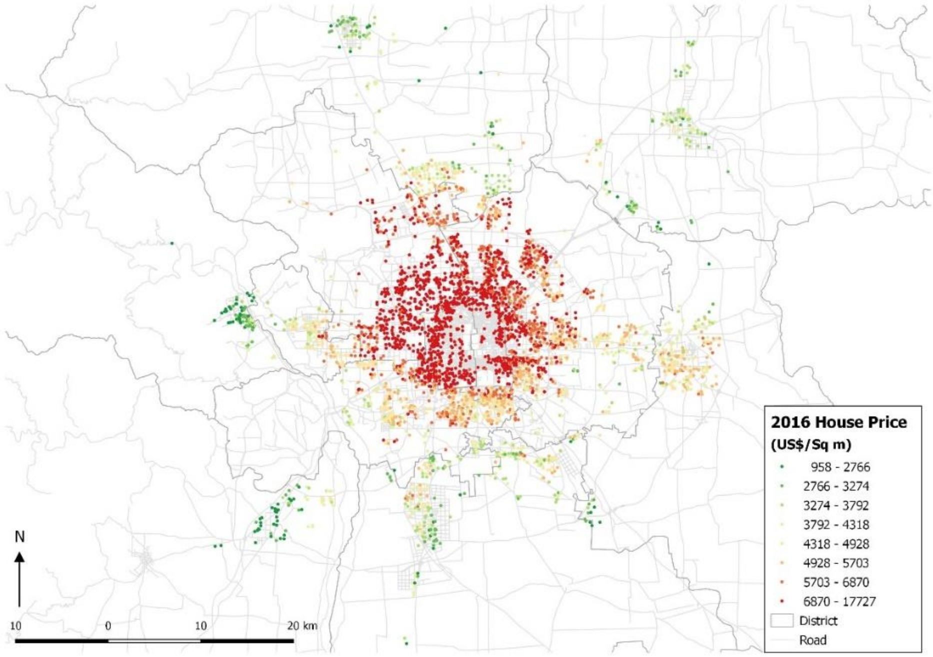


Figure 1.8 Beijing Resale Housing Price (2016 Sample)

The independent variables have four types: structural attributes, building-related attributes, time identifiers, and location identifiers.

Structural attributes contain the number of rooms, including bedrooms (BED), living rooms (LIV), and bathrooms (BATH). The signs of the estimated coefficients of these rooms are expected to be positive. However, the relationship between bedroom and price could be nonlinear. The size of the unit is measured by its floor area (FAREA). The sign of FAREA is uncertain from the literature. However, since this paper uses the unit price, the sign is more likely to be negative. Due to the data limitation, the floor level (FLVL) is in relative levels. There are five categories (1-5) to indicate whether a housing unit is located on the ground floor, low level, middle level, high level, or the top floor of a building respectively. Exposure to the south (SOUTH) is expected to have a positive effect. Averagely, 77% units have at least one window facing south. This percent is significant because not only right south but directions like southwest and southeast are also regarded as an exposure to the south. Facing both south and north (SNEXP) is good for ventilation, which is considered as a premium in Beijing.

For building attributes, the total number of floors (TNFL) represents the height of the building. The effect of a high building is mixed with unexpected signs. Building types (BTYPE) include slab building, tower building and mixed building. Year built (BYEAR) is the year in which the building was completed. The age of buildings has a negative correlation with housing prices in most studies of MRMs (Mok et al., 1995; Bao & Wan, 2004; Jim & Chen, 2009; Jim & Chen, 2010; Shimizu et al., 2010). A negative effect but positive sign is expected. Elevator (ELEV) equals to one if the building is equipped with elevators. The elevator is important for high-rise buildings and thus should be positively correlated with prices. Interaction terms should also be considered. Heating system (HEAT) is available in the dataset. It equals one if there is a central heating system in the building and equals to zero if it is self-heating.

Table 1.1 Descriptive Statistics

Variable	Description	Unit & Type	Mean	St. d.	Min	Median	Max
Dependent Variable							
DTUP	Deflated transaction unit price*	RMB/Square Meter	32564.9	14175.3	4954.5	29632.4	123902.29
Independent Variable							
House-specific Structural Attributes							
BED	Number of bedrooms		1.97	0.71	1	2	5
LIV	Number of living Rooms		1.18	0.50	0	1	2
BATH	Number of bathrooms		1.16	0.38	0	1	3
FAREA	Floor area	Square Meter	81.23	31.09	12.96	74.29	356
FLVL	Floor Level; relative position in the building: 1—Ground Floor 2—Low Position 3—Middle Position 4—High Position 5—Top Floor	Categorical (1-5)	3.10	1.09	1	3	5
SOUTH	Whether the housing unit	Dummy	0.77	0.42	0	1	1

	is exposed to South: 0—No 1—Yes						
SNEXP	Whether the housing unit is exposed to both North and South: 0—No 1—Yes	Dummy	0.47	0.50	0	0	1
Building-related Structural Attributes							
TNFL	Total number of floors		13.30	7.83	2	11	40
BTYPE	Building type: 1—Slab Building 2—Tower Building 3—Mixed Type	Categorical (1-3)	1.63	0.78	1	1	3
BYEAR	Year built		1999.37	8.24	1893	2001	2016
ELEV	Whether equipped with elevator: 0—No 1—Yes	Dummy	0.58	0.49	0	1	1
HEAT	Heating: 0—Self Heating 1—Building Heating	Dummy	0.85	0.35	0	1	1
Time Identifiers							
TYEAR	Transaction year	Categorical (2012-2016)	2014.51	1.43	2012	2015	2016
TMON	Transaction month	Categorical (1-12)	6.61	3.35	1	7	12
Location Identifiers							
LOC	Location	Categorical					

*All prices are deflated into real prices in the year 2009 RMB. CPI indices are from China Statistical Yearbook 2011-2015 and other sources from the government.

1.4.4 Model Specification

This paper applies the hedonic price function in a linear semi-log form[16]: logged housing prices as the dependent variable and no transformation for the independent variables. Different specifications include different variables.

Temporal factors are incorporated as dummies of transaction years and months, or dummies of year-month interactions. The number of bedrooms is treated as in two ways: categorical and continuous. Another categorization is applied to the floor areas. Following

[16] This paper tried Box-Cox transformation of the dependent variable. However, the likelihood-ratio tests cannot reject lamda = 0, i.e., the log specification of the dependent variable. To focus on the comparison of FE and LNE, this paper applies the semi-log form for all models.

Gao and Asami (2011), this paper splits the floor area into 6 categories: [12, 40), [40, 50), [50, 60), [60, 80), [80, 190), and [190, 440). In addition, interactions among floor level, elevator, and building height are explored.

Note that for each model above, both FE and LNE are employed for estimation. Following the recommendation of Solon, Haider, & Wooldridge (2015), both FE (unweighted) and LNE (weighted) estimates and their cluster-robust standard errors (in parentheses) are reported.

1.4.5 Result

Comparing OLS, FE, and LNE

In Table 1.2, the left three columns show the comparison between OLS, FE, and LNE. Some location variables are added for OLS. FE and LNE have no estimates for these location variables but produce consistent estimates for others. Theoretically, FE and LNE should produce similar estimates since both are unbiased and consistent; and the cluster-robust standard errors of FE should be smaller than those of LNE since FE is asymptotically more efficient according to Chapter 2. This is exactly the case in Table 1.2. The estimates of FE are close to LNE but with smaller standard errors. In contrast, the estimates of OLS are very different, suggesting the biasedness from endogeneity issues.

Table 1.2 FE and LNE with Different Time Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	OLS	FE	LNE	FE	LNE	FE	LNE	FE	LNE	FE	LNE
BED	0.0149**	0.0286***	0.0384***	0.0286***	0.0381***	0.0310***	0.0411***	0.0281***	0.0381***	0.0310***	0.0408***
	(0.00563)	(0.00170)	(0.00350)	(0.00173)	(0.00350)	(0.00201)	(0.00362)	(0.00169)	(0.00353)	(0.00203)	(0.00370)
LIV	0.0394***	0.0158***	0.0223***	0.0119***	0.0186***	-0.0120***	-0.00777	0.0168***	0.0241***	-0.0141***	-0.00964*
	(0.00522)	(0.00173)	(0.00484)	(0.00179)	(0.00478)	(0.00227)	(0.00469)	(0.00171)	(0.00492)	(0.00231)	(0.00468)
BATH	0.0459***	0.0124***	0.0124**	0.0126***	0.0119**	0.0120***	0.00759	0.0107***	0.0102*	0.0124***	0.00769
	(0.00778)	(0.00235)	(0.00445)	(0.00241)	(0.00450)	(0.00298)	(0.00426)	(0.00234)	(0.00451)	(0.00302)	(0.00432)
FAREA	-0.00229***	-0.00266***	-0.00322***	-0.00261***	-0.00316***	-0.00230***	-0.00273***	-0.00266***	-0.00322***	-0.00226***	-0.00269***
	(0.000257)	(0.0000854)	(0.000204)	(0.0000861)	(0.000199)	(0.0000915)	(0.000191)	(0.0000852)	(0.000205)	(0.0000923)	(0.000192)
SOUTH	0.0421***	0.0555***	0.0561***	0.0568***	0.0574***	0.0539***	0.0560***	0.0550***	0.0553***	0.0541***	0.0564***
	(0.00421)	(0.00136)	(0.00244)	(0.00144)	(0.00251)	(0.00193)	(0.00327)	(0.00132)	(0.00222)	(0.00198)	(0.00340)
SNEXP	-0.00105	0.00660***	0.00527*	0.00465***	0.00376	0.000362	-0.00226	0.00766***	0.00683**	-0.000928	-0.00317
	(0.00386)	(0.00126)	(0.00244)	(0.00134)	(0.00250)	(0.00181)	(0.00326)	(0.00121)	(0.00232)	(0.00186)	(0.00329)
TNFL	0.000699	-0.00197***	-0.00245	-0.00192***	-0.00224	-0.00186***	-0.00204	-0.00190***	-0.00244	-0.00185***	-0.00194
	(0.00101)	(0.000510)	(0.00133)	(0.000491)	(0.00126)	(0.000510)	(0.00123)	(0.000520)	(0.00137)	(0.000506)	(0.00121)
BYEAR	0.000833	0.00296***	0.00375***	0.00303***	0.00378***	0.00356***	0.00429***	0.00296***	0.00378***	0.00359***	0.00428***
	(0.000637)	(0.000238)	(0.000435)	(0.000234)	(0.000403)	(0.000279)	(0.000438)	(0.000238)	(0.000463)	(0.000284)	(0.000460)
ELEV	0.122***	0.0223***	0.0226*	0.0220***	0.0228**	0.0153***	0.0158*	0.0212***	0.0221*	0.0154***	0.0166*
	(0.0126)	(0.00416)	(0.00888)	(0.00416)	(0.00869)	(0.00454)	(0.00781)	(0.00417)	(0.00894)	(0.00459)	(0.00793)
2.FLVL	-0.0152***	-0.00833***	-0.00861***	-0.00700***	-0.00672**	-0.00954***	-0.00659*	-0.00792***	-0.00857***	-0.00961***	-0.00627
	(0.00266)	(0.00152)	(0.00250)	(0.00160)	(0.00252)	(0.00221)	(0.00332)	(0.00145)	(0.00245)	(0.00223)	(0.00335)
3.FLVL	-0.00747**	0.00127	0.00282	0.00296	0.00544	-0.000482	0.00373	0.00137	0.00264	-0.000222	0.00488
	(0.00259)	(0.00158)	(0.00305)	(0.00166)	(0.00307)	(0.00211)	(0.00348)	(0.00152)	(0.00294)	(0.00212)	(0.00352)
4.FLVL	-0.0188***	-0.00559**	-0.00288	-0.00320	0.000459	-0.00709**	-0.00232	-0.00560**	-0.00334	-0.00648**	-0.000879
	(0.00287)	(0.00181)	(0.00365)	(0.00188)	(0.00374)	(0.00234)	(0.00385)	(0.00172)	(0.00345)	(0.00237)	(0.00395)
5.FLVL	-0.0570***	-0.0494***	-0.0475***	-0.0456***	-0.0434***	-0.0463***	-0.0428***	-0.0492***	-0.0475***	-0.0442***	-0.0406***
	(0.00321)	(0.00208)	(0.00439)	(0.00216)	(0.00437)	(0.00262)	(0.00457)	(0.00200)	(0.00429)	(0.00267)	(0.00462)
2.BTYPE	-0.104***	-0.0164***	-0.0118	-0.0169***	-0.0142	-0.0215***	-0.0223*	-0.0167***	-0.0118	-0.0218***	-0.0239*
	(0.0135)	(0.00432)	(0.0110)	(0.00431)	(0.0105)	(0.00487)	(0.0108)	(0.00437)	(0.0112)	(0.00491)	(0.0108)

3.BTYPE	-0.0446***	-0.0153***	-0.00960	-0.0163***	-0.0115	-0.0189***	-0.0162*	-0.0146***	-0.00853	-0.0191***	-0.0176**
	(0.0116)	(0.00318)	(0.00604)	(0.00323)	(0.00588)	(0.00418)	(0.00645)	(0.00318)	(0.00611)	(0.00422)	(0.00653)
HEAT	0.0521***	-0.00801	-0.00334	-0.00650	-0.00174	-0.00767	-0.0103	-0.00657	-0.000349	-0.00678	-0.00842
	(0.0125)	(0.0108)	(0.0156)	(0.0111)	(0.0159)	(0.0131)	(0.0160)	(0.0110)	(0.0158)	(0.0134)	(0.0165)
Locational	Yes										
TYEAR?	Yes	Yes	Yes	Yes	Yes						
TMON?	Yes	Yes	Yes			Yes	Yes				
TYEAR* TMON?								Yes	Yes		
N	262533	262533	262533	262533	262533	262533	262533	262533	262533	262533	262533
adj. R^2	0.755	0.768	0.766	0.673	0.674	0.086	0.089	0.831	0.831	0.029	0.035
F	1446.4	3375.3	1845.9	3482.1	1765.3	464.5	342.0	2117.5	1805.6	132.7	60.46

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Year Dummies

For the year dummies, all three estimators produce similar results (see Table 1.3 and Figure 1.9). This is largely because the year factor is relatively independent and immune to the endogeneity issues. The year effect on housing prices is very large. Comparing to the year 2012, transactions in 2013 have on average 30% higher prices. Prices in 2014 fall a little bit but are still 27% higher than 2012. Prices in 2015 are at similar level with 2013. In 2016, the housing price almost doubles the level of 2013.

Table 1.3 Estimates of Year Dummies

	(1) OLS	(2) FE	(3) LNE
2013.TYEAR	0.318^{***}	0.321^{***}	0.321^{***}
	(0.00232)	(0.00150)	(0.00270)
2014.TYEAR	0.278^{***}	0.265^{***}	0.267^{***}
	(0.00308)	(0.00198)	(0.00325)
2015.TYEAR	0.318^{***}	0.303^{***}	0.298^{***}
	(0.00314)	(0.00230)	(0.00412)
2016.TYEAR	0.606^{***}	0.602^{***}	0.602^{***}
	(0.00406)	(0.00338)	(0.00683)
N	262533	262533	262533
adj. R^2	0.755	0.768	0.766
F	1446.4	3375.3	1845.9

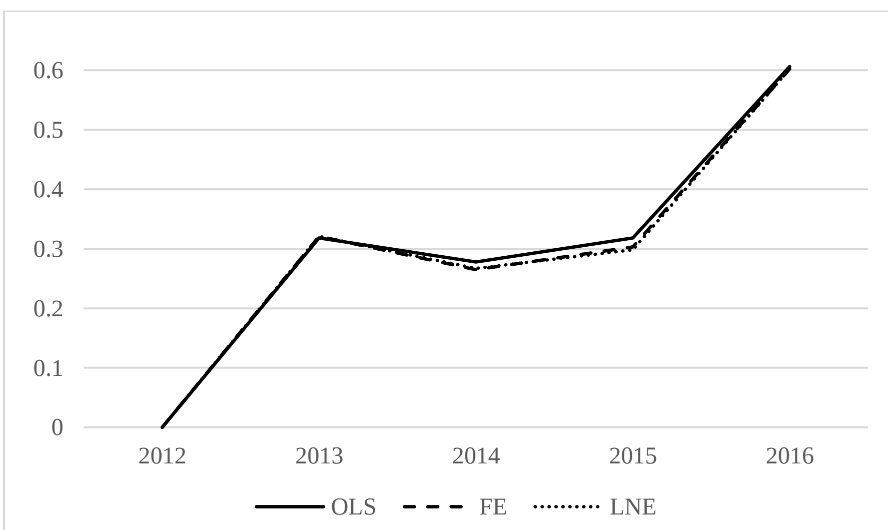


Figure 1.9 Estimates of Year Dummies

Model (4) to (11) in Table 1.2 use different specification of the time variables. Focusing on FE models. When the model excludes all temporal factors as in Model (10), only around 2.9% of the price variation is explained, which could be regarded as the explanatory power of the structural variables. When adding the month dummies (Model 6), the ratio increases to 8.6%; the difference 5.7% is the explanatory power of the month variable. When only adding the year dummies (Model 4), the ratio becomes 67.3%, which indicates a 64.4% of the variation is explained by the year changes. At last, the best model seems to be (8) which use the interaction between years and months. Note that the parameter estimates are stable across different model specifications.

Structural and Building Variables

According to the results of Model (2), for housing-specific variables, one more unit of bedrooms (BED), living rooms (LIV), and bathrooms (BATH) are associated with an increase in housing price per square meter by 2.86%, 1.58%, and 1.24%, respectively. The floor area (FAREA) is negatively associated with housing prices by 0.3%, which is a quite small magnitude as compared to other housing attributes and very similar to the findings of Mok et al. (1995) in Hong Kong.

In terms of the magnitude of the coefficients, the exposure to the south (SOUTH) is the most influential positive attributes by 5.5%. The effect of being exposed to both south and north (SNEXP) is positive but with a smaller value of 0.66%. There is a slight ground premium. Compared to the ground floor (FLVL), low-level position corresponds to a price depreciation by 0.83%, high position by 0.56%, and the top floor by 4.9%. The effect of the middle position is not significantly different from that of the ground floor.

For building-specific variables, the total number of floors (TNFL) have a slightly negative relationship to housing prices by 0.2%. The built year (BYEAR) effect is positive. Housing prices increase by 0.3% if a building is one year newer. Being equipped with elevators (ELEV) has a significantly positive effect by 2.2%. For building types (BTYPE), the slab building has a larger effect of 1.5% than the tower building and the mixed building. The heating type (HEAT) has no significant relationship with housing prices.

Floor Level, Total Floors, and Elevator

In Table 1.2, one striking finding is the ground floor premium, which means residents prefer the ground floor to all other levels. This is a huge contrast to many MRMs where people favor the top floors. This paper further investigates the interaction between floor levels, elevators, and total floors.

Table 1.4 shows the results of FE models with interaction terms. Model (12) includes the interaction term between the floor level and the elevator. Model (13) adds the total number of floors into the interaction. Most interactive terms have significant estimates. Floor level coefficients in different models are summarized in Table 1.5. The two columns of Model (12) in Table 1.5 and Figure 1.10 show that the ground premium found in model (2) gets larger when the building has no elevator but disappears for the buildings with elevators. In the latter case, the high position (not the top floor) is mostly favored. The columns of Model (13) and Figure 1.11 show that for buildings without elevators, the higher the building is, the less the top floor would be valued. In contrast, the low and middle positions are more valued with the building gets higher. For no-elevator buildings higher than 9 floors, both low and middle positions of the building are preferred to the ground level. For buildings with elevators, as the total number of floors increases, the ground level premium decreases comparing to all other floor levels. When the total floor is larger than 13, all other floor levels are preferred to the ground one.

Analysis with interaction terms tells us that the ground level premium exists in Beijing for certain types of buildings. For no-elevator buildings less than 7 floors or with-elevator buildings less than 10 floors, the ground floor has the highest effects on housing prices. The ground level in a higher building is exposed to larger negative externality brought by the denser living environment.

Table 1.4 FE Models with Interaction Terms

	(2) FE	(12) FE with Interaction	(13) FE with Interaction
TNFL	-0.00197***	-0.00195***	-0.00598***
	(0.000510)	(0.000511)	(0.00150)
ELEV	0.0223***	0.0670***	0.0188*
	(0.00416)	(0.00458)	(0.00924)
2.FLVL	-0.00833***	-0.00815***	-0.0551***
	(0.00152)	(0.00180)	(0.00715)
3.FLVL	0.00127	-0.0131***	-0.0432***
	(0.00158)	(0.00173)	(0.00726)
4.FLVL	-0.00559**	-0.0389***	-0.0403***
	(0.00181)	(0.00216)	(0.00755)
5.FLVL	-0.0494***	-0.0786***	-0.0426***
	(0.00208)	(0.00225)	(0.00832)
1.ELEV * 1.FLVL		-0.0967***	
		(0.00350)	
1.ELEV * 2.FLVL		-0.0684***	
		(0.00254)	
1.ELEV * 3.FLVL		-0.0493***	
		(0.00234)	
1.ELEV * 4.FLVL		-0.0207***	
		(0.00242)	
0.ELEV * 2.FLVL * TNFL			0.00776***
			(0.00120)
0.ELEV * 3.FLVL * TNFL			0.00523***
			(0.00123)
0.ELEV * 4.FLVL * TNFL			0.000558
			(0.00131)
0.ELEV * 5.FLVL * TNFL			-0.00594***
			(0.00145)
1.ELEV * 2.FLVL* TNFL			0.00424**
			(0.00156)
1.ELEV * 3.FLVL * TNFL			0.00436**
			(0.00156)
1.ELEV * 4.FLVL * TNFL			0.00432**
			(0.00155)
1.ELEV * 5.FLVL * TNFL			0.00330*
			(0.00156)
N	262533	262533	262533
adj. R^2	0.768	0.771	0.771
F	3375.3	3090.8	2812.4

** Other variables are those in Table 1.2. The interaction models have very similar estimates for them.
Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.5 Interaction among Floor Level, Elevator, and Total Floor

Variable	(2) FE	(12) No Elevator	(12) Elevator	(13) No Elevator	(13) Elevator
Ground Floor	0	0	0	0	0
Low Position	-0.00833	-0.00815	0.0201	-0.0551 + 0.00776 * TNFL	-0.0551 + 0.00424 * TNFL
Middle Position	Not significant	-0.0131	0.0343	-0.0432 + 0.00523 * TNFL	-0.0432 + 0.00436 * TNFL
High Position	-0.00559	-0.0389	0.0371	- 0.0403	-0.0403 + 0.00432 * TNFL
Top Floor	-0.0494	-0.0786	0.0181	-0.0426 - 0.00594 * TNFL	-0.0426 + 0.00330 * TNFL

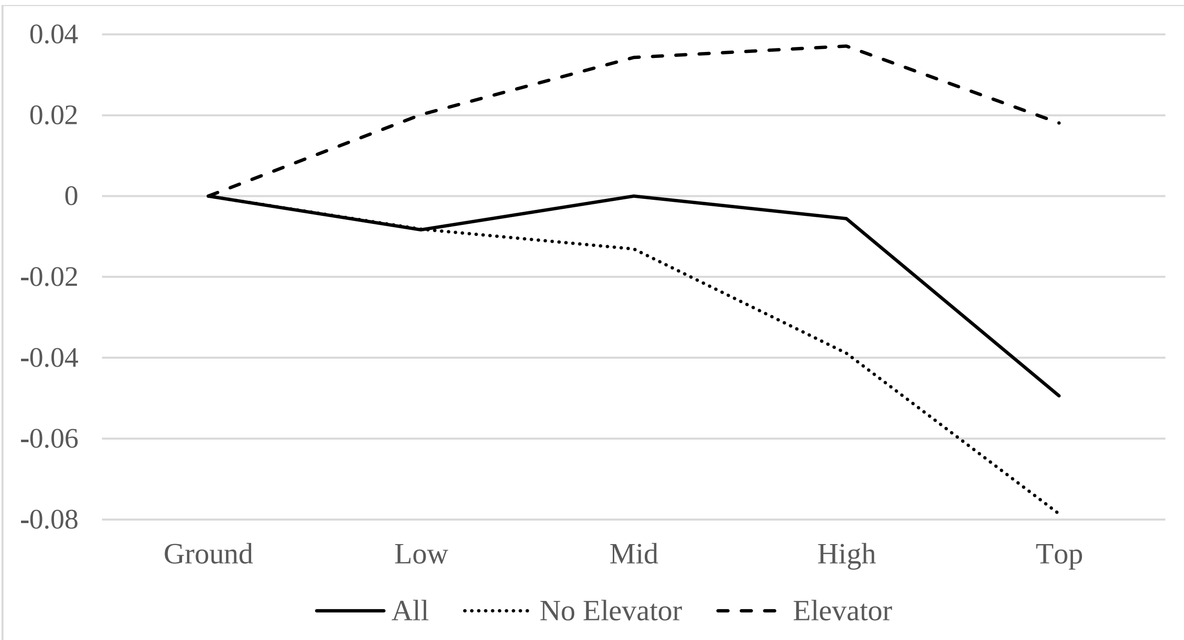


Figure 0.10 Vertical Gradient (with and without Elevator)

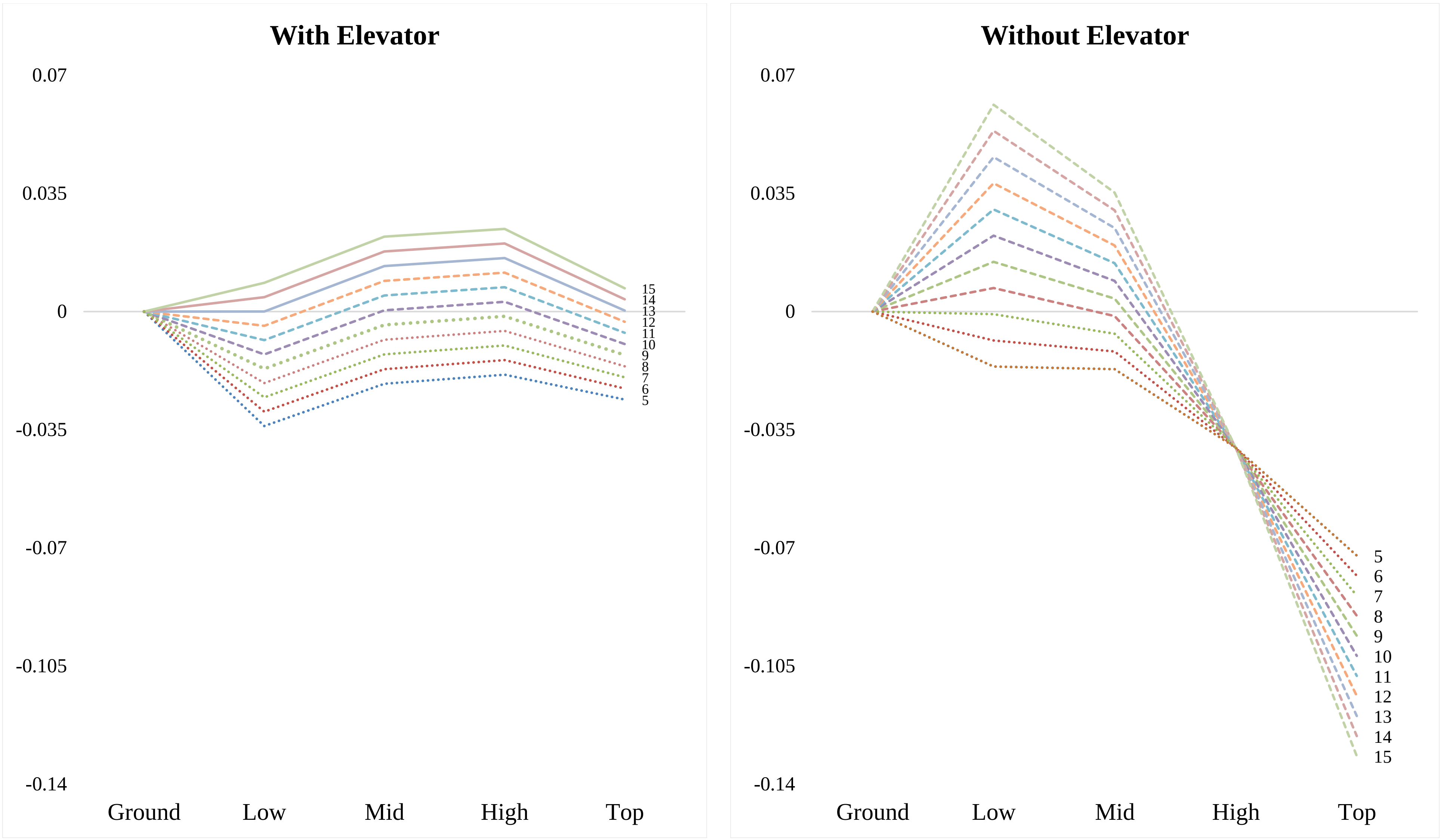


Figure 1.11 Vertical Gradient (with Elevator and Total Floors from 5 to 15)

Categorical Bedroom Numbers and Floor Areas

Lastly, this paper specifies the number of bedrooms as dummy variables and the floor area as a categorical variable. In Table 1.6, Model (2) is listed as a benchmark. Model (14) uses bedroom dummies, model (15) uses floor area categories, model (16) includes both, and model (17) excludes the variable of the bedroom from the regression. First, all other variables, except for other room numbers and the elevator, have stable parameter estimates across all specifications. Second, comparing model (14) and (16), the estimates of the bedroom dummies are inconsistent[17]. Without controlling for floor area categories (model 14), the 1-bedroom house is least favorable, and the 4-bedroom house has the largest effect on prices; while controlling for the floor area categories (model 16), the 2-bedroom house is the favorite, and houses with more bedrooms are less favored. Similar inconsistency can be found in models (2) and (15); the estimate of the original bedroom variable is affected and changes from positive to negative. In contrast, the floor area categories have more stable estimates. In model (17), when the variable of the bedroom is excluded from the regression, the results are still similar. Different from the findings from Gao and Asami (2011), who argues that [50, 60) and [80, 190) are more highly valued, this paper shows the smaller sizes are preferred.

