From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2023

ESSAYS ON MATCHING MARKETS WITH INCOMPLETE INFORMATION

Ferdowsian, Andrew Behrad Araghi

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp018c97kt679

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON MATCHING MARKETS WITH INCOMPLETE INFORMATION

Andrew Ferdowsian

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY
THE DEPARTMENT OF
ECONOMICS
Adviser: Leeat Yariv

May 2023

© Copyright by Andrew Ferdowsian, 2023.

All rights reserved.

Abstract

This dissertation investigates matching markets with incomplete information.

In chapter 1, I introduce a framework for studying transient matching in decentralized markets where workers learn about their preferences through their experiences. Limits on the number of available positions force workers to compete over matches. Each capacity-constrained firm employs workers whose match value exceeds a threshold. Since employment offers both payoff and information benefits, workers effectively face a multi-armed bandit problem. To them each firm acts as a bandit where the probability of "success" at the firm is driven by market competition. In such markets, aggregate demand for firms satisfies the gross substitutes condition which ensures equilibrium existence. The resulting search patterns match a variety of stylized facts from labor market data. High-quality workers search less and tenure increases with age. In general, equilibria are inefficient because competition depresses the level of search. Natural interventions designed to improve efficiency are effective in uncongested markets, but can fail when congestion is several. From a market design perspective, the utilization of headhunters has differential effects depending on workers' quality, conclusively improving both outcomes for low-quality workers and overall efficiency. Reducing congestion through unemployment benefits, can depress search and may ultimately reduce match efficiency.

Chapter 2, joint with Kwok Hao Lee and Luther Yap, considers the optimal allocation of public housing from the perspective of dynamic mechanism design. Novel in our setting: the designer can choose what types of apartments to supply, in response to earlier demand realizations. However, since the designer does not know which apartments are desired by households, she must trade-off learning about preferences with matching current demand. This control radically changes the optimal mechanism. In this mechanism, under-demanded housing is overproduced relative to a benchmark in which apartments are built in proportion to demand. Competition between housing applicants improves average match quality

when the government controls supply. Competition can be artificially generated by batching applications, thus increasing market thickness. If the planner values match quality highly, the optimal mechanism features batching.

Chapter 3, also joint with Lee and Yap, considers the design of a large-scale public housing program where consumers face dynamic tradeoffs over apartments rationed via lotteries and prices. We show that changing rules complements increasing supply. By waiting for "better" developments arriving tomorrow, households forgo mediocre developments available today, raising vacancies. We formulate a dynamic choice model over housing lotteries and estimate it on data from Singapore. Under the existing mechanism, we find that increasing supply fails to lower wait times. However, when a strategyproof mechanism is implemented, vacancies and wait times fall, but prices on the secondary market rise. Under this new mechanism, building more apartments lowers wait times and reduces the upward pricing pressure on the secondary market.

Acknowledgements

To start, I am incredibly grateful to Leeat Yariv for her unwavering guidance and support. I have benefited immensely from her dedication to her students. Without her, I doubt I would have made it through the program and job market season. I intend to use her as a model in my future interactions with students.

I also want to acknowledge the assistance I received from Alessandro Lizzeri and Can Urgun, both of whom went above and beyond to encourage me to think more deeply about the implications of my work.

Many other Princeton faculty members, both within the Dietrich economic theory center and beyond, have taken the time to dispense their wisdom. I thank Sylvain Chassang, Faruk Gul, Adam Kapor, Sofia Moroni, Xiaosheng Mu, Pietro Ortoleva, Wolfgang Pesendorfer, and Richard Rogerson for the feedback they have provided during workshops and their help through meetings.

Importantly, I would like to express my sincere gratitude to Laura Hedden for being an indispensable resource for us as graduate students. Her work is truly impressive, yet often goes unnoticed. Throughout my time here, I have never encountered any administrative issues, thanks to her unwavering efforts. Additionally, she always promptly answered any questions I had.

Of course, this section would not be complete without mentioning my present and future co-authors, all of whom have been a delight to work with. Kwok Hao Lee and Luther Yap have been fantastic collaborators, and friends, who have contributed greatly to this dissertation. I look forward to future work with Yiying Tan in exploring patterns uncovered during research on this dissertation, alongside work with many others. I would also like to thank Don Noh, for joint work on educational loans.

I highlight several friends who have gone out of their way to support me, with mock presentations and proofreading drafts, despite having no obligation to do so. These include:

Dorisz Albrecht, Robert Bell, Jack Deschenes, Lukas Mann, Dan McGee, Kirill Rudov, Joseph Ruggiero, Christine Song, Alan Wei, Tim Wang, and Dan Morrison.

This dissertation has benefited from discussions with many individuals. I would like to thank: Nick Arnosti, Jacob Dorn, Francesco Fabbri, Stephanie Hao, Navin Kartik, Benny Kleinman, Tyler Maxey, Paola Moscariello, Gaetan Nandong, Chris Neilson, Franz Ostrizek, Shumiao Ouyang, Parth Parihar, Sebastian Roelsgaard, Evgenii Safonov, Jeffrey Sun, Christiane Szerman, Rui Tang, Joao Thereze, Martin Vaeth, Darwin Yang, Mu Zhang, Henry Zhao, Qiyuan Zheng, and Minghao Zou among many others. I am also grateful for financial support from the Department of Economics at Princeton University and the Dietrich Economic Theory Center.

Presentation participants at the following institutions have contributed immensely to this work: Princeton, the International Stony Brook Conference on Game Theory, Stony Brook itself, the UChicago WiET conference, Simon Fraser University, Monash University, the University of Exeter, the University of Calgary, and the Eastern Economics Association.

Looking back to my undergraduate studies at Boston College, I would like to express my gratitude to C-K Cheung and William Keane for reaching out to me and convincing me to enroll in the mathematics curriculum. John Baldwin gave me my first taste of research during a summer fellowship, and Maksym Fedorchuk further fueled my interest. Without any of the individuals mentioned above, I would likely be graduating from law school now, on an entirely different path. Attending Erez Yoeli's and Moshe Hoffman's course on social behavior and game theory showed me how game theory could be used to address real-world questions.

I am grateful to my friends and fellow board game players who helped me stay sane throughout this process. I have enjoyed time spent with Anshu Chen, Harrison Ho, Jonah Kaye, Yotam Katznelson, Max Li, Jeff Fournier, Frankie Lam, Thomas MacAuley, Myron Pow, Dee Ruttenberg, Ken Saxton, Alex Tseng, Rick Wytmar, and others.

Undoubtedly, there may be people I have missed, and I apologize in advance for anyone who deserved a mention, but was mistakenly neglected, except for omitting you, Sol Chen, that was intentional :).

Lastly, I would like to thank my family. My parents for pushing me to learn, my brother for being a constant source of joy (and irritation), and my partner Maggie for supporting me through everything. I look forward to our future together!

For mom, dad, Ryan, and Maggie.

Contents

Abstract . . . 3

Acknowledgements . . . 5

List of Tables . . . 14

List of Figures . . . 16

1 Learning through Transient Matching in Congested Markets 19

1.1 Introduction . . . 19

1.1.1 Overview . . . 19

1.1.2 Related Literature . . . 24

1.2 The Model . . . 26

1.3 Incomplete Information Generates Transience . . . 30

1.4 Equilibrium Characterization . . . 33

1.4.1 Hiring Thresholds . . . 34

1.4.2 Equilibrium Worker Strategies . . . 36

1.4.3 Alignment Implies Uniqueness . . . 39

1.4.4 Gross Substitutes in Equilibrium . . . 42

1.4.5 Efficiency and Long-Lived Behavior . . . 46

1.4.6 Tenure . . . 53

1.5 Policy Interventions . . . 54

1.5.1 Headhunters . . . 54

1.5.2 Unemployment Benefits in Congested Markets . . . 58

1.6 Transferable Utility . . . 59
1.6.1 Competitive Equilibrium . . . 59
1.6.2 Non-Existence of Competitive Equilibrium . . . 61
1.6.3 Resumes . . . 63
1.7 Discussion . . . 64

Appendices **66**
1.A Proofs . . . 66
1.B Extensions Satisfy Gross Substitutes . . . 81
1.B.1 Gradual Learning . . . 81
1.B.2 Learning Through Interviews . . . 83
1.B.3 Heterogeneous Discounting . . . 83

2 Build To Order: Endogenous Supply in Centralized Mechanisms **85**
2.1 Introduction . . . 85
2.2 Policy Background . . . 93
2.2.1 The Build To Order Scheme . . . 93
2.2.2 Responsive Apartment Supply . . . 95
2.3 Model . . . 97
2.3.1 Model Discussion . . . 101
2.4 Results . . . 103
2.4.1 Perfect Information Benchmark . . . 103
2.4.2 Implementing the First-Best . . . 104
2.4.3 Exogenous Supply Benchmark . . . 105
2.4.4 When the First-Best Cannot be Implemented . . . 108
2.5 Competition and Market Thickness . . . 116
2.5.1 Persistence . . . 116

2.5.2 Oversubscription . . . 118
2.5.3 Batching . . . 123
2.6 Descriptive Analysis . . . 127
2.6.1 Data . . . 127
2.6.2 Lagged Effects of Supply and Demand . . . 128
2.6.3 Estimating the Upper Bound on Sincere Applications . . . 129
2.7 Discussion . . . 132

Appendices **134**

2.A Appendix—Proofs . . . 134
2.B Queue Hopping . . . 145
2.C Suboptimality of Markovian Mechanisms . . . 149
2.D Labelling . . . 152
2.E M Apartment types . . . 153

3 The Dynamic Allocation of Public Housing: Policy and Spillovers **154**

3.1 Introduction . . . 154
3.2 Recent Literature . . . 158
3.3 Policy Background . . . 161
3.3.1 Housing stock in Singapore . . . 161
3.3.2 The Housing and Development Board and its objective . . . 161
3.3.3 The Build-to-Order Scheme . . . 163
3.3.4 Information available to household at application . . . 165
3.4 Data . . . 166
3.5 Descriptive Analysis . . . 168
3.5.1 Apartments in BTO developments are homogeneous . . . 168
3.5.2 Construction costs shift BTO prices . . . 169

3.5.3 Access to (good) schools and public transit . . . 170
3.5.4 Households are forward-looking . . . 171
3.6 Theoretical Example . . . 172
3.6.1 Setup . . . 173
3.6.2 BTO Equilibrium . . . 173
3.6.3 Under the current rules, building more "B" is ineffective . . . 174
3.6.4 Once we shut down intertemporal risk, adding supply can be effective 175
3.7 Model . . . 176
3.7.1 Developments offered at BTO . . . 177
3.7.2 Model timing . . . 177
3.7.3 Supply in the secondary market . . . 179
3.7.4 Decision problem for eligible BTO applicants . . . 180
3.7.5 Solution concept . . . 183
3.7.6 Model discussion . . . 185
3.8 Estimation . . . 187
3.8.1 Secondary market supply . . . 188
3.8.2 Demand . . . 190
3.9 Parameter Estimates . . . 196
3.9.1 Parameter Estimates — Supply . . . 196
3.9.2 Parameter Estimates — Demand . . . 197
3.9.3 Model fit . . . 199
3.10 Counterfactuals . . . 199
3.10.1 Alternative policies . . . 199
3.10.2 Outcome comparisons . . . 202
3.11 Conclusion . . . 206

Appendices **208**

3.A Additional descriptive statistics . . . 208
3.A.1 Additional correlation plots . . . 208
3.A.2 BTO apartments are oversubscribed . . . 209
3.A.3 Implicit subsidies differ by apartment size and neighborhood . . . 209
3.A.4 Hedonic regressions . . . 210
3.B Under new rules, do BTO applicants pay more for aftermarket apartments? . 213
3.B.1 Our proposed rules could lower average aftermarket prices . . . 213
3.B.2 Our proposed rules could raise average aftermarket prices . . . 215
3.C Motivating our specification of household preferences . . . 216
3.D Estimation algorithm for demand . . . 217
3.E Construction of Differentiation IVs . . . 218
3.F "Burn-in" procedure for initial conditions . . . 218
3.G Counterfactuals: Computational details . . . 220
3.G.1 Rule changes: Computation . . . 220
3.G.2 Building more: Computation . . . 221
3.H Channels driving our results . . . 223
3.H.1 Under our current mechanism, could expanding supply ever "work?" 223
3.I Tables and Figures . . . 224

Bibliography **241**

List of Tables

3.I.1 Summary statistics by development in the BTO mechanism between January 2012 and May 2015. . . . 225

3.I.2 Summary statistics of secondary market transactions by (neighborhood, apartment type, BTO cycle) from August 2010 to February 2021. . . . 225

3.I.3 Government apartments in Singapore by number of rooms and approximate size. . . . 225

3.I.4 Table of results for the secondary market regressions. . . . 226

3.I.5 Table of results from estimating Equation (3.12). . . . 226

3.I.6 Table of willingness to pay (WTP) for several amenities, in equivalent discounts on an apartment in the BTO mechanism. . . . 227

3.I.7 Table of targeted and untargeted moments in the GMM estimation of young household demand. . . . 227

3.I.8 Summary of counterfactual policies. . . . 227

3.I.9 Table of key outcome comparisons between the actual mechanism and three alternate policies in the sample period of January 2012 and May 2015. . . . 228

3.I.10 Application rates (selected) for the May 2015 BTO cycle. . . . 228

3.I.11 Regression results for hedonic (1) and boundary discontinuity (2) regressions for amenity access. . . . 229

3.I.12 (Selected) clearinghouse application rates for the May 2018 Sale of Balance Flats cycle. . . . 230

3.I.13 Household income ceilings over time in nominal Singapore dollars per year. 230

3.I.14 Household income ceilings by apartment size in nominal Singapore dollars per year, as of September 2019. 230
3.I.15 Maximal household explicit subsidies by apartment size and town type over time. 231
3.I.16 Explicit subsidies by monthly income, apartment size and town type over time. 232
3.I.17 Table of first stage regressions of BTO price on the standardized price of concreting sand. 233

List of Figures

1.1 The x-axis represents δ, the probability a worker does not retire. The blue line shows the proportion of uninformed workers that apply to f_r when $\theta_w^{\varnothing} = 0$. The yellow line shows the proportion of uninformed workers that apply to f_r when $\theta_w^{\varnothing} = 1$. 59

2.2.1 Percentage of apartments, per half year by type, built in non-mature neighborhoods. 96

2.2.2 Percentage of apartments, per half year by type, built in mature neighborhoods. 96

2.3.1 Market Timing 98

2.4.1 Finite-state automaton depicting the first-best mechanism, μ_{fb}. Circles indicate states, while arrows depict transitions given an incoming household of type θ. 103

2.4.2 Finite state automaton depicting the pooling mechanism. 110

2.4.3 Finite state automaton depicting the two-state mechanism, μ_q. Arrows denote transition probabilities. 112

2.5.1 The probability that a type-A apartment is built under the optimal mechanism with over subscription. The x-axis depicts the number of households in queue A minus the number in queue B. The y-axis is the probability that the government builds a type-A apartment in the corresponding state. States $(0, N)$ and $(N, 0)$ are omitted, the corresponding y-values are 0 and 1 respectively. 122

2.5.2 Lower bound for γ under endogenous supply and exogenous supply. The x-axis displays the number of excess households (in the original model $x = 1$), while the y-axis displays the lower bound on γ. Endogenous Supply Threshold refers to the lower bound on γ under which the first-best can be implemented. Fixed Supply Threshold refers to the lower bound on γ under which individuals are willing to apply sincerely while no apartments are vacant. 123

2.6.1 4-room apartments are represented in the top figure, while 5-room apartments are represented in the bottom figure. Solid lines indicate the difference between previous oversubscription and current demand, $dd_t - os_{t-1}$. Dashed lines indicate the supply of apartments. 131

3.I.1 Binned scatter plot of BTO prices against prices of concreting sand, residualized on controls including distances to downtown and to the nearest metro stop. 231

3.I.2 Binned scatter plots of oversubscription rates by development against the distance of said development to various amenities. 234

3.I.3 Means and interquartile ranges of percentage vacancies per school by neighborhood by year. 235

3.I.4 Binned scatter plots of oversubscription rates in a BTO cycle against the mean distance to the nearest metro stop of all developments offered in that cycle. 236
3.I.5 Correlation plot of selected variables against application rates (# applications / # quantity). 237
3.I.6 Implicit discount, by apartment type, of the BTO sticker price relative to the five-year-ahead resale price of similar apartments in the same neighborhood. 238
3.I.7 BTO application rates by distance to SAP elementary school. 238
3.I.8 Information available to BTO applicants. 239
3.I.9 Prices and distance to metro stop seem evenly distributed across both sides of the 2-kilometer school priority boundary. 240

Chapter 1

Learning through Transient Matching in Congested Markets

1.1 Introduction

1.1.1 Overview

In many markets, participants learn about potential matches via pairing. In labor markets, employees often learn about a company's culture on the job; in residency markets, aspiring doctors learn about their preferred specialties through apprenticeships; and in marriage markets, individuals learn about prospective spouses through dating. Such markets are often congested: Firms hire a limited number of employees; hospitals have federal funding limits capping the number of residents they can hire; and many relationships are monogamous. Congestion limits the ability of agents to learn: a worker who is not hired cannot learn her match value with a firm. How do agents strategically search in congested matching markets? When agents learn through matching, who ultimately is matched with whom? How do common interventions—hiring intermediaries or increasing unemployment benefits—change the quality of matches?

This chapter develops a novel model of learning through matching when there is a limited number of positions available. I extend techniques from the operations literature on multi-armed bandits to determine workers' equilibrium search patterns, when the rewards from search are endogenous. I characterize the set of equilibria in congested markets with transient matching. In general, equilibria are inefficient due to competition depressing the level of search. In line with empirical work on tenure, in equilibrium, workers with higher unanticipated match values, as well as older workers, search less (Gorry, 2016). I consider the impact of two common policy interventions: introducing informed intermediaries—headhunters—and increasing unemployment benefits. Both interventions unambiguously improve welfare when the market is not congested. I show that revealing information about a firm's match values through an intermediary still improves *total* equilibrium welfare, though benefits are unequally distributed among workers. Increasing unemployment benefits intensifies competition—which can be detrimental to welfare—when there is a commonly known top firm. In contrast, when markets are uncongested, unemployment benefits always improve equilibrium welfare.

In the model, a continuum of workers repeatedly search for jobs at a finite number of firms. Workers differ in their observable characteristics, but do not know their fit at a given firm until they are hired by that firm.[1] Each period, every worker applies to a single firm. Firms interview their set of applicants and hire the workers with the best fit subject to capacity constraints. Hired workers learn about the quality of their match, while rejected workers only learn of their rejections.[2] Workers and firms split the surplus from matching. At the end of the period, workers retire with a fixed probability, exiting the market. At the beginning of the next period, a proportional mass of workers is born.

[1]For instance, a young graduate of computer science would know her grades and the school she attended, but might not know that a position at Google would feature her best fit.

[2]Results in the empirical literature on search motivate the modelling choice that workers only learn upon being matched (Menzio et al., 2016). In the appendix, I show that the qualitative results are similar if workers learn upon applying.

A novel technical contribution of the chapter is showing how workers evaluate the set of firms as a multi-armed bandit problem. I show that firms act as *endogenous* bandits. In the standard multi-armed bandit problem, a single decision-maker explores a fixed set of bandits. An important difference in this setting is that multiple workers simultaneously compete over a limited number of positions at firms. The reward from a given firm depends on the probability of hire at that firm. However, the probability of hire is endogenously determined by workers' strategies. *Hiring thresholds* suffice as a manner of describing competitive forces and the ability to learn in equilibrium. Despite the fact that workers simultaneously learn and optimize, thresholds fill a role similar to that of prices in competitive equilibrium. Allowing thresholds to adjust is enough to guarantee an equilibrium exists. While prices and thresholds fill a similar role, the two are not interchangeable. In particular, a worker's choice to apply to a firm does not correspond to demand for that firm in the event the worker is not qualified enough.

Firms act as endogenous bandits whose rewards are determined in equilibrium. Utilizing the multi-armed bandit characterization, I show that aggregate demand over firms satisfies the gross substitutes condition of Kelso and Crawford (1982). This condition enables the development of a threshold adjustment process, wherein thresholds converge to a fixed point equilibrium in which firms' hiring thresholds are consistent with workers' behavior. To the best of my knowledge, my model is the first to endogenously determine the rewards of experimentation through competition in a market setting.[3] Additionally, the techniques easily extend to other markets. I show in the appendix that gradual learning, heterogeneous discounting, and flexible firm capacities can all be incorporated into the model, without affecting qualitative results, as none cause the gross substitutes condition to be violated.

The model can be fruitfully applied to data on workers' tenure that has been presented in labor markets. Two facts have emerged from studies of employment. Workers' transition

[3]For an example with experimentation in a setting with publicly shared information, see Bergemann and Välimäki (1996).

rates between jobs decrease with age (Menzio et al., 2016), and higher quality workers are more likely to be satisfied with any given match (Network, 2017) compared to lower quality workers. I show that both facts are a natural consequence of transient matching with incomplete preference information. Older workers have had more chances to investigate firms, and so are likely to have found a satisfactory match. Then, an older worker's incentive to test out other firms is lower than that of a young worker with far more unexplored options to choose from. As for the second fact, I show that if two workers share a prior regarding their fits at firms, the worker with higher realized fit will search less. Her matches provide more surplus than the other worker's matches do. If she is not perfectly informed, she may believe her outside options are equivalent to those of the other worker. Then, she wishes to continue searching only if the other worker also wishes to continue searching. As such, she spends weakly less time searching relative to the other worker, and in turn has higher average tenure.

I investigate the impact of commonly suggested policy interventions targeting markets with learning. One such policy is unemployment benefits, as in principle increasing these benefits decreases the cost of search, which allows workers to be more selective. However, when workers have the option of returning to a firm that previously hired them, unemployment benefits also reduce the cost of taking a risk by applying to a competitive firm. When a firm is highly ranked by all workers, increasing unemployment benefits tempers workers' incentives to investigate other firms. In equilibrium, this can lead to underutilization of these firms and decreased surplus.

Next, I characterize the impact of revealing information about a firm's match values. For example, a company such as Google could hold an open house or hire headhunters, informing all workers of their fit at Google. The revelation has two direct effects. Workers with low match values at that firm avoid the firm, and workers with high match values at that firm can target the firm. The workers with lower firm-specific match values strictly benefit from the

revelation, while the effect is ambiguous for workers with higher firm-specific match values. As those who are poor fits avoid the firm, they intensify the level of competition at other firms. When the market is sufficiently congested, this channel harms workers who would have been good fits at the original firm.

Last, my model provides insight regarding the influential empirical literature on search (Chade et al., 2017). Models in this literature typically assume firms can add or subtract positions at no cost: if a worker could productively match with a firm, that firm can always add an extra job.[4] In the long run, markets can be expected to adjust, so this is a reasonable assumption. However, on the shorter timescale at which workers make strategic decisions, firms may be unable to freely add and remove positions. In many scenarios, such as union jobs or during a recession, firms cannot easily react to increased market demand. In markets with transient matchings, like the described examples, my model shows that calibrated search models which omit the effect of competition can yield upwards biased estimates of worker match quality. To see why, note that the incentive to search is decreasing in the quality of the current match and increasing in the expected value of the outside option. When outside options become less competitive, workers must value their current positions more greatly in order to not switch. Thus, when firms can freely add positions, workers must value their current positions more highly in order to not transfer. Understanding the quality of matches present in equilibrium is critical when evaluating the efficacy of potential counterfactual policy changes. As such, systematically biased estimates of match quality can generate incorrect conclusions about the impact of new policies.

The model focuses on markets with non-transferable utility. Despite this, the model applies more broadly to markets where wages are determined independently of match quality. When wages are fixed across realized match values, my qualitative results hold. Many labor markets feature fixed wages. Hall and Krueger (2012) find that less than 35% of job-seekers

[4]For examples of search models where firms have flexible positions, see Christensen et al. (2005), Menzio and Shi (2011), and Postel-Vinay and Turon (2010).

bargain over wage; most job-seekers accept posted wages. As a result of laws or bargaining, government and union jobs have fixed wages. In France, public school teachers are allowed to apply yearly for placement in any region (Combe et al., 2018). Wages are fixed across regions, conditional on total experience, and so the value of a match depends primarily on the teacher's fit.