[17] This suggests a strong correlation between the number of bedroom and the floor area, which is in fact 0.7052.

Table 1.6 Bedroom Dummies and Floor Area Categories

	(2)	(14)	(15)	(16)	(17)
	FE	Bed_D	FAREA_C	Bed_D + FAREA_C	FAREA_C No Bed
BED	0.0286***		-0.00211		
	(0.00170)		(0.00150)		
2.BED		0.0372***		0.0282***	
		(0.00235)		(0.00202)	
3.BED		0.0574***		0.0122***	
		(0.00338)		(0.00283)	
4.BED		0.0856***		-0.0272***	
		(0.00671)		(0.00617)	
5.BED		0.0631***		-0.0911***	
		(0.0189)		(0.0202)	
FAREA	-0.00266***	-0.00267***			
	(0.0000854)	(0.0000849)			
[40, 50)			-0.0534***	-0.0557***	-0.0536***
			(0.00468)	(0.00449)	(0.00465)
[50, 60)			-0.0761***	-0.0915***	-0.0772***
			(0.00501)	(0.00490)	(0.00479)
[60, 80)			-0.0940***	-0.115***	-0.0959***
			(0.00544)	(0.00534)	(0.00500)
[80, 190)			-0.128***	-0.158***	-0.131***
			(0.00640)	(0.00625)	(0.00555)
[190, 440)			-0.207***	-0.211***	-0.212***
			(0.0125)	(0.0134)	(0.0123)
LIV	0.0158***	0.0146***	0.00379*	0.00320	0.00379*
	(0.00173)	(0.00168)	(0.00178)	(0.00175)	(0.00178)
BATH	0.0124***	0.0166***	-0.0430***	-0.0355***	-0.0443***
	(0.00235)	(0.00223)	(0.00286)	(0.00272)	(0.00311)
SOUTH	0.0555***	0.0547***	0.0508***	0.0495***	0.0509***
	(0.00136)	(0.00135)	(0.00145)	(0.00141)	(0.00144)
SNEXP	0.00660***	0.00594***	0.00646***	0.00570***	0.00605***
	(0.00126)	(0.00125)	(0.00133)	(0.00130)	(0.00132)
TNFL	-0.00197***	-0.00198***	-0.00222***	-0.00220***	-0.00223***
	(0.000510)	(0.000508)	(0.000537)	(0.000530)	(0.000537)
BYEAR	0.00296***	0.00298***	0.00218***	0.00247***	0.00222***
	(0.000238)	(0.000238)	(0.000237)	(0.000236)	(0.000236)
ELEV	0.0223***	0.0228***	0.0106*	0.0129**	0.0108*
	(0.00416)	(0.00414)	(0.00458)	(0.00441)	(0.00457)
2.FLVL	-0.00833***	-0.00841***	-0.00906***	-0.00893***	-0.00902***
	(0.00152)	(0.00152)	(0.00156)	(0.00156)	(0.00155)

3.FLVL	0.00127	0.00117	0.000229	0.000341	0.000281
	(0.00158)	(0.00159)	(0.00163)	(0.00163)	(0.00162)
4.FLVL	-0.00559**	-0.00563**	-0.00651***	-0.00615***	-0.00648***
	(0.00181)	(0.00181)	(0.00186)	(0.00185)	(0.00185)
5.FLVL	-0.0494***	-0.0493***	-0.0500***	-0.0494***	-0.0500***
	(0.00208)	(0.00207)	(0.00223)	(0.00220)	(0.00223)
2.BTYPE	-0.0164***	-0.0167***	-0.00915	-0.00956*	-0.00921
	(0.00432)	(0.00432)	(0.00471)	(0.00465)	(0.00471)
3.BTYPE	-0.0153***	-0.0153***	-0.0110**	-0.0110**	-0.0110**
	(0.00318)	(0.00318)	(0.00344)	(0.00338)	(0.00344)
2013.TYEAR	0.321***	0.321***	0.320***	0.320***	0.320***
	(0.00150)	(0.00151)	(0.00152)	(0.00152)	(0.00152)
2014.TYEAR	0.265***	0.265***	0.265***	0.265***	0.265***
	(0.00198)	(0.00198)	(0.00198)	(0.00197)	(0.00198)
2015.TYEAR	0.303***	0.303***	0.302***	0.301***	0.302***
	(0.00230)	(0.00230)	(0.00230)	(0.00230)	(0.00231)
2016.TYEAR	0.602***	0.602***	0.601***	0.600***	0.601***
	(0.00338)	(0.00338)	(0.00339)	(0.00338)	(0.00339)
N	262533	262533	262533	262533	262533
adj. R^2	0.768	0.769	0.758	0.760	0.758
F	3375.3	3098.9	2930.7	2774.7	3016.1

1.5 CONCLUSION AND DISCUSSION

This paper aims to explain the vertical variation in housing prices. The contribution of this paper is two-fold: empirical evidence and methodological exploration.

Empirically, this paper uses a large dataset of resale housing transaction records in Beijing from 2012 to 2016 and shows that temporal factors and the structural attributes together account for around 76% of the variation in housing prices. The structural attributes only account for about 3% of the variation; while the temporal factors, years and months, explain the rest 73%. The most important structural factor in Beijing is the exposure to the south, leading to around 5% increase in housing prices. Beijing features a ground floor premium, which is conditional on the building's total number of floors and whether has an elevator. The floor areas are examined in both continuous and categorical terms. This small size preference contradicts the existing findings of Beijing housing markets in the literature.

Methodologically, this paper develops a least nugget effect estimator (LNE) which is equivalent to the pairwise differences estimator (PD) and generalizes the fixed effects estimator (FE) by weighting group sizes. This paper provides empirical evidence for the efficiency of FE over PD (and thus LNE) given large samples.

There are two limitations of this paper. The first is the assumption that there are no spatial and temporal dependencies of housing prices. This paper controls the within-group correlation by applying cluster-robust standard errors but at the same time assumes no between-group or inter-temporal dependencies. This may not be valid in housing markets where spatial and temporal autocorrelations widely exist. The fixed effects estimator has also been criticized for not fully controlling for spatial dependency (Anselin & Arribas-

Bel, 2013). Scholars have taken both spatial and temporal dependencies into consideration to control for the spatio-temporal autocorrelation effects in MRMs (Sun et al., 2005). The assumption may potentially bias the standard errors and the inference of the parameter estimates. The second limitation is to assume that the numbers of observations at each location are exogenous. This may not be valid since the unobservable characteristics of houses sold frequently could differ from those sold infrequently (Englund, Quigley, & Redfearn, 1999). The resultant selection bias might undermine the findings of this paper.

CHAPTER 2

PAIRWISE DIFFERENCES ESTIMATORS IN LINEAR FIXED EFFECTS MODELS

2.1 INTRODUCTION

Empirical studies with clustered data commonly apply linear fixed effects models which assume that the unobserved cluster effects may be correlated with regressors. Under this favorable assumption, a fixed effects estimator (hereafter FE) is the default choice for estimation due to its consistency achieved by eliminating the unobserved cluster effects via mean-differencing transformation. In the econometric literature, there are two other estimators applying this "elimination-by-transformation" strategy but with different transformation approaches (other than mean-differencing): one is the first differences estimator (hereafter FD), transforming the data by taking successive differences for individual time series (Wooldridge, 2010); the other is called pairwise differences estimator (hereafter PD[18]), which transforms the data by differencing all pairs of observations in each group (Honoré & Powell, 1997). FD has been intensively studied as an alternative to FE in a panel setting, yet it cannot be directly applied to clustered data since within-cluster observations seldom have any orderings. In contrast, PD seems to be a feasible alternative to FE in a clustered setting, however, its property and comparison with FE are still understudied. This paper fills this gap by exploring the properties of PD and its relationship with FE in linear fixed effects models.

The pairwise differences estimator (PD) features "pairwise-differencing" or

[18] In this paper, the abbreviations FE and PD only refer to the least squared estimators in linear fixed effects models under corresponding transformations (mean-differencing and pairwise differencing).

“pairwise comparing” approach, which has been studied in the literature of statistics and econometrics. In statistics, this approach was discussed as an alternative to mean-differencing. Von Neumann, Kent, Bellinson, & Hart (1941) cited that the sum of the squared pairwise differences and the sum of the squared deviations from the mean generate the same estimate for data without group structures. The widely applied spatial statistics variogram and semivariogram (Matheron, 1963; Cressie, 1993) are also based on pairwise comparing the observations. In econometrics, this approach has been applied in the fields of robust estimation (Stromberg, Hössjer, & Hawkins, 2000; Baltagi & Bresson, 2015), truncated and censored regressions (Honoré, 1992; Honoré & Powell, 1994) and nonlinear models (Honoré & Powell, 1997).

In the context of linear fixed effects models, this pairwise differencing approach was firstly described in Honoré and Powell (1997) without further discussion; and it was not until Aquaro and Čížek (2013) that an estimation using PD was documented. Aquaro and Čížek (2013) compared the Monte Carlo performance of the least squared estimator and some robust estimators under mean-differencing and pairwise-differencing transformations. Their results showed the advantage of applying pairwise differencing transformation to robust estimators. However, for the least squared estimators of two transformations (i.e. FE and PD)—which are the focus of this paper, their simulation produced exact same results in both estimation and inference using original data, and only slight differences after adding some outliers.

This paper will show that this equivalency between FE and PD are essentially due to the balancedness of the panel data used by Aquaro and Čížek (2013), i.e. the number of observations for each cluster (or individual in panel) is the same. Adding outliers unevenly made the data slightly unbalanced, and thus led to the slight differences in the results. A main finding of this paper is that PD generalizes FE by weighting each observation by its cluster size, making the balancedness a necessary condition for the equivalency between PD and FE. Another necessary condition is also identified—inspired by the concept of mediant and weighted mediant in the number theory, that is, equal local estimates for each cluster. If the local OLS estimates for each cluster are the same, the global FE and PD estimates would be equal even under severe unbalancedness.

In clustered data, these two necessary conditions (unequal cluster sizes and unequal local estimates) are commonly satisfied, and thus FE and PD always perform differently. One (and, as far as I know, the only one) example is provided by Desai and Begg (2008). They compared FE and PD in studying the association between children's head circumference at birth and intelligence at age 7 years. Differences between FE and PD were documented: the estimate by FE is 1.18 with 95% confidence interval (0.96, 1.4) and the corresponding by PD is 0.69 with (0.41, 0.97). Their data is unbalanced (although slightly), containing 9147 children observations clustered in 4170 families and at least 2 children per family. In addition, the smaller confidence interval of FE suggests its efficiency over PD. It is worth mentioning that the number of clusters in their study is large (e.g. 4170 families) relative to cluster sizes (e.g. 2 or more children), which is not always the case for many clustered data studies, for example a study with unequal number of counties (over hundreds) clustered in a few states (less than 50). This paper will conduct Monte Carlo simulation and compare the performance of FE and PD in different cluster structures, including different numbers of clusters and levels of unbalancedness.

As shown later in this paper, since the asymptotic variance matrices of FE and PD are hard to compare, this paper conducts Monte Carlo simulation to examine their performance in efficiency. The simulation employs two common variance estimators: the default i.i.d. standard errors and the cluster-robust standard errors (CRVE[19]). However, these two estimators are known to be downward biased and lead to over-rejection in certain situations: for example, the i.i.d. standard error is downward biased when the errors are not i.i.d.; and the CRVE over-rejects when the number of clusters is small (Bertrand, Duflo, & Mullainathan, 2004; Cameron et al., 2008; MacKinnon & Webb, 2017; Wooldridge, 2003), both of which are cases of my interest. Thus, rejection rates are additionally reported to validate the findings of PD and FE.

Based on 10,000 simulations, the paper shows that: 1) PD and FE provide similarly accurate estimates, which is as expected since both of them are unbiased and consistent; 2) When the number of clusters is large, FE is a more efficient choice than PD in terms of smaller CRVE; 3) Although PD generates smaller CRVE when the number of clusters is small (less than 10), in most cases, the over-rejection issue of CRVE undermines its validity; 4) PD could provide valid efficiency improvement over FE: by 1.7%-9.4% in CRVE when using heteroscedastic clustered data with few clusters (6 to 10) and with very high within-cluster correlation of regressors; by 1%-5% in default i.i.d. standard errors when using homoscedastic panel data with zero or very high within-cluster error correlations.

In addition to the investigation of PD as an alternative to FE, this paper contributes to the literature in two other aspects. First, I introduce the mediant and weighted mediant

[19] CRVE represents cluster-robust variance estimate. This paper uses CRVE to denote the cluster-robust standard errors following the practice of Cameron, Gelbach, & Miller (2008).

to explain the estimators' relationship. This pair of concepts provide a new perspective to view the relationship between local and global estimates: the global estimate is a mediant of local estimates. Second, the fact that PD is a weighted version of FE provides another answer to the question of Solon, Haider, & Wooldridge (2015)—"What are we weighting for?": weighting by cluster sizes to individual data and then applying fixed effects estimation is equivalent to pairwise comparing within-group individuals. In what follows, Section 2 introduces PD and its properties, Section 3 compares PD and FE from the perspective of the mediant theory, Section 4 conducts Monte Carlo simulation and summarizes the results, and Section 5 discusses and concludes with key findings.

2.2 PAIRWISE DIFFERENCES ESTIMATOR

2.2.1 Model and Estimator

The linear model to be estimated is:

$$y_{ig} = \boldsymbol{x}_{ig}'\boldsymbol{\beta} + \alpha_g + \varepsilon_{ig} \qquad \text{for } i = 1, \ldots, N_g;\ g = 1, \ldots, G \tag{2.1}$$

where g indexes the clusters and i indexes observations in each cluster; G is the number of clusters, N_g is the cluster size, and thus the total number of observation is $N = \sum_g N_g$; $\boldsymbol{x}_{ig}$ is a $1 \times k$ column vector of explanatory variables that vary within cluster; $\boldsymbol{\beta}$ is a $k \times 1$ column vector of parameters; α_g is the unobserved cluster effect and may correlate with $\boldsymbol{x}_{ig}$; and ε_{ig} is the idiosyncratic error term.

Transform (2.1) by taking pairwise differences within each cluster g to eliminate the unobserved cluster effect α_g:

$$y_{ig} - y_{jg} = \left(\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg}\right)'\boldsymbol{\beta} + (\varepsilon_{ig} - \varepsilon_{jg}). \tag{2.2}$$

For each cluster g, the set of all pairs of observations P_g is defined as

$$P_g \equiv \{(i,j)|i \neq j; both\ i,j\ are\ in\ cluster\ g\},$$

and the number of pairs is $|P_g| = N_g(N_g - 1)/2$. Thus, the total number of pairs in the sample is,

$$|P| = \sum_{g=1}^{G} |P_g| = \sum_{g=1}^{G} N_g(N_g - 1)/2. \tag{2.3}$$

The pairwise differences estimator (PD) is defined as the OLS estimator of (2.2), i.e.

$$\widehat{\boldsymbol{\beta}}_{PD} \equiv \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{(i,j)\in P_g} (e_{ig} - e_{jg})^2, \tag{2.4}$$

where $e_{ig} = y_{ig} - \boldsymbol{x}_{ig}'\widehat{\boldsymbol{\beta}}_{PD} - \alpha_g$ and $e_{ig} - e_{jg} = (y_{ig} - y_{jg}) - (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})'\widehat{\boldsymbol{\beta}}_{PD}$. This is first proposed by Honoré and Powell (1994) as a special case when there is neither censoring nor truncation and the loss function is in squared form. The first-order condition yields:

$$\widehat{\boldsymbol{\beta}}_{PD} = \left[\sum_{g=1}^{G} \sum_{(i,j)\in P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})'\right]^{-1} \sum_{g=1}^{G} \sum_{(i,j)\in P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(y_{ig} - y_{jg}). \tag{2.5}$$

The PD estimator is unbiased under the exogeneity condition, i.e. $E[\varepsilon_{ig} - \varepsilon_{jg} | \boldsymbol{x}_{ig} - \boldsymbol{x}_{jg}] = 0$, which is the same for FE estimator:

$$\begin{aligned}
E(\widehat{\boldsymbol{\beta}}_{PD}|\boldsymbol{X}) &= E\left\{\left[\sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})'\right]^{-1} \left[\sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(y_{ig} - y_{jg})\right] |\boldsymbol{X}\right\} \\
&= \boldsymbol{\beta} + E\left\{\left[\sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})'\right]^{-1} \left[\sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\varepsilon_{ig} - \varepsilon_{jg})\right] |\boldsymbol{X}\right\} \\
&= \boldsymbol{\beta} + \left[\sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})'\right]^{-1} \sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg}) E[(\varepsilon_{ig} - \varepsilon_{jg})|\boldsymbol{X}]
\end{aligned}$$

The PD estimator is consistent if $plim|P|^{-1} \sum_g \sum_{P_g} (\boldsymbol{x}_{ig} - \boldsymbol{x}_{jg})(\varepsilon_{ig} - \varepsilon_{jg}) = 0$, which happens if $|P| \to \infty$ and the exogeneity condition holds.

2.2.2 PD Generalizes FE

Recall that a fixed effects estimator (FE) is defined as the OLS estimator of the mean-differenced model

$$y_{ig} - \bar{y}_g = \left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)'\boldsymbol{\beta} + (\varepsilon_{ig} - \bar{\varepsilon}_g),$$

$$\widehat{\boldsymbol{\beta}}_{FE} \equiv \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (e_{ig} - \bar{e}_g)^2, \tag{2.6}$$

$$\widehat{\boldsymbol{\beta}}_{FE} = \left[\sum_g \sum_i \left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)\left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)'\right]^{-1} \sum_g \sum_i \left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)\left(y_{ig} - \bar{y}_g\right),$$

where $\bar{e}_g = \frac{1}{N_g}\sum_i e_{ig} = \bar{y}_g - \overline{\boldsymbol{x}}_g{}'\widehat{\boldsymbol{\beta}}_{PD} - \alpha_g$ with $\bar{y}_g = \frac{1}{N_g}\sum_i y_{ig}$ and $\overline{\boldsymbol{x}}_g = \frac{1}{N_g}\sum_i \boldsymbol{x}_{ig}$.

Using the algebraic relationship between the sum squared differences and the sum squared deviations from means (see Appendix B), it can be shown that the loss function of PD in (2.4) is equivalent to a N_g-weighted version of the loss function of FE in (2.6):

$$\sum_{g=1}^{G} \sum_{(i,j)\in P_g} (e_{ig} - e_{jg})^2 = \sum_{g=1}^{G} N_g \sum_{i=1}^{N_g} (e_{ig} - \bar{e}_g)^2 , \tag{2.7}$$

Thus, $\widehat{\boldsymbol{\beta}}_{PD}$ in (2.5) can be rewritten as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{PD} &= \left[\sum_g N_g \sum_i \left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)\left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)'\right]^{-1} \sum_g N_g \sum_i \left(\boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g\right)\left(y_{ig} - \bar{y}_g\right) \\ &= \left[\sum_g N_g \sum_i \widetilde{\boldsymbol{x}}_{ig}\widetilde{\boldsymbol{x}}_{ig}{}'\right]^{-1} \sum_g N_g \sum_i \widetilde{\boldsymbol{x}}_{ig}\tilde{y}_{ig} \\ &= \left[\sum_g N_g \widetilde{\boldsymbol{X}}_g'\widetilde{\boldsymbol{X}}_g\right]^{-1} \sum_g N_g \widetilde{\boldsymbol{X}}_g'\widetilde{\boldsymbol{y}}_g ,\end{aligned}$$

where $\widetilde{\boldsymbol{x}}_{ig} = \boldsymbol{x}_{ig} - \overline{\boldsymbol{x}}_g$; $\tilde{y}_{ig} = y_{ig} - \bar{y}_g$; $\widetilde{\boldsymbol{X}}_g$ is a $N_g \times k$ matrix; and $\widetilde{\boldsymbol{y}}_g$ is a $N_g \times 1$ column vector.

Hence, PD generalizes FE by weighting each observation by its cluster size. The intuition is that a larger weight is given to the observations in a larger cluster, in contrast to FE which gives each cluster the same weight. If cluster size N_g is the same (i.e. equal cluster size $N_g = N/G$), then $\widehat{\boldsymbol{\beta}}_{PD}$ can be simplified into $\widehat{\boldsymbol{\beta}}_{FE}$:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{PD} &= [\textstyle\sum_g N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_g N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g \\
&= [\textstyle\sum_g (\frac{N}{G}) \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_g (\frac{N}{G}) \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g \\
&= [\textstyle\sum_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g \\
&= \widehat{\boldsymbol{\beta}}_{FE}.
\end{aligned}
$$

From the perspective of weighted least squared estimation, PD is simply the FE for $\sqrt{N_g} y_{ig} = \sqrt{N_g} \boldsymbol{x}_{ig}{}' \boldsymbol{\beta} + \sqrt{N_g} \alpha_g + \sqrt{N_g} \varepsilon_{ig}$.

2.2.3 Variance-Covariance Matrix Estimation

To control both within-cluster correlation and heteroscedasticity (Cameron & Trivedi, 2005; Cameron & Miller, 2015), the cluster-robust estimate of the variance matrix (CRVE) is widely used. Assuming error independence across clusters, the CRVEs for FE has been shown as the "sandwich" estimate (Cameron & Miller, 2015):

$$\hat{V}(\widehat{\boldsymbol{\beta}}_{FE}) = [\textstyle\sum_G \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_G \widetilde{\boldsymbol{X}}_g' \hat{\boldsymbol{\varepsilon}}_g \hat{\boldsymbol{\varepsilon}}_g' \widetilde{\boldsymbol{X}}_g \, [\sum_G \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1}, \tag{2.8}$$

where $\hat{\boldsymbol{\varepsilon}}_g = \widetilde{\boldsymbol{y}}_g - \widetilde{\boldsymbol{X}}_g \widehat{\boldsymbol{\beta}}_{FE}$. For PD, the sandwich CRVE can be derived as

$$\hat{V}(\widehat{\boldsymbol{\beta}}_{PD}) = [\textstyle\sum_G N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_G N_g^2 \widetilde{\boldsymbol{X}}_g' \hat{\boldsymbol{\varepsilon}}_g \hat{\boldsymbol{\varepsilon}}_g' \widetilde{\boldsymbol{X}}_g \, [\sum_G N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1}, \tag{2.9}$$

where $\hat{\boldsymbol{\varepsilon}}_g = \widetilde{\boldsymbol{y}}_g - \widetilde{\boldsymbol{X}}_g \widehat{\boldsymbol{\beta}}_{PD}$. Compared to (2.8), N_g appears in three places of (2.9), making it infeasible to compare $V(\widehat{\boldsymbol{\beta}}_{FE})$and $V(\widehat{\boldsymbol{\beta}}_{PD})$ mathematically. In another word, PD does not necessarily increase the efficiency of FE. It depends on three quantities related to cluster heterogeneity (Carter, Schnepel, & Steigerwald, 2017): (1) cluster sizes; (2) the cluster specific error covariance matrix; and (3) the observed value of the regressors.

2.3 COMPARISON BETWEEN FE AND PD ESTIMATOR

$$\widehat{\boldsymbol{\beta}}_{FE} = [\textstyle\sum_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g;$$

$$\widehat{\boldsymbol{\beta}}_{PD} = [\textstyle\sum_g N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \sum_g N_g \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g,$$

Looking at the formulae of FE and PD, $\widehat{\boldsymbol{\beta}}_{PD}$ imposes density weights N_g to both $\widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g$ (matrix denominator) and $\widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g$ (matrix numerator) of $\widehat{\boldsymbol{\beta}}_{FE}$ for each cluster g. The inverse operation makes it hard to see how the weighting practice deviates $\widehat{\boldsymbol{\beta}}_{PD}$ from $\widehat{\boldsymbol{\beta}}_{FE}$.

This section discusses the two equivalence conditions between FE and PD, and how PD deviates from FE as the conditions are relaxed. The discussion is inspired by a pair of mathematical concepts—mediant and weighted mediant, which are analogous to FE and PD in the number theory. This pair of concepts is introduced in the following section.