I relax the assumption of non-transferable utility in Section 1.6. To do so, I extend the model to a competitive equilibrium setting, where equilibrium wages are strategically chosen by firms. My qualitative results extend to the transferable utility setting. I also show that resumes—the ability of a worker to prove she has been previously hired elsewhere—play an important role in information transmission. Without them, competitive equilibria may fail to exist.

My results have several implications for designing centralized mechanisms. If a mechanism designer does not control the aftermarket, agents may be free to rematch after their initial assignment. Agents that are aware of this may alter their initial applications accordingly, skewing the initial outcome of the market. In particular, when agents have incomplete information about their match values, and that incomplete information is correlated with their potential for success on the aftermarket, standard algorithms such as deferred acceptance may no longer be strategy-proof. Through understanding how agents with heterogeneous incomplete information match in decentralized markets, we can better understand the impact aftermarkets have on centralized settings.

1.1.2 Related Literature

This chapter relates to several distinct literatures: matching with incomplete information, dynamic matching, directed search, and bandits with collisions.

There is a burgeoning literature on matching with incomplete information. Previous work has focused on settings with centralized clearing houses, whereas in this chapter I consider

decentralized markets, to better match settings such as labor markets or dating markets. Immorlica et al. (2020) consider school choice where students have incomplete preference information and dynamically learn through costly inspection. They solve the mechanism design problem of generating "regret-free stable" outcomes, wherein agents never regret their search decisions. There are two key differences between their setup and mine. First, they study a centralized one-shot school choice market, with fixed participants. Second, the cost of inspecting a school in their model is fixed and exogenous. Doval (2022), Liu et al. (2014), and Liu (2020) study stable outcomes in markets with incomplete information. Chen and Hu (2020) provide a dynamic justification for stability with incomplete information, in which firms evaluate potential employees according to their "worst" possible match values. These works focus on one-shot matching markets, in which no new participants enter the market after the game begins. In many dynamic environments, including the examples from the previous section, this fails to be the case.[5]

Similarly, recent work in the matching literature has begun to incorporate dynamics.[6] Akbarpour et al. (2020) consider dynamic markets with networked agents, and solve the designer's problem of choosing which agents to match. Anderson and Smith (2010) examine matchings where agents form reputations regarding their quality over time and show that positive assortative matching emerges over time. Ferdowsian et al. (2022) are at the intersection of decentralized dynamic matching and matching with incomplete information, and study the hurdles to stability that arise, even in one-shot markets. The paper shows that stringent assumptions are required to ensure stable matchings are equilibrium outcomes in markets with incomplete information. Kadam and Kotowski (2018) also consider markets with transience, and treat the problem from a more classical view of stability. They find conditions under which dynamic stability can be generated in a setting where a centralized

[5]There are also several examples of centralized markets with incomplete information. For instance, Fernandez et al. (2022) show that standard predictions of centralized markets are not robust to perturbations of information, while Li et al. (2016) experimentally tests predictions of truth telling.

[6]For a brief survey of the literature on dynamic matching, see Baccara and Yariv (2021).

authority may be necessary for finding the stable matching. I place more structure on agent's preferences, which enables me to study the related problem in a decentralized environment.

The directed search literature has studied labor markets where workers intentionally target firms.[7] Within the search literature, this paper connects two strains, search with marriage matching and search with frictions. Miller (1984); Dagsvik et al. (1985); Jovanovic (1979) consider variations of a single-agent directed-search problem.

Technically, this setting is reminiscent of the multi-armed bandit setting. The solution to the standard multi-armed bandit problem was found by Whittle (1980). Weitzman (1979) studies the mathematical problem where a decision-maker chooses when to stop testing alternatives. Several papers on multi-armed bandits with collisions have recently emerged in the computer science literature, (see Liu et al. 2020 and Liu et al. 2021). These papers assume that agents know how they are ranked by the other side of the market. In many practical situations, agents face uncertainty regarding their acceptance prospects not only because they are unaware of their competition, but also because they do not know how they will be ranked. Another key difference between this paper and the literature on bandits with collisions, is that extant algorithms fail to satisfy incentive compatibility when firms disagree on the rankings of workers.

1.2 The Model

I begin by describing the agents involved. The market matches a continuum of workers with a finite set of firms, $|\mathcal{F}| = F$. Workers are categorized based on observable information, their *class*, and unobservable information, their *type*. A worker knows her class, $c \in \mathcal{C}$, where $\mathcal{C}$ is finite. For instance, a worker knows the school she graduated from, or the grades she received. If a worker had complete information, she would also know her type, $\theta \in c$, describing her match value with each firm. Specifically, θ_w^j is the match value she receives from matching

[7] Chade et al. (2017) provide a useful survey of the extensive literature studying directed search.

with firm j, while θ_f^j is firm j's value from the match. $\theta = \{(\theta_w^j, \theta_f^j)\}_{j\in\mathcal{F}} \in [0, K]^{2F}$, is a type-$\theta$ worker's vector of match values across firms. Instead of knowing their types, workers know each class's distribution over types. When there are a finite number of types, $m_c(\theta)$ denotes the *total* mass of class-c type-θ workers; otherwise, $m_c(\theta)$ denotes class c's density of θ. The total mass of class c, $m_c = \int_{\theta\in c} m_c(\theta)d\theta$, is the summation of mass across all types of class c. I normalize the total mass of workers to 1: $\int_c m_c dc = 1$.

Each firm j has a hiring capacity of $m(j) > 0$, limiting the mass of workers he can hire each period. A *market* is a triplet $\mathcal{M} = (\mathcal{F}, \mathcal{C}, m)$. I assume that a worker's class already encapsulates any correlation in match values across firms. That is, worker i learning her match value at firm j, is not informed regarding her match value at firm j', conditional on worker i's class. Workers learn about their fit at individual firms, as opposed to learning about their preferences for certain fields or industries. Let $m_c(x, y) = \int_{\{\theta\in c|(\theta_w^j,\theta_f^j)=(x,y)\}} m_c(\theta)d\theta$ be the mass of class-c workers with match value (x, y) at firm j. Then, the absence of cross-firm learning can be formally stated as:

$$m_c(\theta) = \prod_{j\in\mathcal{F}} m_c(\theta_w^j, \theta_f^j).$$

The matching process is straightforward. Each period, every worker chooses a single firm to apply to. The choice of worker i to apply to firm j in period t is denoted by $a_i(t) = j$. Then, each firm observes his list of applicants; j observes $a_j(t) = \{i|a_i(t) = j\}$, and learns his match value with each applicant, θ_f^j.[8] Firms cannot distinguish between workers of the same class with equal firm match values. For instance, if two workers of types θ and θ' share a class, apply to firm j in a given period, and $\theta_f^j = \theta_f'^j$; then j must hire workers of the two types with equal probability.

[8]The application process can be thought of as containing an interview stage that informs firms. Because firms have previous experience with hiring workers, they are more informed about the quality of the match. Interview frictions have been previously discussed in the matching literature (Lee and Schwarz, 2017). To refine the focus on transient matchings in this paper, I abstract from these frictions in the interview process.

$A_i(t) \in \mathcal{F} \cup \varnothing$ denotes worker i's match in period t, if $A_i(t) = \varnothing$, i is *rejected*. Should a firm j hire a worker, that worker is *accepted*. A type-θ worker accepted by firm j receives θ_w^j, while the firm receives θ_f^j.[9] A rejected worker receives $\theta_w^{\varnothing} = 0$.[10] The market structure makes all matches transient, a worker that wishes to stay at a firm must apply for it every period. The set of workers hired by firm j in period t, is $A_j(t)$. Let $m(A_j(t), \theta)$ denote the mass of type-θ workers in $A_j(t)$. Then, firm j's period-t profit is:

$$\pi_j(t) = \int_{\theta \in A_j(t)} \theta_f^j dm(A_j(t), \theta)$$

Profits are discounted at a rate of δ. Each firm aims to maximize the sum of discounted profits, π:

$$\pi = \sum_{t=0}^{\infty} \delta^t \pi(t).$$

Workers retire with probability $1 - \delta$ at the end of each period, exiting the game. When a type-θ worker retires, a new type-θ worker enters the market. Critically, the new worker will no longer have the information that her predecessor acquired through play. In total, a mass of $(1 - \delta)m_c(\theta)$ class-c type-θ workers are born each period. The fact that workers stochastically exit has two implications. First, a worker's expected utility remains equivalent to the classical discounting interpretation. Second, the distribution of worker-type masses remains unchanged over time, generating a stationary environment.

When a worker enters the market, she is unaware of her exact type, but knows her class. Upon acceptance by firm j, a type-θ worker learns θ_w^j.[11] Rejected workers only learn of

[9] The main body of the paper focuses on non-transferable utility. As discussed in the introduction, several recent empirical results support this assumption. For instance, Becker (2011) shows that as much as 40% of benefits from employment are non-wage based, implying that proper matching is critical. In Section 1.6, I extend the model to a transferable utility setting where firms choose match-value dependent wages. I show that the key results carry over to the transferable utility case.

[10] In Section 1.5.2, I consider the impact of adding unemployment benefits to the market, modelled by $\theta_w^{\varnothing} > 0$.

[11] Other models of learning on the job, such as Jovanovic (1979), use more gradual learning processes, where a worker receives a noisy signal of the true match value. Since the implications generated by gradual

their rejection. There is no public history, rather workers can only learn through applying to firms. Workers aim to maximize their lifetime expected utility. That is, worker i of class c, chooses how to apply to firms, to maximize:

$$U_i^c = \mathbb{E}_c \left[\sum_{t=0}^{\infty} \delta^t \theta_w^{A_i(t)} \right]$$

To eliminate equilibria that depend upon coordination, or time-specific applications, I refine the set of equilibria to those supported by Markovian strategies. The state variable for a worker is the tuple of her payoff-relevant variables: her class, and her posterior over her type. For a given strategy profile, firms face no payoff-relevant dynamic uncertainty, so their payoff-relevant variable in a given period is the types of their applicants. Let h_t^i denote the private history for a worker i of age t, consisting of her application and realized match values in each period, $h_t^i = ((a_i(1), \theta_w^{A_i(1)}), (a_i(2), \theta_w^{A_i(2)}), \ldots, (a_i(t-1), \theta_w^{A_i(t-1)}))$. For a given history, h_t^i, and strategy profile σ, a class-c worker's posterior regarding her type can be computed using Bayes rule, and is an element of $\Delta\Theta$.

Definition (Markovian Strategies)**.** *A worker i's strategy is Markovian, if in each period, her application is only a function of her class and the posterior over her type. A firm's strategy is Markovian, if in each period, the set of workers hired is a function of the firm's distribution over applicant match values.*

The focus on equilibria in Markovian strategies rules out equilibria that depend on the calendar period. A strategy profile is Markovian if in that profile all agents use Markovian strategies. Similarly, an equilibrium is Markovian if the associated strategy profile is Markovian.

The *state* of the market, is the proportion of workers with each possible posterior over their type. A state is denoted by $\mu = \{\mu_c\}_{c \in \mathcal{C}}$, where $\mu_c \in \Delta\Theta$ denotes the posteriors of

learning have already been discussed in the search literature, I instead assume learning is immediate, to isolate the effect of competition. In Appendix section 1.B.1, I show that the core results of this paper still hold when learning is gradual.

class-c workers. Throughout the paper, I focus on the *steady state* generated by a strategy profile, σ—a state that is self-perpetuating when agents follow strategy profile σ. Namely, let the transition map, $u_\sigma : \Delta\Theta \to \Delta\Theta$, denote the mapping from a current state to next period's state when the strategy profile is σ. If for all c, $u_\sigma(\mu) = \mu$, then μ is a steady state.

A Markovian strategy profile is an equilibrium only if no agent has a profitable deviation in the steady state. This amounts to requiring firms to be unable to commit to a hiring policy. A firm's strategy is part of an equilibrium strategy profile, only if in the steady state of that equilibrium, the firm does not wish to deviate in any period *after* observing his applicants. Where appropriate, I drop time indices. Last, to deal with a trivial source of non-uniqueness, this paper considers the matching outcomes that result in the "long-run" for a Markovian strategy profile that constitutes a Perfect Bayesian Equilibrium.

Definition (Outcomes)**.** *An outcome for market $\mathcal{M}$, of a Markovian strategy profile σ, is the distribution over each types' applications to firms in the steady state of σ.*[12]

An outcome is the payoff-relevant information for a snapshot of the market. Namely, outcomes include the proportion of information each worker has learned, as well as their choice of firms to apply to. Restricting attention to the outcomes of Markovian strategy profiles removes another trivial source of non-uniqueness. Should a worker have two strategies which differ conditional on information regarding a firm j that she never applies to under either strategy, the outcomes of the two strategies are equivalent.

1.3 Incomplete Information Generates Transience

I begin by analyzing the benchmark case where workers are fully informed about their type. I find that market outcomes do not exhibit transience, instead workers follow static application strategies. In this section, I detail how the complete information setting exemplifies the

[12] I show in Section 1.4.1 that every Markovian strategy profile generates a unique steady state, and therefore this notation is well-defined.

interplay between transience and information. Should an outside observer compare outcomes in a complete information market with transient matchings to outcomes in a market with permanent matchings, she would be unable to distinguish them. In the following section, I show that the complete information case is similar to the case where long-lived workers have incomplete preference information. However, several key differences emerge between the two. With incomplete information, worker outcomes exhibit path dependence—the results of a finite number of initial applications determine their long-term earnings. Furthermore, when markets are congested, workers may be persistently matched to firms different from those they would be matched to under complete information.

To begin, I develop a benchmark for the case where workers are fully informed, and workers and firms agree on the value of each match. That is, match values are *aligned*.[13] Match values are aligned if they can be represented by a joint "ordinal potential." This is similar to the notion of potentials in normal-form games (Monderer and Shapley, 1996). Formally, alignment is satisfied if there exists a matrix $\Phi = (\Phi_{\theta j})_{j \in \mathcal{F}} \in \mathbb{R}$, such that for any types $\theta, \theta' \in \mathcal{C}$ and firms j, j':

$$\text{If } \theta_w^j > \theta_w^{j'} \text{ then } \Phi_{\theta j} > \Phi_{\theta j'} \text{ and if } \theta_f^j > \theta_f'^j \text{ then } \Phi_{\theta j} > \Phi_{\theta j'}$$

For instance, if workers and firms split the surplus from any match according to a fixed proportion, match values are aligned: $\exists \alpha$ s.t. $\forall j, \theta : \theta_w^j = \alpha \theta_f^j$.[14] Alignment will imply that a firm cannot reject a worker and trigger a cycle of other firms' rejections, which culminates in a preferred worker's application.[15] When match values are aligned, I will simplify notation

[13] The term aligned comes from Ferdowsian et al. (2022), in which alignment is shown to be critical for the emergence of stability in decentralized markets.

[14] In this case, $\Phi = (\Phi_{\theta j})_{\theta \in \mathcal{C}, j \in \mathcal{F}}$ where $\Phi_{\theta j} = \theta_f^j$ for all θ, j serves as an ordinal potential.

[15] In Voorneveld and Norde (1997), it is shown that potential games cannot have cycles in the payoff matrix. In this setting, their result will imply that markets with aligned match values cannot have rejection cycles. That is, in equilibrium, no firm can reject a worker, and thereby trigger a change in applications such that the firm receives an application from a preferred worker.

by letting, $\forall \theta, j\ \theta_w^j = \theta_f^j = \theta^j$, but the results hold more generally for any aligned match values.

Definition (Complete Information Market)**.** *A market $\mathcal{M}_I$ is a complete information market, if every worker class c contains a single type.*

In a complete information market, all workers are perfectly informed about their match values at each firm. Suppose all match values are strictly ordered. Then, there exists some class-c^* and firm j^*, such that class-c^* workers and j^* generate the maximal match value: $(c^*, j^*) = \arg\max_{c,j \in \mathcal{E}} \theta_c^j$. Class-$c^*$ workers must apply to j^* with positive probability. Otherwise, some class-c^* worker, i, would benefit by deviating and applying to j^*. Because no other class-c^* workers are applying to j^*, i is guaranteed to be hired since j^* prefers i to any other applicant. Furthermore, by switching her application, i will receive a higher match value than she would have previously. Indeed, either all class-c^* workers must apply to j^*, or at least enough must apply to force j^* to begin rejecting class-c^* applicants. If too many class-c^* workers apply to j^*, the resulting competition may leave class-c^* workers preferring their second choice, $j_2 = \arg\max_{j \neq j^*} \theta_{c^*}^j$. Firm j^* is *oversubscripted* if class-c^* workers would prefer guaranteed hire by j_2 to applying to j^* when all other class-c^* workers apply to j^* as well: $\frac{m(j^*)}{m(c^*)} \theta_{c^*}^{j^*} < \theta_{c^*}^{j_2}$. In the event of oversubscription, at least $m(j^*) \frac{\theta_{j^*}^{j^*}}{\theta_{c^*}^{j_2}}$ class-c^* workers must apply to j^*.

If oversubscription occurs, j^* is filled to capacity, otherwise all class-c^* workers apply to j^*. In either event, a submarket can be considered with strictly less agents than $\mathcal{M}_I$. The previous argument can be repeated in the submarket, removing another agent from the market. Extra care must be taken if a class or firm that was previous part of an oversubscription is being considered. I go into detail about how to ensure indifference in such a step, when I formalize the procedure as the "Top-Down" algorithm in the appendix. Because there are a finite number of classes and firms, the process must eventually terminate, in at most $|\mathcal{C}| + |\mathcal{F}|$ steps.

Lemma 1 (Complete Information Equilibrium)**.** *The Top-Down algorithm characterizes the unique equilibrium outcome of $\mathcal{M}_I$ when match values are aligned.*

Matchings are permanent in complete information environments—workers apply to the same firms every period. However, despite information being complete, decentralization inherently generates two types of inefficiencies. First, workers that are qualified for desirable positions over-compete for those positions. This feature of oversubscription generates significant congestion, a negative externality on other workers of their class. Second, workers do not consider their impact on other classes of workers. For instance, suppose there are two firms, $\mathcal{F} = \{j_1, j_2\}$, each with a capacity of $1/2$. There are two classes of workers, $\mathcal{C} = \{c_1, c_2\}$, with match values $\theta_1 = (5, 4)$ and $\theta_2 = (4, 0)$, and masses $m_{c_1} = 1/2, m_{c_2} = 1/2$. Then, in equilibrium, class-c_1 workers will always apply to j_1, blocking class-c_2 workers from being hired there. From a utilitarian standard point, a more efficient outcome would involve class-c_1 workers applying to j_2, while class-c_2 workers apply to j_1. Inherently, workers do not consider the externality they impose on workers of other classes. As will be seen, introducing incomplete information adds a third externality. Workers do not consider how their applications block other workers from searching for good fits. In Section 1.4.5, I show that when workers have incomplete information and become more long-lived, outcomes converge to the complete information outcomes, though marked differences appear for individual workers.

1.4 Equilibrium Characterization

When workers have incomplete information and are not long-lived, they face a non-trivial tradeoff. Workers must decide between exploiting their information—applying to firms with high expected match values; and learning—applying to new firms. When determining whether to explore, workers must take into account other workers' application decisions, which generate competition over firms.

I proceed by characterizing best responses for both firms and workers. First, the following section shows that in any strategy profile where a firm does not hire his most qualified applicants, he has a profitable deviation. Once firms' strategies are thresholds, workers can anticipate those thresholds in equilibrium, and evaluate firms accordingly. Using techniques from the literature on multi-armed bandits, worker behavior can then be determined. Then, because workers evaluate firms as if they were bandits, demand for a firm is increasing in other firms' thresholds. Using this, I develop an algorithm to find a fixed point where the demand for each firm coincides with the firm's capacity and threshold, and show that the algorithm captures the unique equilibrium outcome. Last, I show that worker's search patterns aligns with empirical results.

Before proceeding, Lemma 2 shows that any Markovian strategy profile generates a unique outcome, implying that the focus on steady state outcomes is well-defined.

Lemma 2 (Unique Steady State)**.** *Any Markovian strategy profile has a unique steady state.*

1.4.1 Hiring Thresholds

In Example 3, it was assumed that a firm would always hire his most qualified applicants in every period, without regard for how doing so would impact future applications. I show that firms' equilibrium strategies can be described as threshold strategies.

When a firm's strategy profile dictates he rejects workers more qualified than those he accepts, he has a profitable deviation. Through accepting workers with higher match values, the firm is able to reap the benefit from hiring the qualified worker immediately, instead of delaying the match value from acceptance to a future period. The key insight is that when match values are aligned, a firm which rejects a worker with a high match value can only do so to incentivize that worker to apply again in a future period. However, the rejection simply shifts back the expected gains from matching with that worker, and so the firm would prefer to accept her in the current period.

Proposition 1 (Firms Use Hiring Thresholds)**.** *Suppose match values are aligned. In the steady state of any equilibrium, if firm j hires a worker with a match value of θ_f^j, then firm j also hires all applicants with match values above θ_f^j. Additionally, if firm j hires below its capacity, $m(A_j) < m(j)$, then j hires all applicants.*

A worker's expected utility from a firm inherently depends upon that firm's hiring decision, which in turn is a function of the application strategies of other workers. There is a natural ordering on worker types at firm j, $\theta \gtrsim \theta'$ if $\theta_f^j > \theta_f'^j$. Proposition 1 implies that the firms' hiring decision can be summarized by the minimal hired θ_f^j, and the probability with which type-θ workers are hired. For a worker, a firm's strategy is payoff-relevant only in how it affects the probability of hire at that firm. This motivates a natural method of summarizing a firm's strategy, the threshold in match values below which it rejects applicants.

Definition (Hiring Threshold)**.** *A hiring threshold for some firm, j, is a tuple (v_j, p_j), which consists of a match value and a hiring probability.*

Crucially, in the steady state of an equilibrium, thresholds are time-invariant. To determine firm j's threshold, begin by finding the set of workers that apply to j within a single period. Because firm j uses a threshold strategy, j hires a worker with the minimal match value. Let θ be that worker's type. Set v_j to θ's firm j match value, θ_f^j, and p_j to type-θ's probability of hire at j. If the mass of applicants received by j is below its capacity, $m(a_j) \leq m(j)$, define j's threshold as $(0, 1)$—all applicants to j are hired in equilibrium.

A threshold suffices to describe a firm's behavior in equilibrium. Workers can use v_j and p_j to compute their expected utility from applying to firm j. If, for a worker type θ, θ_f^j is above v_j, then type-θ workers are always hired by j. If θ_f^j is below v_j, type-θ workers are never hired by j. Last, if $\theta_f^j = v_j$, then type-θ workers are hired by j with probability p_j.

It will be convenient to rank thresholds in the lexicographic order. A threshold, (v_j, p_j), is greater than another threshold, (v_j', p_j'), $(v_j, p_j) \succ (v_j', p_j')$ if $v_j > v_j'$, or if $v_j = v_j'$ and $p_j \geq p_j'$. $(v, p) = \{(v_j, p_j)\}_{j \in \mathcal{F}}$ will refer to the vector of thresholds.

The optimality of threshold strategies is a direct consequence of firms' inability to commit to hiring policies. If a firm could commit to hiring workers of a given class, then that firm could encourage certain classes of workers to apply through affirmative action policies. This would enable the firm to attract highly qualified workers of types within those classes through the promise of guaranteed jobs. I expand on this motive in Section 1.5.1.

1.4.2 Equilibrium Worker Strategies

Next, I discuss the characterization of a worker's equilibrium strategy. As previously mentioned, workers trade-off exploring firms where their match value is still unknown, and exploiting firms that are known to be profitable matches. I discuss a technique from the multi-armed bandit literature, which provides an index-based solution method that holds for any market where all firms hire all applicants. I extend the technique to settings where the hiring probabilities are endogenously determined, and show that the index solution carries over.

Definition (Gittins Index)**.** *Gittins (1979) The Gittins index for worker i at firm j given history h_t^i, $GI_i(j, h_t)$, is the solution to the following optimal stopping problem, where τ can depend upon the realization of match values:*

$$GI_i(j, h_t^i) = \sup_{\tau} \frac{\mathbb{E}_{i,h_t^i}\left[\sum_{t=1}^{\tau} \delta^t \theta_w^j\right]}{\mathbb{E}_{i,h_t^i}\left[\sum_{t=1}^{\tau} \delta^t\right]}$$

Gittins indices provide a simple characterization of the benefits from learning. Intuitively, the Gittins index of firm j for a class-c worker is her expected match value at j, weighted by her ability to learn. Namely, the worker can strategically reapply to j conditional on the realized match value. If the worker discovers a high match value at j she can take advantage of the positive realization and reapply to j, instead of leaving immediately as she would after discovering a low match value at j. The ability to strategically reapply biases the Gittins index upward from the standard expected value, and incorporates a benefit from learning.

When describing the Gittins index of a newly arrived worker, I omit the trivial match history to conserve space.

Determining the optimal worker policy can be done tractably, thanks to a result from the operations literature. Rather than a worker worrying about the timing of learning and equilibrium effects, the worker simply compares the equilibrium Gittins indices associated with each firm. Naturally, these Gittins indices are endogenous objects, which depend on the proportion of workers applying to each firm.

The above formulation is insufficient when firms do not hire all applicants. However, declaring firm j's threshold allows for the computation of each worker's initial Gittins index at j. To do so, simply replace each match value with the imputed match value condition on (v_j, p_j). Suppose for an equilibrium strategy profile σ, the hiring thresholds are (v, p). Let $\psi_\theta(v, p)$ be a type-θ worker's realized match values given thresholds (v, p). That is:

$$\psi_\theta^j(v,p) \equiv \begin{cases} \theta_w^j & \theta_f^j > v_j \\ p_j \theta_w^j & \theta_f^j = v_j \\ 0 & \theta_f^j < v_j \end{cases}$$

What stopping strategy τ determines j's Gittins index? It can be shown that the following heuristic determines the value of $GI_i(j, h_t^i)$. Choose an arbitrary value x and apply to j. If the resulting match value is above x, stay with j forever, $\tau = \infty$. If the match value is below x, leave j immediately, $\tau = 1$. This procedure determines a value for the stopping problem if x is equal to the value of the stopping problem. The maximal value of x determines the Gittins index. Lemma 3 formalizes this heuristic computation.

Lemma 3 (Firms' Equilibrium Indices)**.** *$GI_i^\sigma(j, h_t^i)$ for a class-c worker i is characterized by the fixed point solution to the following functional equation:*

$$GI_i^\sigma(j,h_t^i) = \frac{\mathbb{P}\left[\psi_w^j < GI_i^\sigma(j,h_t^i)\right]\mathbb{E}\left[\psi_w^j | \psi_w^j < GI_i^\sigma(j,h_t^i)\right] + \frac{1}{1-\delta}\mathbb{P}\left[\psi_w^j \geq GI_i^\sigma(j,h_t^i)\right]\mathbb{E}\left[\psi_w^j | \psi_w^j > GI_i^\sigma(j,h_t^i)\right]}{\mathbb{P}\left[\psi_w^j < GI_i^\sigma(j,h_t^i)\right] + \frac{1}{1-\delta}\mathbb{P}\left[\psi_w^j > GI_i^\sigma(j,h_t^i)\right]}$$

In Example 3, determining the value of $GI_i(f_s)$ was trivial.[16] Since each worker knows her match value at f_s, for any choice of τ, $GI_i(f_s) = 2$. Computing $GI_i(f_r)$ requires slightly more work. First, consider the case where $m(f_r) = 1$, that is, f_r is not capacity constrained. Then, f_r must hire all applicants in equilibrium. If a worker applies to f_r and receives a match value of 1/2, she immediately knows her type is θ_l. If she applies to f_r a second time, the numerator of $GI_i(f_r)$ is increased by $1/2\delta$, and the denominator is increased by δ. If $GI_i(f_r) > 1/2$, then it can be shown that the solution of the stopping time problem induced by $GI_i(f_r)$ never involves applying a second time to f_r after observing $\theta_w^{f_r} = 1/2$. If the worker had applied to f_r and received a match value of 3, she would know that her type was θ_h. An extension of the previous logic shows that if $GI_i(f_r) < 3$, the solution to the stopping time problem involves applying to f_r a second time upon observing $\theta_w^{f_r} = 3$. Putting everything together, the value of $GI_i(f_r)$ is found by infinitely applying to f_r, if type θ_h, and applying only once if type θ_l, yielding:

$$GI_i(f_r) = \frac{1/2\left(\frac{3}{1-\delta} + 1/2\right)}{1/2\left(\frac{1}{1-\delta} + 1\right)}$$

Notably, when computing $GI_i(f_r)$, the distribution over possible match values at f_s was not utilized. Even when the number of firms, F, increases, the number of computations to determine all Gittins indices also increases linearly in F, one additional step for each additional firm.

[16] As a reminder, both types had a match value of 2 at f_s, while type-θ_h had a match value of 3 at f_r and type-θ_l had a match value of 1/2 at f_r.

Standard results in the operations literature show that, determining the optimal policy takes two steps: in each period compute the Gittins index for each firm, then apply to the firm with the highest Gittins index (Whittle, 1980). Previous results involving directed search settings or Gittins indices have required that only one decision-maker is present or that the decisions made by each decision-maker are independent of one another.[17] Through classifying firms' strategies as threshold strategies, the results are extended to settings where multiple workers compete over matches.

1.4.3 Alignment Implies Uniqueness

In this section, I show that when firms and workers agree on the quality of each match, i.e., match values are *aligned*, there exists a unique equilibrium outcome under a mild condition on payoffs. To do so, I first develop the following example, which illustrates the specificity in payoffs required to generate non-unique equilibrium outcomes.

Example 1 (Non-Unique Outcomes)**.** *Rows represent worker types, while columns represent firms. The mass of each agent is given in the first element of the corresponding row or column. The match values for a worker and firm are given at the intersection of their row and column.*

Firms / *Class-c_1 Workers*	$m(f_r) = 1/2$	$m(f_s) = 1/2$
$m_1(\theta_h) = 1/4$	$(4, 4)$	$(3, 3)$
$m_1(\theta_l) = 1/4$	$(2, 2)$	$(3, 3)$

Firms / *Class-c_2 Workers*	$m(f_r) = 1/2$	$m(f_s) = 1/2$
$m_2(\theta) = 1/2$	$(1, 1)$	$(1, 1)$

[17] For examples in economics, see Papageorgiou (2018), Rothschild (1974), Urgun (2021), and Weitzman (1979).

There are two worker classes, $\mathcal{C} = \{1, 2\}$, two firms, $F = 2$, and $\delta = 0$. When class-c_1 workers are guaranteed to be hired, and indifferent over firms, there exists a multiplicity of equilibrium outcomes.

Class-1 workers are indifferent between the two firms, and furthermore are never rejected in equilibrium. Therefore, incoming class-1 workers are indifferent between the two firms in any equilibrium. There exists a continuum of equilibrium outcomes characterized by the proportion of class-1 workers that applies to f_r.

The non-uniqueness of equilibrium outcomes cannot be trivially fixed by requiring each firm's Gittins index to be unique in the absence of firm hiring capacities. To see why, return to Example 1 and suppose class 1 contained several additional types, such that match values below 1 were possible at both firms. In the absence of firm hiring constraints, class 1's Gittins index at f_r, $GI_1(f_r)$, would be higher than $GI_1(f_s)$. In practice, the difference would not factor into class-1 workers' decision-making, because in any equilibrium, workers with match values below 1 are never hired.

Furthermore, Example 1 did not hinge on the absence of learning. Consider the same market, with some $\delta > 0$. A simple computation can be used to evaluate $GI_1(f_r) = \frac{1/2\frac{4}{1-\delta}+1/2\cdot 2}{1/2\frac{1}{1-\delta}+1/2\cdot 1}$ while $GI_1(f_s) = 3$. For $\delta > 0$, $GI_1(f_r) > GI_1(f_s)$ due to the option value of switching to f_s if a low utility is realized. However, a small adjustment to θ_h^r can set $GI_1(f_r) = GI_2(f_s)$. Replacing $\theta_h^{f_r} = 4$ with $\theta_h^{f_r} = 4 - \delta$ ensures that the two Gittins indices are equal once more. In the resulting market, there again exists a continuum of equilibrium outcomes. As such, a more stringent condition is needed to rule out this trivial form of non-uniqueness.

The *strict dynamic preferences* assumption requires workers to not face indifference between two firms when each firm hires deterministically. This assumption depends on δ, but only rules out a finite number of parameters from an infinite set. The parameter set after strict dynamic preferences are ensured is generic. For instance, the parameter set which generated multiple equilibrium outcomes in Example 1 is eliminated by strict dynamic pref-

erences, preventing class-1 workers from mixing. Intuitively, strict dynamic preferences requires that the equilibrium Gittins indices at any two non-competitive firms cannot be equal. It should be emphasized that this assumption is exogenous, the assumption only depends on utilities and masses.

Assumption 1 (Strict Dynamic Preferences)**.** *Let every type of worker class c have probability of hire* 0 *or* 1 *at firm j and j'. Strict dynamic preferences holds if for every class c there does not exist a type $\theta \in i$ such that the following hold:*

- *θ has probability* 1 *of hire at j*
- *$GI_i(j) = GI_i(j')$, $\theta_w^j = GI_i(j')$, or $\theta_w^j = \theta_w^{j'}$*

First, no strict dynamic preferences implies that two strategy profiles, which induce different equilibrium outcomes, must also generate different hiring thresholds. The claim is formally proved in the appendix. Given this, Proposition 2 shows that equilibrium outcomes must be unique under strict dynamic preferences.

Proposition 2 (Unique Outcome)**.** *When match values are aligned and satisfy strict dynamic preferences, there is a unique equilibrium outcome.*

To provide intuition, suppose σ and σ' are distinct equilibrium outcomes. Then, the thresholds induced by σ and σ' must be distinct. Without loss of generality, let firm j have a higher threshold under σ than σ', $(v_f(\sigma), p_f(\sigma)) \succ (v_f(\sigma'), p_f(\sigma'))$. Then, f must have more high quality applicants under σ relative to σ', otherwise it could not have a higher threshold. Those applicants must come from another firm f' under σ'. For those applicants to switch to f, f' must be less attractive under σ. However, when match values are aligned, thresholds imply that the quality of workers hired is monotonic, $(v_{f'}(\sigma), p_{f'}(\sigma)) \succ (v_{f'}(\sigma'), p_{f'}(\sigma'))$. Iterating this logic implies that there exists a set of firms with higher thresholds. Since all involved firms have binding capacity constraints, weakly more workers must be hired under

σ. Furthermore, the fact that all hiring thresholds are higher, implies that total worker utility is also higher. Then, at least one class, c, must be strictly better off under σ. Since hiring thresholds were lower under σ', class-c workers can profitably deviate to their strategy under σ. Then, σ' could not have been an equilibrium outcome.

The uniqueness of equilibrium transient matchings is useful for the prospective econometrician. When estimating structural parameters, the statistician does not need to worry about the problem of equilibrium selection. Instead, the unique equilibrium can be determined and exploited.

1.4.4 Gross Substitutes in Equilibrium

In this section, I proceed by describing equilibrium strategies, utilizing firm thresholds to characterize equilibria. Importantly, because workers evaluate firms as if they were bandits, their demand for firms is increasing in other firm's thresholds.

First, I show that standard bandit results can be extended from settings where all firms have unbounded capacities to equilibrium settings where firms can only hire a limited subset of workers.

Lemma 4 (Firms are Endogenous Bandits)**.** *For any equilibrium strategy profile σ, for any private history h_t^i, each worker i applies to $j \in \arg\max_j GI_i^\sigma(j, h_t^i)$.*

Lemma 4 shows that the optimal worker strategy for a given set of thresholds can be simply determined. For each worker i, determine each firm's Gittins index, potentially updating the probability of each match value based on h_t^i. Each worker i applies to a firm with maximal Gittins index, $j \in \arg\max_j GI_i^\sigma(j, h_t^i)$. This application strategy prescribes a distribution over learning paths for each worker type. In turn, the mass of informed worker types and therefore the proportion of each type applying to each firm is fixed, as well as their equilibrium hiring rates. Type-θ's demand for firm j given thresholds (v, p) will be

denoted $D_j^\theta(v,p)$. *Aggregate demand*, $D(v,p)$ is defined as the integral over all types' realized demands.

Definition (Aggregate Demand)**.**

$$D(v,p) = \left(\int_\theta D_{j_1}^\theta(v,p)d\theta, \int_\theta D_{j_2}^\theta(v,p)d\theta, \ldots, \int_\theta D_{j_F}^\theta(v,p)d\theta \right)$$

The characterization of optimal index-based worker strategies has an additional feature. When a firm increases its threshold, demand at other firms can only increase. To see why, note that workers apply to the firm with the maximal Gittins index. Increasing a firm's threshold decreases that firm's Gittins index for all types. Furthermore, the independence of the Gittins indices implies that all other firms' Gittins indices remain fixed. The increase in the threshold will not change other firms' demands unless it causes the Gittins index of one type at a firm to exceed their Gittins index at another firm. Doing so, will shift that type's demand to the other firm. It can be shown that the effect on that type's demand in future periods does not exceed the immediate effect. More generally, increasing the thresholds of a subset of firm cannot decrease the demand of firms outside that subset. Namely, aggregate demand satisfies the *gross substitutes* condition, used in Kelso and Crawford (1982) and Gul and Stacchetti (2000), to prove existence of equilibria. Let firm j's demand, $D_j(v,p) = \int_\theta D_j^\theta(v,p)d\theta$, be the demand across all types for firm j. Then, gross substitutes is defined as follows.

Definition (Gross Substitutes)**.** *Let $\boldsymbol{F}$ be a set of firms, $\boldsymbol{F} \subset \mathcal{F}$, and $(v,p),(v',p')$ be two thresholds, such that $(v_j,p_j) = (v'_j,p'_j)$ for $j \in \mathbf{F}$ and $(v_j,p_j) \geq (v'_j,p'_j)$ otherwise. Gross substitutes is satisfied if demand for $\boldsymbol{F}$ is greater under (v_j,p_j) than (v'_j,p'_j):*

$$\forall j \in \boldsymbol{F} : D_j(v,p) \geq D_j(v',p')$$

Proposition 3. *Aggregate demand satisfies gross substitutes.*

Given Proposition 3, the equilibrium can be found through the *threshold adjustment process*. The the threshold adjustment process follows five steps:

Step 1 (Remove Capacities): Treat every firm as if it had no capacity constraint: set all thresholds to $(v_j, p_j) = (0, 1)$.

Step 2 (Compute Demand): For every worker class, compute the Gittins index of each firm. Using these Gittins indices, determine individual demand, and in turn, aggregate demand for each firm.

Step 3 (Select Firm): If all firms' have demand below capacity, $\forall j, D_j(v, p) < m(j)$, terminate. Otherwise, select a firm j such that j's aggregate demand exceeds its capacity, $D_j(v, p) > m(j)$.

Step 4 (Reduce Capacity): Decrease the rate of hire for workers with the lowest match value at firm j continuously, $p_j \to 0$. If p_j reaches zero, repeat with the next lowest match value at firm j. Stop when either enough workers are fired, or decide to apply to other firms, $D_j(v, p) = m(j)$. Importantly, Proposition 3 implies that increasing firm j's threshold never causes a decrease in demand for other firms.[18]

Step 5 (Repeat or Terminate): If there exists another firm j' such that $D_{j'} > m(j')$: repeat step 3. While doing so, whenever a type-θ worker with a history realization that occurs with positive probability, would be rendered indifferent between j' and one or more previously treated firms, $\mathbf{F}$, simultaneously decrease the p_j of all $j \in \mathbf{F} \cup j'$, to ensure that the type's Gittin indices are equal across all such firms, $GI_\theta(j) = GI_\theta(j')$, while reallocating type-θ workers across these firms to equalize demand and capacity. This continues until either all $p_j = 0$, or $D_{j'}(v, p) = m(j')$.[19]

[18] This step requires a technical adjustment when worker types have positive mass to avoid cycles, see the appendix for the details.

[19] This characterization of equilibrium behavior is similar in spirit to the characterization of Azevedo and Leshno (2016). There are two key differences. First, aggregate demand is not only a function of workers'

By design, the procedure begins with all workers applying as if they anticipated firms would not be congested. In practice, this will cause certain firms to receive more applicants than could be hired, due to the firms' capacity constraints. Through continuously firing the worst applicants, a firm decreases the number of applicants who expect to be hired. The decreased rate of hire may cause workers to apply to other firms, but due to the gross substitutes condition, other firms' demands never decrease. Eventually, the firm reaches its capacity. The process then repeats with another firm that is hiring more workers than it can support. However, any time a worker is rendered indifferent between the current firm and a previously treated firm, that worker then randomizes between the two firms so as to not violate the previously treated firm's capacity. This ensures that once a firm has met its capacity, it never again exceeds its capacity. As such, the threshold adjustment process terminates in a finite number of iterations, one for each firm.

Theorem 1 (Equilibrium Characterization)**.** *The threshold adjustment process culminates in an equilibrium strategy profile in at most F steps.*

Standard bandit techniques such as the Gittins index are normally only valid in decision problems. Despite this, the characterization through hiring thresholds provides the necessary step for the proof of Theorem 1. These thresholds reduce the best response problem into worker-specific decision problems, each of which can then be solved through computation of Gittins indices for a given set of thresholds. Furthermore, this procedure only requires a finite number of iterations, making implementation simple.

The threshold adjustment process shares several features with the worker proposing Deferred Acceptance algorithm. Workers begin by proposing to their ideal positions, then reconsider as they are rejected. In settings where match values are not aligned, there may be multiple equilibria. However, the equilibria can be ranked through a weak order in terms of

initial applications, but also must account for applications from workers who have previously applied and learned. Second, my formulation allows for settings where firms are indifferent over a positive mass of workers—for instance, when worker types are discrete.

worker and firm preferences. Consider two equilibrium strategy profiles, σ and σ', where the thresholds induced by σ are greater than those induced by σ', $(v,p)(\sigma) \geq (v,p)(\sigma')$. Workers must prefer the equilibrium under σ' to σ, while firms prefer σ to σ'.

Proposition 4 (Threshold Ranking)**.** *If σ and σ' are equilibrium strategy profiles, and $(v,p)(\sigma) \geq (v,p)(\sigma')$, then for all $c \in \mathcal{C}, U^c(\sigma) \leq U^c(\sigma')$, and for all $j \in \mathcal{F}$, $\pi_j(\sigma) \geq \pi_j(\sigma')$.*

Importantly, if the expected utility under σ was above that under σ', a class c worker could deviate to her strategy under σ, be hired weakly more often, and receive greater utility. Conversely, firms receive higher utility under σ because their thresholds are higher, implying a higher quality for each worker hired.

1.4.5 Efficiency and Long-Lived Behavior

In this section, I show how congestion can lead to inefficiently low levels of search. I then prove that search converges to the efficient benchmark as δ converges to 1. In particular, outcomes converge to the complete information outcome as δ converges to 1. However, in certain markets, for any $\delta < 1$, congestion always prevents a mass of workers from learning their type—for any $\delta < 1$, these workers would strictly benefit from learning their type at the outset.

In Example 2, two classes of workers decide whether to apply to f_r, or take a guaranteed payoff at f_s. I show that in equilibrium, f_r would prefer a certain class of workers to apply, but those workers do not do so, due to the fear of rejection. Had a social planner determined the acceptances of workers, total equilibrium payoff would be higher, and furthermore, f_r would have been better off. Subsequently, I show that through hiring headhunters, f_r can attract these workers, increasing its equilibrium profit. The safe firm has enough capacity to hire all workers, and so hires all applicants in equilibrium. The following payoff matrix depicts the market.

Example 2 (Congestion Restricts Efficient Search)**.**

Class-c_1 Workers \ *Firms*	$m(f_r) = 1/2$	$m(f_s) = 1$
$m(\theta_h) = 1/4$	$(3, 3)$	$(2, 2)$
$m(\theta_l) = 1/4$	$(1/2, 1/2)$	$(2, 2)$

Class-c_2 Workers \ *Firms*	$m(f_r) = 1/2$	$m(f_s) = 1$
$m(\theta) = 1/2$	$(2, 2)$	$(2 - \epsilon, 2 - \epsilon)$

There are two classes of workers, $\mathcal{C} = \{c_1, c_2\}$. For $\delta \in (1/3, 1/2)$, congestion restricts the willingness of c_1 workers to search, resulting in an inefficient outcome.

In Example 2, class-c_1 workers face uncertainty regarding their type, and would prefer to apply to f_r in the absence of congestion for sufficiently high δ. However, the presence of class-c_2 workers means that a class-c_1 type-θ_l worker is not hired at f_r. For $\delta < 1/2$, the only equilibrium is the strategy profile where all class-c_1 workers apply to f_s, while all class-c_2 workers apply to f_r. To see why, consider the payoff a class-c_1 worker receives from deviating and applying to f_r. With probability $1/2$, her type is θ_h. She receives 3 immediately from the match, and can continue applying to f_r indefinitely. Otherwise, her type is θ_l, causing f_r to reject her, after which she applies to f_s indefinitely. In expectation, her utility from deviation is $1/2\frac{3}{1-\delta} + 1/2\frac{2\delta}{1-\delta}$, whereas if she had remained at f_s, she would have received $\frac{2}{1-\delta}$. For $\delta < 1/2$, this deviation is not profitable, and so the original strategy profile is an equilibrium.

However, this equilibrium is inefficient for intermediate values of δ, namely when $\delta \in (1/3, 1/2)$. Suppose a centralized authority required f_r to hire all applicants from class-c_1. Then, class-c_1 workers who apply to f_r receive $1/2\frac{3}{1-\delta} + 1/2(1/2 + \frac{2\delta}{1-\delta})$ and so apply to f_r when $\delta > 1/3$. In the new equilibrium, all class-c_1 workers would initially apply to f_r, then type-θ_h workers would continue applying to f_r while type-θ_l workers would apply to f_s.

This implies that f_r receives a mass of $(1-\delta)(\frac{1/4}{1-\delta}+1/4) = 1/4(2-\delta)$ class-c_1 applicants. Importantly, for any $\delta \in (1/3, 1/2)$, $1/4(2-\delta) < 1/2$, so f_r does not exceed its capacity constraint. In equilibrium, the applications of class-c_2 workers to f_r would then adjust accordingly to render class-c_2 workers indifferent between the two firms. From a utilitarian perspective, when $\delta \in (1/3, 1/2)$, this equilibrium is more efficient than the previous one. Indeed, while the new total worker utility is above $1/4\frac{3}{1-\delta} + 1/4(1/2 + \frac{2\delta}{1-\delta}) + 1/2(2-\epsilon)$. Note that class-$c_1$ workers necessarily receive more than 2 in utility, otherwise they would be better off applying to f_s. Therefore, total worker utility has increased from 2, the previous total worker utility, for sufficiently small values of ϵ.[20] Total welfare is increased through preventing congestion from limiting socially optimal search.

Next, I investigate outcomes when workers are long-lived. As δ increases, the inefficiency found in Example 2 is ameliorated. Indeed, for sufficiently high values of δ, workers are motivated to search for their maximal match values. In fact, the equilibrium outcome converges to the complete information equilibrium outcome as workers become more patient, i.e., when δ converges to 1. The inefficiencies present in the complete information environment carry over to the case where workers are long-lived, and have incomplete preference information. When workers do not know their type, they must also invest time to determine their top possible match. However, workers may be unwilling to do so if δ is sufficiently small. In the limit, this source of inefficiency converges to 0, as δ goes to 1. However, for any $\delta < 1$, the inefficiency always remains strictly positive in markets with oversubscription. Even as δ goes to 1, certain worker types always benefit from learning their match values, i.e., there exists path dependence despite vanishing time frictions. Path dependence arises due to the possibility that an unlucky hiring realization can leave a worker uninformed regarding her type. Example 3, with a single class of workers, $\mathcal{C} = 1$, illustrates this form of path dependence.