2.3.1 Mediant and Weighted Mediant

A *mediant* m for a sequence of n fractions $\frac{a_1}{b_1}, \frac{a_2}{b_2}, \dots, \frac{a_n}{b_n}$ is defined as

$$m = \frac{a_1 + a_2 + \cdots + a_n}{b_1 + b_2 + \cdots + b_n} = (\textstyle\sum_i b_i)^{-1} \sum_i a_i, \quad \text{for } i = 1 \dots n,$$

where a_i is a nonnegative real number, and b_i a is positive real number. If $w_1, w_2, \dots, w_n$ are n positive real numbers, then a *weighted median* is defined as

$$m_w = \frac{w_1 a_1 + w_2 a_2 + \cdots + w_n a_n}{w_1 b_1 + w_2 b_2 + \cdots + w_n b_n} = (\textstyle\sum_i w_i b_i)^{-1} \sum_i w_i a_i, \quad \text{for } i = 1 \dots n.$$

It is interesting to see that $\widehat{\boldsymbol{\beta}}_{FE}$ and $\widehat{\boldsymbol{\beta}}_{PD}$ can be viewed as, in matrix form, the mediant and the weighted mediant of the sequence of $[\widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{X}}_g]^{-1} \widetilde{\boldsymbol{X}}_g' \widetilde{\boldsymbol{y}}_g$. Unfortunately, the literature of the mediant is based on numbers, not vectors or matrices. To avoid groundless application of the mediant theory to the comparison between FE and PD, this paper chooses relevant theories and then conducts Monte Carlo simulation to verify their applicability to the matrix setting.

2.3.2 Relevant Theories from Mediant Theory

In general, there are two relevant conclusions from the mediant theory: the first (equivalence conditions) states two sufficient conditions for the equivalency between mediant and weighted mediant; and the second (deviation conditions) informs how the difference between them is correlated to the association between weights and values of the fractions.

Equivalence Conditions: the mediant equals to the weighted mediant if any of the following conditions holds (see proofs in Appendix C):

a. (*Equal-weight Condition*) all weights are equal: $w_i = w$ for all $i = 1 \ldots n$ and $w \in \mathbb{R}_{>0}$;

b. (*Equal-fraction Condition*) all fractions are equal: $\frac{a_i}{b_i} = \frac{a}{b}$ for all $i = 1 \ldots n$ and $a \in \mathbb{R}_{\geq 0}, b \in \mathbb{R}_{>0}$.

Deviation Condition: If a relatively larger fraction $\frac{a_n}{b_n}$ is associated with a relatively larger weight w_n, then $m_w > m$. In another word, if the covariance between $\frac{a_n}{b_n}$ and w_n is positive, then $m_w > m$; otherwise, $m_w < m$. A simple case when $n = 2$ and a simulation with $n > 2$ are shown in Appendix C.

2.3.3 Extension to Vector and Matrix

For the matrix setting, the mediant and weighted mediant could be defined as

$$M = (B_1 + B_2 + \cdots + B_n)^{-1}(A_1 + A_2 + \cdots + A_n) = \left(\sum_i B_i\right)^{-1} \sum_i A_i,$$

$$M_w = (w_1B_1 + w_2B_2 + \cdots + w_nB_n)^{-1}(w_1A_1 + w_2A_2 + \cdots + w_nA_n)$$
$$= (\textstyle\sum_i w_iB_i)^{-1} \textstyle\sum_i w_iA_i,$$

in which the "fraction" is $\boldsymbol{f}_i = {B_i}^{-1}A_i$. Since FE and PD are vectors, we only consider the situations when M, M_w and $\boldsymbol{f}_i$ are vectors. Thus, if B_i is, say a $K \times K$ invertible matrix, then A_i is a $K \times 1$ vector and the resulting M, M_w and $\boldsymbol{f}_i$ are all $K \times 1$ vectors.

It can be shown that both equivalence conditions hold, meaning that the mediant equals to the weighted median in the matrix setting when either all weights are equal or all matrix fractions are equal. For the equal-weight condition, if $w_i = w$ for all $i = 1 \ldots n$, $M_w = (\sum_i w_iB_i)^{-1} \sum_i w_iA_i = (\sum_i wB_i)^{-1} \sum_i wA_i = (\sum_i B_i)^{-1} \sum_i A_i = M$. This condition generalizes the case (in section 2.2) in which FE and PD are equivalent when cluster sizes are equal. For the equal-fraction condition, if ${B_i}^{-1}A_i = B^{-1}A$ for all i, then $A_i = B_iB^{-1}A$, and $M_w = (\sum_i w_iB_i)^{-1} \sum_i w_iA_i = (\sum_i w_iB_i)^{-1}(\sum_i w_iB_i)B^{-1}A = B^{-1}A = (\sum_i B_i)^{-1}(\sum_i B_i\, B^{-1}A) = (\sum_i B_i)^{-1}(\sum_i A_i) = M$. This condition suggests that if $\boldsymbol{f}_i = {B_i}^{-1}A_i$ is the same for all i, the weighted mediant does not deviate from the mediant even though the weights are very different. In the case of $\widehat{\boldsymbol{\beta}}_{FE}$ and $\widehat{\boldsymbol{\beta}}_{PD}$, the fraction is $\widehat{\boldsymbol{\beta}}_g = [\widetilde{\boldsymbol{X}}_g'\widetilde{\boldsymbol{X}}_g]^{-1}\widetilde{\boldsymbol{X}}_g'\widetilde{\boldsymbol{y}}_g$, which is the local OLS estimate of $\boldsymbol{\beta}$ using observations only in cluster g. The equal-fraction condition informs us that if the local estimates are the same, $\widehat{\boldsymbol{\beta}}_{FE}$ equals to $\widehat{\boldsymbol{\beta}}_{PD}$ no matter how different the cluster sizes are.

Unfortunately, the *deviation condition* no longer holds. Measuring the vector deviations with either element-wise comparison or the Euclidean distance between vectors, no correlation can be found between weighting larger weights to "larger" matrix fractions and “larger” changes from mediant vector to the weighted mediant vector[20].

In regression analysis, even under stationary assumption of $\boldsymbol{\beta}$, local estimates $\widehat{\boldsymbol{\beta}}_g$ could still deviate from the true $\boldsymbol{\beta}$ and vary across clusters due to differences in error structures and cluster sizes.

For the variance-covariance matrices in (2.8) and (2.9), they are not in the form of matrix mediant and weighted mediant. Together with $\widehat{\boldsymbol{\beta}}_{FE}$ and $\widehat{\boldsymbol{\beta}}_{PD}$, the comparison between the variance estimators are examined by a Monte Carlo simulation in the next section.

2.4 A MONTE CARLO APPROACH

When weak statistical theory exists for an estimator, Monte Carlo methods are often applied to investigate the finite-sample properties of estimator and test statistics (Davidson & MacKinnon, 1993). This section introduces the Monte Carlo design for comparing the performance of FE and PD in linear fixed effects models.

2.4.1 Performance Tests Used in this Paper

The performance tests used in this paper include biasedness, efficiency and validity of inference. For the biasedness, it is measured by mean absolute errors in parameter estimates. Both FE and PD are expected to generate small mean squared errors, since they are both unbiased estimators. For the efficiency, it is measured by mean standard errors using either default i.i.d. standard error or cluster-robust standard errors (CRVE). The default i.i.d. standard error assumes the homoscedasticity and no serial correlation in error

[20] The simulation results are not shown in the paper.

terms, and serves as a benchmark (Cameron et al., 2008). The cluster-robust standard errors (CRVE) are more popular since it allows both heteroscedastic and serial correlated error structures. Even in fixed effects model, the within-cluster correlation is not fully controlled and the CRVE should still be used (Cameron & Miller, 2015). See Kezdi (2003) for an excellent review on the relationship among different standard errors. For the validity of inference, it is measured by the rejection rate against a normal critical value at a certain confidence level.

2.4.2 Monte Carlo Design

The Monte Carlo design includes generating processes of the dependent variable, independent variables, and error terms, the number of clusters and total observations, the cluster sizes and unbalancedness, and the number of replication. Among these aspects, the unbalancedness is essential to distinct FE and PD. Ahren and Pincus (1981) provides a measure for the unbalancedness of an N_g-pattern, which has been used in the study of incomplete panels (e.g., Baltigi & Chang, 1994). The measure is defined as

$$\tau = G/(\bar{L} \sum_g {N_g}^{-1}) \, , \tag{2.10}$$

where G is the number of clusters, N_g is the cluster size of g, $\bar{L} = G^{-1} \sum_g N_g$, and τ ranges in $(0, 1]$. Note that τ equals one when the pattern is balanced and gets smaller as the pattern becomes more unbalanced.

Table 2.1 Summary of Monte Carlo Simulation Designs*

CATEGORY	ITEM	Baltagi & Chang (1994)	Cameron, Gelbach, & Miller (2008)**	MacKinnon & Webb (2017)
Model	Specification $g = 1, \ldots, G;\ i = 1, \ldots, N_g$	$y_{ig} = \beta_0 + \beta x_{ig} + \mu_{ig}$	$y_{ig} = \beta_0 + \beta x_{ig} + \mu_{ig}$	$y_{ig} = \beta_0 + \beta x_{ig} + \mu_{ig}$
	Parameter	$\beta_0 = 5;\ \beta = 0.5$	$\beta_0 = 1;\ \beta = 1$	$\beta = 0$
Regressor $\boldsymbol{x_{ig}}$	Generation Process	$x_{ig} = 0.3i + 0.8x_{ig-1} + \omega_{ig}$ $\omega_{ig} \sim \mathbb{U}[-0.5, 0.5]$ $x_{g0} = 100 + 250\omega_{ig}$	$x_{ig} = z_g + z_{ig}$ $z_g \sim \mathbb{IIN}(0, 1)$ $z_{ig} \sim \mathbb{IIN}(0, 1)$	$x_{ig} \sim \mathbb{N}(0, 1)$ Independent across g
	Correlation of Regressors $\boldsymbol{\rho_x}$	0.8*	0.5*	{0, 0.2, 0.4, 0.6, 0.8, 1}
Error $\boldsymbol{\mu_{ig}}$	Generation Process	$\mu_{ig} = \alpha_g + \varepsilon_{ig}$ $\alpha_g \sim \mathbb{IIN}(0, \sigma_\alpha^2)\ \varepsilon_{ig} \sim \mathbb{IIN}(0, \sigma_\varepsilon^2)$ $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 20$	$\mu_{ig} = \alpha_g + \varepsilon_{ig}$ $\alpha_g \sim \mathbb{IIN}(0, 1)$ $\varepsilon_{ig} \sim \mathbb{IIN}(0, 9 \times (z_g + z_{ig})^2)$	$\mu_{ig} \sim \mathbb{N}(0, 1)$ Independent across g
	Correlation of Errors $\boldsymbol{\rho_\mu}$	{0, 0.2, 1/3, 0.5, 2/3, 0.8, 0.9}	Heteroscedastic	{0, 0.1, …, 0.8, 0.9}
Sample	Number of Clusters	30/60/90/240	10	50/100
	Number of Total Obs.	210/420/630/1680	300	2000
Unbalance	Design	8 different patterns	5 clusters with 10 obs and 5 clusters with 50 obs	proportional to US state sizes
	Measure	{0.918, 0.841, 0.813, 0.754, 0.519, 0.490, 0.209, 0.207}	0.55556*	0.3802* (50 clusters) 0.3718* (100 clusters)
Replication	Per Experiment	1000	1000	400,000
	Bootstrap Sample	N/A	399	399

* The notations in the table are unified for comparison and may be different from the original literature. The number with * are calculated in this paper and are not reported in the original literature.
** Only the heteroscedastic clustered errors case is shown here.

The Monte Carlo study in this paper follows three studies in the literature involving unbalanced data design, summarized in Table 2.1. Baltigi and Chang (1994) focus on incomplete panels while Cameron et al. (2008) and MacKinnon and Webb (2017) on the clustered settings. Cameron et al. (2008) also introduce the case of heteroscedastic errors. These lead to different ways to generate regressors and error terms. The summary shows the parameter of interests: the within-cluster correlation of the regressor, the within-cluster correlation of the error, the number of clusters, the number of total observations, and the unbalancedness. The Monte Carlo design of this study is shown in Table 2.2.

Following the three papers, the model used in this study is specified as:

$$y_{ig} = \beta_0 + \beta x_{ig} + \mu_{ig}, \qquad g = 1, \dots, G, \quad i = 1, \dots, N_g,$$

where $\beta_0 = 1$, and $\beta = 1$.

For the homoscedastic case, the regressors x_{ig} is set to be one-dimensional and generated in two different ways: a panel setting following Baltigi and Chang (1994) and a clustered setting following Cameron et al. (2008). In the panel setting, $x_{ig} = 0.3i + \rho_x x_{ig-1} + \omega_{ig}$, where ω_{ig} is uniformly distributed on the interval [-0.5,0.5]. Note that a serial correlation ρ_x is imposed to the units in one cluster based on a random order. x_{g0} is set as $100 + 250\omega_{ig}$ and excluded from the simulated data. In the clustered setting, $x_{ig} = z_g + z_{ig}$, in which $z_g \sim \mathbb{IIN}\left(0, \sigma_{zg}^2\right)$, $z_{ig} \sim \mathbb{IIN}(0, \sigma_{zi}^2)$, and $\sigma_{zg}^2 + \sigma_{zi}^2 = 20$. In this case, the within-cluster correlation is $\rho_x = \sigma_{zg}^2/(\sigma_{zg}^2 + \sigma_{zi}^2)$. For both settings, ρ_x takes a value from {0, 1/3, 0.5, 2/3, 0.9} in each experiment. For the error terms, it is assumed that $\mu_{ig} = \alpha_g + \varepsilon_{ig}$, and $\alpha_g \sim \text{IIN}(0, \sigma_\alpha^2)$ and $\varepsilon_{ig} \sim \text{IIN}(0, \sigma_\varepsilon^2)$. Let $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 20$, and the within-cluster correlation of errors is $\rho_\mu = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$. ρ_μ will take values in {0, 1/3, 0.5, 2/3, 0.9}, same with ρ_x.

For the heteroscedastic case, following Cameron et al. (2008) in Table 2.1 except that the within-cluster correlation of regressors is allowed to take different values. In contrast to the homoscedastic case, the sum of variances of regressor components is changed to 2. The idiosyncratic term is generated as $\varepsilon_{ig} \sim \mathbb{IIN}(0, 9 \times x_{ig}^2)$, through which the heteroscedasticity is introduced. As a result, the within-cluster correlation of errors is undetermined. All other settings are the same for all cases.

Table 2.2 Monte Carlo Simulation Design

CATEGORY	ITEM	DESIGN		
Model	Specification	$y_{ig} = \beta_0 + \beta x_{ig} + \mu_{ig}$ $g = 1, \dots, G;\ i = 1, \dots, N_g$		
	Parameter	$\beta_0 = 1;\ \ \beta = 1$		
Regressor $\boldsymbol{x_{ig}}$		**Homoscedasticity**		**Heteroscedasticity**
	Generation Process	Panel Setting $x_{ig} = 0.3i + \rho_x x_{ig-1} + \omega_{ig}$ $\omega_{ig} \sim \mathbb{U}[-0.5, 0.5]$ $x_{g0} = 100 + 250\omega_{ig}$	Clustered Setting $x_{ig} = z_g + z_{ig}$ $z_g \sim \mathbb{IIN}(0, \sigma_{zg}^2)$ $z_{ig} \sim \mathbb{IIN}(0, \sigma_{zi}^2)$ $\sigma_{zg}^2 + \sigma_{zi}^2 = 20$	Clustered Setting $x_{ig} = z_g + z_{ig}$ $z_g \sim \mathbb{IIN}(0, \sigma_{zg}^2)$ $z_{ig} \sim \mathbb{IIN}(0, \sigma_{zi}^2)$ $\sigma_{zg}^2 + \sigma_{zi}^2 = 2$
	Correlation of Regressors $\boldsymbol{\rho_x}$	$\boldsymbol{\rho_x} \in \{0, 1/3, 0.5, 2/3, 0.9\}$		
Error $\boldsymbol{\mu_{ig}}$	Generation Process	$\mu_{ig} = \alpha_g + \varepsilon_{ig}$ $\alpha_g \sim \mathbb{IIN}(0, \sigma_\alpha^2)\ \varepsilon_{ig} \sim \mathbb{IIN}(0, \sigma_\varepsilon^2)$ $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 20$		$\mu_{ig} = \alpha_g + \varepsilon_{ig}$ $\alpha_g \sim \mathbb{IIN}(0, 1)$ $\varepsilon_{ig} \sim \mathbb{IIN}(0, 9 \times x_{ig}^2)$
	Correlation of Errors $\boldsymbol{\rho_\mu}$	$\boldsymbol{\rho_\mu} \in \{0, 1/3, 0.5, 2/3, 0.9\}$		N/A
Sample	Number of Clusters G	$G \in \{2, 4, 6, \dots, 36, 38, 40\}$		
	Number of Total Obs.	$\{40, 80, 120, \dots, 320, 360, 400\}$		
Unbalance	Design	3 base patterns $P_1 = 2(1), 38(1);$ $P_2 = 6(1), 34(1);$ $P_3 = 14(1), 26(1)$		
	Measure τ	$\tau \in \{0.19, 0.51, 0.91\}$		
Replication	Per Experiment R	10000		

For the sample design, it is worth noting that the number of clusters and clusters sizes will determine the unbalancedness. An ideal design could fix the unbalanceness while the number of clusters varies. This cannot be achieved by randomly choosing cluster sizes at each given number of clusters. All three papers apply fixed design of cluster sizes to control for the unbalancedness. This paper follows this method and choose three different unbalanced levels {0.19, 0.51, 0.91}. The details of cluster-size design are discussion as follows.

First, I introduce the concept of a base design. A key property of the unbalancedness measure τ in (10) is that replicating a sample pattern, the unbalancedness is unchanged. Take an example from Baltigi and Chang (1994). Let 5(15) represent a sample pattern with 15 clusters, each with 5 observations. A 30-cluster sample pattern $P = 5(15), 9(15)$ has an unbalanced level of 0.918. If one replicates P once and gets $P' = 5(30), 9(30)$, the unbalanced level of this 60-cluster sample P' is still 0.918. One could reduce the number of clusters to 2 and get sample $P^b = 5(1), 9(1)$ with the same unbalanced level. I called this smallest sample as a base design for this pattern (half sample with size 5, the other half with size 9). Any integer replications of the base design have the same unbalanced level, which allows us to control for the unbalancedness as the number of clusters grows.

In this paper, I do not directly apply Baltigi and Chang (1994)'s designs for three reasons. First, many of their designs contain clusters with only one observation, which contribute nothing to within-cluster variations. Second, given a base design with 2 clusters, I could examine all even numbers of clusters. However, many designs in Baltigi and Chang (1994) are based on 5 or more clusters, for example $P_4 = 3(9), 5(6), 9(6), 11(9)$ has a base design $3(3), 5(2), 9(2), 11(3)$ with 10 clusters; and $P_8 = 1(20), 10(5), 28(5)$ has a base design $1(4), 10(1), 28(1)$ with 6 clusters. In these cases, we could only examine the numbers of clusters {10, 20, 30, …} or {6, 12, 18, …}, with many cases being skipped in simulations. Third, the 2-cluster base designs in their study contain only 14 number of observations, which is too small for valid inferences with few clusters.

As a result, this study takes three different base designs: $P_1 = 2(1), 38(1)$; $P_2 = 6(1), 34(1)$; and $P_3 = 14(1), 26(1)$. The unbalanced levels are {0.19, 0.51, 0.91}, representing the cases of severely unbalanced, middle-level unbalanced, and almost balanced. Given the number of clusters, the total number of observations is the same across different unbalanced designs (see details in Table 2.3).

Table 2.3 Sample and Unbalancedness Design

Design	$\mathbf{G = 2}$ (Base)	$\mathbf{G = 4}$	$\mathbf{G = 6}$	…	$\mathbf{G = 38}$	$\mathbf{G = 40}$	**Unbalance** $\boldsymbol{\tau}$
$\boldsymbol{P_1}$	2(1), 38(1)	2(2), 38(2)	2(3), 38(3)	…	2(19), 38(19)	2(20), 38(20)	0.19
$\boldsymbol{P_2}$	6(1), 34(1)	6(2), 34(2)	6(3), 34(3)	…	6(19), 34(19)	6(20), 34(20)	0.51
$\boldsymbol{P_3}$	14(1), 26(1)	14(2), 26(2)	14(3), 26(3)	…	14(19), 26(19)	14(20), 26(20)	0.91
Number of Observations	40	80	120	…	360	800	

For each simulation, the process is proceeded as in MacKinnon and Webb (2017):

1. Specify the generation process of the regressor to be panel setting or clustered setting; and if clustered setting, whether homoscedastic or heteroscedastic.
2. Specify $\tau \in \{0.19, 0.51, 0.91\}$ and corresponding sample base design;
3. Specify $\rho_x \in \{0, 1/3, 0.5, 2/3, 0.9\}$ and $\rho_\mu \in \{0, 1/3, 0.5, 2/3, 0.9\}$;
4. Specify the number of clusters $G \in \{2, 4, 6, \dots, 36, 38, 40\}$;
5. Based on the configuration above, a sample data (x_{ig}, μ_{ig}, and y_{ig}) is generated through the data generation process. Both FE and PD are applied to produce $\hat{\beta}_{FE}$ and $\hat{\beta}_{PD}$ as estimates for $\beta = 1$, and to calculate default i.i.d. standard errors and CRVE.

6. The absolute errors $|\hat{\beta} - \beta|$ are calculated each time for both FE and PD estimates. Hypothesis testing is conducted using *t* test with i.i.d. standard error or CRVE against the normal critical value 1.96 at 95% confidence level.
7. Repeat the two preceding steps 10,000 times, and calculate the mean absolute error over replications, and the mean of the 10,000 standard error estimates (both i.i.d. and CRVE), and estimate the rejection frequency of 10,000 replications.

2.4.3 Simulation Results

Unbiasedness

For the unbiasedness of FE and PD, the mean absolute error (MEA) and relative mean absolute error (RMAE) in parameter estimation are shown in the Table 2.4 to Table 2.8. Generally, since FE and PD are unbiased and consistent estimators, their MAEs are small in most cases as expected—less than 0.01 given that the true value β_0 equals to 1.

For the homoscedastic case, in the panel setting (Table 2.4), the largest errors (0.01893 for FE and 0.01961 for PD) occur when the data is almost balanced and the within-cluster correlations of both regressors and errors are zero. In the clustered setting (Table 2.5), the largest errors (0.03925 for FE and 0.02807 for PD) appear in the same situation except that the correlation of the error are very high. The relative mean absolute errors are calculated by the MSA ratios of PD to FE and summarized in Table 2.6. Most values are around 1, indicating similar results of FE and PD. For the heteroscedastic case, the results are similar. There seem to be no specific settings in which FE or PD outperforms the other. Considering the small values of MAEs, the differences are trivial.

Efficiency

To compare the estimation efficiency of FE and PD, the change rates from FE to PD are calculated for both default i.i.d. standard errors and CRVE as

$$change\ rate(\%) = \frac{se_{PD} - se_{FE}}{se_{FE}} \times 100.$$

Note that a negative change rate indicates an efficiency improvement of PD over FE; while a positive one means an efficiency loss. The results of the homo- and hetero-scedastic cases are reported in Table 2.9-2.10 and Table 2.11, respectively.

<u>Homoscedastic Case</u>

For i.i.d. standard errors, FE and PD are very similar in the clustered setting. In contrast, in the panel setting, the differences between FE and PD vary with $\boldsymbol{\rho_\mu}$ and the unbalanced level τ. First, when there is no correlation among within-cluster errors ($\boldsymbol{\rho_\mu} = 0$), PD always shows an efficiency improvement over FE by more than 5% when the unbalanced level is middle or low ($\tau = 0.51$ or 0.91), or around 1.2% when the data is severely unbalanced ($\tau = 0.19$). Second, when the correlation among within-cluster errors is 0.5, PD exhibits an efficiency loss from FE by around 15% when the unbalanced level is middle, or around 3%-5% when the data is either severely unbalanced or almost balanced. Third, when within-cluster errors are highly correlated ($\boldsymbol{\rho_\mu} = 0.9$), the situation may differ: PD improves FE by 1%-2% when the data is severely or middle-level unbalanced; PD worsens FE by less than 1% given an almost balanced data. While the within-cluster correlations of the regressors $\boldsymbol{\rho_x}$ make no differences for any comparison, the number of clusters G does influence the magnitude of the differences to some extent. With G increases, all differences tend to shrink by a small amount (0%-3%).

For CRVE, the story changes. It is the number of clusters G that determines the comparison.

In the clustered setting, for small G (less than 10), PD has smaller CRVE than FE. Given $G = 2$, this improvement reaches the largest 94% when the data is severely unbalanced, around 76% when middle-level unbalanced, and 22% when almost balanced. This advantage shrinks very fast as G increases. At $G = 6$, only around 1.6%, 4.5% and 3% improvements are kept at three unbalanced levels. When G is larger than 10, the loss starts and gets larger. As G grows, the increase in loss slows down and CRVEs for FE and PD converge to their own limits. In this process, neither $\boldsymbol{\rho_{\mu}}$ nor $\boldsymbol{\rho_{x}}$ is significantly relevant. In the panel setting, the situation is very similar except that $\boldsymbol{\rho_{\mu}}$ matters. When $\boldsymbol{\rho_{\mu}} = 0.5$, the advantage of PD over FE disappears faster given middle-level unbalanced or almost balanced data, starting an efficiency loss by 1%-2% at $G = 6$. Generally, in these two cases ($\boldsymbol{\rho_{\mu}} = 0.5$ with $\tau = 0.51$ and $\boldsymbol{\rho_{\mu}} = 0.5$ with $\tau = 0.91$), PD is less advantageous in the panel setting than in the clustered one in three aspects: smaller efficiency gain for small G, larger decrease rate in efficiency gain as G increases, and larger efficiency loss for large G. When the data is severely unbalanced, $\boldsymbol{\rho_{\mu}} = 0.5$ also makes the efficiency loss larger than the level in the clustered setting.

Figure 2.1 to 2.2 show how the change rates vary with the increases in the number of clusters in different data settings given $\boldsymbol{\rho_{x}} = 0.5$. For CRVE in all cases, PD does exhibit a faster convergence rate than FE, however, with the number of clusters increases, FE converges to a lower level than PD.

<u>Heteroscedastic Case</u>

Under heteroscedastic assumption, the comparison using CRVE seems to be the same with the homoscedastic case, except that a higher within-cluster correlation of regressors favors PD over FE. For the default i.i.d. standard errors, PD and FE lead to similar results except that when the within-cluster correlation of regressors is high and the number of clusters is less than 10, PD generates smaller values by 1%-6%.

Rejection Rate

The rejection rate matters since it validates the usage of standard errors to make inference. The results of the rejection rates for both i.i.d. standard errors and CRVE in the panel and the clustered settings are shown in Table 2.11-2.14; and for the heteroscedastic case in Table 2.15-2.16.

With homoscedastic errors, for both FE and PD, the i.i.d. standard errors have rejection rates very close to the nominal level 0.05 (Table 2.11-2.12). This is because after mean-differencing, the within-cluster correlations in both regressors and errors are eliminated. The transformed data is basically i.i.d., in which the default standard errors are valid. However, under heteroscedasticity, the i.i.d. standard errors always over-reject. In contrary, the CRVE shows over-rejection issue in both homo- and hetero-scedastic cases, and the over-rejection becomes less severe as the number of clusters G grows, similar to the findings of Cameron et al. (2008). In the heteroscedastic case, the faster convergence rate is found to be associated with a very high within-cluster correlation of regressors.