[20] It is also worth noting that f_r benefits from this change as well. Thus if f_r could commit at the beginning of each period to hiring all class-c_1 workers, he would prefer to do so.

Example 3 (Path Dependence)**.**

Workers \ *Firms*	$m(f_r) = 1/4$	$m(f_s) = 1$
$m(\theta_h) = 1/2$	$(3, 3)$	$(2, 2)$
$m(\theta_l) = 1/2$	$(1/2, 1/2)$	$(2, 2)$

Even when workers are long-lived, path dependence emerges in equilibrium.

The set of possible equilibrium strategy profiles can be categorized by the number of failed applications to f_r after which workers cease applying to f_r, and instead apply to f_s forever. Naturally, the equilibrium strategy profile depends upon δ. When δ is low, workers are unwilling to invest into learning, and all apply to f_s. However, even when δ is high, any strategy profile where all type-θ_h workers learn their type with arbitrarily high precision through repeated applications to f_r cannot be an equilibrium. To see why, consider the strategy profile where all workers apply to f_r repeatedly. Then, if all informed type-θ_h workers continued applying to f_r, each would be hired with probability $\frac{1/4}{1/2} = 1/2$. This cannot be an equilibrium, as type-θ_h workers would deviate and apply to f_s instead. Suppose informed type-θ_h workers would randomize between f_r and f_s such that informed type-θ_h workers are indifferent between the two firms, namely a total of $3/8$ type-θ_h workers apply to f_s so that the expected utility from f_r is 2. However, this would imply that uninformed workers expect a match value less than 2 from applying to f_r, because of the non-zero probability that they are type-θ_l. Uninformed workers would then deviate and apply to f_s. Therefore, all workers applying to f_r indefinitely could not have been an equilibrium.

When δ is sufficiently close to one, all workers will apply at least once to f_r. Since workers are not informed of their type when they are rejected, a portion of the workers that fail their initial application will reapply to f_r. The more workers that reapply, the lower an uninformed worker's expected payoff from f_r will be, due to the competition type-θ_h workers

face. As will be shown, when $\delta < 1$, in any equilibrium, uninformed workers with two failed applications to f_r never apply to f_r.

Suppose it was optimal for an uninformed worker to apply more than twice to f_r. Uninformed workers with two failed applications place a lower weight on their probability of being type θ_h than uninformed worker with zero or one failed application. As such, uninformed workers with zero or one failed applications must strictly prefer to reapply to f_r. Then, for δ sufficiently close to 1, almost all workers will have applied at least twice to f_r. Since each type-θ_h worker is hired with a minimum probability of 1/2, a minimum of 3/4 of type-θ_h workers will have been hired at least once at f_r and therefore know their type with certainty.

Informed type-θ_h workers must continue applying to f_r, otherwise applying to f_r while uninformed could not have been profitable. Then, the average rate of hire for type-θ_h workers at f_r will be below 2/3, implying informed workers would prefer to apply to f_s. This proves that any strategy profile where uninformed workers apply to f_r after more than one rejection cannot be an equilibrium. Therefore, all uninformed workers with two failed applications apply to f_s. The reapplication rate will be such that workers with a single failed application are indifferent between reapplying and applying to f_s. In equilibrium, the probability a type-θ_h worker is hired at f_r, denoted by p, and the probability an uninformed worker with a single failed application to f_r reapplies to f_r, denoted by x, must satisfy two equations. First, p must equal $m(f_r)$ divided by the proportion of type-θ_h workers applying to f_r:

$$p = \frac{1/4}{(1-\delta)1/2(1+\frac{\delta}{1-\delta}p + \delta x(1-p)(1+p\frac{\delta}{1-\delta}))}.$$

Let $\gamma_h = \frac{1/2(1-p)}{1/2(1-p)+1/2}$ denote an uninformed worker's posterior probability that she is of type θ_h after a single rejection from f_r. In equilibrium, uninformed workers with a single failed application to f_r must be indifferent between applying to f_r and f_s:

$$\frac{\gamma_h p(3 + \frac{\delta 3p}{1-\delta})}{\frac{\gamma_h p}{1-\delta} + (1 - \gamma_h p)} = 2.$$

This completes the characterization of equilibrium for Example 3. As δ goes to 1, the proportion of uninformed agents applying to f_r is bounded above, since uninformed agents do not apply to f_r more than twice. However, as δ increases, there are less uninformed agents present each period, and so the mass of agents applying to f_s must increase in turn. In the limit, the equilibrium outcome converges to that of the corresponding market with complete information. The implications of Example 3 can be shown to generalize.

The equilibrium outcomes of aligned markets converge to the corresponding complete information outcomes as $\delta \to 1$. As δ increases, the opportunity cost of applying to firms with uncertain match values decreases. However, when competition is present, the gain from learning one's type also decreases as more workers attempt to learn their type. As $\delta \to 1$, workers who know they are in the class with the maximal match value must attempt to learn if they have access to that match value. In any market, after a finite number of applications to firms, any worker can learn about their type to arbitrary precision. Then, on a type by type basis, as $\delta \to 1$, outcomes converge to the corresponding outcomes in the complete information case.

A market $\mathcal{M}$ with incomplete preference information can be translated to one with complete preference information. For every worker class c, generate $|c|$ individual classes, in the natural manner. Each type θ belongs to a new class with a degenerate distribution of match values δ_θ and mass $m_c(\theta)$.

Lemma 5 (Equilibrium Outcomes with Long-Lived Workers)**.** *Suppose there are a finite number of worker types, and match values are aligned. For each type-θ, as $\delta \to 1$, the equilibrium probability with which type-θ workers apply to any given firm converges in dis-*

tribution to the probability class-θ workers apply to that firm in the corresponding complete information market.[21]

Despite the fact that the long-lived outcome converges to the complete information outcome, there is a marked difference in worker's information in the two economies. For instance, in Example 3, a minimum of 1/4 type-θ_h workers will never learn their type in equilibrium, for *any* $\delta < 1$. Furthermore, in any equilibrium, a type-θ_h worker that was informed of her type upon arriving would receive strictly higher expected utility than she would if she was uninformed. Both features result directly from oversubscription. For a corresponding complete information market, had the Top-Down algorithm ever encountered oversubscription, workers must apply to firms that are not their most preferred firms with strictly positive probability. This occurs in equilibrium only if workers are not perfectly informed, when $\delta < 1$. Indeed, note that if all workers knew their type, then all would apply to their top choice until the utility from that firm was decreased to that of the next alternative. However, in such an equilibrium in which learning a match value has no benefit, incoming workers would have no incentive to learn their type. Applying to a firm and being rejected, yields a direct loss—not being hired in that given period. By comparison, succeeding fails to change their expected utility, due to the indifference condition. It follows that there must be path dependence in equilibrium for any market with over-subscription.

Lemma 6 (Equilibrium Path Dependence)**.** *In any market with over-subscription, there exists path dependence in equilibrium for any $\delta < 1$.*

Thus even though market outcomes look similar in complete information markets and markets with long-lived workers, the search patterns of workers vary drastically.

[21] Formally, for any $\epsilon > 0$, there exists $\underline{\delta} < 1$ such that for any $\delta > \underline{\delta}$, any type θ, and any firm $j \in \mathcal{F}$, the equilibrium probability a type-θ worker applies to j in any period of the equilibrium with δ is within ϵ of the equilibrium probability that a type-θ worker applies to j in the corresponding complete information market.

1.4.6 Tenure

On the individual level, the effects of incomplete preference information also naturally capture several of the empirical facts mentioned in the introduction. One consistent trend in the data is that higher earners or those who attended better schools are less likely to state that given the chance they would have changed their major or the school they attend (Network, 2017). For instance, while 50% of US adults reported that they would change an important education decision, that number drops to 40% when considering adults at top schools and further drops to 23% when considering adults with incomes above $250,000. Lemma 7 shows that despite having potentially higher outside options, those with higher current matches are more likely to be satisfied with any given match. Let s_θ denote the average stopping age for type-θ workers.

Definition (Stopping Age)**.** *Let $t_s(i)$ be the first period t such that $A_i(t) = A_i(\tau)$ for some $\tau < t$, then $s_\theta \equiv \mathbb{E}_\theta[t_s(i)|i$ is a type-θ worker$]$.*

Lemma 7 (High-Quality Workers Search Less)**.** *Suppose θ_h and θ_l are both types in class c, and θ_h has uniformly higher match values than θ_l; that is, $\theta_h^j \geq \theta_l^j \forall\ f$.*

Then, in equilibrium, $s_{\theta_h} \leq s_{\theta_l}$.

Lemma 7 is a natural consequence of the optimal search policy. Workers apply for the firm with the maximal Gittins index. Since types θ_h and θ_l are members of the same class, c, type-θ_h workers and type-θ_l workers would have identical Gittins indices across firms they have not yet applied to. That is, their option value is identical across firms. However, type-θ_h workers always realize a higher match value than type-θ_l workers. Therefore, if a type-θ_l worker is satisfied with a firm j, then a type-θ_h worker would be satisfied with j as well. Therefore, whenever type-θ_l workers are willing to stop searching at a given firm, type-θ_h workers would also stop searching at that firm. When matches are permanent, this

would manifest as type-θ_h workers reporting higher levels of satisfaction with their current matches, in line with the Gallup report.

1.5 Policy Interventions

I proceed by evaluating the impact of two natural policy interventions designed for markets with incomplete information. I show that congestion alters workers' incentives to search, changing equilibrium outcomes in markets with headhunters and unemployment benefits. While there are many instruments that could be used to increase the quality of matches, I focus on headhunters and unemployment benefits as they are commonly used tools in practice. Furthermore, it is often claimed that these interventions increase workers' propensity to search.

1.5.1 Headhunters

A natural question is the impact of hidden information, are all workers affected equally by the presence of uncertainty? When firms are congested, incomplete preference information benefits workers with high match values. Under incomplete information, workers must apply to multiple firms to find their best matches. This search comes at the cost of time and firings for workers with low match values. Effectively, the ability of workers with low match values to apply to other firms is restricted, leaving more spots open to other workers. Furthermore, even when preferences are aligned, firms' incentives may be in line with certain workers' incentives but not others.

Suppose firm f_r hired a headhunter to target workers with high match values at firm f_r. The headhunter then could send out offers to desirable workers, informing them of their match values at f_r, $(\theta_w^{f_r}, \theta_f^{f_r})$. Workers that were not approached would be able to infer that they had a low match value at firm f_r. To formalize this extension, I define the market with

revelation at firm j as the market induced by all workers learning their match value at firm j.

Definition (Revelation at Firm j)**.** *Let market $\mathcal{M} = (\mathcal{F}, \mathcal{C}, m)$ be given.*

Then, market $\mathcal{M}_j = (\mathcal{F}', \mathcal{C}', m')$ is induced by revelation at firm j if:

1. $\mathcal{F}' = \mathcal{F}$,
2. $\mathcal{C}' = \cup_{c \in \mathcal{C}} \left\{ \cup_{x \in \mathbb{R}^+} \{\theta | ((\theta_w^j, \theta_f^j), \theta_{-j}) = \theta \in c\} \right\}$,
3. $m'_{c'}(\theta') = m_c(\theta')$, $m'(f) = m(f)$.

Condition 2 splits each worker class c into k_c worker classes, where each new class involves a unique match value at firm j. Condition 3 simply requires that the mass function is proportionally distributed across the new classes.

Returning to Example 2, in the market with revelation at firm f_r, the equilibrium is simple: all type-θ_h workers apply to f_r, while class-c_2 workers randomize between the two firms and type-θ_l workers apply to f_s. Firm f_r benefits from this change in outcome, as it receives higher profit than in the original equilibrium. However, revealing information does not always benefit workers with high match values at that firm. Example 4 illustrates how headhunters can benefit workers with low match values to the detriment of other workers.

Example 4 (Revelation)**.**

Workers \ *Firms*	$m(f_r) = 1/4$	$m(f_s) = 1/2$
$m(\theta_h) = 1/2$	$(3, 3)$	$(2, 2)$
$m(\theta_l) = 1/2$	$(1/2, 1/2)$	$(2, 2)$

Revealing information at f_r benefits type-θ_l workers through inducing them to avoid f_r. Type-θ_h workers are harmed through increasing competition at f_s.

When $\delta = 1/2$, equilibrium payoffs for workers of type θ_h and θ_l are $\frac{9}{4}$ and $\frac{5}{4}$, respectively. The equilibrium strategy profile is simple. Workers randomize between the two firms when uninformed, and apply to f_r if type θ_h or to f_s if rejected initially from f_r. In equilibrium, when type-θ_l workers are rejected after applying to f_r, the mass of workers that can apply to f_s is reduced. The low capacity of f_r requires type-θ_h workers to apply to f_s in equilibrium, and so they benefit from the rejections suffered by type-θ_l workers.

Now consider $\mathcal{M}_r$. Type-θ_l workers know they will never be hired by f_r. As such, all type-θ_l workers apply to f_s. In doing so, they decrease the hiring probability at f_s. This weakens the outside option for type-θ_h workers who previously benefited from access to f_s, but now have a reduced safety valve for congestion at f_r. In market $\mathcal{M}_r$, both types' equilibrium payoffs are $\frac{7}{4}$. Type-θ_l workers are made better off by the revelation, while type-θ_h workers are harmed. In this stylized example, complete information harms type-θ_h workers relative to type-θ_l workers because the information itself provides no aggregate increase in efficiency.

In general, firm j hiring a headhunter has two effects; a sorting effect, where workers with high match values at j can immediately apply to j, and a congestion effect, where workers with low match values at j apply to safer options. When j is sufficiently congested, the former effect vanishes, leaving only the latter, benefiting workers with low match values at j on the whole. If j is not sufficiently congested, the result of this policy is ambiguous. Proposition 5 characterizes these two effects, and provides a sufficient condition for less competitive types to be better off than more competitive types.

Two definitions will be useful. First, a market is *congested* if each firm fully utilizes its capacity in equilibrium.

Definition (Congestion)**.** *Firm j is congested, if in equilibrium, $D^j = m(j)$. The market is congested if every firm is congested.*

Second, to economize on notation, let $U^j_{\mathcal{M}}(\theta)$ denote type-θ workers' payoff from firm j in market $\mathcal{M}$:

$$U^j_{\mathcal{M}}(\theta) = \mathbb{E}_{\mathcal{M}}\left[\sum_{t=0}^{\infty} \delta^t \mathbf{1}_{w(t)=j}\theta^j_w\right]$$

Proposition 5 (Impact of Headhunters)**.** *Suppose $\mathcal{M}$ is congested and θ_h and θ_l are members of the same class, but $\theta^j_h > \theta^j_l$.*

1. *The increase in type-θ_h workers' matches with firm j is more than type-θ_l worker's increase:*

$$U^j_{\mathcal{M}_j}(\theta_h) - U^j_{\mathcal{M}}(\theta_h) \geq U^j_{\mathcal{M}_j}(\theta_l) - U^j_{\mathcal{M}}(\theta_l).$$

2. *If $m(\theta^j_h) > m(j)$ and $|C| = 1$, type-θ_h is harmed relative to type-θ_l:*

$$U_{\mathcal{M}_j}(\theta_h) - U_{\mathcal{M}}(\theta_h) \leq U_{\mathcal{M}_j}(\theta_l) - U_{\mathcal{M}}(\theta_l).$$

Part 1 of Proposition 5 shows that workers with high match values at firm j receive higher net utility from firm j. Part 2, whose condition was satisfied in Example 4, states that on net, a worker with high match values at firm j gains less from headhunters when j is heavily congested. There are two forces behind the second result: the nature of the outside option and the independence of match values across firms. In decentralized markets, the cost of a poor match is embedded in the opportunity cost. A centralized market, wherein a rejected worker could immediately apply to another firm within the same period, would eliminate this force. In such a setting, this result would reverse, headhunters would solely benefit workers with high match values at j. Similarly, if types were correlated across firms, that is if $\theta^j_h > \theta^j_l$ implied $\theta^{j'}_h > \theta^{j'}_l$, then the competition effect would vanish. This follows because θ_h would not be effected by the increased level of congestion at f_s.

1.5.2 Unemployment Benefits in Congested Markets

A common intuition holds that increasing unemployment benefits further encourages search, because it reduces the loss from rejection. In turn, unemployment benefits could be expected to improve equilibrium efficiency as workers and firms are better matched. $\theta_w^{\varnothing}$ has a straightforward interpretation in this context, $\theta_w^{\varnothing}$ is the benefit for rejected workers, or the "unemployment benefits." This section considers the impact when the utility from being rejected in a given period is positive. Example 5 shows that unemployment benefits can have a nuanced effect on equilibrium efficiency. *Negative learning*, a worker's incentive to learn about a less desirable firm, is suppressed by unemployment benefits.

Example 5 (Unemployment Benefits)**.**

Workers \ *Firms*	$m(f_r) = 1/4$	$m(f_s) = 1/4$
$m(\theta_h) = 1/2$	$(3, 3)$	$(3, 3)$
$m(\theta_l) = 1/2$	$(1, 1)$	$(3, 3)$

As depicted in Figure 1.1, increasing unemployment benefits when there is an agreed upon top firm can decrease equilibrium welfare.

Safety nets can reduce the overall efficiency of equilibrium. Indeed, when the rate of applications for uninformed workers to f_r drops below $1/4$ in Example 5, the total equilibrium utility, net of $\theta_w^{\varnothing}$, is reduced. The key feature in Example 5 is that f_s is commonly known to be more attractive than f_r. Prior to the introduction of unemployment benefits, workers refrain from applying to f_s due to the high level of competition and resulting low probability of acceptance. As $\theta_w^{\varnothing}$ increases, the loss from rejection decreases. The lowered loss incentivizes more workers to apply to f_s, in order to equilibriate the probability of rejection. However, this reduces equilibrium efficiency when δ is low. The increase in applicants to f_s fails to improve matching, while too few type-θ_h workers are hired by f_r.

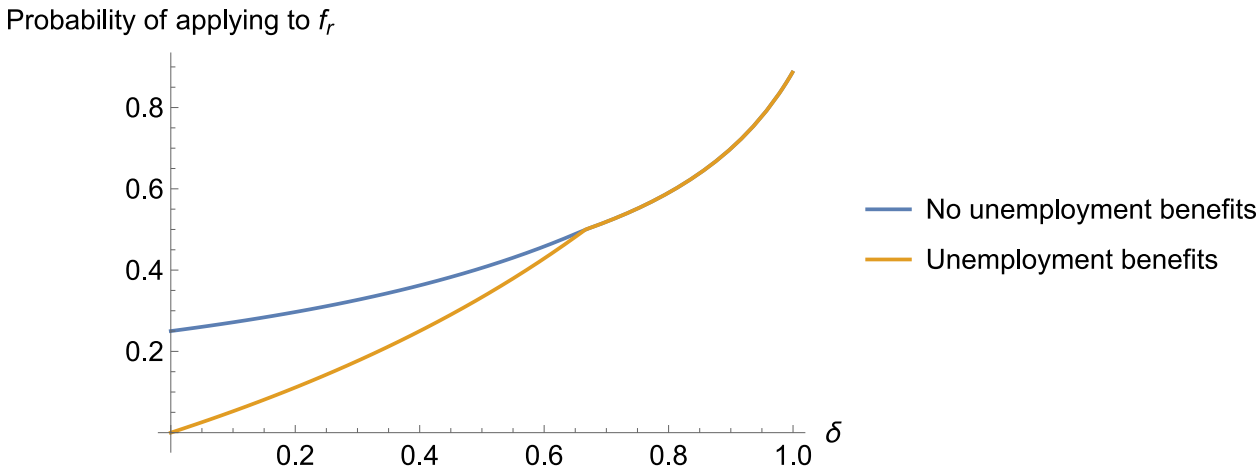


Figure 1.1: The x-axis represents δ, the probability a worker does not retire. The blue line shows the proportion of uninformed workers that apply to f_r when $\theta_w^{\varnothing} = 0$. The yellow line shows the proportion of uninformed workers that apply to f_r when $\theta_w^{\varnothing} = 1$.

1.6 Transferable Utility

I show transferable utility can be simply incorporated into the transient matching model. Furthermore, doing so does not cause worker demand to violate gross substitutes, and therefore does not change the qualitative results. I consider a competitive equilibrium environment where firms strategically choose wages, before workers apply to firms.

1.6.1 Competitive Equilibrium

In a competitive equilibrium, firms choose wages based on observables—match values and classes—to maximize profits. Formally, I consider a separate "wage game," which occurs before the market resolves in period 0. In the wage game, all firms simultaneously announce a *payment function* $\phi_j : \mathcal{C} \times \mathbb{R}_+ \to \mathbb{R}$, denoting the payment from firm j to a worker conditional on that worker's class, c, and firm match value, θ_f^j. Importantly, workers of a given class and equal match values must receive the same payment, even if their type differs.

To maintain the connection with the previous sections, I assume the prenumeration match values remain unchanged. If a class-c type-θ worker matches to firm j, the two agents receive

θ_w^j and θ_f^j, respectively, in addition to the transfer. That is, they receive $\theta_w^j + \phi_j(c, \theta)$ and $\theta_f^j - \phi_j(c, \theta)$. After period 0, wages remain fixed, and workers apply to firms as before, with match values altered accordingly.

Importantly, firm j evaluates his profit through his per-period expected equilibrium profit in the market game, denoted by $\pi_j(\{\phi_f\}_{f \in \mathcal{F}})$. This is in contrast to the previous sections, where firms evaluated profits through their discounted stream. To see why this is necessary, recall the previous discussion of alignment. Alignment was used to rule out the possibility of rejection cycles, wherein a firm might reject one worker in order to receive an application from a preferred worker. Here, with arbitrary transferable utility, there is no assurance that the resulting market will be aligned. To retain the focus on workers' search strategies, firms are required to be myopic in evaluating utility.

I assume firms are myopic—when hiring, firms maximize profit in the current period—and firms randomize when indifferent. This rules out firms incentivizing more or less experimentation through selectively hiring certain worker classes. When each firm is made up of many small teams, each of which must make a hiring decision, this assumption is natural. Each individual team has a small impact on the overall matching and so focuses on maximizing their own utility. Alternatively, if each "firm" were a unique industry comprised of many firms, similar logic would imply that each firm optimizes by maximizing the current period utility.

Assumption (Myopic Firms)**.** *Firms maximize current period profit. Furthermore, if a firm is indifferent among a subset of its applicants, and cannot hire the entire subset, it uniformly randomizes over that subset.*

Wages are required to satisfy a limited liability condition—the size of any transfer cannot exceed the gain from the matching for either party:

Assumption (Limited Liability)**.**

$$-\theta_w^j \leq \phi_j(\theta) \leq \theta_f^j$$

I proceed by defining *competitive equilibrium* in the standard manner. A profile of payment functions constitutes a competitive equilibrium if no firm can change his payment function to increase his profit:

Definition (Competitive Equilibrium)**.** *A profile of payment functions* $\{\phi_j\}_{j\in\mathcal{F}}$ *constitutes a competitive equilibrium if for all* j, ϕ*:*

$$\pi_j\left(\{\phi_k\}_{k\in\mathcal{F}}\right) \geq \pi_j\left(\phi, \{\phi_{j'}\}_{j'\neq j}\right)$$

This equilibrium characterization implicitly prohibits dynamic punishments. Allowing for dynamic punishments would enable a folk-theorem style argument which could rationalize all wages when firms are sufficiently patient.

Importantly, for any profile of payment functions, the resulting market is effectively a new market with non-transferable utility. Then, workers still evaluate firms as if they were endogenous bandits, albeit with differing rewards. This implies that aggregate demand satisfies gross substitutes, and therefore an equilibrium can be characterized using the algorithm in Theorem 1.

Proposition 6 (Wages Satisfies Gross Substitutes)**.** *For any profile of wages* $\{\phi_j\}_{j\in\mathcal{F}}$*, aggregate demand satisfies gross substitutes.*

1.6.2 Non-Existence of Competitive Equilibrium

Even in such a simple environment, existence of competitive equilibria is not guaranteed. Two issues arise: a type's demand for a firm is not necessarily continuous in wage, as workers

can apply multiple times; and firms cannot condition wages on worker's outside options, as they do not observe worker's types.

Example 6 (Competitive Equilibria Need Not Exist)**.**

Workers \ *Firms*	$m(f_r) = 1/2$	$m(f_s) = 2$
$m(\theta_h) = 1$	$(3, 3)$	$(2, 2)$
$m(\theta_l) = 1$	$(1/2, 1/2)$	$(2, 2)$

First, note that the minimum proportion of workers f_s can hire is given by the strategy profile where all workers initially apply to f_r, then type-θ_l workers subsequently apply to f_s. Note that if a type-θ_l worker is willing to remain at f_r after learning her type, then all type-θ_h workers must strictly prefer to remain at f_r. Furthermore, this implies that an uninformed worker must strictly prefer to work at f_r. However, then the mass of workers applying to f_r is greater than $1/2$, so type-θ_l workers are hired with probability 0. Type-θ_l workers would then prefer to apply to f_s, unless f_s's wage is exactly -2, in which case type-θ_l workers are indifferent. By standard bargaining arguments, this implies that $\phi_s = -2$ and all type-θ_l workers apply to f_s. Then, a lower bound for π_s in equilibrium is given by $1/2\delta \cdot 4 = 2\delta$.

To show no competitive equilibrium exists in Example 6, suppose for contradiction a competitive equilibrium exists. There are three possible categories of strategies for workers. 1) All workers apply to f_s forever; 2) all workers initially apply to f_r, then type-θ_h workers remain at f_r and type-θ_l workers migrate to f_s forever; and 3) workers randomize between the previous two options.

First, suppose in the competitive equilibrium, all workers applied to f_s forever. Then, f_r would hire no workers, and so receive zero profit. f_r would deviate and set payment for θ_h equal to $6 - \epsilon$, since this would generate positive profit for f_r. However, f_s would never offer a wage higher than $2\delta = (2 - w)2$ or $w = 2(1 - \delta/2)$ in equilibrium, as otherwise f_s could

revert to a wage of -2. Since $2(1-\delta/2) < 6$ at least one of f_r or f_s must have a profitable deviation.

Next, suppose all workers initially apply to f_r. Then, f_s must set its wage to -2 as listed above. However, if f_r's wage was above -3, f_r could deviate to a wage of $-3+\epsilon$ while still attracting type-θ_h workers. This implies f_s could set a wage just above ϕ_r, a marginal increase in payment that earns f_s a strict increase in workers of $1-1/2\delta$.

Last, workers randomized between the two firms. Then, workers are exactly indifferent between the two firms. However, a firm could increase its wage marginally to capture all workers. This generates a strict increase in profit, unless that firm was earning zero profit from each worker. For f_r this implies $\phi_r = 3$, while for f_s this implies $\phi_s = 2$. However, then f_s could deviate to $\phi_s = -2$ as detailed above.

The driving force behind the non-existence argument is that firms cannot distinguish workers with different outside options. This reduction in the dimension of possible wages restricts the ability to generate an equilibrium.[22]

1.6.3 Resumes

The non-existence in the previous example can be resolved by the inclusion of resumes. In this context, a resume allows a worker to prove she had been previously hired by a separate firm, in return for higher wages. Resumes allow for wages to be conditioned not only upon the direct match value, but also upon a worker's previous history. Formally, a *resume-dependent payment function* is a payment function defined on a larger domain $\phi_j : \mathcal{C} \times \mathbb{R}_+ \times \{A_i^j(\tau)\}_{\tau=0}^t \to \mathbb{R}$.

[22]It is worth noting that while on the surface this may appear similar to issues such as the non-existence that arises from adverse selection in Rothschild and Stiglitz (1978), the non-existence here comes from a different source. In their work, non-existence results from free firm entry, as firms currently in the market are harmed by entry. In transient markets, the firms already present in the market face incentives to change their pricing structures, in order to manipulate the learning behavior of workers in the market.

To see how the inclusion of resumes generates a competitive equilibrium in Example 6, note that now f_s could choose a wage that incorporates the worker's outside option. The following wages form equilibrium for sufficiently high δ: type-θ_l workers receive 0 from f_r and -2 from f_s, while type-θ_h workers receive 1 from f_r and 2 from f_s upon providing a resume and -2 otherwise. In equilibrium, all workers apply to f_r initially, and type-θ_l workers transition to working at f_s.

The equilibrium non-existence arose due to the requirement that firms treat workers of a single class with equal match values identically. Through introducing resumes, workers could prove that their outside options were distinct. Thus, resumes play an important role in information transmission.

1.7 Discussion

This paper develops a framework for analyzing transient matching when workers learn through experimentation. Firm capacity constraints force workers to anticipate other workers' application decisions. I show that, firms can be evaluated as endogenous bandits. Once firms' hiring decisions are described as thresholds, techniques from the multi-armed bandit literature allow for a simple description of the optimal worker policy. Importantly, aggregate demand satisfies the gross substitutes condition, which enables the characterization of equilibrium. Workers' search patterns match data from labor markets, high-quality workers report higher satisfaction—despite not having better information.

I show that the nature of both learning and competition are critical to understanding the impact of policy interventions. Commonly used interventions, such as hiring headhunters or increasing unemployment benefits, may generate unintended effects in congested markets. For instance, headhunters may redistribute the benefits from matching, while increasing unemployment benefits can intensify competition.

The results also imply careful consideration should be taken before changing centralized mechanisms. Such mechanisms often feature decentralized aftermarkets, where the incoming information from the original centralized market can radically shift the final outcome. This paper provides a first step towards better understanding the effects changes of mechanism rules can have on the aftermarkets of centralized mechanisms.

Appendix

1.A Proofs

I begin by showing that standard results from the operations literature apply when firms hire all applicants. Lemma 8 provides the basis for the index-related arguments in the rest of the appendix.

Lemma 8. *Let σ be an equilibrium, such that, in any subgame, all firms hire all applicants each period. In any period, each worker i applies to some $j \in \arg\max_j GI_i^\sigma(j, h_t^i)$.*

Proof of Lemma 8

This proof follows directly from Theorem 2.1 of Gittins et al. (2011) which states:

Theorem 2.1: *A policy for a simple family of alternative bandit processes is optimal if it is an index policy with respect to the Gittins index of each bandit process.*

Then, it must be shown that the set of firms acts as a simple family of alternative bandit processes for each worker. First, since firms hire all applicants, each worker i faces a fixed decision problem for any strategy profile σ, independent of the other workers' strategies. For any possible worker strategy profile, each worker is always hired, no matter which firm she applies to. As such, the payoff for a type-θ worker from applying to firm j, is simply θ_w^j in any equilibrium.

Each firm is then a bandit process that can either be activated or frozen. A frozen firm provides no payoff, while an activated firm provides θ_w^j. Each worker must activate exactly

one firm each period. The independence of firm match values—conditional on a worker's class—implies that the set of firms is a simple family of alternative bandit processes. Theorem 2.1 from Gittins et al. (2011) then applies, and the result follows. ■

Proof of Lemma 1

I begin by describing the Top-Down algorithm in detail. For any complete-information market $\mathcal{M}_I = (\mathcal{F}, \mathcal{C}, m)$, the Top-Down algorithm proceeds as follows:

Top-Down Algorithm:

Let $\mathcal{E} = \mathcal{F} \cup \mathcal{C}$ be the set of agents present in the market.

1. Find the largest worker-firm match value: $(c^*, j^*) = \arg\max_{c,j \in \mathcal{E}} \theta_c^j$, the finiteness of $\mathcal{C}$ implies this is well-defined.

2. Find class-c^*'s second-highest match value: $j_2 = \arg\max_{j \neq j^*} \theta_{c^*}^j$.

3. Match c^* to j^*, which leads to one of two possible outcomes:

   (a) Undersubscription—When all class-c^* workers apply to j^*, j^* is still preferred by class-c^* workers relative to j_2:

   $$\frac{m(j^*)}{m(c^*)} \theta_{c^*}^{j^*} \geq \theta_{c^*}^{j_2}$$

      i. All class-c^* workers apply to j^* forever.

      ii. Firm j^*'s capacity is reduced by the mass of applicants: $m(j^*) = \max\{0, m(j^*) - m(c^*)$.

      iii. Reduce to the submarket without class-c^*, $\mathcal{E} = \mathcal{E} \setminus \{c^*\}$. Remove j^* from the market as well if $m(j^*) = 0$.

   (b) Oversubscription—When too many class-c^* workers apply to j^*, j_2 becomes preferable:

   $$\frac{m(j^*)}{m(c^*)} \theta_{c^*}^{j^*} < \theta_{c^*}^{j_2}$$

i. c^* randomizes between j^* and j_2 to generate indifference: c^* applies to j^* with probability $m(j^*)\frac{\theta_{i^*}^{j^*}}{\theta_{c^*}^{j_2}}$. The remaining probability, $1 - m(j^*)\frac{\theta_{i^*}^{j^*}}{\theta_{c^*}^{j_2}}$, will be accounted for in a future step.

ii. Reduce to the submarket without j^*, $\mathcal{E} = \mathcal{E} \setminus \{j^*\}$.

4. In the new submarket $\mathcal{E}$, again select the highest match value: $(c^*, j^*) = \arg\max_{c,j \in \mathcal{E}} \theta_c^j$.

   (a) If j^* was not previously selected, repeat steps 2 and 3.

   (b) Otherwise, if c^* was previously oversubscribed:

      i. The proportion of c^* applying to the original firm must be increased to keep c^*'s payoffs from both firms equal.

      ii. Redistribute workers between the two (or more) firms keeping c^*'s payoff from each firm equal, continuing until all of c^* has been allocated, or the payoffs from each firm are equal to those of the next top choice of c^*.

   (c) If j^* was previously selected as j_2, increase the proportion of the previous class-c^* applying to the previous j^* to equalize the previous c^*'s utility between the previous j^* and the current j^*.

5. Repeat the above steps until either $\mathcal{E} \subset \mathcal{C}$ or $\mathcal{E} \subset \mathcal{F}$

6. Any remaining workers apply to any firm and are rejected. All remaining firms operate below capacity.

Now I prove the claim. Suppose there exists a strategy profile σ' whose outcome does not coincide with the Top-Down algorithm. The Top-Down can be used to construct a strategy profile σ. Because σ' and σ differ in their outcome, there is a minimum k such that in the kth iteration, there exists a class c, such that c's strategy profile differs between σ and σ'. By construction, this implies that some member of class c has payoff under σ' below its payoff under σ. Furthermore, all matches corresponding to higher match values are equivalent

between σ and σ' by construction. Then, under σ', a class c worker could deviate to their firm under σ and receive higher expected utility every period. Therefore, σ' could not have been an equilibrium. ■

Proof of Proposition 1

Consider a firm j that is hiring below capacity, namely j rejects applicants while $m(A_j) < m(j)$. Importantly, rejecting some such applicant, i, has two impacts that could potentially benefit j: the rejection could cause i to re-apply to j in a later period, and the rejection could trigger a rejection cycle wherein a more preferable worker i' is rejected by another firm causing i' to apply to j.

Hiring i immediately front-loads the match value from i, avoiding the loss from i possibly retiring before returning to j. Note that it cannot be the case that i eventually returns to j and applies more times to j than i would if j had accepted immediately. Eventually i's information sets under the original strategy and the deviation must coincide, at which point the Markovian nature of i's strategy forces i to apply to the same firm under both strategy profiles. Then, it follows that j's payoff from that point forward is unchanged. Since i was accepted by j at most once before i's information converged under the original strategy profile, j strictly benefits from the deviation.

By assumption, match values are aligned. It is known that aligned markets have no rejection cycles (Voorneveld and Norde, 1997). Namely, it cannot be the case that $\theta_i^{j'} > \theta_{i'}^{j'}, \theta_i^j > \theta_i^{j'}$, and $\theta_{i'}^{j'} > \theta_{i'}^j$. Then, j can not benefit from rejecting i in an attempt to attract other workers. Since j cannot benefit regardless from rejecting i, j must accept as many applicants as possible in equilibrium.

Last, if j is congested, then it must have excess applications with match values equal to its lowest match value. In order to accept a high ranked applicant, j would then need to reject one of its lowest ranked applicant. However, even rejecting that lowest ranked applicant dissuades her from applying in future periods, another worker with equivalent match value

will be available to replace it. Then, the previous arguments imply that j has a profitable deviation.

■

Proof of Lemma 2

A Markovian strategy for firm j, is a mapping from the types of workers, a_j, to acceptances. Notably, because worker's strategies are Markovian, worker's payoff-relevant information is summarized by her belief regarding her type. For a given worker j, and private history h_t^j, let j's posterior regarding her type be given by $p \in \Delta\Theta$. Then, the payoff-relevant state space is a distribution over p, Ψ.

Consider the Markov chain over Ψ. Workers' applications follow the same initial distribution over applications. Furthermore, there are a finite number of firms, and as such workers must eventually converge in belief, independent of the order in which they apply to firms, implying that their final applications are a fixed quantity. Then, there exists a subset $V \subset \Psi$, such that V is irreducible. Furthermore, because new workers enter the market every period, the Markov chain is aperiodic. Then, by the Steady State Theorem, there exists a steady state. ■

Proof of Lemma 3

Suppose for worker i, under an arbitrary strategy profile σ, for some firm j, $GI_i^\sigma(j, h_t^i) = x$. I show that this value of $GI_i^\sigma(j, h_t^i)$ is achieved by the strategy described in the statement of the lemma: if the realization of ϕ_i^j is above x set $\tau = \infty$, otherwise set $\tau = 1$. If exactly x was realized, any stopping time yields an equivalent outcome. Assume throughout that the realization is given by $y > x$. It will be helpful to refer to the value of the stopping time problem that characterizes $GI_i^\sigma(j, h_t^i)$. I let $gi(f, h_t^i, \tau)$ be the value of the stopping time problem that characterizes $GI_i^\sigma(j, h_t^i)$, given a possibly suboptimal stopping time, τ. Formally:

$$gi(f, h_t^i, \tau) \equiv \frac{\mathbb{E}_i \left[\sum_{t=1}^{\tau} \delta^t \phi_i^j\right]}{\mathbb{E}\left[\sum_{t=1}^{\tau} \delta^t\right]}$$

I use induction to show that, for any stopping time $\tau < \infty$, $gi(f, h_t^i, \tau + 1) > gi(f, h_t^i, \tau)$.

To begin, I consider the case where $\tau = 1$:

Let $x = p/q$ where $q > 0$, p and q are not necessarily integers, no rationality assumption is made.

$$\begin{aligned}
\frac{p + \delta y}{q + \delta} &> p/q \\
p + \delta y &> \frac{p(q + \delta)}{q} \\
\delta q y &> \delta p \\
q y &> p \\
y &> p/q = x
\end{aligned}$$

Next, suppose that the claim holds for $\tau \in \{1, 2, \ldots, k\}$. A similar computation implies the inductive step also holds for $\tau + 1$.

$$\begin{aligned}
\frac{p + \delta^{k+1} y}{q + \delta^{k+1}} &> p/q \\
\implies y &> p/q = x
\end{aligned}$$

Then, for any k, $\frac{p+\delta^k y}{q+\delta^k} > p/q$, therefore $gi^\sigma(f, h_t^i, \tau) = GI_i^\sigma(j, h_t^i)$ only if $\tau \in \{1, \infty\}$.

Conversely, the argument also shows that when $y < x$, setting $\tau = 1$ is optimal.

Then, the claim is proven, $GI_i^\sigma(j, h_t^i)$ is characterized by the strategy described in the statement of the lemma. ■

Before proving Proposition 2, I prove a useful lemma. Strict dynamic preferences implies that two equilibrium strategy profiles that induce different outcomes must also generate different hiring thresholds.

Lemma 9. *Under strict dynamic preferences, if σ and σ' generate distinct equilibrium outcomes, then there must exist some firm j whose hiring threshold under σ differs from his hiring threshold under σ'.*

Proof of Lemma 9

Suppose not. That is, σ and σ' induce different outcomes, but every firm has identical thresholds under σ and σ'. Then, every worker faces identical Gittins indices under σ and σ', at every firm, under any informational partition. In particular, since σ and σ' have distinct outcomes, there exists a type θ and history h_t^i such that type-θ workers with history h_t^i make different choices under σ and σ', and type-θ workers are hired with positive probability. Without loss of generality, suppose type-θ workers apply for firm j with greater probability under σ than under σ'. Similarly, type-θ workers must apply to another firm j' with greater probability under σ' than σ. Since forward induction policies are optimal, this implies that $GI_i^{\sigma}(j) = GI_i^{\sigma'}(j')$. Then, strict dynamic preferences implies that type-θ workers must be hired with intermediate probability at either firm j or firm j'. Namely, the threshold for either j or j' is set at a match value of θ_f^j or $\theta_f^{j'}$.

First, suppose type-θ workers were hired with intermediate probability at j. As type-θ workers leave j, if no other workers start applying to j, the probability of being hired increases for remaining type-θ workers due to the reduced competition, thereby increasing $GI_\theta^{\sigma'}(j)$ contradicting the claim. To prevent this, there must be another type θ', whose workers apply to firm j in increased numbers. However, type-θ' workers willingness to do so implies that there exists another firm j'' such that $GI_{\theta'}^{\sigma}(j) = GI_{\theta'}^{\sigma}(j'')$. Again, by assumption 1 workers of this type must be hired with intermediate probability at one of these firms. Since type-θ workers were hired with intermediate probability at firm j, type-θ' workers

must be hired with intermediate probability at firm j''. Repeating this line of logic implies that there exists a cycle of workers, each facing equal Gittins indices at least two firms and hired with intermediate probability at one such firm. However, inherently the total mass of workers is fixed, for every mass of workers that leave a firm, an equal mass must take their place. But marginal workers have a probability less than 1 of being hired, while those taking their place do not. Then, such a cycle cannot keep the total amount of workers hired at involved firms equal. Therefore, the thresholds at those firms must either increase or decrease, contradicting the original assumption. ■

Proof of Proposition 2

Suppose σ and σ' are both equilibria, with distinct outcomes. Lemma 9 then implies that there exists some firm j such that firm j's threshold under σ is distinct from its threshold under σ'.

Without loss of generality, suppose firm j's threshold is higher under σ than under σ'. Then, consider worker i, where i previously applied to a firm j' but now applies to j. A positive mass of such workers must exist, otherwise j's threshold could not have increased. One of two cases must have occurred to generate this change in application, either worker i's original firm's Gittins index decreased, $GI_i^{\sigma'}(j', h_t^i) < GI_i^{\sigma}(j', h_t^i)$, or worker i's Gittin index at their new firm increased, $GI_i^{\sigma'}(j, h_t^i) > GI_i^{\sigma}(j, h_t^i)$. However, $GI_i^{\sigma'}(j, h_t^i)$ cannot be larger than $GI_i^{\sigma}(j, h_t^i)$ because firm j's threshold has increased, therefore every worker has a weakly lower firm j Gittins index than before. Then, $GI_i^{\sigma'}(j', h_t^i) < GI_i^{\sigma}(j', h_t^i)$, implying that the threshold at firm j' has increased.

Then, the same line of logic as before implies that there exists another non-zero mass of workers applying to firm j', and their original firm must have increased its hiring thresholds. Since there are a finite number of firms, eventually this chain must return to a firm it has previously visited, generating a set of firms with higher thresholds under σ' relative to σ.

Consider the implications of such a set, it cannot be the case that a smaller total of workers are being hired at these firms under σ, otherwise thresholds would fall accordingly. Furthermore, since thresholds are higher, the total utility of hired workers must be higher under σ relative to σ'. Then, there must be at least one worker with strictly greater utility than before. Then, such a worker could have profitably deviated in σ' to its strategy under σ. However, this contradicts the fact that σ' was an equilibrium. ■

Proof of Lemma 4

Lemma 8 showed that in equilibrium workers must apply to a firm with maximal Gittins index when capacities were unconstrained. Repeating the proof of Lemma 8, updating the rewards from firm j, replacing θ_w^j with $\psi_\theta^j(v,p)$ immediately implies that the set of capacity constrained firms is still a simple family of multi-armed bandits. The result follows. ■

Proof of Proposition 3

I begin by proving a stronger claim for individual demand. I show that when whenever any type-θ worker faces a multi-armed bandit problem, and type-θ worker's Gittins index for any firm j, $GI_\theta(j)$, is decreasing in j's threshold, (v_j, p_j), type-θ's demand will be decreasing in j's threshold as well. Gross substitutes is satisfied in multi-armed bandit problems because the Gittins index for each individual firm is only a function of the rewards from that firm. Then, increasing the threshold of firm j_2 cannot affect the Gittins index of firm j_1. Furthermore, the realized demand over each firm is weakly increasing in that firm's Gittins index. As such, raising the thresholds of a set of firms, decreases the Gittins indices of those firms, but fails to change the Gittins indices of other firms. Then, applications to the original set of firms must weakly decrease.

Theorem 2 (Bandits imply Gross Substitutes)**.** *Suppose for all θ, type-θ workers face a simple multi-armed bandit problem, and for any firm j, $GI_\theta(j, h_t^i)$ is decreasing in (v_j, p_j). Then, type-θ's demand satisfies gross substitutes.*

Proof of Theorem 2

Lemma 8 implies that, when firms hire all workers, workers apply to a firm with maximal Gittins index each period. However, for any firm j and threshold (v_j, p_j), an auxiliary firm j' can be defined, where type-θ's match values are given by $\psi_\theta(v_j, p_j)$. Then, j' hires all workers, and so Lemma 8 applies.

Then, let $\mathbf{F}$ be a set of firms, $\mathbf{F} \subset \mathcal{F}$, and $(v, p), (v', p')$ be two vectors of thresholds, such that $(v_j, p_j) = (v'_j, p'_j)$ for $j \in \mathbf{F}$ and $(v_j, p_j) \geq (v'_j, p'_j)$ otherwise. Because the thresholds for firms in $\mathbf{F}$ remain unchanged, their Gittins indices are also equal under (v, p) and (v', p'). However, the Gittins indices for firms outside of $\mathbf{F}$, must be weakly lower than before under (v, p). Then, the total demand for firms in $\mathbf{F}$ must weakly increase, as the corresponding Gittins indices are always higher in relative terms.

■

Returning to the proof of Proposition 3, it must be shown that for a given set of thresholds, aggregating across workers preserves changes in demand. This immediately follows from monotonicity of integration. By the definition of aggregate demand, for a given firm $j \in \mathbf{F}$, $D_j(v, p) = \int_\theta D_j^\theta(v, p) d\theta$. Then, if for all θ, $D_j^\theta(v, p) \geq D_j^\theta(v', p')$, it must be that $D_j(v, p) \geq D_j(v', p')$. As such, when every individual type's demand satisfies gross substitutes, so must aggregate demand. The result follows. ■

Before addressing Theorem 1, I prove the following helpful lemma:

Lemma 10. *Suppose after some private history, h_t^θ, which occurs with positive probability for type-θ workers; type-θ applies to j with positive probability, $v_j = \theta^j$, and $GI_\theta(j, h_t^\theta) = GI_\theta(j', h_t^\theta)$. Then, a marginal decrease in p_j yields a discontinuous decrease in D^j.*

Proof of Lemma 10

Forward induction implies a form of independence of irrelevant alternatives, reducing $GI_\theta(j)$ below $GI_\theta(j')$ only affects type-θ workers' choices regarding whether to apply to j or

j'. In particular, there is exactly one change, type-θ workers that would have applied to j apply to j' instead. Afterwards, type-θ workers will either remain at j' or apply next to j.