2.4.4 Result Summary

In summary, PD shows efficiency improvement over FE in some cases. This section inspects the validity of these cases with the rejection rates.

The most significant efficiency improvements of PD over FE occur when using CRVE given the number of clusters less than 10 (Table 2.9, 2.10 and 2.11). However, results in Table 2.14-2.16 show the severe over-rejection issue of CRVE when G is small. One exception is under heteroscedasticity, when G ranges from 6 to 10 and $\boldsymbol{\rho_x} = 0.9$, the rejection rates of CRVE are from 0.059 to 0.071, which are close to the nominal level.

Other improvements happen to the default i.i.d. standard errors in homoscedastic panel settings and in heteroscedastic clustered settings with few clusters. In the former setting, the improvements are about 1.2% when data is severely unbalanced and $\boldsymbol{\rho_\mu} = 0$ or $\boldsymbol{\rho_\mu} = 0.9$, about 2% when data is middle-level unbalanced and $\boldsymbol{\rho_\mu} = 0.9$, and about 5.2% when the data is either middle-level unbalanced or almost balanced and $\boldsymbol{\rho_\mu} = 0$. The rejection rates are very close to the nominal size 0.05 in all these cases. In the latter setting, the improvement is 1%-6% when the number of clusters is less than 10. However, over-rejection occurs to i.i.d. standard errors in this setting.

Overall, there are two cases in which PD has valid efficiency advantage over FE: (1) for i.i.d. standard errors: panel setting + homoscedasticity + $\boldsymbol{\rho_\mu} = 0$ or 0.9; (2) for CRVE: clustered setting + heteroscedasticity + $\boldsymbol{\rho_x} = 0.9$ + $G \in [6, 10]$. In the former case, although the rejection rates are very close to the nominal level 0.05, the assumptions especially the homoscedasticity seem to be restrictive. In contrast, the latter one allows heteroscedasticity. The main restrictions are on $\boldsymbol{\rho_x} = 0.9$, which means the regressors are very similar within clusters, and on the number of clusters ranging from 6 to 10.

2.5 DISCUSSION AND CONCLUSION

This paper explores the properties of the pairwise differences estimator (PD) in comparison with the fixed effects estimator (FE) in linear fixed effects models. The main finding is that PD is a generalized version of FE by weighting each observation by its cluster size. There are two conditions for them to be equivalent: 1) the cluster sizes are equal (balanced data); 2) the local estimates using each cluster data are equal. The question addressed in this paper is: when FE and PD are distinct, i.e. given unbalanced data and unequal local estimates, is PD a good alternative to

FE in linear fixed effects models?

Since both FE and PD are unbiased and consistent estimator under regular assumptions and the efficiency comparison is undetermined theoretically, this paper designs a Monte Carlo simulation with different data settings. The results show that: 1) both FE and PD provide accurate estimation in the parameter of interest; 2) valid efficiency improvements of PD over FE occur when: using the default i.i.d. standard errors in the homoscedastic panel data with the within-cluster error correlations being zero or very high; or using the cluster-robust standard errors in heteroscedastic clustered data with very high within-cluster regressor correlations and the number of clusters ranging from 6 to 10. Considering in the empirical works nowadays, CRVE is very popular and big data sets are more available, in most cases, the fixed effects estimator (FE) would still be the more efficient choice than the pairwise differences estimators (PD).

Table 2.4 Mean Absolute Errors (Panel Setting)

			P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.00238	0.00551	0.00299	0.00716	0.00068	0.00195	0.01893	0.00128	0.00073
		PD	0.00225	0.00539	0.00301	0.00627	0.00089	0.00187	0.01961	0.00049	0.00073
	0.5	FE	0.00484	0.00086	0.00009	0.00561	0.00159	0.00061	0.00275	0.00207	0.00034
		PD	0.00584	0.00061	0.00028	0.00410	0.00039	0.00087	0.00338	0.00244	0.00011
	0.9	FE	0.00259	0.00043	0.00021	0.00257	0.00118	0.00003	0.00028	0.00298	0.00004
		PD	0.00253	0.00044	0.00020	0.00335	0.00120	0.00003	0.00065	0.00242	0.00001
6	0	FE	0.00217	0.00038	0.00025	0.00381	0.00321	0.00059	0.00141	0.00102	0.00078
		PD	0.00199	0.00041	0.00025	0.00288	0.00276	0.00063	0.00223	0.00050	0.00048
	0.5	FE	0.00016	0.00226	0.00006	0.00121	0.00059	0.00002	0.00281	0.00025	0.00009
		PD	0.00100	0.00227	0.00007	0.00115	0.00081	0.00022	0.00311	0.00005	0.00012
	0.9	FE	0.00095	0.00046	0.00002	0.00189	0.00022	0.00017	0.00234	0.00038	0.00007
		PD	0.00095	0.00046	0.00002	0.00176	0.00010	0.00016	0.00253	0.00051	0.00009
10	0	FE	0.00153	0.00159	0.00035	0.00796	0.00080	0.00025	0.00162	0.00676	0.00023
		PD	0.00154	0.00163	0.00034	0.00804	0.00086	0.00026	0.00103	0.00670	0.00034
	0.5	FE	0.00037	0.00004	0.00001	0.00084	0.00081	0.00015	0.00086	0.00037	0.00001
		PD	0.00005	0.00004	0.00004	0.00095	0.00137	0.00025	0.00056	0.00025	0.00003
	0.9	FE	0.00032	0.00002	0.00008	0.00094	0.00044	0.00002	0.00021	0.00007	0.00002
		PD	0.00026	0.00003	0.00009	0.00062	0.00058	0.00004	0.00011	0.00011	0.00003
20	0	FE	0.00282	0.00109	0.00010	0.00134	0.00294	0.00028	0.00592	0.00351	0.00001
		PD	0.00282	0.00108	0.00011	0.00158	0.00300	0.00030	0.00657	0.00383	0.00006
	0.5	FE	0.00013	0.00011	0.00005	0.00042	0.00031	0.00006	0.00003	0.00006	0.00003
		PD	0.00046	0.00010	0.00004	0.00031	0.00057	0.00000	0.00014	0.00004	0.00002
	0.9	FE	0.00001	0.00004	0.00010	0.00094	0.00018	0.00003	0.00042	0.00002	0.00002
		PD	0.00003	0.00003	0.00010	0.00105	0.00023	0.00003	0.00045	0.00001	0.00003
30	0	FE	0.00184	0.00031	0.00030	0.00243	0.00146	0.00017	0.00605	0.00257	0.00043
		PD	0.00177	0.00031	0.00029	0.00226	0.00155	0.00017	0.00494	0.00246	0.00050
	0.5	FE	0.00042	0.00026	0.00006	0.00024	0.00060	0.00004	0.00087	0.00006	0.00004
		PD	0.00047	0.00026	0.00008	0.00082	0.00097	0.00006	0.00094	0.00001	0.00004
	0.9	FE	0.00006	0.00011	0.00002	0.00063	0.00019	0.00001	0.00017	0.00029	0.00008
		PD	0.00007	0.00012	0.00002	0.00047	0.00023	0.00001	0.00021	0.00018	0.00007
40	0	FE	0.00214	0.00081	0.00068	0.00036	0.00087	0.00032	0.00193	0.00065	0.00036
		PD	0.00212	0.00078	0.00067	0.00037	0.00097	0.00031	0.00235	0.00065	0.00039
	0.5	FE	0.00034	0.00059	0.00010	0.00014	0.00009	0.00007	0.00073	0.00037	0.00012
		PD	0.00012	0.00059	0.00012	0.00006	0.00008	0.00002	0.00067	0.00035	0.00011
	0.9	FE	0.00070	0.00009	0.00001	0.00045	0.00005	0.00003	0.00016	0.00003	0.00001
		PD	0.00067	0.00009	0.00001	0.00044	0.00002	0.00003	0.00017	0.00000	0.00002

Table 2.5 Mean Absolute Errors (Clustered Setting)

			$P_1\ (\tau = 0.19)$			$P_2\ (\tau = 0.51)$			$P_3\ (\tau = 0.91)$		
G	ρ_μ		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.00007	0.00049	0.00006	0.00076	0.00046	0.00005	0.00243	0.00061	0.00027
		PD	0.00022	0.00055	0.00006	0.00023	0.00081	0.00010	0.00239	0.00062	0.00027
	0.5	FE	0.00473	0.00113	0.00023	0.00301	0.00052	0.00006	0.00104	0.00144	0.00001
		PD	0.00507	0.00157	0.00021	0.00359	0.00024	0.00003	0.00120	0.00159	0.00007
	0.9	FE	0.03622	0.00105	0.00133	0.00304	0.00431	0.00115	0.03925	0.01677	0.00124
		PD	0.04023	0.00084	0.00133	0.00227	0.00720	0.00073	0.02807	0.01713	0.00095
6	0	FE	0.00047	0.00022	0.00008	0.00009	0.00021	0.00009	0.00053	0.00004	0.00011
		PD	0.00041	0.00026	0.00010	0.00011	0.00007	0.00005	0.00008	0.00009	0.00011
	0.5	FE	0.00148	0.00077	0.00020	0.00235	0.00145	0.00016	0.00114	0.00208	0.00020
		PD	0.00147	0.00068	0.00023	0.00266	0.00124	0.00012	0.00103	0.00191	0.00029
	0.9	FE	0.00066	0.00152	0.00114	0.00202	0.00566	0.00060	0.00649	0.01092	0.00064
		PD	0.00031	0.00253	0.00096	0.00163	0.00537	0.00041	0.00863	0.01022	0.00041
10	0	FE	0.00027	0.00035	0.00002	0.00059	0.00012	0.00003	0.00126	0.00011	0.00010
		PD	0.00005	0.00040	0.00002	0.00081	0.00014	0.00000	0.00143	0.00021	0.00009
	0.5	FE	0.00117	0.00073	0.00000	0.00144	0.00019	0.00026	0.00045	0.00176	0.00022
		PD	0.00124	0.00070	0.00002	0.00193	0.00026	0.00028	0.00139	0.00182	0.00020
	0.9	FE	0.00446	0.00025	0.00172	0.00542	0.00737	0.00127	0.00244	0.00287	0.00043
		PD	0.00522	0.00042	0.00170	0.00442	0.00646	0.00131	0.00441	0.00538	0.00034
20	0	FE	0.00033	0.00025	0.00003	0.00067	0.00017	0.00005	0.00024	0.00027	0.00002
		PD	0.00038	0.00026	0.00003	0.00057	0.00025	0.00010	0.00028	0.00032	0.00002
	0.5	FE	0.00012	0.00040	0.00011	0.00074	0.00070	0.00005	0.00100	0.00014	0.00009
		PD	0.00021	0.00031	0.00010	0.00107	0.00058	0.00004	0.00116	0.00024	0.00007
	0.9	FE	0.01635	0.00122	0.00004	0.00002	0.00151	0.00020	0.00520	0.00147	0.00044
		PD	0.01542	0.00108	0.00012	0.00081	0.00112	0.00012	0.00524	0.00267	0.00057
30	0	FE	0.00033	0.00008	0.00001	0.00063	0.00018	0.00001	0.00026	0.00027	0.00006
		PD	0.00039	0.00013	0.00001	0.00064	0.00015	0.00001	0.00012	0.00035	0.00004
	0.5	FE	0.00048	0.00012	0.00010	0.00093	0.00006	0.00011	0.00125	0.00032	0.00009
		PD	0.00044	0.00011	0.00011	0.00092	0.00009	0.00009	0.00103	0.00038	0.00005
	0.9	FE	0.00598	0.00177	0.00048	0.00346	0.00010	0.00004	0.00295	0.00206	0.00046
		PD	0.00637	0.00188	0.00050	0.00441	0.00022	0.00002	0.00286	0.00128	0.00045
40	0	FE	0.00024	0.00014	0.00007	0.00021	0.00026	0.00001	0.00071	0.00022	0.00005
		PD	0.00027	0.00016	0.00007	0.00023	0.00021	0.00002	0.00083	0.00013	0.00005
	0.5	FE	0.00079	0.00010	0.00005	0.00006	0.00042	0.00004	0.00098	0.00017	0.00001
		PD	0.00090	0.00016	0.00002	0.00001	0.00050	0.00002	0.00104	0.00033	0.00001
	0.9	FE	0.00131	0.00041	0.00016	0.00663	0.00175	0.00061	0.00047	0.00016	0.00023
		PD	0.00100	0.00064	0.00011	0.00720	0.00111	0.00086	0.00031	0.00015	0.00030

Table 2.6 Relative Mean Absolute Errors (PD/ FE)

Panel Setting

		P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
	0	0.945	0.978	1.007	0.876	1.309	0.959	1.036	0.383	1.000
2	**0.5**	1.207	0.709	3.111	0.731	0.245	1.426	1.229	1.179	0.324
	0.9	0.977	1.023	0.952	1.304	1.017	1.000	2.321	0.812	0.250
	0	0.917	1.079	1.000	0.756	0.860	1.068	1.582	0.490	0.615
6	**0.5**	6.250	1.004	1.167	0.950	1.373	11.000	1.107	0.200	1.333
	0.9	1.000	1.000	1.000	0.931	0.455	0.941	1.081	1.342	1.286
	0	1.007	1.025	0.971	1.010	1.075	1.040	0.636	0.991	1.478
10	**0.5**	0.135	1.000	4.000	1.131	1.691	1.667	0.651	0.676	3.000
	0.9	0.813	1.500	1.125	0.660	1.318	2.000	0.524	1.571	1.500
	0	1.000	0.991	1.100	1.179	1.020	1.071	1.110	1.091	6.000
20	**0.5**	3.538	0.909	0.800	0.738	1.839	0.000	4.667	0.667	0.667
	0.9	3.000	0.750	1.000	1.117	1.278	1.000	1.071	0.500	1.500
	0	0.962	1.000	0.967	0.930	1.062	1.000	0.817	0.957	1.163
30	**0.5**	1.119	1.000	1.333	3.417	1.617	1.500	1.080	0.167	1.000
	0.9	1.167	1.091	1.000	0.746	1.211	1.000	1.235	0.621	0.875
	0	0.991	0.963	0.985	1.028	1.115	0.969	1.218	1.000	1.083
40	**0.5**	0.353	1.000	1.200	0.429	0.889	0.286	0.918	0.946	0.917
	0.9	0.957	1.000	1.000	0.978	0.400	1.000	1.063	0.000	2.000

Clustered Setting

		P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
	0	3.143	1.122	1.000	0.303	1.761	2.000	0.984	1.016	1.000
2	**0.5**	1.072	1.389	0.913	1.193	0.462	0.500	1.154	1.104	7.000
	0.9	1.111	0.800	1.000	0.747	1.671	0.635	0.715	1.021	0.766
	0	0.872	1.182	1.250	1.222	0.333	0.556	0.151	2.250	1.000
6	**0.5**	0.993	0.883	1.150	1.132	0.855	0.750	0.904	0.918	1.450
	0.9	0.470	1.664	0.842	0.807	0.949	0.683	1.330	0.936	0.641
	0	0.185	1.143	1.000	1.373	1.167	0.000	1.135	1.909	0.900
10	**0.5**	1.060	0.959	1.000	1.340	1.368	1.077	3.089	1.034	0.909
	0.9	1.170	1.680	0.988	0.815	0.877	1.031	1.807	1.875	0.791
	0	1.152	1.040	1.000	0.851	1.471	2.000	1.167	1.185	1.000
20	**0.5**	1.750	0.775	0.909	1.446	0.829	0.800	1.160	1.714	0.778
	0.9	0.943	0.885	3.000	40.500	0.742	0.600	1.008	1.816	1.295
	0	1.182	1.625	1.000	1.016	0.833	1.000	0.462	1.296	0.667
30	**0.5**	0.917	0.917	1.100	0.989	1.500	0.818	0.824	1.188	0.556
	0.9	1.065	1.062	1.042	1.275	2.200	0.500	0.969	0.621	0.978
	0	1.125	1.143	1.000	1.095	0.808	2.000	1.169	0.591	1.000
40	**0.5**	1.139	1.600	0.400	0.167	1.190	0.500	1.061	1.941	1.000
	0.9	0.763	1.561	0.688	1.086	0.634	1.410	0.660	0.938	1.304

Table 2.7 Mean Absolute Errors (Heteroscedasticity)

G		P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	FE	0.01262	0.00827	0.04510	0.00382	0.00536	0.04114	0.00506	0.01345	0.07049
	PD	0.01038	0.00747	0.04517	0.00521	0.00788	0.06711	0.00255	0.01194	0.08642
6	FE	0.00327	0.00978	0.02599	0.00316	0.00337	0.00142	0.00356	0.00320	0.03129
	PD	0.00313	0.00855	0.02792	0.00443	0.00290	0.00364	0.00427	0.00034	0.03163
10	FE	0.00051	0.00074	0.00901	0.00984	0.00115	0.03228	0.01564	0.00144	0.02738
	PD	0.00045	0.00104	0.01013	0.01114	0.00221	0.04977	0.01606	0.00114	0.02727
20	FE	0.00108	0.00324	0.01166	0.00523	0.00148	0.00612	0.00210	0.00139	0.00220
	PD	0.00074	0.00343	0.01389	0.00668	0.00163	0.01700	0.00074	0.00052	0.00433
30	FE	0.00227	0.00157	0.00176	0.00228	0.00495	0.00864	0.00191	0.00277	0.01514
	PD	0.00168	0.00182	0.00238	0.00165	0.00475	0.00468	0.00200	0.00286	0.01466
40	FE	0.00132	0.00037	0.01927	0.00255	0.00242	0.01558	0.00497	0.00150	0.01386
	PD	0.00163	0.00043	0.02009	0.00124	0.00169	0.01739	0.00653	0.00083	0.00915

Table 2.8 Relative Mean Absolute Errors (Heteroscedasticity; PD/ FE)

	P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0.822	0.903	1.002	1.363	1.470	1.631	0.505	0.887	1.226
6	0.956	0.874	1.074	1.403	0.858	2.572	1.199	0.105	1.011
10	0.891	1.416	1.124	1.133	1.918	1.542	1.027	0.793	0.996
20	0.687	1.058	1.191	1.278	1.102	2.777	0.353	0.372	1.971
30	0.744	1.163	1.354	0.723	0.959	0.541	1.045	1.034	0.968
40	1.239	1.169	1.042	0.485	0.700	1.116	1.314	0.554	0.660

Table 2.9 Change Rate % in Standard Error (Panel Setting)

G	ρ_μ		P_1 ($\tau = 0.19$) $\rho_x = 0$	P_1 $\rho_x = 0.5$	P_1 $\rho_x = 0.9$	P_2 ($\tau = 0.51$) $\rho_x = 0$	P_2 $\rho_x = 0.5$	P_2 $\rho_x = 0.9$	P_3 ($\tau = 0.91$) $\rho_x = 0$	P_3 $\rho_x = 0.5$	P_3 $\rho_x = 0.9$
2	0	i.i.d.	-1.278	-1.329	-1.264	-5.535	-5.561	-5.565	-5.383	-5.375	-5.360
		CRVE	-94.731	-94.731	-94.731	-82.144	-82.144	-82.143	-38.467	-38.457	-38.468
	0.5	i.i.d.	6.845	6.833	7.081	15.694	15.579	15.969	2.875	2.895	2.892
		CRVE	-90.773	-90.730	-90.735	-39.591	-39.617	-39.048	-8.246	-8.196	-8.272
	0.9	i.i.d.	-1.105	-1.102	-1.102	-1.457	-1.428	-1.346	0.220	0.487	0.278
		CRVE	-94.659	-94.657	-94.655	-75.276	-75.383	-75.298	-14.756	-13.807	-14.704
6	0	i.i.d.	-1.248	-1.245	-1.239	-5.373	-5.368	-5.396	-5.225	-5.242	-5.218
		CRVE	-0.029	-0.034	-0.028	-0.586	-0.570	-0.605	-4.444	-4.385	-4.355
	0.5	i.i.d.	5.234	5.186	5.138	16.687	16.520	16.664	4.647	4.456	4.499
		CRVE	-4.534	-4.758	-4.294	1.787	1.834	1.593	1.504	1.055	1.357
	0.9	i.i.d.	-1.113	-1.105	-1.106	-2.074	-2.142	-2.127	0.797	0.867	0.839
		CRVE	-0.365	-0.381	-0.389	-4.389	-4.627	-4.399	-1.656	-1.717	-1.843
10	0	i.i.d.	-1.231	-1.249	-1.259	-5.336	-5.375	-5.346	-5.188	-5.186	-5.196
		CRVE	0.000	-0.002	-0.001	-0.094	-0.094	-0.082	-1.661	-1.420	-1.492
	0.5	i.i.d.	4.518	4.570	4.661	15.779	15.627	15.656	4.472	4.436	4.361
		CRVE	-0.019	0.047	0.214	6.446	6.025	6.069	1.632	1.575	1.802
	0.9	i.i.d.	-1.122	-1.107	-1.120	-2.338	-2.338	-2.361	0.560	0.602	0.633
		CRVE	-0.047	-0.054	-0.049	-1.085	-1.055	-0.894	-0.273	-0.535	-0.639
20	0	i.i.d.	-1.248	-1.246	-1.241	-5.338	-5.333	-5.335	-5.174	-5.152	-5.163
		CRVE	0.008	0.008	0.008	0.107	0.108	0.106	0.105	0.144	0.176
	0.5	i.i.d.	4.089	4.119	4.147	14.747	14.820	14.732	4.227	4.273	4.251
		CRVE	2.575	2.672	2.597	9.964	9.996	10.642	2.598	2.438	2.765
	0.9	i.i.d.	-1.118	-1.111	-1.117	-2.544	-2.518	-2.531	0.343	0.326	0.393
		CRVE	0.064	0.062	0.075	1.010	0.977	0.954	0.936	0.977	0.932
30	0	i.i.d.	-1.248	-1.249	-1.239	-5.328	-5.334	-5.315	-5.157	-5.164	-5.153
		CRVE	0.010	0.010	0.010	0.150	0.154	0.150	0.593	0.623	0.582
	0.5	i.i.d.	3.990	3.971	3.976	14.629	14.474	14.553	4.190	4.239	4.151
		CRVE	3.444	3.362	3.411	12.357	11.860	12.063	2.995	3.089	3.086
	0.9	i.i.d.	-1.118	-1.115	-1.108	-2.588	-2.586	-2.599	0.261	0.209	0.199
		CRVE	0.089	0.095	0.098	1.526	1.544	1.569	1.708	1.586	1.627
40	0	i.i.d.	-1.241	-1.245	-1.242	-5.321	-5.318	-5.316	-5.146	-5.155	-5.150
		CRVE	0.011	0.011	0.011	0.170	0.172	0.173	0.834	0.876	0.811
	0.5	i.i.d.	3.888	3.877	3.888	14.448	14.406	14.501	4.205	4.161	4.194
		CRVE	3.716	3.707	3.648	13.322	13.221	13.033	3.347	3.318	3.316
	0.9	i.i.d.	-1.108	-1.117	-1.117	-2.604	-2.611	-2.624	0.185	0.173	0.195
		CRVE	0.102	0.101	0.103	1.785	1.757	1.755	2.111	1.980	1.991

Table 2.10 Change Rate % in Standard Error (Clustered Setting)

			P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	i.i.d.	0.023	0.015	-0.006	-0.004	0.050	0.014	-0.118	-0.012	-0.017
		CRVE	-94.165	-94.122	-94.150	-76.652	-76.528	-76.652	-22.754	-22.751	-22.747
	0.5	i.i.d.	0.010	0.024	-0.034	0.001	0.036	0.072	-0.048	0.060	0.027
		CRVE	-94.175	-94.149	-94.161	-76.661	-76.612	-76.644	-22.912	-22.687	-22.857
	0.9	i.i.d.	0.033	0.032	-0.009	0.095	-0.029	0.045	0.005	-0.032	0.064
		CRVE	-94.161	-94.146	-94.178	-76.564	-76.597	-76.544	-22.510	-22.729	-22.487
6	0	i.i.d.	-0.003	0.006	0.043	0.003	-0.002	-0.018	-0.007	-0.032	-0.037
		CRVE	-1.621	-1.440	-1.374	-4.827	-4.635	-4.662	-2.904	-2.804	-2.757
	0.5	i.i.d.	0.007	-0.023	0.010	-0.027	0.021	-0.028	-0.010	-0.021	0.007
		CRVE	-1.591	-1.526	-1.519	-4.707	-4.491	-4.385	-2.886	-3.096	-3.128
	0.9	i.i.d.	0.017	-0.019	-0.007	-0.033	0.000	0.011	0.061	-0.023	0.007
		CRVE	-1.573	-1.569	-1.660	-4.439	-4.784	-4.578	-2.852	-2.985	-3.045
10	0	i.i.d.	-0.006	0.005	-0.010	0.004	-0.013	-0.021	0.013	0.024	0.029
		CRVE	0.018	-0.009	-0.021	0.289	0.187	0.139	-0.528	-0.247	-0.198
	0.5	i.i.d.	0.000	-0.010	0.005	0.013	0.009	-0.017	0.029	-0.024	-0.008
		CRVE	0.018	0.008	0.066	0.012	0.096	-0.100	-0.139	-0.477	-0.263
	0.9	i.i.d.	0.009	0.017	-0.002	0.011	-0.017	0.030	0.011	0.002	-0.019
		CRVE	0.022	0.037	0.051	0.063	0.090	0.115	-0.456	-0.427	-0.212
20	0	i.i.d.	-0.003	0.004	-0.004	0.028	0.012	-0.014	-0.024	0.010	0.010
		CRVE	0.720	0.713	0.725	2.857	2.781	2.749	1.556	1.570	1.637
	0.5	i.i.d.	-0.008	0.007	0.008	0.001	-0.002	-0.002	0.031	-0.001	-0.027
		CRVE	0.718	0.713	0.728	2.794	2.822	2.637	1.468	1.597	1.507
	0.9	i.i.d.	-0.002	-0.009	-0.014	-0.013	0.001	0.007	-0.009	0.029	-0.007
		CRVE	0.712	0.742	0.726	2.835	2.713	2.789	1.467	1.587	1.627
30	0	i.i.d.	0.003	-0.008	0.008	0.014	0.007	-0.021	0.020	-0.001	0.000
		CRVE	0.919	0.919	0.934	3.514	3.457	3.522	2.244	2.110	2.149
	0.5	i.i.d.	-0.007	-0.007	-0.004	0.000	-0.006	0.015	0.001	0.027	-0.021
		CRVE	0.899	0.901	0.903	3.483	3.549	3.491	2.215	2.173	2.171
	0.9	i.i.d.	0.003	0.011	0.000	-0.007	-0.003	-0.001	-0.003	0.009	-0.015
		CRVE	0.901	0.899	0.890	3.459	3.479	3.456	2.169	2.263	2.129
40	0	i.i.d.	0.007	0.000	0.005	-0.019	-0.001	0.016	0.012	0.019	0.004
		CRVE	0.993	0.983	0.993	3.685	3.809	3.794	2.489	2.546	2.445
	0.5	i.i.d.	0.007	-0.010	-0.004	-0.016	0.021	0.014	-0.013	-0.001	0.018
		CRVE	0.999	0.971	0.985	3.815	3.839	3.810	2.501	2.481	2.570
	0.9	i.i.d.	-0.003	-0.002	-0.005	0.007	0.010	-0.008	0.010	0.017	0.010
		CRVE	0.976	0.984	0.985	3.805	3.812	3.837	2.401	2.484	2.452

Figure 2.1 Change Rate in Mean Standard Error (Panel Setting, $\rho_x = 0.5$)

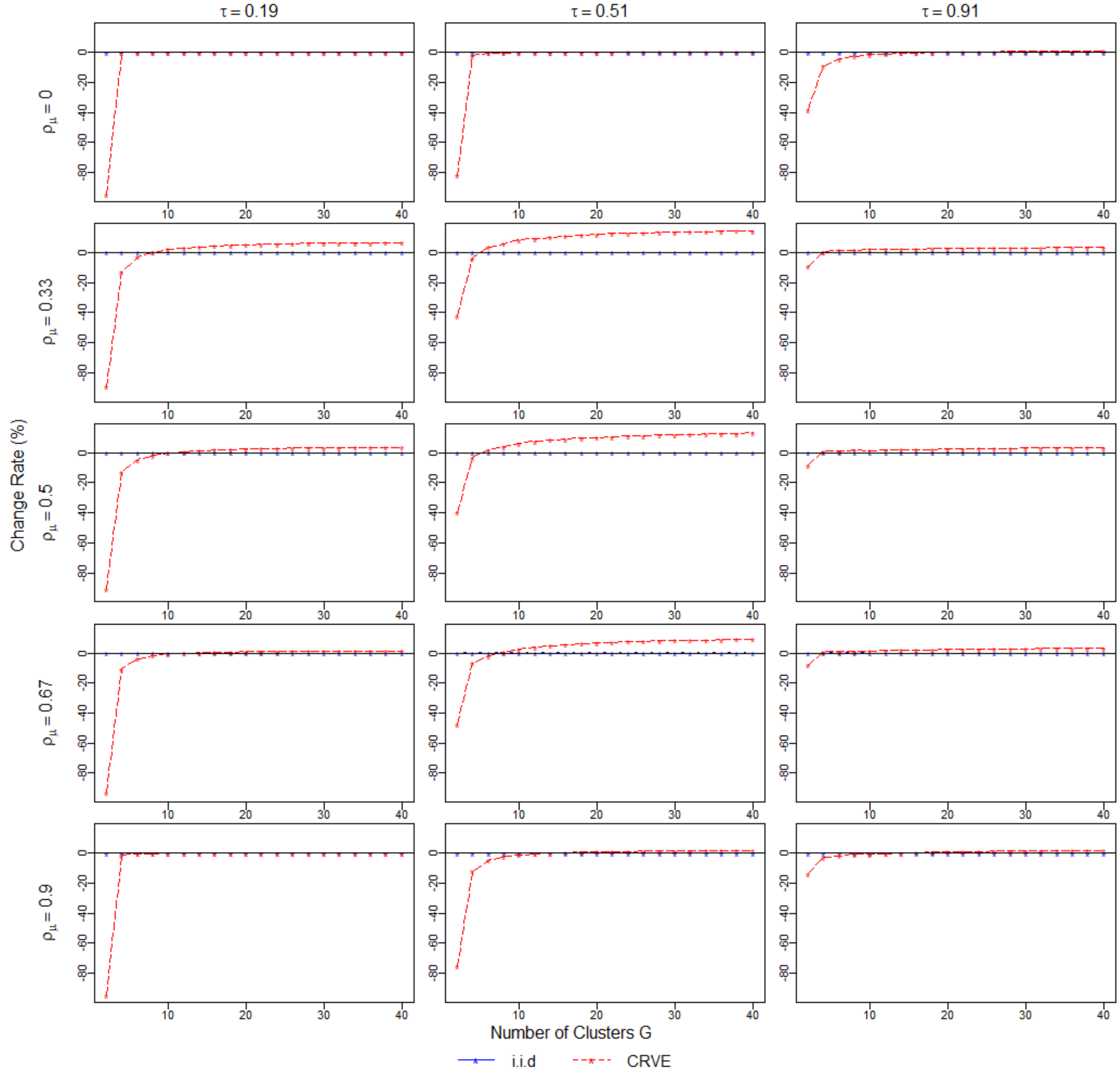


Figure 2.2 Change Rate in Mean Standard Error (Clustered Setting, $\rho_x = 0.5$)

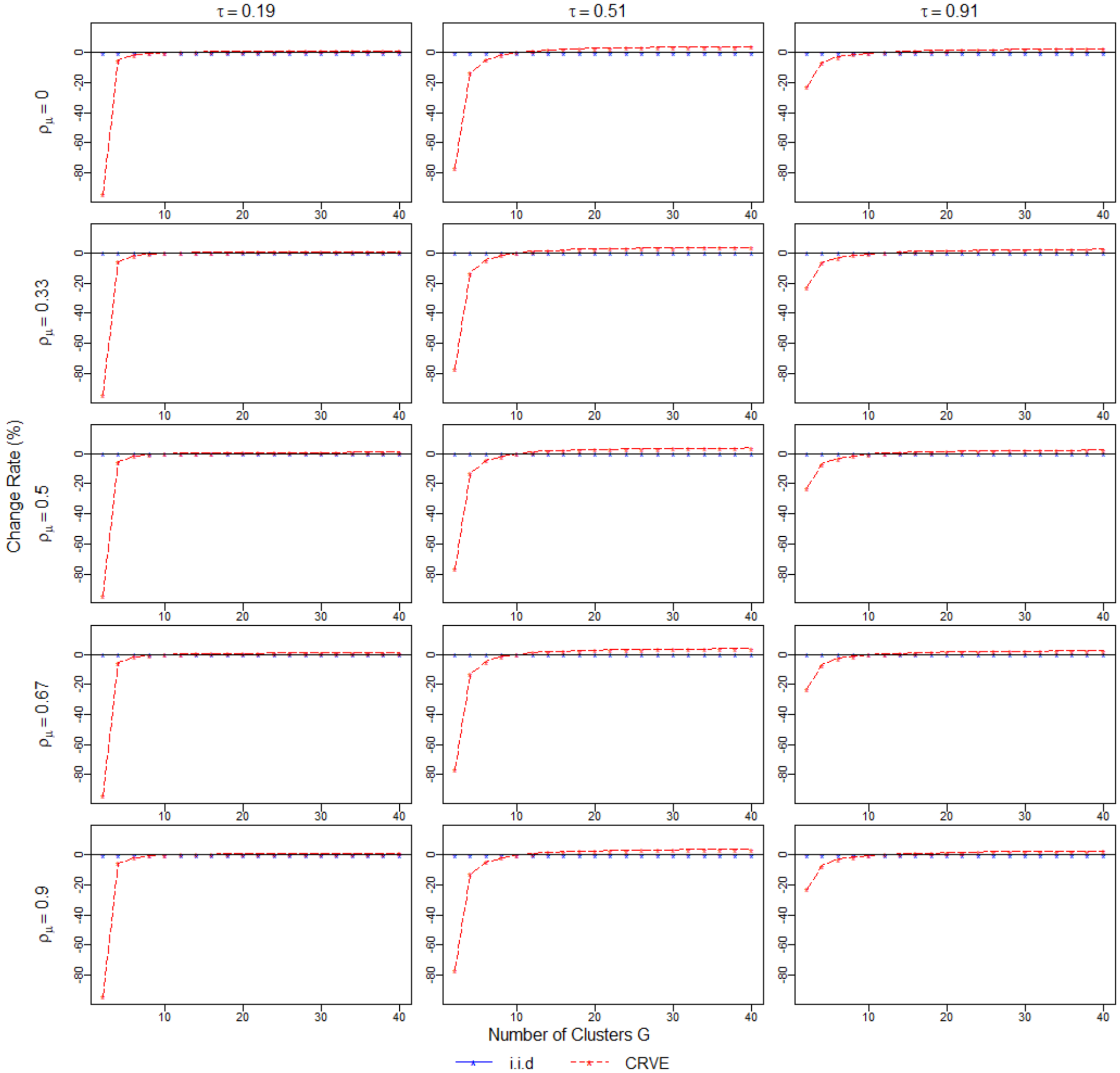


Table 2.11 Change Rate % in Standard Error (Heteroscedasticity)

G		P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	i.i.d.	-0.508	-0.885	-3.702	-1.293	-2.540	-6.578	-0.584	-0.971	-1.947
	CRVE	-93.850	-93.948	-94.106	-75.251	-75.747	-76.496	-21.764	-21.983	-22.362
6	i.i.d.	-0.161	-0.480	-1.567	-0.511	-1.216	-3.828	-0.308	-0.699	-1.424
	CRVE	-0.255	-0.765	-3.929	-3.588	-4.023	-9.403	-2.767	-3.159	-4.346
10	i.i.d.	-0.083	-0.295	-1.075	-0.265	-0.765	-3.147	-0.263	-0.386	-1.101
	CRVE	0.847	0.536	-1.755	0.859	0.375	-4.590	-0.429	-0.332	-1.972
20	i.i.d.	-0.045	-0.146	-0.533	-0.173	-0.462	-1.673	-0.111	-0.247	-0.719
	CRVE	1.313	1.119	-0.203	3.159	2.696	-0.718	1.509	1.197	-0.117
30	i.i.d.	-0.033	-0.091	-0.352	-0.120	-0.346	-1.173	-0.117	-0.174	-0.608
	CRVE	1.444	1.281	0.217	4.010	3.651	0.790	2.081	1.999	0.561
40	i.i.d.	-0.026	-0.067	-0.235	-0.098	-0.275	-0.996	-0.049	-0.134	-0.546
	CRVE	1.519	1.379	0.479	4.308	4.044	1.574	2.426	2.348	0.888

Figure 2.3 Change Rate in Mean Standard Error (Heteroscedasticity, $\rho_x = 0.5$)

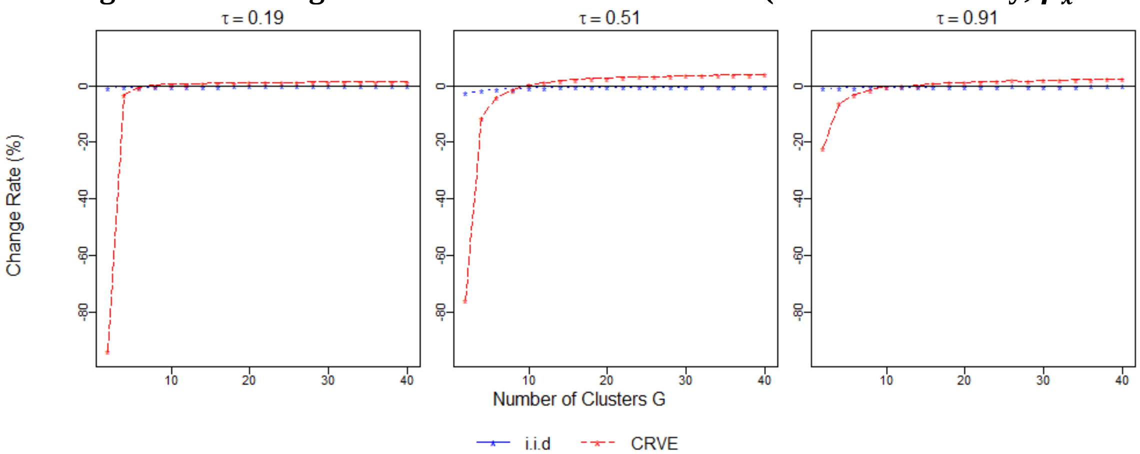


Table 2.12 Rejection Rate (Default I.I.D., Nominal Size 0.05, Panel Setting)

			P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.056	0.059	0.053	0.059	0.056	0.054	0.059	0.060	0.056
		PD	0.060	0.063	0.055	0.076	0.074	0.071	0.080	0.079	0.076
	0.5	FE	0.060	0.059	0.054	0.058	0.055	0.060	0.056	0.058	0.058
		PD	0.063	0.061	0.060	0.058	0.053	0.058	0.053	0.056	0.058
	0.9	FE	0.056	0.058	0.056	0.055	0.060	0.057	0.055	0.062	0.056
		PD	0.060	0.061	0.058	0.072	0.073	0.070	0.060	0.067	0.058
6	0	FE	0.053	0.051	0.050	0.052	0.054	0.049	0.051	0.053	0.051
		PD	0.055	0.054	0.052	0.066	0.070	0.065	0.066	0.072	0.070
	0.5	FE	0.056	0.052	0.051	0.051	0.052	0.048	0.056	0.054	0.050
		PD	0.058	0.054	0.052	0.051	0.057	0.058	0.055	0.052	0.051
	0.9	FE	0.052	0.051	0.051	0.052	0.049	0.050	0.053	0.055	0.055
		PD	0.055	0.053	0.054	0.065	0.066	0.064	0.061	0.064	0.057
10	0	FE	0.051	0.050	0.048	0.052	0.054	0.050	0.054	0.053	0.051
		PD	0.054	0.053	0.051	0.065	0.067	0.064	0.074	0.069	0.067
	0.5	FE	0.053	0.051	0.050	0.057	0.052	0.055	0.055	0.052	0.048
		PD	0.053	0.051	0.054	0.055	0.054	0.055	0.052	0.056	0.049
	0.9	FE	0.051	0.049	0.051	0.050	0.052	0.048	0.054	0.049	0.050
		PD	0.054	0.051	0.054	0.065	0.067	0.061	0.061	0.058	0.058
20	0	FE	0.049	0.050	0.050	0.051	0.049	0.048	0.050	0.054	0.050
		PD	0.052	0.053	0.053	0.065	0.064	0.063	0.067	0.070	0.068
	0.5	FE	0.050	0.050	0.049	0.050	0.049	0.044	0.050	0.051	0.051
		PD	0.051	0.053	0.055	0.055	0.053	0.049	0.051	0.051	0.048
	0.9	FE	0.049	0.049	0.049	0.051	0.050	0.053	0.047	0.049	0.048
		PD	0.052	0.052	0.053	0.064	0.062	0.065	0.056	0.057	0.056
30	0	FE	0.051	0.051	0.049	0.054	0.046	0.052	0.050	0.049	0.053
		PD	0.054	0.054	0.051	0.068	0.060	0.066	0.068	0.065	0.069
	0.5	FE	0.051	0.054	0.051	0.049	0.050	0.046	0.051	0.054	0.049
		PD	0.057	0.057	0.054	0.054	0.053	0.051	0.049	0.053	0.052
	0.9	FE	0.054	0.052	0.051	0.048	0.050	0.050	0.049	0.052	0.050
		PD	0.057	0.054	0.053	0.063	0.061	0.061	0.057	0.060	0.058
40	0	FE	0.049	0.054	0.053	0.051	0.051	0.052	0.050	0.048	0.051
		PD	0.053	0.058	0.055	0.065	0.063	0.065	0.066	0.064	0.068
	0.5	FE	0.050	0.051	0.051	0.053	0.053	0.052	0.052	0.052	0.048
		PD	0.053	0.053	0.054	0.056	0.057	0.056	0.051	0.052	0.048
	0.9	FE	0.050	0.052	0.052	0.052	0.053	0.049	0.049	0.049	0.053
		PD	0.052	0.055	0.056	0.065	0.065	0.062	0.055	0.057	0.059

Table 2.13 Rejection Rate (Default I.I.D., Nominal Size 0.05, Clustered Setting)

			P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_μ		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.059	0.057	0.060	0.055	0.058	0.058	0.059	0.055	0.054
		PD	0.061	0.060	0.064	0.069	0.073	0.069	0.068	0.065	0.062
	0.5	FE	0.055	0.060	0.056	0.057	0.058	0.059	0.058	0.056	0.057
		PD	0.060	0.062	0.059	0.070	0.071	0.070	0.066	0.062	0.066
	0.9	FE	0.056	0.059	0.058	0.056	0.059	0.059	0.058	0.058	0.058
		PD	0.060	0.062	0.062	0.070	0.072	0.071	0.065	0.068	0.063
6	0	FE	0.048	0.055	0.052	0.050	0.055	0.051	0.053	0.054	0.056
		PD	0.051	0.058	0.057	0.062	0.066	0.064	0.059	0.064	0.061
	0.5	FE	0.054	0.052	0.051	0.052	0.052	0.052	0.051	0.052	0.053
		PD	0.057	0.055	0.054	0.067	0.066	0.064	0.056	0.062	0.063
	0.9	FE	0.054	0.052	0.052	0.051	0.048	0.050	0.051	0.054	0.054
		PD	0.056	0.052	0.055	0.063	0.062	0.062	0.058	0.063	0.063
10	0	FE	0.052	0.050	0.054	0.051	0.054	0.053	0.051	0.053	0.053
		PD	0.056	0.055	0.057	0.065	0.065	0.067	0.058	0.062	0.060
	0.5	FE	0.052	0.053	0.050	0.053	0.056	0.055	0.051	0.055	0.050
		PD	0.055	0.055	0.053	0.065	0.071	0.064	0.060	0.063	0.060
	0.9	FE	0.050	0.052	0.052	0.053	0.053	0.050	0.050	0.050	0.052
		PD	0.054	0.055	0.055	0.063	0.063	0.063	0.060	0.057	0.061
20	0	FE	0.051	0.049	0.051	0.051	0.053	0.050	0.050	0.054	0.050
		PD	0.054	0.051	0.054	0.063	0.065	0.063	0.056	0.063	0.057
	0.5	FE	0.051	0.053	0.051	0.054	0.052	0.052	0.055	0.046	0.048
		PD	0.055	0.054	0.055	0.065	0.062	0.062	0.061	0.056	0.056
	0.9	FE	0.050	0.054	0.050	0.049	0.054	0.048	0.051	0.050	0.053
		PD	0.052	0.055	0.052	0.062	0.065	0.060	0.059	0.059	0.061
30	0	FE	0.052	0.049	0.047	0.054	0.047	0.052	0.053	0.052	0.050
		PD	0.053	0.052	0.048	0.064	0.058	0.062	0.060	0.058	0.057
	0.5	FE	0.050	0.051	0.050	0.051	0.049	0.049	0.050	0.050	0.050
		PD	0.053	0.054	0.054	0.064	0.059	0.062	0.059	0.058	0.057
	0.9	FE	0.050	0.052	0.053	0.049	0.053	0.050	0.048	0.050	0.055
		PD	0.055	0.055	0.056	0.061	0.064	0.062	0.056	0.058	0.062
40	0	FE	0.053	0.048	0.051	0.055	0.053	0.051	0.050	0.051	0.055
		PD	0.056	0.051	0.054	0.068	0.065	0.061	0.059	0.060	0.060
	0.5	FE	0.054	0.053	0.050	0.049	0.052	0.047	0.049	0.049	0.053
		PD	0.054	0.057	0.054	0.064	0.063	0.060	0.061	0.057	0.058
	0.9	FE	0.051	0.050	0.051	0.048	0.050	0.049	0.052	0.052	0.052
		PD	0.053	0.051	0.055	0.061	0.064	0.058	0.057	0.059	0.061

Table 2.14 Rejection Rate of CRVE (Nominal Size 0.05, Panel Setting)

G	ρ_μ		P_1 $(\tau = 0.19)$			P_2 $(\tau = 0.51)$			P_3 $(\tau = 0.91)$		
			$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.963	0.967	0.963	0.803	0.809	0.808	0.407	0.415	0.406
		PD	0.998	0.998	0.998	0.964	0.963	0.965	0.570	0.574	0.566
	0.5	FE	0.561	0.553	0.559	0.454	0.445	0.449	0.425	0.412	0.419
		PD	0.954	0.956	0.953	0.669	0.671	0.666	0.458	0.450	0.449
	0.9	FE	0.883	0.875	0.880	0.598	0.606	0.595	0.433	0.426	0.434
		PD	0.994	0.992	0.993	0.882	0.885	0.881	0.504	0.490	0.496
6	0	FE	0.216	0.222	0.212	0.216	0.214	0.211	0.145	0.145	0.144
		PD	0.217	0.222	0.212	0.221	0.220	0.222	0.185	0.185	0.182
	0.5	FE	0.195	0.188	0.190	0.162	0.161	0.159	0.179	0.172	0.174
		PD	0.264	0.255	0.255	0.237	0.240	0.239	0.191	0.181	0.183
	0.9	FE	0.236	0.237	0.237	0.195	0.197	0.196	0.168	0.172	0.179
		PD	0.243	0.241	0.241	0.247	0.247	0.242	0.197	0.199	0.206
10	0	FE	0.134	0.140	0.137	0.136	0.143	0.138	0.112	0.104	0.104
		PD	0.135	0.140	0.138	0.138	0.146	0.139	0.130	0.121	0.125
	0.5	FE	0.134	0.135	0.130	0.127	0.114	0.119	0.129	0.128	0.122
		PD	0.172	0.166	0.166	0.172	0.170	0.170	0.135	0.138	0.131
	0.9	FE	0.153	0.150	0.150	0.139	0.138	0.136	0.122	0.119	0.122
		PD	0.155	0.149	0.150	0.165	0.162	0.162	0.141	0.135	0.142
20	0	FE	0.090	0.087	0.091	0.087	0.090	0.086	0.073	0.077	0.074
		PD	0.091	0.086	0.090	0.088	0.091	0.087	0.078	0.084	0.080
	0.5	FE	0.097	0.095	0.092	0.084	0.083	0.080	0.086	0.086	0.083
		PD	0.110	0.110	0.108	0.108	0.110	0.104	0.092	0.093	0.092
	0.9	FE	0.098	0.097	0.095	0.089	0.094	0.098	0.085	0.081	0.082
		PD	0.098	0.095	0.094	0.097	0.103	0.105	0.095	0.093	0.092
30	0	FE	0.072	0.074	0.077	0.076	0.070	0.075	0.065	0.062	0.069
		PD	0.072	0.074	0.077	0.078	0.069	0.077	0.071	0.065	0.073
	0.5	FE	0.082	0.083	0.082	0.068	0.070	0.070	0.072	0.078	0.074
		PD	0.091	0.091	0.090	0.087	0.086	0.084	0.077	0.079	0.076
	0.9	FE	0.082	0.083	0.080	0.075	0.079	0.075	0.072	0.072	0.077
		PD	0.082	0.084	0.080	0.082	0.083	0.083	0.079	0.083	0.085
40	0	FE	0.070	0.074	0.068	0.070	0.067	0.071	0.060	0.061	0.062
		PD	0.070	0.073	0.068	0.071	0.065	0.071	0.065	0.064	0.065
	0.5	FE	0.075	0.075	0.072	0.068	0.069	0.068	0.068	0.071	0.066
		PD	0.081	0.081	0.078	0.078	0.080	0.085	0.071	0.072	0.071
	0.9	FE	0.073	0.075	0.075	0.071	0.072	0.070	0.069	0.068	0.072
		PD	0.074	0.075	0.075	0.076	0.078	0.073	0.072	0.073	0.077

Table 2.15 Rejection Rate of CRVE (Nominal Size 0.05, Clustered Setting)

			P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
G	ρ_u		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
2	0	FE	0.740	0.735	0.735	0.437	0.448	0.446	0.322	0.320	0.323
		PD	0.984	0.984	0.983	0.839	0.835	0.838	0.413	0.408	0.411
	0.5	FE	0.732	0.729	0.733	0.441	0.442	0.440	0.320	0.320	0.324
		PD	0.983	0.983	0.984	0.834	0.833	0.833	0.407	0.409	0.409
	0.9	FE	0.733	0.730	0.732	0.443	0.449	0.444	0.325	0.321	0.326
		PD	0.983	0.983	0.982	0.837	0.838	0.838	0.406	0.408	0.413
6	0	FE	0.202	0.203	0.203	0.152	0.156	0.160	0.123	0.123	0.116
		PD	0.229	0.224	0.223	0.219	0.221	0.224	0.151	0.157	0.148
	0.5	FE	0.207	0.204	0.204	0.155	0.158	0.155	0.112	0.119	0.118
		PD	0.228	0.222	0.227	0.225	0.224	0.222	0.144	0.154	0.148
	0.9	FE	0.205	0.199	0.205	0.153	0.148	0.152	0.116	0.120	0.122
		PD	0.228	0.222	0.229	0.218	0.219	0.218	0.149	0.150	0.152
10	0	FE	0.131	0.135	0.132	0.109	0.108	0.111	0.089	0.092	0.086
		PD	0.138	0.144	0.143	0.136	0.136	0.138	0.107	0.111	0.103
	0.5	FE	0.137	0.134	0.135	0.116	0.111	0.113	0.089	0.091	0.085
		PD	0.143	0.143	0.142	0.149	0.142	0.143	0.106	0.106	0.105
	0.9	FE	0.132	0.138	0.129	0.106	0.102	0.111	0.092	0.089	0.088
		PD	0.144	0.140	0.139	0.137	0.134	0.142	0.108	0.106	0.107
20	0	FE	0.086	0.087	0.086	0.076	0.084	0.082	0.069	0.069	0.069
		PD	0.089	0.092	0.090	0.090	0.099	0.092	0.074	0.082	0.079
	0.5	FE	0.089	0.091	0.090	0.079	0.080	0.079	0.074	0.066	0.067
		PD	0.092	0.093	0.091	0.092	0.090	0.091	0.083	0.074	0.076
	0.9	FE	0.090	0.089	0.085	0.076	0.082	0.077	0.069	0.069	0.070
		PD	0.093	0.089	0.087	0.083	0.096	0.089	0.079	0.079	0.076
30	0	FE	0.074	0.070	0.070	0.072	0.065	0.069	0.066	0.064	0.057
		PD	0.078	0.073	0.072	0.077	0.072	0.074	0.069	0.068	0.063
	0.5	FE	0.074	0.074	0.068	0.064	0.066	0.067	0.063	0.062	0.062
		PD	0.076	0.077	0.072	0.075	0.073	0.079	0.068	0.067	0.069
	0.9	FE	0.075	0.075	0.074	0.070	0.071	0.067	0.060	0.061	0.066
		PD	0.076	0.076	0.076	0.075	0.079	0.074	0.066	0.071	0.070
40	0	FE	0.072	0.063	0.069	0.068	0.065	0.064	0.058	0.060	0.064
		PD	0.073	0.064	0.070	0.073	0.069	0.067	0.061	0.063	0.069
	0.5	FE	0.069	0.066	0.068	0.060	0.063	0.063	0.058	0.057	0.059
		PD	0.071	0.068	0.070	0.069	0.068	0.070	0.065	0.060	0.063
	0.9	FE	0.071	0.068	0.069	0.062	0.064	0.062	0.059	0.060	0.064
		PD	0.072	0.069	0.071	0.065	0.070	0.066	0.062	0.067	0.066

Table 2.16 Rejection Rate when Heteroscedasticity

G		P_1 ($\tau = 0.19$)			P_2 ($\tau = 0.51$)			P_3 ($\tau = 0.91$)		
		$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$	$\rho_x = 0$	$\rho_x = 0.5$	$\rho_x = 0.9$
						i.i.d.				
2	FE	0.124	0.104	0.598	0.437	0.350	1.626	0.711	0.578	2.524
	PD	0.008	0.006	0.035	0.108	0.085	0.382	0.556	0.451	1.960
6	FE	0.522	0.401	1.682	0.555	0.429	1.882	0.598	0.463	2.017
	PD	0.521	0.398	1.616	0.535	0.412	1.705	0.581	0.449	1.930
10	FE	0.460	0.352	1.501	0.476	0.368	1.615	0.497	0.383	1.700
	PD	0.464	0.353	1.475	0.480	0.369	1.541	0.495	0.382	1.666
20	FE	0.354	0.273	1.227	0.362	0.279	1.261	0.367	0.284	1.315
	PD	0.359	0.276	1.225	0.373	0.286	1.252	0.372	0.288	1.314
30	FE	0.298	0.231	1.041	0.301	0.233	1.080	0.305	0.237	1.108
	PD	0.302	0.234	1.043	0.313	0.241	1.089	0.312	0.242	1.115
40	FE	0.264	0.204	0.940	0.265	0.204	0.949	0.267	0.207	0.977
	PD	0.268	0.207	0.945	0.276	0.212	0.964	0.273	0.212	0.986
						CRVE				
2	FE	0.387	0.301	0.106	0.374	0.290	0.090	0.379	0.290	0.086
	PD	0.398	0.313	0.115	0.399	0.320	0.092	0.389	0.301	0.087
6	FE	0.369	0.276	0.064	0.374	0.266	0.066	0.376	0.262	0.062
	PD	0.382	0.287	0.069	0.396	0.296	0.069	0.395	0.278	0.065
10	FE	0.370	0.268	0.065	0.373	0.267	0.057	0.368	0.264	0.057
	PD	0.375	0.279	0.071	0.404	0.288	0.070	0.379	0.277	0.059
20	FE	0.379	0.261	0.056	0.365	0.265	0.058	0.375	0.259	0.058
	PD	0.387	0.267	0.061	0.388	0.288	0.067	0.388	0.280	0.064
30	FE	0.374	0.259	0.059	0.368	0.245	0.057	0.373	0.254	0.057
	PD	0.383	0.265	0.061	0.393	0.271	0.069	0.387	0.272	0.065
40	FE	0.372	0.254	0.061	0.372	0.252	0.054	0.360	0.261	0.057
	PD	0.384	0.257	0.064	0.399	0.276	0.064	0.383	0.277	0.067

CHAPTER 3
MODELING BUY-SELLER IN HOUSING MARKET—A BIDDING NETWORK PERSPECTIVE TO UNDERSTAND MARKET MECHANISM

3.1 INTRODUCTION

3.1.1 Define Buy-sellers

This paper studies “buy-sellers” in housing markets. A buy-seller refers to a household or individual who is making both buying and selling decisions which are interdependent in certain ways. A typical example is an intra-urban residential mover. Imagine in a city an owner-occupier wants to buy a new house. Due to some reasons, he may have to sell the current unit to achieve the purchase. There are two possible reasons for this: 1) budget constraints: a household has to cash the owning unit to afford the new purchase; 2) policy constraints: a household may not be allowed to own more than a certain number of units due to housing policies such as purchase restrictions. In either case, a purchase decision is contingent on a successful sale. This paper calls this dependency a “must-sell” restriction and the household with such restriction a “buy-seller”.