Even if type-θ workers later return to j, a period by period comparison shows that type-θ workers' demand for j has decreased. In the first period, the number applying to j is zero, since all have applied to j'. Then, type-θ workers either remain at j' forever, in which case the claim follows, or they proceed by applying to j. However, even should type-θ workers apply to j, upon having applied to both firms, the impact of decreasing $GI_\theta(j)$ has washed out. Type-θ workers now have applied to both firms. Subsequently, type-θ workers will be weakly more likely to apply to firm j'. Last, since $\delta < 1$ front loading applications to firm j' decreases the total number of type-θ workers applying to firm j. ■

Proof of Theorem 1

I begin by describing the threshold adjustment process in more detail:

Threshold adjustment process:

1. Begin by setting all thresholds to $(v_j, p_j) = (0, 1)$.
2. Determine the aggregate demand vector, D.
3. If no firms are hiring beyond their capacities; $\forall j, D_j \leq m(j)$, terminate.
4. Otherwise, take an arbitrary firm j that is over capacity, and impose its capacity continuously:
   (a) Select the worst match quality θ^j for firm j that is hired with positive probability. Let $\underline{v}_j = \min_{\{\theta|\theta^j \geq v_j\}} \theta^j$ and let c be the class of $\theta \in \arg\min_{\{\theta|\theta \geq v_j\}} \theta^j$.
   (b) Continuously decrease θ's hire probability at j, $p_j \to 0$.
   (c) Three events could result:
      i. Lowering p_j equates demand and capacity, $D^j = m(j)$, continue to step 5.

ii. Type-$\underline{v}_j$ workers are never hired, $p_j = 0$, return to step 4a and repeat.

iii. Lowering p_j equalizes c's Gittins indices at j and at least one another firm j', $GI_i(j) = GI_i(j')$. Increase type-θ workers' demand for j' at this information set while lowering their demand for j until the demand for j is 0 or $D^j = m(j)$. Proceed to step 4a or 5 accordingly.

5. Again select a firm j such that $D^j > m(j)$, if no such j exists, terminate. Otherwise, repeat the process in step 4 with the following change. In step 4(c)iii, if j' is a firm that was not selected in a previous step, then repeat step 4(c)iii without alteration. However, if j' was selected in a previous step, the above method leads to cycles. To see why, note that Lemma 10 implies any non-zero decrease in p_j may cause demand for j and j' to oscillate. Modify the procedure by selecting all firms j' with equal Gittins indices that were previously selected, and simultaneously lower all corresponding $p_{j'}$s along with p_j in proportions such that $GI_i(j) = GI_i(j') \forall j'$ while also reallocating demand between the set of firms accordingly such that no firm in the set is left with excess capacity.

   This process must end either through all selected firms equating demand and capacity or with some $p_{\tilde{f}} = 0$. In the first event, return to the beginning of step 5, in the latter event, step 4(c)iii can be resumed as normal.

By design, the threshold adjustment process always *increases* the thresholds of various firms. Then, the gross substitutes condition implies that unselected firms never have their demand decrease under the process. Furthermore, step 5 ensures that previously selected firms never face over-demand. Firm j's step only concludes when $D^j = m(j)$. The only unselected firms are those such that $D^j(0, 1) \leq m(j)$. Then, the procedure must conclude within a finite number of iterations, at most one per firm. Upon termination, all firms have demand equal to supply. By construction, worker incentives are incorporated through the

Gittins index and threshold characterization, therefore the procedure finds an equilibrium, concluding the argument. ■

It is worth emphasizing that the threshold adjustment process converges due to the optimality of forward induction. In particular, had type values been correlated between firms, increasing the threshold of one firm could *decrease* demand at another firm. Alternatively phrased, independence of types is a key component of gross substitutes. Gross substitutes requires that the demand for a firm is weakly increasing in the prices (thresholds) of other firms. Gross substitutes proves to be a weaker sufficient condition in order for firm thresholds to characterize equilibrium. Notably, as shown in Appendix section 1.B.1, a model of gradual learning—where instead of receiving θ_w^j from a match, a worker received a noisy signal of θ_w^j—would also converge to equilibrium under the threshold adjustment process as the gross substitutes condition would still be satisfied.

Proof of Lemma 5

To prove the claim, I first characterize the equilibrium strategy profile in the limit as $\delta \to 1$. I then show that the resulting outcome coincides with the equilibrium outcome in the market with complete information. As $\delta \to 1$, the maximum possible match value must be fully utilized; either the associated firm reaches capacity, or by the associated worker type fully applies to the firm. Otherwise, for sufficiently high δ, a worker of the class with the maximal match value would benefit from applying immediately to the firm, and learning their match value. Then, an extension of the Top-Down algorithm can be used to characterize long-lived workers' strategies.

One key alteration is necessary. Since, workers have incomplete information, rather than specifying the application strategy of a given match value, the algorithm now specifies the application strategy for an entire class. Then, agents no longer behave identically every period. Instead, workers of a given class follow a descending chain of applications, hunting for their top match value. When doing so, the mass of each type within a class must decrease

by $1-\delta$ every step as workers retire along the way. However, as $\delta \to 1$, this decrease in the mass of the overall class collapses to zero, and therefore does not affect the limiting outcome.

To proceed, consider the previous Top-Down algorithm, where workers now follow the strategy for their entire class, stopping searching once they have found their proscribed match value. Notably, when a firm's capacity is not exhausted in a given iteration, workers immediately learn if they would achieve the relevant match value upon applying once. Either they are accepted and learn their actual match value, or they are rejected and learn that they do not have the match value specified in that step of the procedure. If a firm's capacity were to be exhausted for a given iteration, it would be possible for a rejected worker to have access to the associated match value, but simply be unlucky. Then, equilibrium behavior may require that workers of such a class apply multiple times to the same firm. Similar to the complete information setting, one of two cases results, either oversubscription or undersubscription.

Type-convergence holds inductively. Note that there are a finite number of steps, and each step differs by a continuous function of δ. First, the top type converges as $\delta \to 1$. Then, the distance between the outcomes of $\mathcal{M}$ and $\mathcal{M}^I$ is a continuous function of δ. When types are discrete, the rest converge immediately as well. ■

Proof of Lemma 6

At the first step in which oversubscription occurs in the algorithm of Lemma 1, select the associated type-θ^* and firms j^* and j_2. Type-i^* workers then either apply to j^* or j_2. Those that fail to apply to j^* will never return, while those who apply to j^* are rejected with positive probability. This generates path dependence in equilibrium. ■

Proof of Proposition 5

To begin, note that congestion implies that the total number of workers hired does not change under information revelation. In particular, market congestion implies that several

workers who applied to firm j under $\mathcal{M}$ were rejected. Furthermore, the rejected workers must have weakly lower match values than the workers who are hired. If θ_h was maximally hired at firm j, then the total payoff for θ_h at firm j cannot increase. However, since θ_l must decrease its level of applications to firm j relative to θ_h, θ_l must apply in larger quantities to firms other than j. $|\mathcal{C}| = 1$ requires θ_h and θ_l to share a class. In expectation, their match values at other firms are equivalent. As such, the net effect is a reduction in the proportion of θ_l rejected at j, increasing the congestion at the remaining set of firms. This causes the average payoff of workers at those firms to decrease. This argument proves parts 1 and 2 of the claim. When $m(j) < m(\theta_h)$ this can be the only effect on θ_h type workers, and so their payoff decreases, however type-θ_l workers see a commensurate increase in payoff since they are no longer being rejected from firm j. Therefore, the change in payoffs must be greater for θ_l relative to θ_h. ■

Proof of Lemma 7

Since types θ_h and θ_l are in class c, workers of either type have Gittins indices that are equal across every firm. As a reminder, the optimal strategy every period for a worker was to apply to the firm with the highest Gittins index. In particular, a worker stops searching if her match value at a given firm is above the Gittins index of any other firm. That is, whenever θ_ls stop searching at some firm j, $GI_{\theta_l}(j) \geq GI_{\theta_l}(j')\forall j'$. Since both worker types are in class c, this implies that both types follow the same probability distribution over initial applications. Furthermore, $\theta_h^j \geq \theta_l^j$ by assumption, and so $\theta_h^j \geq GI_{\theta_h}(j')$. Therefore, $s_{\theta_h} \leq s_{\theta_l}$. ■

Proof of Proposition 6

For any market, $\mathcal{M}$, and profile of payment functions $\{\phi_j\}_{j\in\mathcal{F}}$, observe that a new market, $\mathcal{M}_\phi$, can be defined to incorporate payment functions into match values.

Definition. *Let market $\mathcal{M} = (\mathcal{F}, \mathcal{C}, m)$ be given.*

Then, market $\mathcal{M}_\phi = (\mathcal{F}', \mathcal{C}', m')$*, where:*

1. $\mathcal{F}' = \mathcal{F}$,

2. $\mathcal{C}' = \cup_{c \in \mathcal{C}} \{\theta' = (\theta_w^1 + \phi_1(\theta), \theta_f^1 - \phi_1(\theta), \ldots) | \theta \in c\}$,

3. $m'(\theta') = m(\theta)$.

Then, observe that $\mathcal{M}_\phi$ satisfies all the assumptions of the model. Limited liability implies that match values are always above zero. Furthermore, because wages could only be conditioned on match value, not type, the types in $\mathcal{M}_\phi$ satisfy independence conditional on a worker's class. Then, Proposition 3 shows that aggregate demand in $\mathcal{M}_\phi$ satisfies gross substitutes. ■

1.B Extensions Satisfy Gross Substitutes

This section shows that the bandit structure ensures that gross substitutes is satisfied in more general settings. To do so, I build on Lemma 4 to extend the model in a natural manner. In each instance, the set of firms remains a simple family of multi-armed bandits, and furthermore, worker's Gittins indices are decreasing in each firms' thresholds. Then, Lemma 4 implies that gross substitutes is satisfied, and the threshold adjustment process of Theorem 1 can be used to characterize an equilibrium.

1.B.1 Gradual Learning

Throughout the paper so far, it has been assumed that workers learn immediately—a type-θ worker that is hired by firm j learns θ_w^j. In many scenarios, learning requires time, or is noisy. Here, I show that in markets where workers receive a noisy signal of their match value, aggregate demand still satisfies gross substitutes.

Noise is parametrized through a normal distribution. When a type-θ worker, i, is hired by firm j in period t, i observes her utility from the match, which is given by $\theta_w^j + \epsilon_i(t)$. $\epsilon_i(t)$ is drawn iid each period from a normal distribution, $\epsilon_i(t) \sim N(0, \xi)$, where $\xi > 0$.

Lemma 11. *Aggregate demand satisfies gross substitutes when learning is gradual.*

Proof of Lemma 11

To begin, I note that workers face a simple family of multi-armed bandit processes. For any firm j, associated thresholds, (v_j, p_j), and type-θ; the expected payoff for type-θ from applying to j is equivalent to type-θ's expected payoff from applying to j when learning was instant.

Then, it remains to be shown that $GI_\theta(j, h_t^\theta)$ is decreasing in (v_j, p_j). A simple interchange argument proves the point. Suppose the optimal stopping solution to $GI_\theta(j, h_t^\theta)$ yielded a larger value for some $(v_j', p_j') > (v_j, p_j)$. Then, utilize the stopping solution for (v_j', p_j') in place of the original stopping solution for (v_j, p_j), with one key difference. Observe that the value from the new solution, under (v_j, p_j) can be decomposed into two terms, one corresponding to matches would have been received under (v_j', p_j'), and a second corresponding to the value for the additional matches due to the difference between (v_j, p_j) and (v_j', p_j'). Then, possibly through garbling the original stopping solution, a new stopping solution can be characterized that mimics the original stopping solution for matches above (v_j', p_j') on average, while treating matches between (v_j', p_j') and (v_j, p_j) as if they had led to rejection. This new stopping solution's value can then be decomposed into two terms, one of which is equal to the original stopping solution's value, and the second which includes the benefit from matching for a single period, and is therefore positive.

Last, Lemma 4 implies that individual demand satisfies gross substitutes, and therefore aggregate demand satisfies gross substitutes. ■

1.B.2 Learning Through Interviews

Similar to the previous result, when workers learn at the interview stage, the gross substitutes condition holds. Formally, when a type-θ worker applies to firm j, the worker learns θ_w^j, regardless of whether she is hired.

Lemma 12. *Aggregate demand satisfies gross substitutes when workers learn through interviewing.*

Proof of Lemma 12

Learning through interviewing simplifies the previous constructions. Previously, when a type-θ worker applied to a firm with threshold (v_j, p_j), if $\theta_w^j > v_j$ she learned as much. If instead $\theta_w^j \leq v_j$, then she may have been unable to tell whether $\theta_w^j = v_j$ and she was unlucky or whether $\theta_w^j < v_j$. Now, she faces a simple family of multi-armed bandit processes where the reward from each bandit j is the realization of Ψ_θ^j in expectation. Then, by construction, $GI_\theta(j, h_t^\theta)$ is decreasing in (v_j, p_j) and therefore Lemma 4 implies that individual demand satisfies gross substitutes. It follows that aggregate demand satisfies gross substitutes as well. ■

1.B.3 Heterogeneous Discounting

In practice, different classes of workers may not only have different priors regarding their type, but they may also retire from the market at different rates. For instance, if different classes also correspond to different ages of workers, those classes may retire at individual rates. Suppose each class c retires from the market with probability $\delta_c \in [0, 1)$.

Lemma 13. *Aggregate demand satisfies gross substitutes when different classes have distinct discount factors.*

Proof of Lemma 13

For each class of workers, the set of firms acts as a multi-armed bandit problem, albeit with differing values of δ_c. Furthermore, the monotonicity of the Gittins index in rewards implies that $GI_\theta(j, h_t^i)$ is decreasing in (v_j, p_j). As such, Theorem 2 implies that each type's demand satisfies gross substitutes. By the previous arguments it follows that aggregate demand satisfies gross substitutes. ■

Chapter 2

Build To Order: Endogenous Supply in Centralized Mechanisms

2.1 Introduction

Because land and funding are limited, public housing is rationed. Much attention has been placed on how to best allocate public housing, but an equally important question is *what type* of housing to build.[1] These decisions are not made arbitrarily: governments can infer which developments are more desirable using past realizations of demand. When deciding what apartments to build and how to allocate them, a public housing authority must contend with two objectives: minimizing vacancies and matching households to the apartments they desire. In this setting, how can a government better build and allocate public housing? To answer this question, we form a dynamic queuing model specialized to the Singaporean public housing system, which houses 80% of the resident population.[2] We characterize

[1] See e.g., Waldinger (2021), Van Dijk (2019) for empirical work pertaining to Cambridge, MA and Amsterdam; and Arnosti and Shi (2020) for a theoretical treatment.

[2] Singapore is an important case study because of the size of its public housing market. Over 80% of Singaporeans live in government-built housing. In 2019, there were over 15,000 apartments transacted; each with a sticker price of at least US$200,000. This implies that at least US$3 billion were transacted in apartment value of government-built housing, suggesting large potential gains to improvements in efficiency.

the government's strategy when supply is endogenous and evaluate the welfare loss when supply is exogenous. We show that building underdemanded apartments is crucial to ensure incoming households report their types truthfully. Furthermore, higher demand can improve allocative efficiency when supply is endogenous. In contrast, when supply is exogenous, allocative efficiency is reduced when demand increases. A key takeaway of our results is that batching multiple applications together is highly desirable in terms of reducing the aggregate uncertainty households face, thereby increasing their willingness to apply sincerely, improving the overall allocation.

The first contribution of this chapter is to develop a model that explores the link between revealed demand and hidden preferences, when the demand responds to supply. Due to the wait times inherent to the housing process, households may choose to take a less desirable apartment today in lieu of their preferred option in the future. We examine the circumstances under which *insincere* applications are prevalent—applications where households apply to a queue that does not match their type. We show that under exogenous supply, insincere applications occur when one queue "overflows" and households of that type apply for the underdemanded apartment. In our solution to the government's dynamic problem under endogenous supply, the optimal mechanism is limited by the reverse issue; the designer's binding constraint comes from households who desire underdemanded apartments. The government would prefer to only build apartments in high demand; but if the government does so, it cannot motivate households to apply sincerely. Hence, the government trades off exploiting its knowledge about the current stock of households against learning about the preferences of incoming households.

To further motivate why the optimal mechanism prescribes supplying less-demanded apartments, consider an environment where households care both about match quality and about match timing. If a household knows only apartments in high demand will be built, it will be tempted to misreport if it prefers an apartment type that is rarely demanded and

so built infrequently. Were that household to enter the queue for its desired apartment, it would need to wait for at least one allocation period before having the chance to receive an apartment. Then, unless the cost of waiting is small, the household prefers to enter the queue for apartments that are more likely to be built.[3]

In Section 2.3, we propose a dynamic model of public housing allocation in which supply can be adjusted over time. At the beginning of each period, the government observes previous household applications, then decides how to allocate new housing across apartment types, where types reflect both apartment size and location. Simultaneously, a household arrives with a private type, corresponding to the apartment type that the household prefers. Newly arrived households select one type of apartment to apply for.[4] When an apartment type has more applicants than available apartments, the available apartments of that type are randomly allocated. If a household is allocated an apartment, both exit the market. Otherwise, the household pays a waiting cost and applies again in the next period. The government aims to minimize a convex combination of two types of inefficiency: allocation and unassignment. Allocation inefficiency captures the government's desire to assign households to apartments of their type. Unassignment inefficiency reflects the government's aim to minimize the number of households that remain unmatched in a given period.[5]

Using data on Singaporean housing applications, we undertake descriptive analysis in Section 2.6 to show that our focus on strategic behavior is empirically justified. We have suggestive evidence that households strategically apply to queues, and the government takes

[3]Importantly, in Singapore, most applicants' alternatives to receiving an apartment are living with family or renting. While non-trivial, the cost of being unmatched for an additional period is far less than it would be in other public housing markets. For instance, in the US, public housing is primarily used as an alternative to homelessness. We show that in such a setting, where the cost of waiting is high, a pooling equilibrium where all applicants receive the same housing is optimal. This analysis corroborates results in Arnosti and Shi (2019) with the fact that many public housing programs ignore household preferences.

[4]Unlike the standard literature on mechanism design where the designer observes types within a truthful mechanism, here, in line with the real-world BTO mechanism, the government can only observe applications to queues.

[5]Allocation inefficiency is standard in the literature, and unassignment generates vacancies which are expensive to maintain. We elaborate on the government's historical incentive to minimize unassignment in Section 2.2.

past demand into account when determining where to build new developments. Our descriptive findings aid us in characterizing the nature of the government's optimal strategy in this mechanism, but leave us short on the specifics. In view of the government's responsiveness to past information and households' strategic applications, we propose a dynamic model to deepen our understanding of the contrast between settings with exogenous and endogenous supply. In such a model, governments respond optimally to applicant behavior, and vice versa.

We focus our attention on the Singaporean mechanism, the Build-To-Order scheme (BTO), which we describe in Section 2.2. Our model is carefully tailored to fit the Singaporean setting, which is of interest to policymakers and academics alike because the market for public housing is large. We show in Section 2.3.1 that the optimal unconstrained mechanism is a first-in-first-out mechanism, and detail why the government stopped using a first-in-first-out mechanism because of historical and other policy concerns.[6] Accordingly, our model differs from standard mechanism design. We restrict the space of policies, only allowing the government to utilize lottery mechanisms, focusing our attention instead on the optimal supply.

As alluded to beforehand, to assess the importance of endogenous supply, we also consider an analogous setting in which supply is exogenous. With exogenous supply, vacancies and inefficient allocation occur even under arbitrarily high levels of patience. We analyze this setting and show that, compared to the setting with endogenous supply, there is strictly higher allocative inefficiency in all parameter regions.

Our second contribution is to add to a growing literature on thickness in dynamic markets through characterizing when batching multiple application cycles is optimal. We examine

[6]In practice, a household can be given priority in a few specific instances. A mature neighborhood is one that has been a residential area for over 20 years. Non-mature neighborhoods are typically farther away from the city center and have fewer completed apartments. For example, if a household has previously been rejected twice, then applies for a flat in a non-mature neighborhood, they are more likely to receive an early queue number. Since non-mature queues are likely to be undersubscribed in the first place, we abstract from non-constant priorities in our model.

the effect of various forms of competition, both when supply is endogenous or when it is exogenous.[7] We show that when the government can control the supply of housing, oversubscription improves the ability of the government to generate responsive mechanisms.[8] While oversubscription increases both market thickness and competition, the increase in thickness enables the government to manipulate the expected wait times between different queues. This policy lever increases the willingness of households to apply sincerely, improving allocative efficiency. In contrast, when the apartment supply is exogenous, the overwhelming impact of competition is to increase expected wait times. Thus, competition exacerbates the inefficiencies of the exogenous setting, increasing the disparities between the exogenous setting and the endogenous setting. We utilize this insight to show that, in the optimal mechanism, thickness is artificially generated in the market by batching applications.

For our final contribution, we leverage novel data on historical BTO applications to develop a lower bound for the proportion of insincere applications. To do so, we compare fluctuations in applications to apartment types to the outflow of successful applicants. We find consistent evidence of households switching their applications between quarters. Such rapid application shifts are difficult to explain through preference changes, supporting our claim that they are the result of strategic applications. Notably, these switches are associated with the relative oversubscription of apartments. This provides evidence for our focus on the household trade-off between quality of match and wait times.

While we focus our attention on the BTO mechanism in this chapter, we note there exist several other instances wherein a centralized planner must choose the supply of a good with incomplete preference information. For instance, consider class schedules. Electives are often substitutable for students, and faculty may be reallocated to address demand spikes. Thus,

[7] We refer to environments where households anticipate low odds of success in the queuing lottery as competitive. Oversubscription refers to environments where the average ratio of households to available apartments is high.

[8] A responsive mechanism is one where different preferences of incoming households lead to different allocations, i.e., the government responds to preferences.

schools face a year-over-year decision regarding which electives to offer and the number of students to cap each course at. The trade-offs in class scheduling are similar to those considered in the BTO mechanism. Similarly, Budish and Cantillon (2012) explore class selection at the Harvard Business School. They show that students strategically report their preferences in response to the course allocation mechanism.

Another surprising example can be found in the area of food donations. Prendergast (2016) and Altmann (2023) examine an innovation in the allocation of food donations by Feeding America, the second-largest American charity by revenue. In 2005, a market was established with "Monopoly money" to improve the distribution of food among food banks across the US. Prior to the introduction of the market, food banks had a fixed food need by weight, and received "take it or leave it" offers commensurate with their need levels. These offers made no allowance for the type of food, whether it be produce or pasta. However, these food types featured real differences in storage needs and often individual food banks received food donations outside the Feeding America system. One primary goal of the new market was to ensure that food banks could bid on the types of food they actually wanted when they wanted. Equally important, upon observing the relative pricing of foods, Feeding America was able to then structure its fundraising requests to increase the quantity of highly demanded food types.

One key difference between the Singaporean public housing system and Feeding America is that the Singaporean government does not allocate apartments by an applicant's willingness to pay, because they believe that housing assistance should not necessarily be disbursed to the highest bidder. In this chapter, we will take it for granted that a market equilibrium will not achieve the government objective. Instead, we focus on finding the mechanism that minimizes inefficiency subject to the government's outside constraints.

Related Literature

A large literature, focusing on the optimal allocation of scarce resources, has improved the design of markets ranging from kidney exchange to school choice. Standard models in this literature have centered primarily on markets where the supply of the scarce resource is exogenous. For instance, a designer of a kidney allocation scheme cannot choose the blood types of the organs entering the system. Importantly, under traditional allocation mechanisms, the supply remains fixed and independent of agent preferences. However, in many markets, a centralized agency may control both the incoming supply of goods and the allocation of goods to agents. For instance, in public housing, the government can control the type of apartments built and the allocation of apartments.

This chapter is closely related to the theoretical literature in matching, much of which stems from studying the problem of optimal student assignment to schools (Abdulkadiroğlu and Sönmez, 2003).[9] Within this literature, our work is most closely related to papers on optimal dynamic matching. In this context, agents are "born" in sequence and face a trade-off between taking their best option at birth and waiting for a better match (Baccara et al., 2020). Akbarpour et al. (2020) shows that a mechanism designer may wish to focus on increasing market thickness, over matching agents myopically. We show that this insight carries over to our setting, even when the good is produced endogenously.