Also, for a buy-seller, the final decision on sale may reversely depend on a successful purchase, or at least an accepted bid. This commonly happens to single-house owner-occupiers. They may be reluctant to sign a sale contract without a guaranteed purchase; otherwise, they may end up in a rental unit. This paper calls this dependency a “must-buy” restriction. Note that this restriction may not exist for multi-house owners or those single-house owners who do not care to rent. In all, a buy-seller in this paper must be constrained by the “must-sell” restriction and may additionally be subject to the “must-buy” restriction. The two restrictions and the resultant buy-sell interdependency distinguish the buy-sellers from all household participants that have been modeled in the literature.

3.1.2 Resale Behavior and Dichotomic Tradition

Models of resale housing markets have discussed both buying and selling behaviors of households extensively. However, most modeling efforts have constrained a household in a dichotomic manner, as either a buyer or a seller. In a few studies, being a buyer and seller at the same time has been acknowledged and such "joint buyer-seller problem" (Anenberg & Bayer, 2013) has been modeled to study the volatility of transaction volume and prices in urban housing markets (Moen, Nenov, & Sniekers, 2015; Stein, 1995). In addition, the literature on intra-urban mobility, housing adjustment, and moving and transaction costs have shed light on the relationship between the owning and the desired units (e.g., Rosen, 1986; Winstanley et al., 2002; Brasington, 2014). However, neither "must-buy" and "must-sell" restrictions nor the interdependency feature of buy-sellers has been explored. One possible reason is that the dichotomic framework would regard a buy-seller as two independent roles in two different transactions: a buyer in one and a seller in another, treating the interdependency as nonexistent or trivial. In fact, the interdependency feature of buy-sellers not only exists but also provides rich insights into housing market mechanisms. In recent literature of economic matching theory, Fleiner et al. (2017) explicitly argue, "*houses are highly differentiated, and agents might act as both buyers and sellers, making the housing market an interconnected trading network*." This paper shows this point by illustrating how buy-sellers build connections in bidding networks in an agent-based model (ABM).

3.1.3 Bidding Network Insights

In essence, the "must-buy" and "must-sell" restrictions not only change the decision process of households, but also make buy-sellers influential "nodes" in trading networks. From the perspective of the graph theory, at any given time, buyers and sellers are vertices (nodes), and the bids made by buyers for sellers are edges (links). Without buy-sellers, all vertices are touched only by edges from one side (bidding or bidden); and the whole market comprises disconnected and independent edges between each buyer-seller pair. If the bid is accepted, the edge becomes a transaction; otherwise, the edge is eliminated, and the connected buyer has to make a new bid in the next step. For buy-sellers, since they bid for others' housing units at the same time receive bids from others, they are touched by edges from both sides and thus connect the market as a network.

Beyond the connectedness, what is more interesting is the dependency in the network. The "must-buy" and "must-sell" restrictions make two edges connecting a buy-seller depend on each other. Any success or failure of any bids (links) will influence the results of other bids via the buy-seller nodes throughout the network. For example, a sale contract of a buy-seller is breached by his buyer for some reasons. Being constrained by "must-sell" restriction, this buy-seller has to withdraw his successful bid for a new house, which is a breach of another sale contract directly affecting the owner of that house. If that owner is also a buy-seller, this "torch of the breach" will be passed on to the end of the chain, which is typically a pure-seller.

3.1.4 Existence of Buy-sellers

Buy-sellers exist in housing markets which have factors reinforcing both restrictions. In markets where houses are less affordable; and where housing purchase restrictions are harsh, the "must-sell" restriction is strong. In markets where single-house owners prevail, and the notion of home ownership is strong, the "must-buy" restriction would be strong. For example, in Beijing, the past decade witnessed a housing market boom. The government of Beijing has issued a series of policies to cool down the market, including the housing purchase restrictions (HPR). Also, Chinese people have a particularly strong notion of ownership, which makes rental an inferior choice. The author conducted a housing consumer survey in Beijing in summer 2016, involving 1129 housing buyers. One finding is that almost half (45%) of the housing buyers are "exchanging" housing, that means they sold or are selling their houses before the current purchase. Although an exchange buyer is not necessarily a buyer-seller (BS) since the “must-sell” and “must-buy” restrictions may not exist, this high ratio shows the potential of buy-sellers representing almost half of the market demand.

3.1.5 Agent-based Models

Agent-based modeling (ABM) is commonly employed to simulate the dynamics of urban and housing markets, aiming at reproducing the emergence of an observed pattern of social phenomena. It has advantages in modeling land and property market from a button-up perspective and demonstrates flexibility and implicative power for public policies. Traditional ABMs in housing and land market mainly employ urban economic theories as theoretical foundations to model the behaviors of household agents (Filatova, 2015), and have developed the behavioral designs of both buyers and sellers for decades (Parker and Filatova, 2008).

Among all urban-residential ABMs, 59% of them involves household relocation process. See Huang et al. (2014) for an excellent review. However, the conventional practice is to assume sellers are either 1) landlords who charge rents or 2) households who are leaving the market. The first assumption follows the traditions of classic location models (Alonso, 1960) and ABMs (Schelling, 1971). The market supply is only from landlords, and the agents are merely renters rather than owner-occupiers and move without transactions. The second assumption involves housing transactions but assumes that the sellers quit the market without re-purchase behaviors, suggesting that they are either leaving the region or staying in other properties of their own. Both assumptions are restrictive for the development of empirical ABMs of the regions with dominant resale housing markets.

3.2 A SIMPLE FRAMEWORK

Imagine a market with eight households, households 1-3 are pure-buyers (PB), households 4-7 are buy-sellers (BS), and household 8 is a pure-seller (PS). In this market, each buyer (PB + BS) makes a bid for a house owning by a seller (BS+PS). After bidding, a directed graph is formed with nodes representing households (PB, BS, and PS) and directed edges representing bids from buyers to sellers. (Figure 3.1)

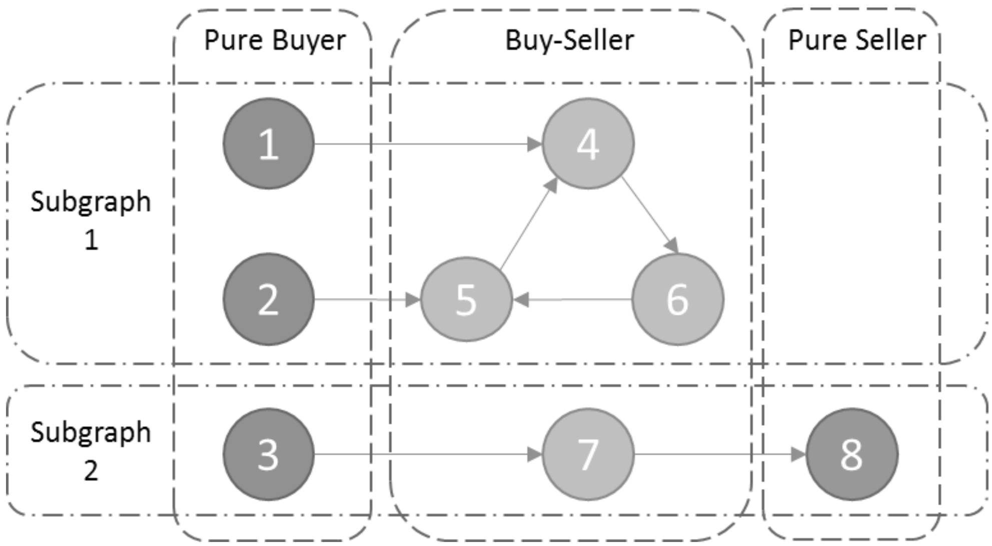


Figure 3.1 A Bidding Network Example

Traditionally, a bid is either accepted or declined. The accepted bid becomes a trade; the declined buyer will search and bid in the next round. With the presence of buy-sellers, this situation is more complicated. For example, BS household 4 receives two bids from 1 and 5, respectively. For any reason, household 4 accepts household 1's bid and declines household 5's. If household 5 is a BS with both "must-buy" and "must-sell" restrictions, the failure in bidding forces him to decline household 2 and 6. Similarly, household 6, as a BS, has to decline household 4's bid. Consequently, household 4 fails and has to decline 1's bid. At last, the entire subgraph 1 fails. In contrast, subgraph 2 could be successful. After being accepted by household 8, the household 7 could accept 3's bid; and both bids become successful trades.

In fact, with the presence of the buy-sellers who are restricted by the “must-buy” and “must-sell” conditions, there are only two successful patterns: a loop pattern connecting BS and a chain pattern connecting PB, BS, and PS. In reality, every final trade between a home buyer and a home seller should be a component of patterns in Figure 3.2.

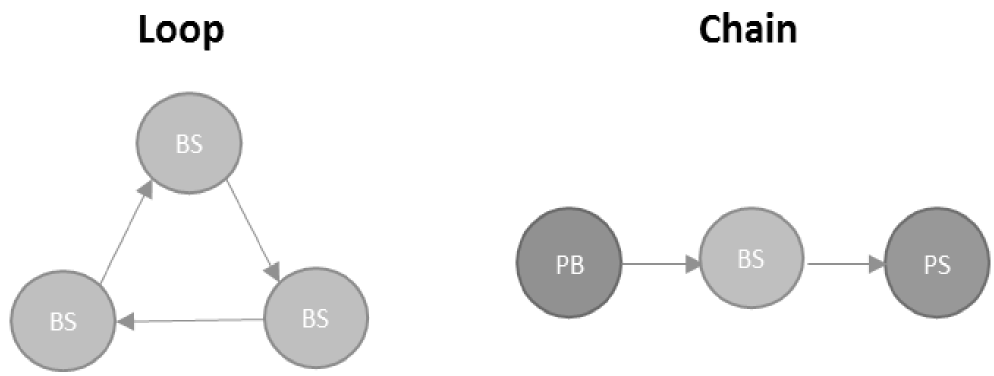


Figure 3.2 Successful Patterns with Buy-sellers

One may argue that once being declined, the BS does not need to decline all their bids. The BS could choose to accept the highest bid and ask the buyer to wait until the BS finds a new house. If this is the case, the subgraph 1 will not fall apart immediately. Household 5 could still accept, say 2, and ask household 2 for a contract which holds the bid for a term. Now, instead of failing, both household 5 and 2 are pending on the market. Since household 2 is a PB, if household 5 could succeed in bidding a house of a PS, they three could form a chain pattern and successfully finish the transactions. However, if household 5 bids a house of another BS, the destiny of both 5 and 2 depend on that of the new BS. This is how BS connects and affects others on the bidding network.

3.3 AGENT-BASED MODEL DESIGN

3.3.1 Agents

There four types of agents in this ABM: buyers, sellers, bids, and patterns. Their generation, interaction, and elimination form the flow of the model.

Buyer, Seller, and Buy-seller

The buyers and sellers refer to households. This ABM models a buy-seller (BS) as two identities—buyer and seller. At each round, the ABM generates two groups of agents: buyers and sellers. A buy-seller (BS) is a pair of buyer and seller. Thus, the buyer group comprises the pure-buyers (PB) and the buyer side of BS (denoted as BS^b); the seller group comprises the pure-sellers (PS) and the seller side of BS (denoted as BS^s). In this ABM, this is realized by allocating ID to the agents in the following way.

- In a certain round, among the existing buyers and sellers, the largest ID number is ID_{max} (if in the first round, ID_{max}=0). For the new added N_B^{new} new buyers and N_S^{new} new sellers, index their IDs as $ID_{max}+1,\dots,ID_{max}+N_B^{new}$ and $ID_{max}+1,\dots,ID_{max}+N_S^{new}$.
- Since the number of the new buyer-sellers N_{BS}^{new} cannot be larger than either N_B^{new} or N_S^{new}. Thus, specify a share parameter α_{BS} such that $N_{BS}^{new}=\alpha_{bs}\times\min(N_B^{new},N_S^{new})$. And the first N_{BS}^{new} of the new IDs of the buyers and sellers are denoted as buyer-sellers. Then, the number of new pure-buyers (PB) is N_{PB}^{new} is $N_{PB}^{new}=N_B^{new}-N_{BS}^{new}$. The number of new pure-sellers (PS) is N_{PS}^{new} is $N_{PS}^{new}=N_S^{new}-N_{BS}^{new}$. The share of the BS in the new population is $\frac{N_{BS}^{new}}{N_{PB}^{new}+N_{PS}^{new}+N_{BS}^{new}}$.

An example is shown in Figure 3.3. Note that the BS^b and BS^s have the same ID number.

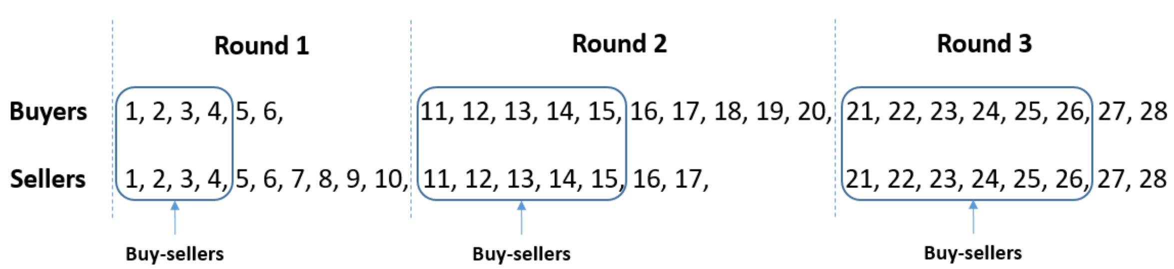


Figure 3.3 ABM ID Allocation

Behaviors and Status

Buyers and sellers have three statuses: active, success, and pending. The new added buyers and sellers are active.

An active buyer will sort (randomly) all available houses (sellers) on the market and make a bid for the favorite. If the bid is not accepted, this buyer stays active and does the same things in the next round. If the bid is accepted, this buyer could be in a certain bidding pattern. If the pattern is a successful one, the buyer's status changes to success and quit the market. If the pattern is a pending one, the buyer's status changes to pending and sign a contract with a maximum waiting time T rounds. In each of the following T rounds, the pending buyer will check whether the pending pattern changes to a successful one. If it does, the buyer's status changes to success; otherwise, the buyer becomes active and restarts from the very beginning.

An active seller waits for bids from buyers. Without receiving any bids in a certain round, the seller stays active and keeps waiting until the next round. Once receiving at least one bid, the seller will choose the highest bid to accept and lie in a pattern. Similar to the buyer, depending on the pattern being successful or pending, the seller's status changes to success or pending; the pending seller will wait and change into success or back to active.

The status of a BS depends on the status of both buyer and seller sides. Each side would behave in the corresponding way. If both sides are in the same status, the BS would be in that status. Due to the must-buy and must-sell restrictions, no single side could be successful alone. So, the complexity emerges when two sides have different status, one active and one pending (Table 3.1). In these situations, only one side—either BS^b or BS^s—of the BS would be active and act as a PB or PS in bidding patterns. The details of the behavior flow are shown in Figure 3.4.

Table 3.1 Buy-seller Status

Seller Side \ Buyer Side	**Active**	**Pending**	**Success**
Active	Active BS	BS^s	
Pending	BS^b	Pending BS	
Success			Success BS

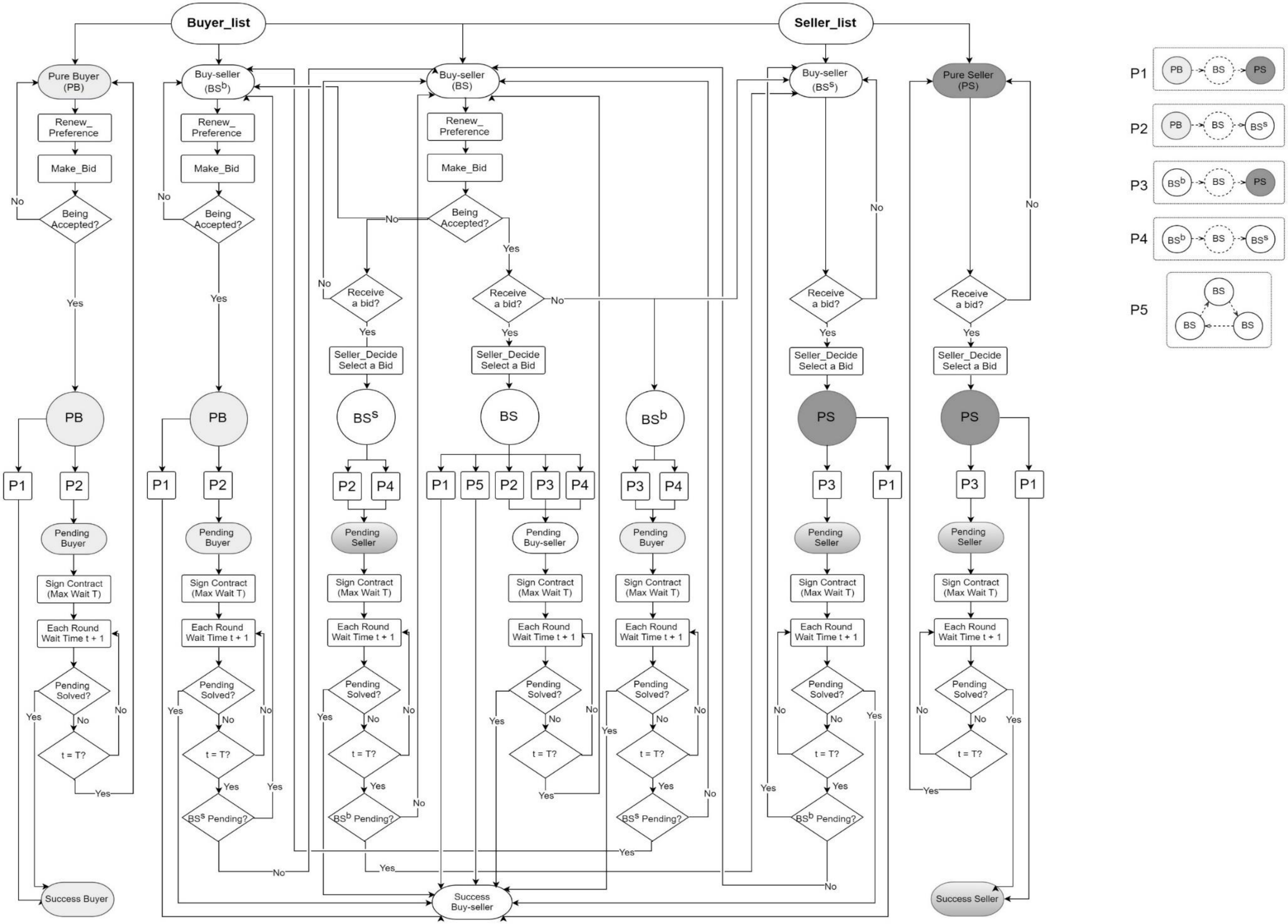


Figure 3.4 Flow Chart of Agents

3.3.2 Bids

A bid is a link connecting a buyer and a seller. A bid is formed when a buyer makes a bid to a seller. A status of a bid could be accepted, declined, successful, and pending. After being accepted, a bid could be in a certain pattern. If the pattern is pending, the bid's status changes to pending with a maximum waiting time T; if the pattern is successful, the bid is successful. For a pending bid, it stays in a pending status until either the pending pattern changes into a successful one or the waiting time is up. In the former case, the bid becomes successful; while in the latter, the bid becomes a declined one. This flow is illustrated in the Figure 3.5.

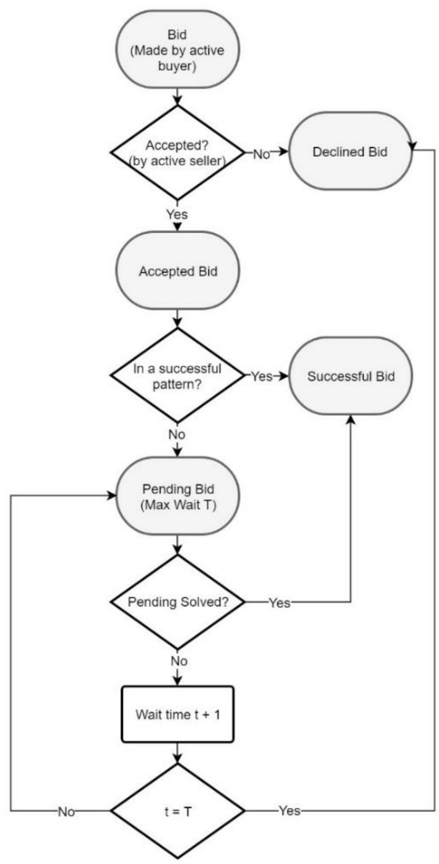


Figure 3.5 Flow Chart of Bids

3.3.3 Patterns

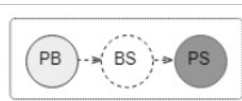
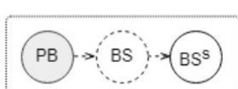
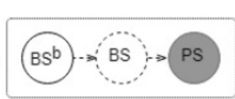
Five Basic Types

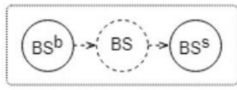
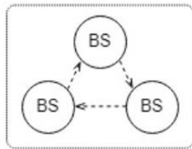
A pattern is a list of accepted bids. It is formed after sellers make the decision of accepting or declining bids. There are five basic types of patterns (Table 3.2).

The first and the fifth are successful patterns. The first pattern is the successful chain pattern, starting from a PB and ending by a PS with any number of BS in between; the fifth pattern is the only loop pattern, connecting at least two BS in a closed loop. All buyers, sellers, and bids in these two patterns are successful.

The rest three patterns are all pending chains. They are pending because at least one of two endings of the chain are BS. In the second pattern, the right ending BS fails as a buyer but succeeds as a seller. Thus, only the seller side BS^s enters this pattern and becomes pending; while the buyer side of this BS—BS^b—is still active and act as a PB in the next round. Similarly, in the third pattern, the left ending BS has a pending BS^b but active BS^s. The fourth pattern has the similar situation at both ends. In all three patterns, the BS in between the two ends are pending BS.

Table 3.2 Five Basic Types of Patterns

Pattern	Form	Status
P1 	Chain	Success
P2 	Chain	Pending
P3 	Chain	Pending

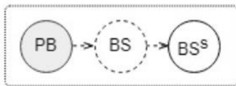
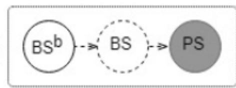
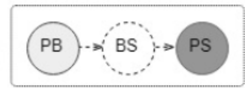
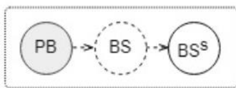
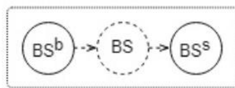
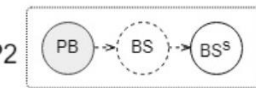
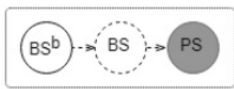
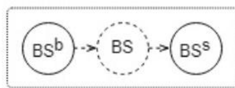
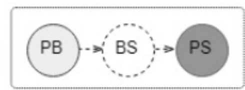
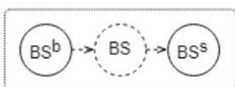
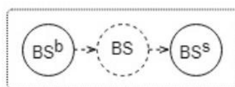
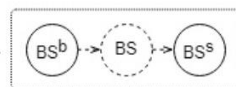
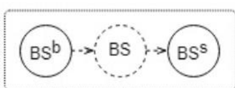
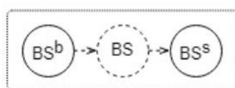
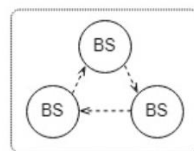
P4		Chain	Pending
P5		Loop	Success

Growth and Decay

A pattern grows by combining with other patterns and decays when any of the bids in the pattern expires.

In each round, the market generates new pending patterns. These patterns could be combined with existing patterns. There are five basic combinations between two pending patterns (Table 3.3). The combinations could be successful patterns or new pending patterns. For example, in the first round, a BS household signed a contract with a PB but failed in bidding for the target house. Thus, this BS is pending on the seller side (i.e., BS^s in the Pattern 2) but active on the buyer side BS^b. In the second round, this BS successfully bid a house of a PS, thus his buyer side BS^b falls in the pattern 3. Since both buyer and seller sides are successful, this BS, together with the PB and PS, becomes successful. In the graph, this means that Pattern 2 and Pattern 3 are combined into a Pattern 1, the successful chain pattern (the first row in the Table3.3). If in the second the BS bid for a house of another BS but that BS fails in bidding, as a result, the former BS becomes pending BS (on both side) and has to wait for the latter one to find in the next round. This is the combination between a Pattern 2 and a Pattern 4 (the second row in the Table 3.3), and the result is still a pending Pattern 2. Note that the combination could happen among more than two patterns. The result could be thought as multiple steps of combinations of two patterns.