In particular, we offer a new take on the queuing literature. The vast majority of this literature focuses on the allocation of a fixed supply of goods, such as organs. Recent work in this literature include Shi (2022), which examines the optimal priority system to allocate agents to objects; and Agarwal et al. (2019), which develops a new organ allocation mechanism. Thakral (2019) shows that there may never exist an ex-post efficient mechanism when supply is stochastic. In all of these papers, the supply remains exogenous; the mechanism designer cannot control or alter in the inflow of goods. In this chapter, we consider the im-

[9] See Abdulkadiroğlu and Sönmez (2013) for a survey.

pact of relaxing the assumption of exogenous supply and allow the designer to freely control the types of goods that arrive.

Closely related to our work, Leshno (2022) characterizes the optimal mechanism when goods arrive according to an exogenous process. We study a different class of markets in which the designer can not only control the allocation procedure, but also the arrival rate of each good. We show that several of his insights are due to this exogenous process, giving rise to different policy prescriptions for social planners with endogenous supply. For instance, while increased household demand decreases allocative efficiency in the exogenous supply setting, it actually improves the government's ability to manipulate wait times in the endogenous setting. We explore these phenomena in Section 2.5.

Several other recent papers consider dynamic allocation problems with private information when monetary transfers are not permissible. Verdier and Reeling (2022) examine the allocation of bear hunting licenses and show that a dynamic mechanism which repeatedly allocates licenses to the same individuals can improve matching over a static mechanism. In our context, individuals only require one apartment, making this approach impractical. Guo and Hörner (2020) consider repeated good allocation to a single agent whose valuation fluctuates over time. Galichon and Hsieh (2018) shows that as long as money burning is permissible, stability can be achieved in many settings with private information.

Last, in our companion work, Lee et al. (2023), we utilize the same dataset to construct a dynamic choice model over housing lotteries and estimate it. Using this model, we were able to answer a separate question: what is the impact of increasing the *total* supply of housing on wait times, vacancies, and prices on the aftermarket for government housing? We find that simply increasing the housing supply can fail to reduce wait times, due to the demand response eclipsing the supply increase. However, improving the allocation procedure can complement increasing supply. Specifically, in combination with switching to a strategyproof

mechanism, increasing supply can keep wait times low and reduce upward pricing pressure on the aftermarket.

2.2 Policy Background

2.2.1 The Build To Order Scheme

Over 80% of resident households in Singapore live in government housing, which makes up 80% of the housing stock in Singapore. These apartments, numbering over 1 million, are administered and maintained by the Housing and Development Board (HDB). In Singapore, many government apartments are first introduced into the housing stock via the BTO scheme.

The BTO scheme superseded the previous Registration for Flats System (RFS). Under RFS, homebuyers first chose the broad geographical area in which they wanted to live, then were informed of the cost and exact location of the apartment when their queue number was called. Not only did buyers not know when they could move in to their apartment, but they only had to pay the down payment for any home loan when their apartment was completed; if their apartment was finished early, some of these buyers could not raise enough funds for their down payment. The Asian Financial Crisis in the late 1990s exacerbated this issue. The government suddenly found itself with a surplus of vacant housing, and incurred heavy maintenance fees. Soon after the crisis, the government switched from RFS to BTO. The key difference between the two mechanisms is that under RFS a first-in-first-out queue is used where households need only apply once, while under BTO households must reapply every quarter. BTO ensured that current applications were an accurate representation of current demand. This motivates our modelling restriction that precludes the government from using allocation mechanisms that reward seniority.[10]

[10]Currently, the HDB allocates its apartments through a complex system of allocation processes with transfers, from which we abstract. This is because we want to keep our model tractable, and moreover, BTO is the scheme through which most government apartments in Singapore are initially allocated. For instance,

Introduced in April 2001 and taking place each quarter, the BTO exercise allows potential homeowners to ballot for their preferred neighborhood and apartment size.[11] Each such ballot, termed a "booking", is secured by a down payment due on application. After the application phase, the HDB assigns each applicant a queue number, indicating the number of other applicants ahead of her in the queue.[12] If, when her turn arrives, she does not like any of her choices (or has none), she may withdraw her application, after which she may participate in a future cycle. Upon withdrawing, her position in the queue is lost and not preserved for future applications. Furthermore, households that have withdrawn more than once incur severe penalties to their priority in future applications. Given that apartments of a similar size are similarly furnished, this motivates our assumption that households who apply to an apartment size and are successful always accept.

After all applicants have either selected an apartment or withdrawn their application, the HDB begins building the apartments if 70% of all properties in a development have been allocated. In practice, BTO apartments of all sizes are oversubscribed, most by at least 2-3 times. To our knowledge, all BTO exercises have successfully reached the construction stage. These apartments are typically ready for homebuyers to move in within 3 years of the corresponding BTO exercise.[13] To prevent immediate arbitrage, apartments may not be sold on the secondary market before 5 years have elapsed after the initial move-in.

---

in Financial Year 2013/2014, there were 86,298 BTO residential units under construction. Under the next largest comparable scheme "Design, Build and Sell", a scheme targeting households with higher incomes and with more extensive private developer involvement, only 3,893 units were under construction (Housing and Development Board, 2014). By Financial Year 2018/2019, all residential units under construction were BTO apartments Housing and Development Board (2019). See https://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/modes-of-sale for more details.

[11]For more on the historical context for the BTO, see the government archives: http://eresources.nlb.gov.sg/history/events/d33acabb-a341-460c-8fde-99cf0a9270f4.

[12]To enforce social mixing, there are ethnic quotas for each housing development. I.e., there is a cap on the number of Chinese, Malay, and Indian households permitted in each government apartment building. These quotas are enforced at both the BTO stage and on subsequent resale. This aspect of the housing system has been extensively studied in previous work (see Wong 2013 and Wong 2014), so we will abstract from these concerns.

[13]Wait times recently have lengthened to 5 years because of labor shortages caused by the COVID-19 pandemic.

These apartments are often oversubscribed because they are sold at highly subsidized prices, resulting in selection into BTO from the private market. These subsidies cause purchases on the private market to feature a selection effect. Buyers with greater wealth levels default towards the private market as a result of their inability to receive subsidies, biasing the housing purchases observed in the market. Thus, due to the selection of buyers, it is difficult to utilize private market prices as indicative of actual preferences over housing.[14]

2.2.2 Responsive Apartment Supply

Under BTO, each booking made by a household is a signal of housing demand. This allows the HDB to adjust future housing supply to meet expected demand. In Figures 2.2.1 and 2.2.2 we compare non-mature and mature neighborhood apartment allocations by type. The larger level of volatility in mature neighborhood apartment supply reflects the government's willingness to adjust the supply of housing when it has access to demand data.

Figure 2.2.1 displays the relative quantities of housing types supplied in non-mature neighborhoods. We observe that while the total quantity supplied fluctuates over time, the relative proportions of each type do not. That is, in neighborhoods where the government has less information, it opts to avoid adjusting the supply of housing.

In contrast, Figure 2.2.2 shows that housing in mature neighborhoods changes sharply over time. For instance, while 3- and 5-room apartments initially are built in similar proportions, their relative ratios change dramatically. Comparing Figure 2.2.2 and Figure 2.2.1, we find a categorical difference between how the government allocates apartments in mature and non-mature neighborhoods. We believe that this difference is due to the government accounting for neighborhood level demand, and using it to determine future offerings within

[14]Despite the hefty subsidies given to successful BTO applicants, a government official we spoke with noted that only about 10% of BTO buyers sell their apartments within 5 to 10 years of purchase. While there is an incentive for arbitrage, this opportunity may only be available to sufficiently wealthy households that are not capital constrained.

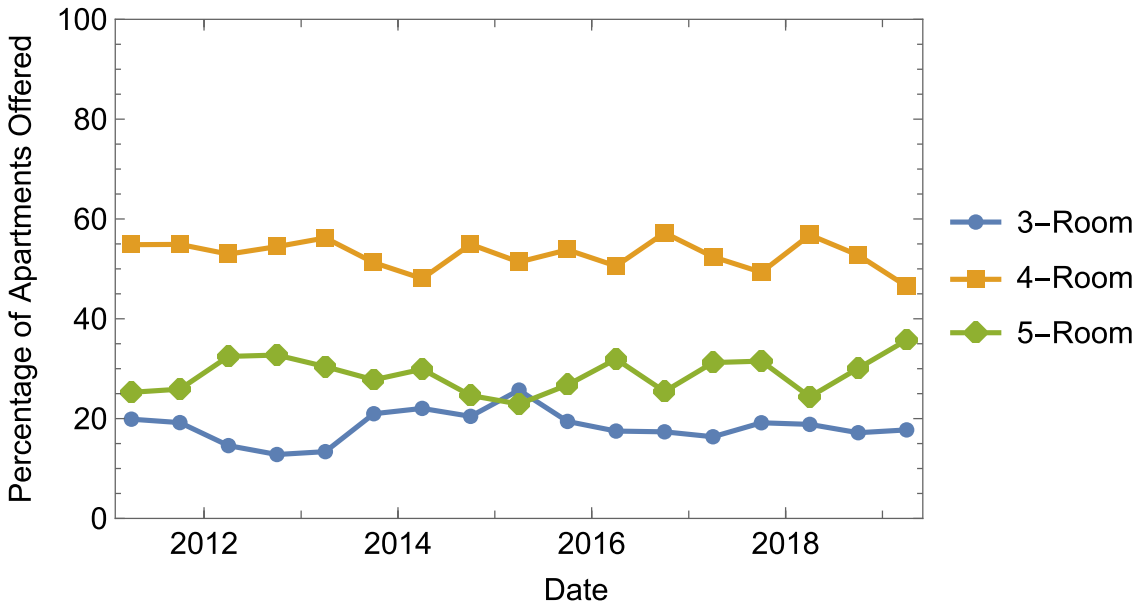


Figure 2.2.1: Percentage of apartments, per half year by type, built in non-mature neighborhoods.

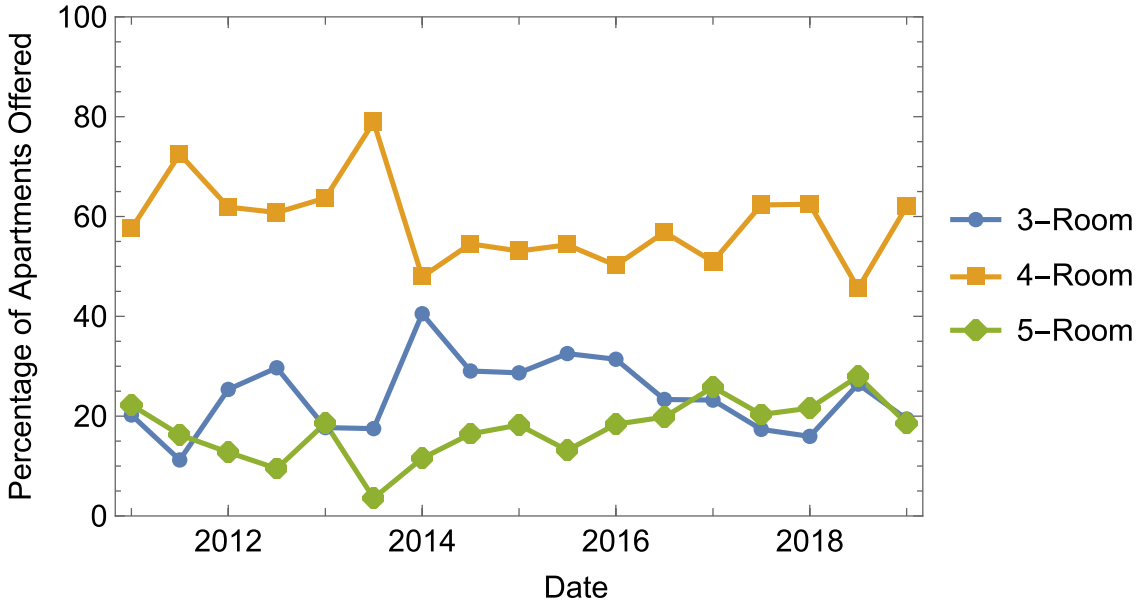


Figure 2.2.2: Percentage of apartments, per half year by type, built in mature neighborhoods.

that neighborhood. In Section 2.6, we conduct a more detailed analysis using BTO data to show that the government takes household demand into consideration.

2.3 Model

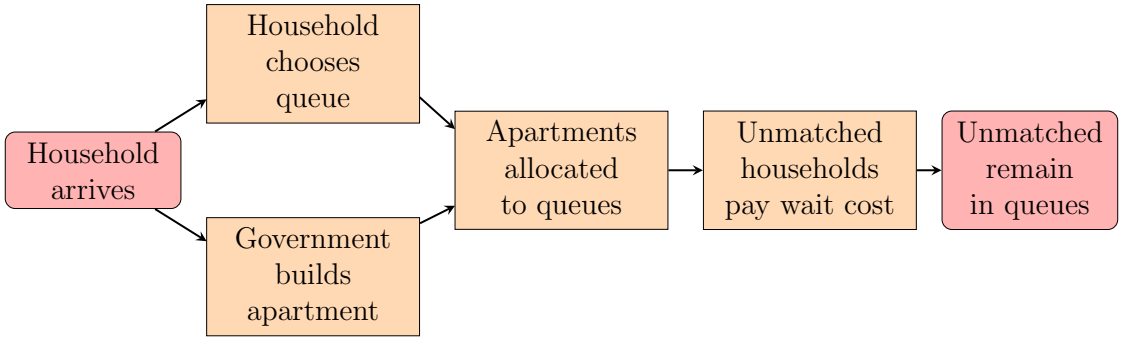
The descriptive facts suggest that the Singaporean government is responsive, but leave us short on prescriptions of how to optimally design a responsive mechanism when supply is endogenous. We develop a model to fill in this gap. We model an allocation mechanism with endogenous supply of goods, akin to the Singaporean BTO mechanism. Time is discrete with an infinite horizon, $t \in \{0, 1, 2, \ldots\}$. The agents are young households born with preferences over the goods, apartments. Each period, one household arrives. The government has no apartments available at $t = 0$, but builds one apartment in every subsequent period. Apartments and households can be of $|\Theta| = 2$ types, $\Theta = \{A, B\}$.[15] We use θ_t to denote the type of the arriving household in period t, and ϕ_t to denote the apartment built in that period. Household types are private information, unknown to the government. A household that is matched to an apartment of the same type receives utility h. A household that is matched to an apartment with a different type receives utility $l < h$. A household that is not matched incurs a flow cost of waiting, c, and remains in its queue.[16]

There is one queue for each type of apartment. The two queues will be referred to as queue A and queue B. At the beginning of each period, all agents are informed of the number of households and apartments in each queue. An incoming household chooses the queue it wishes to enter. We denote the queue choice of the period-t household by $d_t \in \{A, B\}$.

[15] In the appendix, we consider the model with three types. We show that the optimal mechanism is qualitatively similar to the case with two types.

[16] One worry is that agents can change the queues they have entered in BTO. We consider an alternative mechanism in the appendix. Households are able to choose the queue they wish to enter every period. We show that the natural adaption of the mechanisms presented in the body of the chapter remain optimal. See Section 2.B of the Appendix for further details.

Figure 2.3.1: Market Timing



Before observing the incoming household's choice, the government chooses $\Phi_t^A \in [0,1]$, the probability with which $\phi_t = A$.[17] The timing of the market is summarized in Figure 2.3.1.

The government's strategy must satisfy *queue-anonymity*: it must treat households within a single queue identically, without regard to their seniority in the queue. If an apartment is available and at least one household is in the corresponding queue, that apartment is randomly allocated through a uniform distribution to one of the households in the corresponding queue. For instance, if k households are in queue ϕ, and there are $m \leq k$ type-ϕ apartments, each household in queue ϕ has the same m/k probability of receiving an apartment.[18]

We use $s^t = (s_A^t, s_B^t) \in \mathbb{Z}^2$ to denote the net demand of the market in the beginning of period t. For $s_\phi^t > 0$, s_ϕ^t denotes the number of households in queue ϕ. Otherwise, queue ϕ has no households and $|s_\phi^t|$ denotes the number of vacant type-ϕ apartments. The public history at the beginning of time t is $H_t = (x_0, x_1, \phi_1, \ldots, x_{t-1}, \phi_{t-1}) \in \{A, B\}^{2t-1}$.

Definition. *A **mechanism** μ is a sequence $\{\Phi_t^A\}_{t=1}^\infty$, where each Φ_t^A maps public histories to an assignment probability: $\Phi_t^A : H_t \to [0,1]$.*

[17] In practice, the HDB chooses the location of upcoming developments at least two cycles in advance, justifying our assumption that the government must choose Φ_t^A before observing household applications for the current cycle.

[18] We discuss our rationale for modelling the allocation of apartments as a uniform distribution in Section 2.3.1.

We assume the government has commitment power and declares the mechanism μ at the beginning of time. We will focus on Markovian mechanisms that condition only on payoff relevant variables, summarized by the state. Accordingly, where appropriate, we drop time-scripts. The payoff-relevant information in the public history can be summarized by the state s^t. Then, a strategy for the government can be summarized by $\Phi^A : \mathbb{Z}^2 \to [0, 1]$.[19]

We restrict the government's mechanism to be Markovian as motivated by our setting. In interviews with Singaporean housing applicants and anonymous officials from HDB, we learned that the BTO mechanism hews closely to being Markovian to disincentivize the non-needy from gambling for an apartment with high resale potential.[20]

In state s, $W^{\mu}_{\phi}(s)$ will be used to denote the expected wait time for an incoming household that enters queue ϕ. Similarly, $w^{\mu}_{\phi}(s)$ denotes the expected wait time at the beginning of the period for a household already in queue ϕ when the state is s. We will drop the reference to the mechanism when the context poses no confusion.

The expected utility for a household that enters queue ϕ is the match benefit from a type-ϕ apartment minus expected waiting costs from queue ϕ, such a household's expected utility can be written as:

$$U(d_t, \theta_t, s_t) = \mathbb{E}_{\mu}\left[l + \mathbb{1}_{\{d_t=\theta_t\}}(h-l) - c\left[d_t \mathbb{1}_{\{s_A \geq 0\}} \frac{s_A}{s_A+1}(1+w_A(s_{t+1})) + (1-d_t)\mathbb{1}_{\{s_B \geq 0\}} \frac{s_B}{s_B+1}(1+w_B(s_{t+1}))\right]\right].$$

When a household enters the queue of its type, we will say it has *applied sincerely*. A household only applies sincerely if doing so maximizes its utility:

[19] A brief comment on our approach is warranted. Rather than consider the general mechanism design problem, with associated individual rationality and incentive compatibility constraints, we consider the induced market game wherein households take the government's strategy as given and play against one another. In Section 2.4.4 we show that given the government's choice of mechanism, the government's outcome parameters of interest are well-defined and unique.

[20] The focus on Markovian mechanisms is with loss of generality. In Appendix Section 2.C, we show that in settings with higher degrees of oversubscription, a non-Markovian mechanism outperforms the optimal Markovian mechanism.

$$\theta_t \in \arg\max_{d_t} \; U(d_t, \theta_t, s_t).$$

The government's objective depends on two elements: the quality of matches and the frequency of unassignment.[21] Since there is always one more household than available apartments, there will always be a minimum of one unassigned household in every period. To normalize the measure of inefficiency, we only consider unassignment above the baseline of one. In this context, counting the number of vacant apartments is equivalent to counting the number of excess unassigned households. We will use the two measures interchangeably, except in Section 2.5.3 where the difference between the two measures is relevant.

Because we do not know exactly how the Singaporean government ranks these two sources of inefficiency in practice, we characterize the solution to the government's problem for all possible weightings of allocation and unassignment inefficiency. The government aims to minimize total inefficiency, with weights α and $1-\alpha$ placed on allocation and unassignment inefficiency respectively. If the period-t household and the period-τ apartment are matched, the match generates $\mathbb{1}_{\theta_t \neq \phi_\tau}$ allocation inefficiency. Then, m_t, the level of allocation inefficiency in period t, is simply the number of households that did not apply sincerely in some period that were matched in period t. Similarly, we define the unassignment inefficiency in state $s^t = (s_A, s_B)$ as $v_t = \max\{s_A, s_B\} - 1$, the normalized number of unassigned apartments. A mechanism is evaluated by the *average inefficiency* it creates:

$$V(\mu) = \lim_{T\to\infty} \sup \frac{1}{T} \mathbb{E}_\mu \left[\sum_{t=1}^{T} \alpha m_t(\mu) + (1-\alpha) v_t(\mu) \right]. \tag{2.1}$$

Definition. *A mechanism μ^* is **optimal** if it minimizes average inefficiency, $V(\mu^*) = \inf_\mu V(\mu)$.*

[21] Unmatched households need not be without a home. In Singapore, most of those applying under the BTO mechanism have outside options, such as rentals or living in their parents' homes.

In Section 2.4.4, we show that in the limit as $T \to \infty$, every mechanism generates at least one steady state. Furthermore, even when an optimal mechanism generates multiple steady states, those steady states feature equivalent values of U. This allows us to simplify Equation 2.1. Let $m(\mu)$ and $v(\mu)$ refer to allocation and unassignment inefficiency in some steady state of μ. Then, the government's problem can be rewritten as:

$$\min_{\mu} \ \alpha m(\mu) + (1-\alpha)v(\mu).$$

2.3.1 Model Discussion

Here we detail the rationale behind several important modelling decisions.

Random Uniform Lottery: We restrict the government to mechanisms that cannot offer priority. In particular, the government cannot utilize a first-in-first-out mechanism under this assumption. Indeed, in the setting presented in this model, were the government permitted to choose an arbitrary mechanism, a first-in-first-out mechanism would always achieve the first-best outcome under any parameter region.[22]

As previously mentioned, the RFS system previously utilized a first-in-first-out style mechanism which led to a surplus stock during the Asian financial crisis, draining the HDB's wealth through maintenance and holding expenses. Additionally, abstracted from our model is an additional government concern: family formation. At 1.14 children per woman in 2018, Singapore has one of the lowest reproductive rates in the world. The Singaporean government has publicly stated that increasing their reproductive rate is one of their major objectives. In order to achieve this objective, the government gives married couples with children an extra draw from the housing lottery. If a first-in-first-out mechanism were utilized, the benefit to family formation would disappear as soon as a period had passed and the household was

[22]Consider the following mechanism. Let every incoming household be allocated to a single waiting queue, independent of their type. In every period $t \neq 0$, the government builds an apartment to match the type of the household at the beginning of the queue. Then, every household is incentivized to report truthfully, since their report does not change their expected waiting time.

placed in the queue. By contrast, through running the weighted lottery every period, the incentive to form a family persists. This motivates our decision to focus on the specific mechanism utilized in Singapore. Even when the allocation method is constrained, welfare is substantially higher under endogenous supply relative to exogenous supply.

Private Information: The model implicitly assumes that the government cannot elicit household preferences through means outside the allocation mechanism. In particular, the government cannot use a Becker–DeGroot–Marschak (BDM) style mechanism to encourage households to report truthfully (Becker et al., 1964). We draw on recent literature that has shown that BDM mechanisms may not accurately capture willingness to pay. For instance, Lehmann (2015) and Müller and Voigt (2010) show that BDM mechanisms may produce biased estimates of willingness to pay. Relatedly, Cason and Plott (2014) show that BDM mechanisms can confuse subjects due to its complexity, which in turn results in noisy or biased outcomes. A similar line of reasoning motivated the Singaporean government to primarily use household applications to estimate demand (Mah, 2010).[23]

Linear Waiting Costs: We assume waiting costs are linear (as in related work on dynamic matching, e.g., Leshno (2022), Baccara et al. (2020), Ashlagi et al. (2018)). The first reason is normative: exponential waiting costs would imply that an incoming household would be a higher priority for the mechanism designer than a household that has failed to match multiple times. Linear waiting costs take an agnostic stance on this front. The second reason is practical: if waiting costs were exponential, the state space would grow rapidly. Not only would the number of households of each type need to be tracked, but so would their time of arrival.