Table 3.3 Basic Combination of Two Pending Patterns

Index	First Pattern	Second Pattern	Combined Pattern	Status
1	P2 	P3 	P1 	Success
2	P2 	P4 	P2 	Pending
3	P3 	P4 	P1 	Pending
4	P4 	P4 	P4 	Pending
5	P4 	P4 	P5 	Success

At the end of each round, each pending bid checks whether the contract expires. If the contract does, the bid becomes declined, the statuses of the connecting buyer and seller switch from pending to active, and the pattern is broken at this link into pieces. The pieces are new patterns. If all bids in a pattern expire at the same time, the pattern disappears.

3.3.4 Model Flow

There is a clock in the world and counting as rounds t. It is the time unit for agents to bid, to wait, and to decide. One unit of t could be thought as a month. The model proceeds in the following steps shown in the Figure 3.6.

1. ***Adding New Agents***. At the beginning of each round, new buyers and sellers are added to the existing active ones.

2. ***Buyer rank preference***. Each active seller is assumed to sell only one house on the market. Each active buyer is assumed to know all houses listed on the market (maybe through an online system) and form a preference list over them.

3. ***Buyer make a bid***. Each active buyer bids for the first available house in his preference list. Each active buyer is allowed to make only one bid each round.

4. ***Seller accept & decline bids***. If there is no bid for a seller, that seller finishes this round and stay active for the next round; otherwise, the seller screens all received bids, accepts one and declines the others. Every bid has the same probability to be chosen, except that the probability for the bids from buyer-sellers may be $\gamma_{bs} \geq 1$ times those from pure-buyers. This advantage represents the buy-sellers' higher bidding power than pure buyers who have no assets. $\gamma_{bs} = 1$ means equal power between PB and BS. All declined buyers stay active for the next round and all declined bids are eliminated.

Each accepted bid is identified in one of the five basic patterns. Successful chain (Pattern 1) and loop patterns (Pattern 5) and the involving buyers, sellers, and bids are moved to the successful stock.

5. ***New pending sign contract T and Existing pending wait time t + 1***. In each new pending pattern, the buyer and the seller for each component bid sign a contract for a maximum waiting time T. For the existing pending patterns, the waiting time of each component bid and involving buyer and seller accumulates one.

6. ***Combine pending patterns***. Each new pattern is check for combination. If its combination with any of the existing pending patterns is a successful pattern, both new and existing component patterns are moved from pending to successful. If the combined pattern is still pending or there is no combination, the new pending pattern moves into the pending pattern stock.

7. ***Bid declined and pattern decays.*** For each bid in the pending patterns, if its contract expires, the bid is declined, its connecting buyer and seller become active again, and its superior pattern decays.

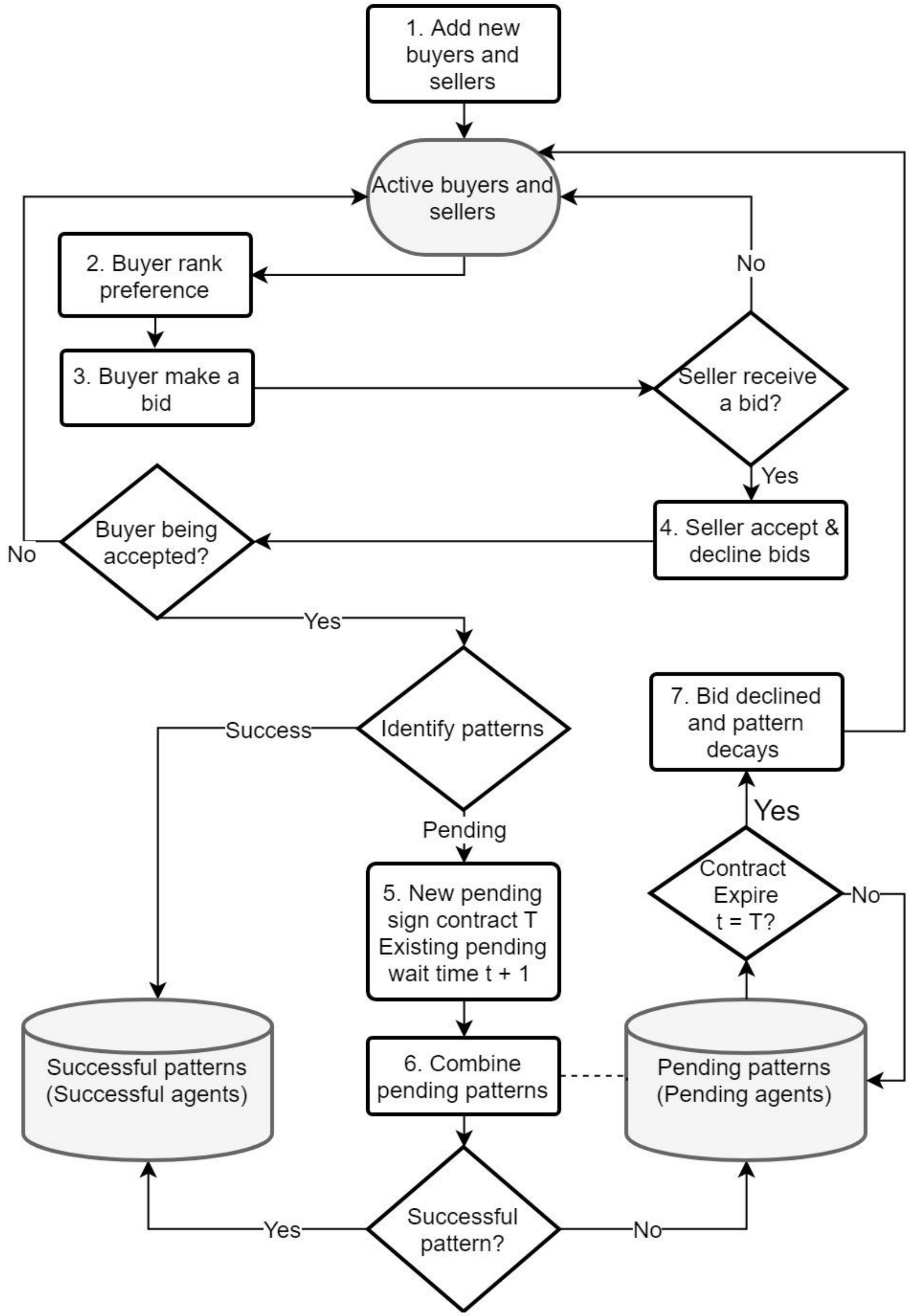


Figure 3.6 Flow Chart of ABM

3.3.5 Algorithm to Identify Patterns

This paper develops an algorithm to identify all patterns—both successful and pending. Traditional ABM algorithms can only be applied to identify Pattern 1 between PB and PS. For chain patterns (Pattern 1 to 4), the key is to find buyers or sellers at either end of the chain, which are touched by one accepted bid on one side. Starting from the agent at the end, one could find all connected buyers and sellers in the same chain pattern along bids. However, in the loop pattern (Pattern 5), there are no ends. The best way to identify Pattern 5 is that after identifying all four chain patterns, the left unidentified bids must be in loop patterns. The algorithm contains following steps.

1. Starting from each PS, keep checking the buyer along connected accepted bids until:
    a. the buyer is a PB. All preceding bids form a Pattern 1 (PB→PS);
    b. the buyer is a BS without an accepted bid at the seller side. All preceding bids form a Pattern 3 (BS→PS).
2. For each PB who has not been checked in the first step, the seller of the accepted bid must be a BS (otherwise the PB should be checked in the first step). Keep checking the BS until the buyer side has no accepted bid. All preceding bids form a Pattern 2 (PB→BS).
3. For each BS who is linked by only one bid, keep checking the BS on the other side of the bid until the latter BS has only one bid. All preceding bids form a Pattern 4.
4. For those BS who have not been checked yet, their bids form loop patterns.

3.4 HOUSING PURCHASE RESTRICTION AND HOME BROKERAGE

This section uses the ABM with buy-sellers to study the effects of housing purchase restrictions. Housing purchase restrictions (HPRs) in China have drawn a lot of attentions on its huge influence on markets and households. Current literature focuses on the effectiveness of this policy instrument using longitudinal housing price data in multiple cities and employing econometric method such as difference-in-differences analysis (Du & Zhang, 2015; Sun et al., 2017). These studies provide evidences on how HPR has reshaped the housing price growth. However, the micro mechanisms of HPR affecting intraurban housing market are still understudied.

Intuitively, HPRs impose purchase amount constraints on some households so that their housing demand cannot be fulfilled. For households with necessary and urgent demand for new housing units, they would sell the current units and become buy-sellers (BS). Thus, one direct consequence of HPRs is the increase in the share of the buy-sellers (BS) on the market. This brings in new market supply as well as trading barriers due to the must-buy and must-sell restrictions of BS. There could be a considerable indirect impact on households who are not directly subject to the HPRs but connected to the BS households through the trading network. It would be interesting to know that with an increase in the share of BS whether the total transaction volume at each period would decrease and whether the average time-on-market (TOM) of households would increases. If they would, these are potential social costs of HPRs which have not been (and cannot be) studied in the tradition market framework.

In addition, this ABM could also provide insights of roles of real estate brokerage in housing transactions. Real estate agents facilitate trades by providing information and third-party guarantees to both buyers and sellers. This paper focuses on their function of providing binding contracts which would reduce frictions caused by BS. For example, contracts with home sale contingencies. These types of contracts are costly and less likely to be made without agents. In the areas with the dominant dual agencies, the effect is more influential. One would expect the real estate brokerage would offset the negative effects of HPRs by providing binding contracts to prevent trades involving BS from falling apart immediately.

Therefore, in this section, one on hand, the effect of HPRs is regarded as an exogenous shock to the share of the BS on the market. Thus, the change in the BS market share represents the magnitude of HPRs. On the other hand, the term of contracts is used to proxy the power of real estate brokerage. The longer the term is, the more powerful the agents are.

3.4.1 Simulation Scenarios

This section conducts simulation in nine scenarios. There could be no HPR, mild HPR, or strong HPR; and at the same time weak, normal or strong brokerage. Correspondingly, the share of BS and the term of contracts are specified at different levels. The low, middle, and high shares of BS are set at 0.2, 0.4, and 0.6, respectively to represent the result from no HPR, mild HPR, and strong HPR. The existence of the 20% BS with the absence of HPRs is due to financial restrictions and reluctance to renting. The term of contracts is set to be 1 round, 3 rounds, and 6 rounds, suggesting the length of simulation rounds within which a pending contract could hold.

Table 3.4 Simulation Scenarios

		Term of Contract		
		Short (1 Round)	**Middle (3 Round)**	**Long (6 Rounds)**
Share of BS	**Low (0.2)**	No HPR; Weak Brokerage	No HPR; Normal Brokerage	No HPR; Strong Brokerage
	Middle (0.4)	Mild HPR; Weak Brokerage	Mild HPR; Normal Brokerage	Mild HPR; Strong Brokerage
	High (0.6)	Strong HPR; Weak Brokerage	Strong HPR; Normal Brokerage	Strong HPR; Strong Brokerage

3.4.2 Outcome Measure

This simulation is essentially a comparative analysis. The effects of HPRs and home brokerage are obtained by comparing the outcomes of the nine scenarios. The section uses two measurements to evaluate the simulation outcome: total number of trades (TNT), i.e., the total number of successful trades; and average time on market (ATOM), i.e., the average time spent on searching or pending on the markets by the traded households.

3.4.3 Model Setting

For each scenario, specify the share of BS and the term of contracts. Initially, generate 500 buyers and 500 sellers on the market with the BS share. In each following round, add 200 new buyers and 200 new sellers to the simulation with the same BS share. The model runs 48 rounds and each round follows the process introduced in the previous section.

3.4.4 Result

The results are shown in Table 3.5 & 3.6. Stronger HPRs or weaker brokerage would decrease the total number of trades and increase the average time-on-market. If there is no HPR and the brokerage is very strong, the market generates the largest number of trades and shortest time-on-market for each household. The change in HPR makes larger differences than the change in brokerage. Specifically, the market outcomes between 3-round and 6-round term of contracts are very similar.

Table 3.5 Total Number of Trades

		Brokerage		
		Weak (1 Round)	**Normal (3 Round)**	**Strong (6 Rounds)**
HPR	**No (0.2)**	7947	7971	7976
	Mild (0.4)	5884	5962	5956
	Strong (0.6)	3729	3950	3960

Table 3.6 Average Time-on-market

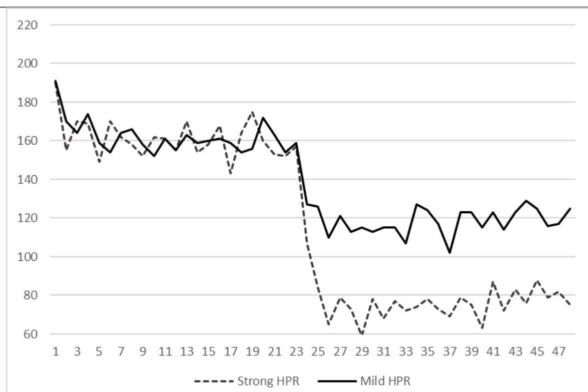
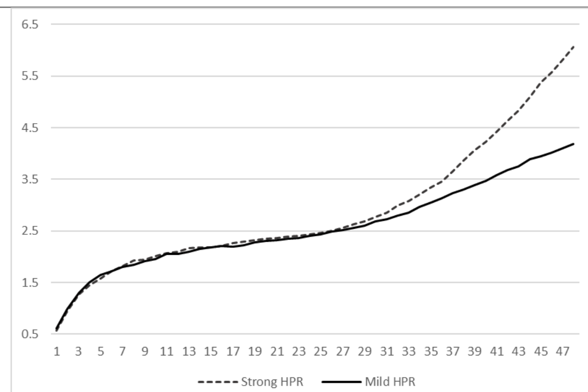
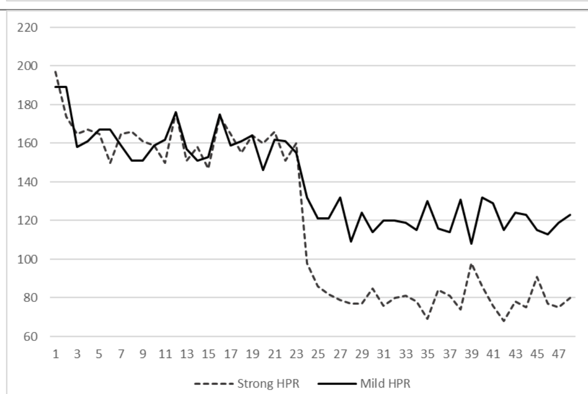
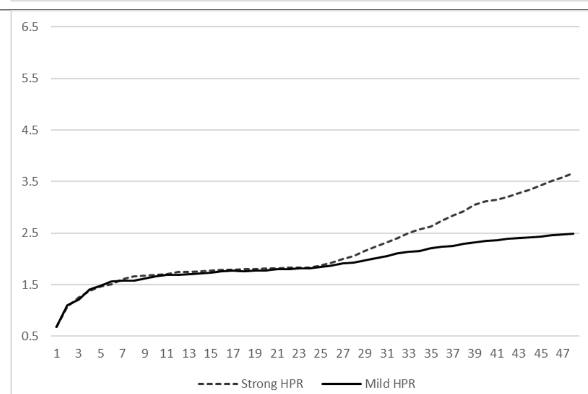
		Brokerage		
		Weak (1 Round)	**Normal (3 Round)**	**Strong (6 Rounds)**
HPR	**No (0.2)**	2.623	1.871	1.864
	Mild (0.4)	8.263	3.429	3.483
	Strong (0.6)	22.294	7.946	7.298

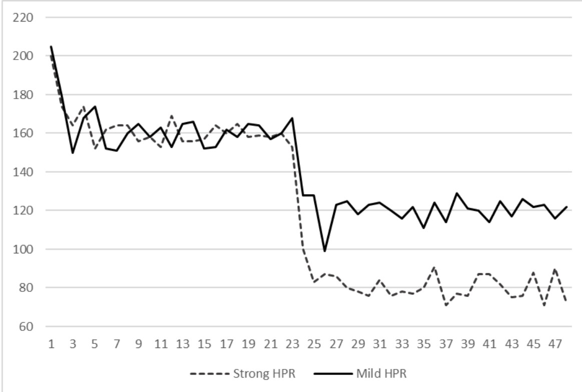
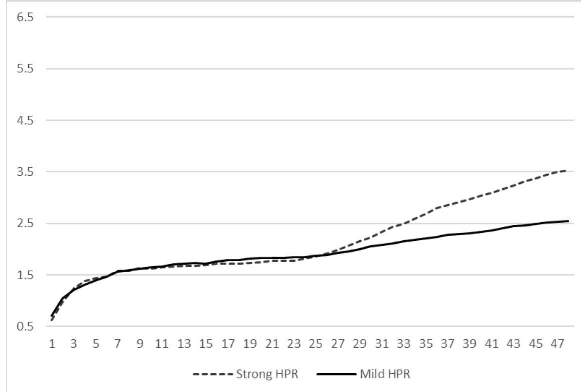
3.4.5 A Dynamic View

The dynamic feature of our ABM allows us to examine the impact of HPR along time. Initially, there is no HPR. The model generates 500 buyers and 500 sellers with a low BS share equal to 0.2. Each following round adds 200 new buyers and 200 new sellers still with this low 0.2 share. A shock of HPR, either mild or strong, occurs in the 24th round, and from then on, the BS share increases to the corresponding level, 0.4 or 0.6. A total of 48 rounds are executed. The whole process is conducted for three levels of brokerage.

The result is show in Table 3.7. Apparently, the HPR shocks decrease the total number of trades and increase the average time-on-market of households. Strong HPRs have much stronger effects on both measurements. As the brokerage gets stronger, the extension of the term of contracts helps ameliorate the negative situations brought by HPRs.

Table 3.7 Simulation Results for Dynamic View

	Total Number of Trades (TNT)	**Average Time-on-market (ATOM)**
Term = 1		
Term = 3		

Term = 6	 220 200 180 160 140 120 100 80 60 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 Strong HPR Mild HPR	 6.5 5.5 4.5 3.5 2.5 1.5 0.5 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 Strong HPR Mild HPR

3.5 DISCUSSION AND CONCLUSION

This paper investigates the behavior of the buy-sellers (BS) in urban housing markets. The role of BS is discussed in a network framework and further studied in an agent-based model (ABM). The existence of BS influences market outcomes significantly. BS provides a new perspective to analyze housing market mechanisms and related policies.

The major contribution of this paper is the development of an ABM for the buy-sellers. The features of BS are incorporated and algorithms to identify trading network patterns are developed. This ABM is useful for empirical applications to regions where resale housing markets are dominant.

In addition, the ABM is employed to analyze the impacts of housing purchase restriction (HPR) policies on market outcomes and the counter-effect from home brokerage. This paper views the HPR as an exogenous shock that changes pure-buyers into buy-sellers and represents the power of brokerage by the term length of contracts between buyers and sellers. Simulation results from different scenarios suggests HPR have negative impacts on housing markets in terms of decreasing total number of trades and increasing the average time-on-market. In contrast, a stronger brokerage would lead to better market performance. From a dynamic view, a HPR shock causes a sudden drop in the total number of trades, and a stronger brokerage would ameliorate the situation to some extent.

APPENDIX A DATA CLEANING AND OUTLIER DETECTION

The first step is to clean missing data and irrelevant housing types. Missing data (coded as missing, unknown or blank) occurs in many variables: listed total price, floor level, building type, year built, elevator, and heating. Irrelevant housing types are also deleted, involving underground housing units, one-story building (called "*Ping Fang*" in Chinese), non-commodity housing (such as affordable housing), non-condominium (such as apartment, detached house, and Chinese courtyard house), non-70-year right of use land (such as 40 or 50 years which infer commercial or mixed land use), and non-flat housing type (such as duplex or multi-story dwelling units). The amount of observations being deleted in each step are shown in Table A1. Data distributions across years before and after cleaning are summarized in Table A2.

Table A1 Summary of Cleaning Variable

Variable	Deleted Value	Deleted Amount	Ratio to Total (386790)
Repeated Records		1452	0.38%
Listed Total Price	Missing	2019	0.52%
Floor Level	Unknown	1389	0.36%
	Underground	3239	0.84%
Building Type	Blank	4104	1.06%
	One-storey (*Ping Fang*)	187	0.05%
Year Built	Unknown	7877	2.04%
Elevator	Unknown	10767	2.78%
Owner-occupied Sector	Not commodity housing	4800	1.24%
Property Type	Not condominium	18422	4.76%
Right of Use Land	Not 70 years	21142	5.47%
Heating	Missing	15021	3.88%
Housing Type	Not flat	3071	0.79%
	Total Deleted (With Repeating)	93490	
	Total Deleted (Without Repeating)	49446	12.78%
	Left Count (Without Repeating)	337344	87.22%

Table A2 Data Summary before and after Cleaning in Stage One

Year	Before		After		Deleted
	Count	Percentage	Count	Percentage	
2012	52901	13.68%	46990	13.93%	-5911
2013	56141	14.51%	48659	14.42%	-7482
2014	45654	11.80%	40554	12.02%	-5100
2015	99583	25.75%	89038	26.39%	-10545
2016	132511	34.26%	112103	33.23%	-20408
Total	**386790**		**337344**		**-49446**

The second step is to detect and clean outliers. Some outliers are obvious typos with unrealistic values—for example it is impossible to sell a housing unit at 10 RMB (about 1.5 US Dollar) per square meter in Beijing from 2012 to 2016. Other outliers require statistical methods to detect. Typically, one may simply apply common outlier detection methods to the whole sample, such as the mean plus or minus three standard deviations and the interquartile range (IQR). However, this would be problematic for the case of this study. The data in this paper involves all districts of Beijing from 2012 to 2016, there exist huge intrinsic variations across locations and years. Hence, it is possible that the transaction prices in a peripheral area are statistically lower than those in the center and are wrongly identified as "outliers" of the whole sample. To avoid this, outliers should be detected in each small group of observations, in this paper, each location-year group. However, some groups have a small sample, which may result in over- or under-identification of outliers using the common approaches (Cousineau and Chartier, 2010). Recent studies suggest using median absolute deviation (MAD)[21] to detect outliers (Leys et al., 2013). This paper applies this method and cleans the outliers in following steps and the data summary is in

[21] MAD has also been used as the measure of the precision of estimators (see example in Honoré & Powell, 1994).

Table A3:

1. Define and delete obvious typo observations. The upper bounds of floor area, listed total price, transacted total price, and transacted unit price are chosen as 500, 5000, 5000, and 500000 respectively; lower bounds as 10, 10, 10, and 1000 respectively. These bounds are conservatively selected and leave most data to the next step.

2. In each year, delete locations with less than 5 observations.

3. Check each observation in its location-year group using the MAD method. Specifically, for each group, calculate the median absolute deviation (MAD) for the kth variable as $MAD_k = b * median(|x_{ik} - median(\mathbf{x}_k)|)$, where $b = 1.4826$, x_{ik} is the kth variable value of the ith observation in the group, and $\mathbf{x}_k$ is the vector of the kth variable value in the group (Leys et al., 2013); delete the ith observation from this groups if for any k, x_{ik} is not in the interval $[\,median(\mathbf{x}_k) - 3 * MAD_k,\ median(\mathbf{x}_k) + 3 * MAD_k]$. The variables are floor area, listed total price, transacted total price, and transacted unit price.

4. Repeat the previous two steps until there is no new deletion.

Table A3 Data Summary for Outlier Detection

Year	**Before**		**Step 1**	**Step 2**	**Step 3**	**After**	
	Count	**Percentage**				**Count**	**Percentage**
2012	46990	13.93%	-514 (together)	-3283	-6676	36823	13.62%
2013	48659	14.42%		-3411	-7111	38031	14.07%
2014	40554	12.02%		-3665	-7333	29463	10.90%
2015	89038	26.39%		-3004	-11761	74200	27.45%
2016	112103	33.23%		-2814	-17499	91756	33.95%
Total	**337344**		**336830**	**-16177**	**-50380**	**270273**	

The third step deletes observations based on empirical and estimation consideration. First, to avoid additional heterogeneity, this paper removes 5725 records of Yanjiaozhen from the cleaned sample. Yanjiaozhen is a town in Hebei Province, bordering Beijing to the west. Its adjacency to Beijing, less restrictive housing policies, and low housing prices attract housing buyers and agencies like Lianjia to make transactions in this alternative market. However, despite of its connectedness to Beijing, Yanjiaozhen has very different market structures and bring in heterogeneity. Second, buildings/locations with only one observation are removed because they make no contribution to within building variations. In the cleaned sample, there are only 6 buildings with only one observation.

Table A4 Data Summary for Third Step

Year	Before		Yanjiaozhen	Buildings with only one observation	After	
	Count	Percentage			Count	Percentage
2012	36823	13.62%	-5725 (together)	-6 (together)	36823	13.92%
2013	38031	14.07%			38031	14.38%
2014	29463	10.90%			29349	11.09%
2015	74200	27.45%			72058	27.24%
2016	91756	33.95%			88281	33.37%
Total	**270273**				**264542**	

The last step is to look at the variable statistics and further clean the variables with suspicious values. There are several variables worth noticing. First, the number of kitchens. There are 1224 observations with the number of kitchen different from one. Most of them are typos. To avoid noises, we delete all 1224 observations. Second, the number of other rooms. There are 433 observations with either more than five bedrooms (53 in total including 47 with six and 6 with seven) or more than two living rooms (320 in total including 313 with three and 7 with seven) or more than three bathrooms (103 in total including 97 with four and 6 with five). These observations are possibly multi-storey units but wrongly tagged as flats. So this paper excludes them from consideration. Last, for further use of the location variables, we have some locations (complexes) with no information of school districts. These locations only involve 348 observations. We exclude these observations from further analyses.

Table A5 Data Summary for Final Round

Year	Before		Number of kitchen unequal to one	Number of rooms abnormal	Corresponding location has no school information	Buildings with only one observation	After	
	Count	Percentage					Count	Percentage
2012	36823	13.92%	-1224 (together)	-433 (together)	-348 (together)	-4 (together)	36571	13.93%
2013	38031	14.38%					37784	14.39%
2014	29349	11.09%					29159	11.11%
2015	72058	27.24%					71573	27.26%
2016	88281	33.37%					87446	33.31%
Total	**264542**						**262533**	

APPENDIX B RELATIONSHIP BETWEEN SUM SQUARED DIFFERENCES AND SUM SQUARED DEVIATIONS FROM THE MEAN

In this appendix, I compare the sum squared differences ($\widehat{SSD}$) and the traditional measurement of deviation, i.e. the sum squared deviations from the mean ($\widehat{SSDM}$).

Scenario One: *N* Observations in One Cluster

First, I consider the situation that N observations in one cluster (or without cluster structure). For observation i, its observed value is denoted as x_i. Thus,

$$\widehat{SSD} = \sum_{(i,j)\in P}(x_i - x_j)^2, \tag{A1}$$

where $P \equiv \{(i,j): i \neq j\}$ denotes all pairs of observations. Using $\bar{x} = \frac{1}{N}\sum_N x_i$, I rewrite (A1) as

$$\widehat{SSD} = \sum_P \left(x_i - \bar{x} + \bar{x} - x_j\right)^2 \tag{A2}$$

$$= \sum_P[(x_i - \bar{x})^2 + (\bar{x} - x_j)^2 + 2(x_i - \bar{x})(\bar{x} - x_j)]$$

$$= \sum_P[(x_i - \bar{x})^2 + (\bar{x} - x_j)^2] + \sum_P[2(x_i - \bar{x})(\bar{x} - x_j)]$$

$$= (N-1)\sum_N(x_i - \bar{x})^2 + 2\sum_P(x_i - \bar{x})(\bar{x} - x_j) \tag{A3}$$

Note that the first term on the RHS is the variance of all N observations. For the second term on the RHS,

$$2\sum_P(x_i - \bar{x})(\bar{x} - x_j) \tag{A4}$$

$$= 2\sum_P(x_i\bar{x} + \bar{x}x_j - \bar{x}^2 - x_i x_j)$$

$$= 2\left[\sum_P\left(x_i\bar{x} + \bar{x}x_j\right) - \sum_P \bar{x}^2 - \sum_P x_i x_j\right]$$

$$= 2[(N-1)\sum_N(x_i\bar{x}) - \frac{N(N-1)}{2}\bar{x}^2 - \sum_P x_i x_j]$$

$$= 2[N(N-1)\bar{x}\frac{1}{N}\sum_N x_i - \frac{N(N-1)}{2}\bar{x}^2 - \sum_P x_i x_j]$$

$$= 2[N(N-1)\bar{x}^2 - \frac{N(N-1)}{2}\bar{x}^2 - \sum_P x_i x_j]$$

$$= 2[\frac{N(N-1)}{2}\bar{x}^2 - \sum_P x_i x_j]$$

$$= 2[\frac{N(N-1)}{2}\frac{1}{N^2}(\sum_N x_i)^2 - \sum_P x_i x_j]$$

$$= 2[\frac{N-1}{2N}(\sum_N x_i^2 + 2\sum_P x_i x_j) - \sum_P x_i x_j]$$

$$= 2[\frac{N-1}{2N}\sum_N x_i^2 + \frac{N-1}{N}\sum_P x_i x_j - \sum_P x_i x_j]$$

$$= 2[\frac{N-1}{2N}\sum_N x_i^2 - \frac{1}{N}\sum_P x_i x_j]$$

$$= 2[\frac{1}{2N}\sum_P \left(x_i^2 + x_j^2\right) - \frac{1}{2N}\sum_P 2x_i x_j]$$

$$= 2[\frac{1}{2N}\sum_P \left(x_i^2 - 2x_i x_j + x_j^2\right)]$$

$$= 2[\frac{1}{2N}\sum_P \left(x_i - x_j\right)^2]$$

$$= \frac{1}{N}\sum_P \left(x_i - x_j\right)^2 \tag{A5}$$

Take (A1) into (A5), then (A4) can be rewritten as

$$2\sum_P (x_i - \bar{x})(\bar{x} - x_j) = \frac{1}{N}\ \widehat{SSD}. \tag{A6}$$

Take (A6) back into (A3)

$$\widehat{SSD} = (N-1)\sum_N (x_i - \bar{x})^2 + \frac{1}{N}\ \widehat{SSD}. \tag{A7}$$

From (A7), I could get

$$\widehat{SSD} = N\sum_N (x_i - \bar{x})^2 = N * \widehat{SSDM}. \tag{A9}$$

Thus, $\widehat{SSD}$ equals to N times $\widehat{SSDM}$ when all observations are in the same cluster.

Scenario Two: N Observations in G Clusters

Now, considering N observations in G clusters, the sum squared differences ($\widehat{SSD}$)

for all within-cluster pairs is

$$\widehat{SSD} = \sum_{g=1}^{G} \sum_{(i,j)\in P_g} (x_{ig} - x_{jg})^2, \tag{A10}$$

where x_{ig} is the ith observation in cluster g; and $P_g \equiv \{(i,j): i, j\ in\ cluster\ g\ and\ i \neq j\}$ is the set of pairs of observations of cluster g. From scenario one, for each cluster g,

$$\sum_{P_g} (x_{ig} - x_{jg})^2 = N_g \sum_{i=1}^{N_g} (x_{ig} - \bar{x}_g)^2, \tag{A11}$$

where N_g is number of observations in cluster g.

Take (A11) into (A10),

$$\widehat{SSD} = \sum_g N_g \sum_i (x_{ig} - \bar{x}_g)^2. \tag{A12}$$

If every cluster has the same size $N_g = N/G$ for all g, then (A12) becomes

$$\widehat{SSD} = \frac{N}{G} \sum_g \sum_i (x_{ig} - \bar{x}_g)^2.$$

APPENDIX C EQUIVALENCE AND DEVIATION CONDITIONS

Equivalence Conditions: The mediant equals to the weighted mediant if any of the following condition holds:

a. (equal-weight condition) all weights are equal: $w_i = w$ for all $i = 1 \dots n$ and $w \in \mathbb{R}_{>0}$;

b. (equal-fraction condition) all fractions are equal: $\frac{a_i}{b_i} = \frac{a}{b}$ for all $i = 1 \dots n$ and $a \in \mathbb{R}_{\geq 0}, b \in \mathbb{R}_{>0}$.

Proof (a): If $w_i = w$ for all $i = 1 \dots n$,

$$m_w = (\sum_i w_i b_i)^{-1} \sum_i w_i a_i = (\sum_i w b_i)^{-1} \sum_i w a_i = (\sum_i b_i)^{-1} \sum_i a_i = m$$

Q.E.D.

Proof (b): If $\frac{a_i}{b_i} = \frac{a}{b}$ for all $i = 1 \dots n$, then $a_i = \frac{a}{b} b_i$,

$$m_w = (\sum_i w_i b_i)^{-1} \sum_i w_i a_i = (\sum_i w_i b_i)^{-1} \sum_i w_i \frac{a}{b} b_i = \frac{a}{b} (\sum_i w_i b_i)^{-1} \sum_i w_i b_i = \frac{a}{b}$$

$$m = (\sum_i b_i)^{-1} \sum_i a_i = (\sum_i b_i)^{-1} \sum_i \frac{a}{b} b_i = \frac{a}{b} (\sum_i b_i)^{-1} \sum_i b_i = \frac{a}{b}$$

Q.E.D.

Deviation Condition: If a relatively larger fraction $\frac{a_n}{b_n}$ is associated with a relatively larger weight w_n, then $m_w > m$. In another word, if the covariance between $\frac{a_n}{b_n}$ and w_n is positive, then $m_w > m$; otherwise, $m_w < m$. I will show a simple case when $n = 2$, and a simulation with $n > 2$.

For $n = 2$, suppose $a_1 = 3$, $a_2 = 2$, $b_1 = 1$, $b_2 = 4$, then $\frac{a_1}{b_1} = \frac{3}{1}$, $\frac{a_2}{b_2} = \frac{2}{4}$, and $\frac{a_1}{b_1} > \frac{a_2}{b_2}$. The mediant $m = \frac{a_1 + a_2}{b_1 + b_2}$ equals to $\frac{3+2}{1+4} = \frac{5}{5}$, which is between the two fraction values. In a vector graph, viewing fractions as slopes, the mediant m can be viewed as the slope of the sum vector $\vec{m}$ of $\overrightarrow{v_1} = (a_1, b_1)$ and $\overrightarrow{v_2} = (a_2, b_2)$, and by parallelogram law, $\frac{a_2}{b_2} < m < \frac{a_1}{b_1}$.

Now consider two weighting scenarios: $w' = (w_1', w_2') = (2, 1)$ and $w'' = (w_1'', w_2'') = (1, 2)$. The first scenario assigns larger weight to the larger fraction, while the second does the opposite. As a result, the weighted mediants are $m_w' = \frac{2a_1+a_2}{2b_1+b_2} = \frac{8}{6}$ and $m_w'' = \frac{a_1+2a_2}{b_1+2b_2} = \frac{7}{9}$. Hence,

$$m_w'' < m < m_w'.$$

This case provides much of the intuition for the more general case. It concludes that if a larger fraction is weighted with larger weights, the weighted mediant is larger than the unweighted mediant, and vice versa. Figure C1 provides good illustration for this point as well. If the steeper vector $\overrightarrow{v_1}$ (larger fraction) receives larger weights than the flatter one $\overrightarrow{v_2}$ (smaller fraction) as in scenario one, the mediant vector $\vec{m}$ will be "lifted up" and become m_w'; otherwise (as in scenario two), the mediant vector $\vec{m}$ will be "dragged down" and become m_w''.

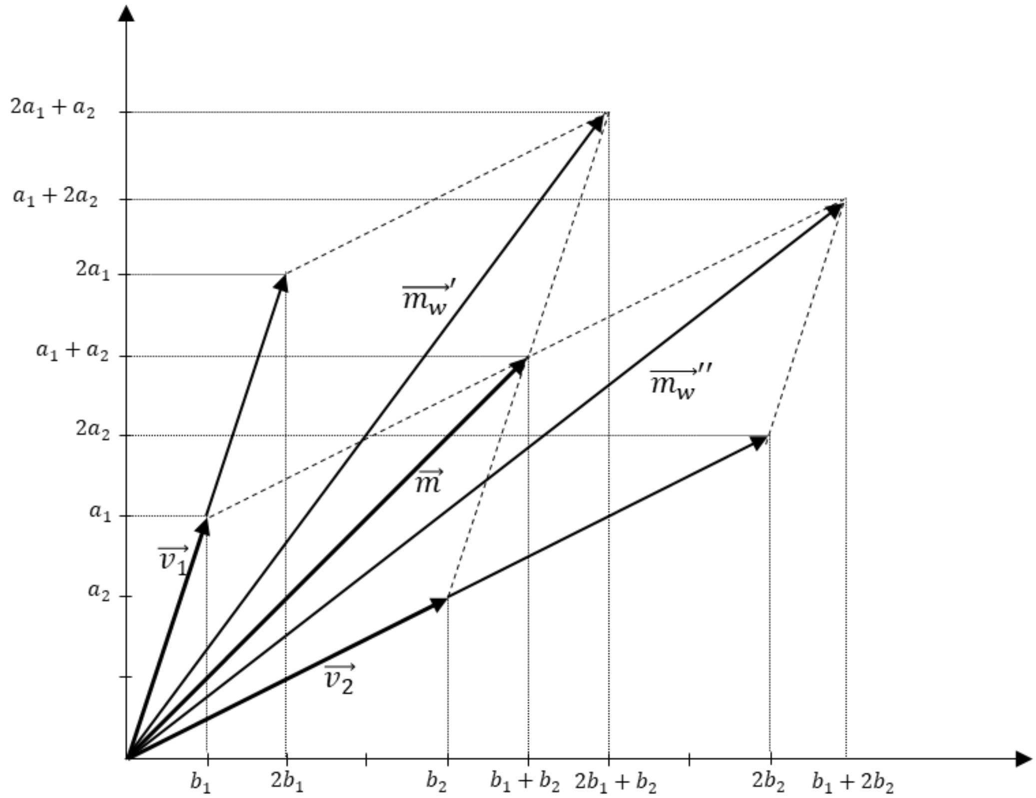
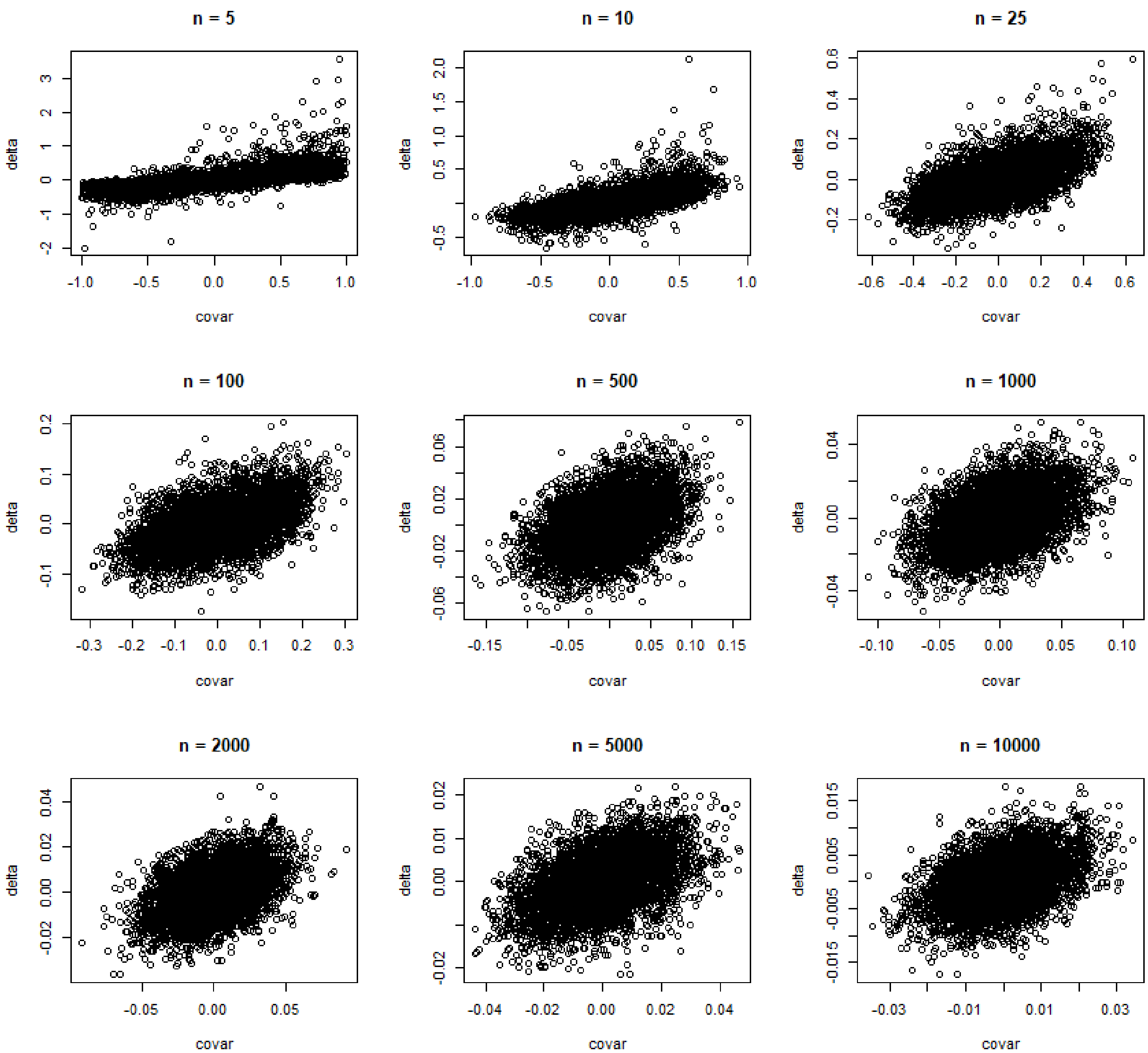


Figure C1 Mediant and Weighted Mediant

For $n > 2$, I hypothesize that if larger weights are assigned to larger fractions (i.e. positive correlation between weights and fractions), the change from mediant to weighted median $m_w - m$ is larger. A simulation is conducted to examine this hypothesis. In each round, sequences $\{a_n\}$, $\{b_n\}$, and $\{w_n\}$ are drawn from a uniform distribution over $[1:100]$ with replacement. Correlation between w_n and a_n/b_n is calculated and denoted as $covar$. Then, mediant m and weighted mediant m_w are calculated and their difference is denoted as $delta = m_w - m$. Let $n = (5, 10, 25, 100, 500, 1000, 2000, 5000, 10000)$. For each value of n, repeat 5000 rounds. The results (Figure C2) show a significant positive correlation between the weight-fraction covariance and change from mediant to weighted mediant.

Figure C2 Correlation between Weight-fraction Covariance and Mediant Change



REFERENCES

CHAPTER ONE

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. Papers in Regional Science, 92(1), 3-17.

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, *58*(304), 933-942.

Bao, H. X., & Wan, A. T. (2004). On the use of spline smoothing in estimating hedonic housing price models: empirical evidence using Hong Kong data. *Real estate economics*, *32*(3), 487-507.

Basu, S., & Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics*, 17(1), 61-85.

Borukhov, E., Ginsberg, Y., & Werczberger, E. (1978). Housing prices and housing preferences in Israel. *Urban Studies*, *15*(2), 187-200.

Butler, R. V. (1982). The specification of hedonic indexes for urban housing. *Land Economics*, *58*(1), 96-108.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional science and urban economics*, 22(3), 453-474.

Chen, Y., Clapp, J. M., & Tirtiroglu, D. (2011). Hedonic estimation of housing demand elasticity with a markup over marginal costs. *Journal of Housing Economics*, *20*(4), 233-248.

Cressie, N. (1993). Statistics for spatial data: Wiley series in probability and statistics. Wiley-Interscience New York, 15, 16.

Danton, J., & Himbert, A. (2017). Residential Vertical Rent Gradients in the City. working paper.

Dong, F., Ding, H., & Zhao, A. P. (2009). Commuting Cost and Space Distribution of Housing Price in Areas near Rail Transit: An Empirical Analysis of Residential Market near No. 5 Line Subway in Beijing [J]. *Journal of Beijing Normal University (Social Sciences)*, *4*, 026.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional science and urban economics*, 22(3), 433-452.

Ekblad, S., & Werne, F. (1990). Housing and health in Beijing: Implications of high-rise housing on children and the aged. *J. Soc. & Soc. Welfare*, *17*, 51.

Englund, P., Quigley, J. M., & Redfearn, C. L. (1999). Do Housing Transactions Provide Misleading Evidence about the Course of Housing Values?.

Freeman, A. M. (1981). Hedonic prices, property values and measuring environmental benefits: a survey of the issues. In *Measurement in public choice* (pp. 13-32). Palgrave Macmillan, London.

Frenkel, A. (2007). Spatial distribution of high-rise buildings within urban areas: The case of the Tel-Aviv metropolitan region. *Urban Studies*, *44*(10), 1973-1996.

Fu, Y., David, K. T., & Zhou, N. (2000). Housing choice behavior of urban workers in China's transition to a housing market. *Journal of Urban Economics*, *47*(1), 61-87.

Gao, X., & Asami, Y. (2011). Preferential size of housing in Beijing. *Habitat International*, *35*(2), 206-213.
Gong, J.H. (2009). Analysis of Determining Factors of Beijing Housing Price and Related Policy Proposals [J]. Social Science of Beijing,1, 009.
Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of urban economics*, *5*(4), 471-484.
Jim, C. Y., & Chen, W. Y. (2007). Consumption preferences and environmental externalities: A hedonic analysis of the housing market in Guangzhou. *Geoforum*, *38*(2), 414-431.
Jim, C. Y., & Chen, W. Y. (2009). Value of scenic views: Hedonic assessment of private housing in Hong Kong. *Landscape and urban planning*, *91*(4), 226-234.
Jim, C. Y., & Chen, W. Y. (2010). External effects of neighbourhood parks and landscape elements on high-rise residential value. *Land Use Policy*, *27*(2), 662-670.
Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American statistical association*, *65*(330), 532-548.
Kanemoto, Y., & Nakamura, R. (1986). A new approach to the estimation of structural equations in hedonic models. *Journal of Urban Economics*, *19*(2), 218-233.
Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?. *Journal of Environmental Economics and Management*, *60*(3), 145-160.
Li, S., Yang, J., Qin, P., & Chonabayashi, S. (2016). Wheels of fortune: Subway expansion and property values in Beijing. *Journal of Regional Science*, *56*(5), 792-813.
Liu, C. H., Rosenthal, S. S., & Strange, W. C. (2018). The Vertical City: Rent Gradients, Spatial Structure, and Agglomeration Economies. *Journal of Urban Economics*.
Ma S. X., Li A. (2003). House price and its determinations in Beijing based on hedonic model. China Civil Engineering Journal, 36(9): 59-64.
Mason, C., & Quigley, J. M. (1996). Non-parametric hedonic housing prices. *Housing studies*, *11*(3), 373-385.
Matheron, G. (1963). Principles of geostatistics. Economic geology, 58(8), 1246-1266.
Meese, R. A., & Wallace, N. E. (1997). The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics*, *14*(1-2), 51-73.
Mitrany, M. (2005). High-density neighborhoods: Who enjoys them?.*GeoJournal*, *64*(2), 131-140.
Mok, H. M., Chan, P. P., & Cho, Y. S. (1995). A hedonic price model for private properties in Hong Kong. *The Journal of Real Estate Finance and Economics*, *10*(1), 37-48.
Ong, S. E., & Koh, Y. C. (2000). Time on-market and price trade-offs in high-rise housing sub-markets. *Urban Studies*, *37*(11), 2057-2071
Quigley, J. M. (1995). A simple hybrid model for estimating real estate price indexes. *Journal of Housing Economics*, *4*(1), 1-12.
Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82(1), 34-55.
Rosen, G., & Walks, A. (2013). Rising cities: Condominium development and the private transformation of the metropolis. *Geoforum*, *49*, 160-172.
Sheppard, S. (1999). Hedonic analysis of housing markets. *Handbook of regional and*

urban economics, *3*, 1595-1635.
Shimizu, C., Takatsuji, H., Ono, H., & Nishimura, K. G. (2010). Structural and temporal changes in the housing market and hedonic housing price indices: a case of the previously owned condominium market in the Tokyo metropolitan area. *International Journal of Housing Markets and Analysis*, *3*(4), 351-368.
Sing, T. F. (2001). Dynamics of the condominium market in Singapore. *International Real Estate Review*, *4*(1), 135-158.
Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, *13*(1), 1-44.
Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for?. *Journal of Human resources*, 50(2), 301-316.
Sun, H., Tu, Y., & Yu, S. M. (2005). A spatio-temporal autoregressive model for multi-unit residential market analysis. The Journal of Real Estate Finance and Economics, 31(2), 155-187.
Turkington, R., Van Kempen, R. & Wassenberg, F. (2004). High-rise housing in Europe: current trends and future prospects. Delft University Press.
Tu, Y. (1997). The local housing sub-market structure and its properties. *Urban Studies*, *34*(2), 337-353.
Wang, Y. P. (2001). Urban housing reform and finance in China: a case study of Beijing. *Urban Affairs Review*, *36*(5), 620-645.
Wong, K. C., So, A. T., & Hung, Y. C. (2002). Neural network vs. hedonic price model: appraisal of high-density condominiums. *Real Estate Valuation Theory*, 181-198.
Wong, S. K., Chau, K. W., Yau, Y., & Cheung, A. K. C. (2011). Property price gradients: the vertical dimension. *Journal of Housing and the Built Environment*, *26*(1), 33-45.
Wooldridge, J. M. (2010). Econometric *analysis* of cross section and panel data. MIT press.
Xu, T. (2008). Heterogeneity in housing attribute prices: a study of the interaction behaviour between property specifics, location coordinates and buyers' characteristics. *International Journal of Housing Markets and Analysis*, *1*(2), 166-181.
Yu, S. M., Han, S. S., & Chai, C. H. (2007). Modeling the value of view in high-rise apartments: a 3D GIS approach. *Environment and Planning B: Planning and Design*, *34*(1), 139-153.
Zheng, S., Hu, W., & Wang, R. (2016). How much is a good school worth in Beijing? Identifying price premium with paired resale and rental data. *The Journal of Real Estate Finance and Economics*, *53*(2), 184-199.

CHAPTER TWO

Ahrens, H., & Pincus, R. (1981). On two measures of unbalancedness in a one-way model and their relation to efficiency. *Biometrical Journal*, *23*(3), 227-235.
Aquaro, M., & Čížek, P. (2013). One-step robust estimation of fixed-effects panel data models. *Computational Statistics & Data Analysis*, 57(1), 536-548.
Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *The Quarterly journal of economics*, *119*(1), 249-275.
Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for

inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414-427.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. *Cambridge university press*.

Carter, A. V., Schnepel, K. T., & Steigerwald, D. G. (2017). Asymptotic Behavior of at-Test Robust to Cluster Heterogeneity. *Review of Economics and Statistics*, 99(4), 698-709.

Cressie, N. (1993). Statistics for spatial data: *Wiley series in probability and statistics. Wiley-Interscience* New York, 15, 16.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. OUP Catalogue.

Desai, M., & Begg, M. D. (2008). A comparison of regression approaches for analyzing clustered data. *American journal of public health*, 98(8), 1425-1429.

Honoré, B. E. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica: *journal of the Econometric Society*, 533-565.

Honoré, B. E., & Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64(1-2), 241-278.

Honoré, B. E., & Powell, J. L. (1997). Pairwise Difference Estimators for Nonlinear Models. University of California, Berkeley.

Kezdi, G. (2003). Robust Standard Error Estimation in Fixed-Effects Panel Models. Available at SSRN: https://ssrn.com/abstract=596988 or http://dx.doi.org/10.2139/ssrn.596988

MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, *32*(2), 233-254.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8), 1246-1266.

Stromberg, A. J., Hössjer, O., & Hawkins, D. M. (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association*, *95*(451), 853-864.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for?. *Journal of Human resources*, 50(2), 301-316.

Von Neumann, J., Kent, R. H., Bellinson, H. R., & Hart, B. T. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, 12(2), 153-162.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*(2), 133-138.

Wooldridge, J. M. (2010). Econometric *analysis* of cross section and panel data. MIT press.

CHAPTER THREE

Alonso, W. (1960). A theory of the urban land market. *Papers in Regional Science*, *6*(1), 149-157.

Anenberg, E., & Bayer, P. (2013). *Endogenous sources of volatility in housing markets: The joint buyer-seller problem*(No. w18980). National Bureau of Economic

Research.

Brasington, D. M. (2014). Housing choice, residential mobility, and hedonic approaches. In *Handbook of Regional Science*(pp. 147-165). Springer Berlin Heidelberg.

Du, Z., & Zhang, L. (2015). Home-purchase restriction, property tax and housing price in China: A counterfactual analysis. *Journal of Econometrics*, *188*(2), 558-568.

Filatova, T. (2015). Empirical agent-based land market: Integrating adaptive economic behavior in urban land-use models. *Computers, environment and urban systems*, *54*, 397-413.

Fleiner, T., Jagadeesan, R., Jankó, Z., & Teytelboym, A. (2017). *Trading networks with frictions*. Oxford University Working Paper.

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, *41*(4), 661-689.

Moen, E., Nenov, P., & Sniekers, F. (2015). Buying first or selling first in housing markets.

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, *32*(6), 454-463.

Rosen, K. T., & Smith, L. B. (1986). The resale housing market. *Real Estate Economics*, *14*(4), 510-524.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, *1*(2), 143-186.

Stein, J. C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *The Quarterly Journal of Economics*, *110*(2), 379-406.

Sun, W., Zheng, S., Geltner, D. M., & Wang, R. (2017). The housing market effects of local home purchase restrictions: evidence from Beijing. *The Journal of Real Estate Finance and Economics*, *55*(3), 288-312.

Winstanley, A., Thorns, D. C., & Perkins, H. C. (2002). Moving house, creating home: Exploring residential mobility. *Housing studies*, *17*(6), 813-832.