[23]The following quotation is from the HDB website in response to a question regarding the possible introduction of a registry to track household preferences. "MND and HDB have considered the Member's suggestion to introduce a register for Build-To-Order (BTO) flat applicants. However, there is no assurance that doing so will improve the planning of BTO launches to meet demand since an indication of interest may not accurately reflect actual demand, as there is no commitment to buy."

2.4 Results

2.4.1 Perfect Information Benchmark

To begin, we analyze the benchmark case where household types are common knowledge and the government can choose the queue a household enters. Formally, rather than households choosing d_t, the government learns θ_t and also selects $d_t(\theta_t)$.

The government can easily ensure that allocative inefficiency is zero. In order to do so, the government allocates households to queues of their types. That is, to avoid allocating an apartment to a household with a different type, the government sets $d_t(\theta_t) = \theta_t$.

Consider state $(1, 0)$. If the incoming household is of type-A, it will be allocated to queue A. Had the government built an apartment of type-B, a vacancy would result. As such, to avoid the possibility of a vacancy, the government must build a type-A apartment to fill the non-empty queue. Therefore $\Phi^A(1, 0) = 1$, and for similar reasons, $\Phi^A(0, 1) = 0$. Under this strategy, the only possible states on the equilibrium path are $(1, 0)$ and $(0, 1)$, therefore defining $\Phi^A(1, 0), \Phi^A(0, 1)$, and $d_t(\theta_t)$ defines a mechanism that we will refer to as the "first-best" mechanism, μ_{fb}. For illustrative purposes, the mechanism can be depicted by the finite state automaton in Figure 2.4.1.

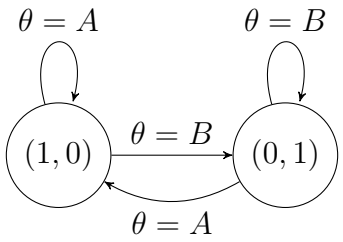


Figure 2.4.1: Finite-state automaton depicting the first-best mechanism, μ_{fb}. Circles indicate states, while arrows depict transitions given an incoming household of type θ.

Lemma 14. *When the government knows θ_t and can select $d_t(\theta_t)$, μ_{fb} generates 0 inefficiency.*

It is worth noting that the government perfectly responds to demand when information is public. If the previous household is of type θ, then a type-θ apartment is built. Additionally, complete information implies the absence of vacancies in an optimal mechanism. We will later show that vacancies occur with positive probability when the government prioritizes minimizing allocation inefficiency in an optimal mechanism. Vacancies are present in non-mature neighborhoods in the real world, implying that the government lacks the ability to perfectly predict household preferences and that the government prioritizes reducing allocation inefficiency over minimizing vacancies.

2.4.2 Implementing the First-Best

For the remainder of this chapter, we assume household types are not public information. To begin, we prove a simple lemma that restricts the state space. A type-A household prefers queue A if and only if the expected wait from entering queue A is no greater than the expected wait of queue B plus $\frac{h-l}{c}$. To see why, suppose the state is $s = (s_A, s_B)$ and $s_A > 0$. Then, a simple calculation shows that the type-A household prefers to enter queue A if the following holds:

$$\begin{aligned} &u(d_t = A, \theta_t = A, s) \geq u(d_t = B, \theta_t = A, s), \\ &\qquad \implies \frac{h-l}{c} \geq \left(\Phi^A(s)\frac{s_A}{s_A+1} + 1 - \Phi^A(s)\right)(1 + w_A(s_{t+1})) - \left(\Phi^A(s)\right)(1 + w_B(s_{t+1})). \end{aligned}$$

Observe that the right-hand side is simply the difference in the expected wait times. Similar computations can be done when $s_B > 0$ or for a type-B household considering queues B and A respectively.

Lemma 15. *A type-θ household prefers to apply sincerely if and only if $\frac{h-l}{c} \geq W_\theta(s) - W_{\theta'}(s)$.*

The proof of Lemma 15 and all future proofs are relegated to the appendix. The left-hand side of the constraint compares the benefit from sincere matching to the cost of waiting. As

households care more about matching to an apartment of their type, they are more willing to accept increases in wait times. Similarly, as the cost of waiting increases, their focus shifts, placing a lower weight on sincerely applying and a higher weight on receiving an apartment immediately. To simplify notation, let $\gamma \equiv \frac{h-l}{c}$ represent the ratio of the benefit from sincerely applying to the loss from waiting an additional period.

Now we return to the government's problem under private information, and ask if the first-best can be implemented. μ_{fb} is still the only mechanism that could achieve 0 inefficiency. Lemma 15 lets us determine if households are willing to select $d_t = \theta_t$ in all states by computing the difference in expected wait times for each state and comparing to γ.

We only need to check the incentives for a type-B household in state $(1, 0)$. The symmetry of the mechanism implies that the incentives are the same for a type-A household in state $(0, 1)$. A simple calculation shows that the wait times from queue A and queue B under the first-best mechanism are $\frac{2}{3}$ and $\frac{4}{3}$ respectively in state $(1, 0)$. Then, households are willing to apply sincerely only if $\frac{4}{3} - \frac{2}{3} \leq \gamma$:

Proposition 7. *A mechanism can achieve $m = v = 0$ if and only if $\gamma \geq \frac{2}{3}$.*

For any value of α and $\gamma \geq \frac{2}{3}$, there is a single optimal mechanism. In state $(1, 0)$, the government always builds an apartment of type A. Similarly, in state $(0, 1)$, the government always builds an apartment of type B. Incoming households always apply sincerely. An outside observer would see the government responding in a manner commensurate to demand. This government response is slightly lagged; it occurs after the demand shock. Since households are more concerned about correct matching than wait times, the government can maximize efficiency through building apartments that are in high current demand.

2.4.3 Exogenous Supply Benchmark

In this section, we expand on the importance of endogenous supply. To do so, we illustrate what happens when the supply is exogenous. That is, we assume the government cannot

choose the type of the incoming apartment. We restrict the government's choice to $\Phi^A(s) = \frac{1}{2}$. Note that this setting is equivalent to what Leshno (2022) called a "balanced" setting. The proportion of incoming type-A households and apartments are equal, implying that in the long run, perfect allocation efficiency is still feasible, if not necessarily incentive compatible.

The first household will always wish to apply sincerely in equilibrium, since both queues have an equal wait time. The strategy of subsequent households will depend upon the current state. When queue B has no households, i.e., the state is (s_A, s_B), for $s_A > 0$, the wait time in queue A is strictly greater than that of queue B. Therefore, incoming type-B households always strictly prefer to apply sincerely. The choice of a type-A household will depend upon the difference in expected wait times. Importantly, the wait time for queue A is increasing in the number of households in queue A, because supply is exogenous. A strategy is a **threshold** strategy if, for some κ, it dictates that a type-θ household, in state $(s_\theta, s_{\theta'})$, enter queue θ if and only if $s_\theta < \kappa$.

Lemma 16. *Any equilibrium generates an outcome equivalent to that generated by some threshold strategy profile.*

When all households use the same κ as their threshold, we denote the expected wait time at the start of the period in state s for a household in queue ϕ by $w^\kappa_\phi(s)$. In equilibrium, the threshold, κ, must be large enough to ensure that incoming households no longer wish to apply sincerely when the state reaches $(\kappa, -(\kappa-1))$. A type-A household that enters queue A in state $(\kappa, -(\kappa-1))$ gives up the chance to immediately receive l, but gains h minus the expected wait cost. When $\kappa > 1$, in equilibrium two constraints must hold. When there are κ households in either queue, incoming households must prefer the empty queue. Second, when there are $\kappa - 1$ households in either queue, incoming households must prefer to apply sincerely. Lemma 15 then implies the following constraints:

$$\text{IC}_\kappa(\kappa) : \gamma \leq \frac{\kappa}{2(\kappa+1)}\Big[1 + w_A^\kappa(\kappa, -(\kappa-1))\Big] + \frac{1 + w_A^\kappa(\kappa+1, -\kappa)}{2},$$
$$\text{IC}_\kappa(\kappa-1) : \gamma \geq \frac{\kappa-1}{2\kappa}\Big[1 + w_A^\kappa(\kappa-1, -(\kappa-2))\Big] + \frac{1 + w_A^\kappa(\kappa, -(\kappa-1))}{2}.$$

The wait times $w_A^\kappa(s)$ can be computed as the solution to a linear system of $\kappa + 1$ equations.[24] We focus on the case $\kappa = 2$ to illustrate a point of comparison with the first-best mechanism. To begin, we compute the solution to the system of equations. This yields wait times of $w_A^2(2,-1) = \frac{5}{2}$, $w_A^2(1,0) = 2$, and $w_A^2(3,-2) = \frac{10}{3}$. Substituting these values into the above equations implies the following:

Lemma 17. *If $\gamma \in [\frac{5}{2}, \frac{10}{3}]$ and supply is exogenous, the strategy profile where all households use a threshold of 2 is an equilibrium.*

As κ, the threshold, increases, the lower bound for γ increases. To see why, observe that as the number of households in a queue increases, each household expects to wait for a longer period of time in that queue. Critically, when the state is $(\kappa - 1, -(\kappa - 2))$, the incentives of an incoming type-A household determine the binding lower bound on γ.

Proposition 8. *As κ increases, the minimum γ under which the threshold κ strategy profile is an equilibrium also increases.*

We focus on $\kappa = 2$, because it is the minimal informative threshold. Suppose $\kappa = 1$ was the threshold in some equilibrium. Then, in state $(1, 0)$, households enter queue B, no matter their type. In effect, households never apply sincerely. Then, the steady state under the $\kappa = 1$ strategy profile would have $\frac{1}{2}$ allocation inefficiency and 0 vacancy inefficiency. Such an equilibrium is not responsive, and is equivalent to allocating apartments independently of type.

[24] The general system of equations is listed in the appendix.

For contrast, if the government had controlled the supply of apartments, the first-best could have been implemented when $\gamma > \frac{2}{3}$. Furthermore, the level of allocation inefficiency is higher for equilibria with $\kappa \geq 2$ relative to the first-best implementation. We compute the level of allocation inefficiency under exogenous supply when $\kappa = 2$. The resulting steady state[25] has a frequency in states $((1,0),(2,-1))$ of $(\frac{2}{3},\frac{1}{3})$ generating an inefficiency level of $\frac{\alpha}{6} + \frac{1-\alpha}{3}$. This inefficiency is directly increasing in the proportion of time spent in state $(2,-1)$. State $(2,-1)$ inherently contains a vacancy and furthermore, households do not apply sincerely. By comparison, when $\gamma > \frac{2}{3}$, there is 0 inefficiency when supply is endogenous.

This suggests that the ability to manipulate the supply of apartments is incredibly important for the government. Even when only a single household is in a queue, households are tempted to apply insincerely. In the remainder of this chapter, we explore the optimal mechanism in various situations with endogenous supply.

2.4.4 When the First-Best Cannot be Implemented

Returning to the setting with endogenous supply, we proceed by assuming that households are unwilling to apply sincerely under μ_{fb}.

Assumption 2. $\gamma < \frac{2}{3}$.

Now, households weigh the cost of waiting more highly relative to the utility gain from matching to the most desirable apartment type. Importantly, Assumption 2 sharply restricts household behavior in state $(2,-1)$. In state $(2,-1)$, all households prefer to enter queue B. To see why, consider the expected wait times for each queue under any mechanism. The maximum wait for a household that enters queue B is 0, because a type-B apartment is available. In contrast, the minimum wait for a household that enters queue A arises when the government always builds type-A apartments and all incoming households apply to queue

[25] We treat symmetric states as one state, i.e., $(1,0)$ and $(0,1)$ are reduced to $(1,0)$.

B. Then, the minimum wait is $\frac{2}{3} + \frac{2}{3} \cdot \frac{1}{2} = 1$. Combined with Lemma 15, a type-A household prefers to apply sincerely only if, $\gamma \geq 1$, which violates Assumption 2. Therefore, in state $(2, -1)$ under Assumption 2 all households will always enter queue B. This statement also holds for any state with more than two households in queue A. The expected wait from queue A in such a state is strictly larger than in state $(2, -1)$, while the expected wait from queue B remains 0.

Lemma 18. *Suppose Assumption 2 holds. If* $\max\{s_A, s_B\} > 1$*, in state* (s_A, s_B)*, either type-A or type-B households do not apply sincerely.*

It follows that the state space of any mechanism's steady state is finite. In particular, any optimal mechanism generates a steady state with a finite state space. Furthermore, since it cannot be optimal to remain permanently in state $(2, -1)$ or state $(-1, 2)$,[26] the steady-state must be recurrent unless it never transitions between $(1, 0)$ or $(0, 1)$. Any mechanisms that fail to transition between $(1, 0)$ and $(0, 1)$ has equal inefficiency. Then, the finiteness of the steady-state combined with the recurrence of the state space implies there is a unique steady state.

Before stating the result, we define *queue symmetric*. Informally, two steady states are queue symmetric if they are equivalent up to relabelling queue A as queue B, and vice versa. Formally, a steady state, $\mathbb{S}$, is queue symmetric if there exists a permutation $\pi : \Theta \to \Theta$ such that the probabilities of any two states, $s, s' \in \mathbb{S}$, are equal under the permutation $\pi(s) = s'$. We next show that any optimal mechanism generates at least one steady state, and furthermore, that outcomes are effectively unique under optimal mechanisms.

Lemma 19.

1. *If* μ *is an optimal mechanism, then there exists at least one steady-state associated with* μ*.*

[26] Such a mechanism would be dominated by a mechanism that always remains in state $(1, 0)$.

2. *If an optimal mechanism μ generates multiple steady states, those steady states are queue symmetric.*

Lemma 19 implies that the average level of inefficiency is well defined. Either the steady state is unique, or the two possible steady states feature equivalent levels of inefficiency. As such, we will proceed by evaluating mechanisms using the average level of inefficiency in any steady state.

Importantly, when Assumption 2 holds, behavior in state $(1, 0)$ is tightly regulated. Suppose the government attempted to avoid vacancy inefficiency through always building an apartment of type A, i.e., $\Phi^A(1, 0) = 1$. Then, incoming households optimally respond by entering queue A irrespective of their type. Since in every period one type-A apartment is built and one household enters queue A, the state remains in $(1, 0)$ indefinitely. We will refer to the described mechanism as the *pooling mechanism*, μ_p.[27] Under the *pooling mechanism*, $\Phi^A(s) = 1$ and $d_t = A$.

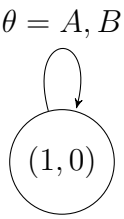


Figure 2.4.2: Finite state automaton depicting the pooling mechanism.

Note that half of the households were of type A. It then immediately follows that the level of allocation inefficiency is $\frac{1}{2}$ and the level of vacancy inefficiency is 0.

Remark 1. *Under μ_p, the level of allocation inefficiency is $\frac{1}{2}$ and the level of vacancy inefficiency is* 0.

The above logic further implies that under Assumption 2, a mechanism that involves always building an apartment of the type matching the non-empty queue causes all households

[27] There are technically several mechanisms that result in similar allocations and equivalent levels of efficiency. For the sake of exposition, we will focus on the pooling mechanism described in the main text.

to prefer allocation to the non-empty queue. Then, the state will never change, because every period a new household enters the queue matching the one that the apartment is built for. To prevent this phenomenon, the government must build the less desirable apartment with positive probability. This insight provides a cautionary tale for endogenous supply policy that responds myopically to demand. Doing so provides counterproductive incentives for households, and leads to households opting to not apply sincerely. Furthermore, a naïve estimation of public demand through observing queuing decisions overestimates how satisfied households are with the current policy regime.[28]

We proceed by searching for the optimal mechanism with allocation efficiency below $\frac{1}{2}$. In order to improve allocation, we must have $\Phi^A(1,0) \neq 1$. In particular, $\Phi^A(1,0)$ must be low enough to incentivize type-B households to apply sincerely. In state $(2,-1)$, the government always builds type-A apartments to minimize vacancy inefficiency, since it cannot increase the incentive for households to apply sincerely regardless.

Consider the following mechanism. In state $(1,0)$, the government builds a type-B apartment with probability q. In state $(2,-1)$, the government always builds a type-A apartment. There are never more households in queue A than in state $(2,-1)$, since households always enter queue B when $s = (2,-1)$ according to Lemma 18. Similarly, $\Phi^A(0,1) = q$ and $\Phi^A(-1,2) = 0$. We refer to this mechanism as the **two-state mechanism with parameter** q. Formally, the *two-state mechanism with parameter* q, μ_q, sets $\Phi^A(1,0) = 1-q$ and $\Phi^A(2,-1) = 1$.[29]

Proposition 9. *Under Assumption 2, the optimal mechanism takes on one of two forms. Either the pooling mechanism is optimal, or there exists* q^* *such that* μ_{q^*} *is optimal.*

In order to determine the value of q^*, we compute all possible on-path wait times. Let q denote the probability with which the government builds a type-B apartment in state $(1,0)$.

[28] In Lee et al. (2023), households trade off higher chances of success against being closer to amenities.

[29] As an aside, we note that a mechanism could potentially sometimes build a type-B apartment when the state is $(2,-1)$. In the appendix, we formally show that doing so never increases the extent to which households apply sincerely. As such, the optimal mechanism never does so.

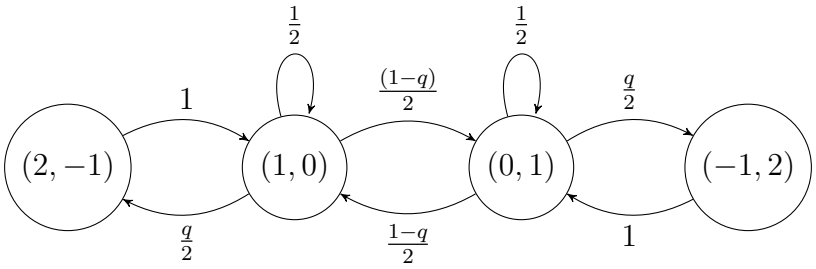


Figure 2.4.3: Finite state automaton depicting the two-state mechanism, μ_q. Arrows denote transition probabilities.

Raising q increases $w_A(1,0)$, the expected wait time for a household in queue A in state $(1,0)$. In return, raising q increases the incentive for an incoming type-B household to apply sincerely.

Given q, we can compute the queue specific wait times for incoming households in state $(1,0)$. Solving yields $w_A(1,0) = \frac{4q+1}{3-2q}$ and $w_A(2,-1) = \frac{2+q}{3-2q}$. Lemma 15 places a bound on the difference in wait times, if the bound is exceeded, households will not apply sincerely. In order to minimize the probability of entering state $(2,-1)$, $q^*(\gamma)$ is implicitly defined as the solution to:

$$\gamma = \frac{1-q}{2}(1 + w_A(1,0)) - q(1 + w_A(2,-1)). \tag{2.2}$$

The right-hand side of Equation 2.2 is the difference in expected wait times. Inputting the values of $w_A(1,0)$ and $w_A(2,-1)$ implies:

$$q^*(\gamma) = \frac{3\gamma - 2}{2(\gamma - 3)}.$$

In the remainder of this chapter, when we refer to μ_q without specifying q, it is understood that $q = q^*(\gamma)$.

When a type-B household is indifferent between entering queue A and queue B, a type-A household in state $(1,0)$ will strictly prefer to enter queue A. This immediately follows from Lemma 15, since the difference in wait times between queues A and B is simply -1 times

the difference in wait times between queues B to A, and therefore is less than 0. By the symmetry of the mechanism, households also apply sincerely in state $(0, 1)$.

Given the symmetry of the optimal mechanism, we again refer to states (s_A, s_B) and (s_B, s_A) as (s_A, s_B) when displaying inefficiencies. This can be thought of as identifying two states that have equal numbers of households in the long queue. We proceed by computing the steady state probabilities up to queue symmetric steady states. In the steady state, the transition probabilities at the beginning of a period are given by:

Origin \ Destination	$(1,0)$	$(2,-1)$
$(1,0)$	$1-\frac{q^*}{2}$	$\frac{q^*}{2}$
$(2,-1)$	1	0

We use P_q to denote the steady state measure of μ_q. That is, $P_q(s)$ denotes the "amount of time" the steady state spends in state s. Let $M(q)$ denote the transition matrix generated by μ_q, then $P_q = P_q M(q)$. Inverting and solving for P_q yields $P_q(1,0) = \frac{2}{2+q^*}$ and $P_q(2,-1) = \frac{q^*}{2+q^*}$.

The level of inefficiency in the steady state is directly proportional to $P_q(2,-1)$. In state $(2,-1)$ there is a $\frac{1}{2}$ probability that the new household is of type A, while all households enter queue B. State $(2,-1)$ is the only source of allocation inefficiency in equilibrium, in state $(1,0)$ households always apply sincerely. Then, the average level of allocative inefficiency is proportional to the fraction of time that the steady state is in state $(2,-1)$, and equals $\frac{q^*}{2(2+q^*)}$. Similarly, the level of unassignment is equal to the proportion of time that the steady state is in state $(2,-1)$—in this case $\frac{q^*}{2+q^*}$.

In principle, the choice of q does not need exactly render type-B households indifferent in state $(1,0)$, indeed larger values of q can also convince households to apply sincerely. The upper limit for q is the point at which type-A households in state $(1,0)$ prefer to enter queue B. This constraint is given by:

$$\gamma \geq q(1 + w_A(2, -1)) - \frac{1-q}{2}(1 + w_A(1, 0)). \tag{2.3}$$

Solving Equation 2.3 for q implies that $q = \frac{3\gamma+2}{2(\gamma+3)}$. Then, the range of values for q that ensures that households apply sincerely in state $(1, 0)$ is $q \in \left[\frac{3\gamma-2}{2(\gamma-3)}, \frac{3\gamma+2}{2(\gamma+3)}\right]$. We utilize the previously calculated values of inefficiency by taking the derivative of inefficiencies with respect to q. Unsurprisingly, both derivatives are positive: increasing q increases inefficiency. Since, inefficiency is increasing in q and therefore minimized by q^*, we continue to focus on the two-state mechanism with parameter q^*.

The above logic implies that there exist two possibly optimal mechanisms, the pooling mechanism and the two-state mechanism. We determine which of the two is optimal, conditional on α, the social planner's preference parameter over allocation inefficiency versus unassignment inefficiency.

Theorem 3. *Let Assumption 2 hold.*

For $\alpha < \frac{2-3\gamma}{8-5\gamma}$, μ_p is the optimal mechanism. For $\alpha > \frac{2-3\gamma}{8-5\gamma}$, the optimal mechanism is μ_q. Finally, if $\alpha = \frac{2-3\gamma}{8-5\gamma}$, μ_p and μ_q are the two optimal mechanisms.

Theorem 3 shows that either μ_p or μ_q must be optimal. Theorem 3 then allows us to consider the impact of an increase in selectivity of households. Formally, we show that μ_q improves in efficiency in response to a decrease in the ratio of wait cost to relative gain from applying sincerely.

As expected, the threshold at which μ_q is optimal decreases when the relative gain from applying sincerely increases. The intuition for this comparative static is simple. As the relative gain from the correct match increases, households become more willing to apply sincerely, improving the ability of the government to match households properly. Since μ_q takes advantage of sincere applications while the pooling mechanism does not, the inefficiency of μ_q is decreasing in γ.

For a direct welfare comparison, suppose the government were utilitarian. The government would then place a weight of $h - l$ on allocation inefficiency and a weight of c on unassignment inefficiency. Without loss of generality, and to provide consistency with the previous results, we normalize the government's objective.

Corollary 1. *Let Assumption 2 hold, and* $\alpha = \frac{h-l}{h-l+c}$. *That is, the government aims to minimize* $\frac{h-l}{h-l+c}m + \frac{c}{h-l+c}v$.

Then, the pooling mechanism is optimal when $\gamma < \frac{9-\sqrt{65}}{4}$. *Otherwise,* μ_q *is optimal.*

The implications of Corollary 1 are similar to those of Theorem 3. When γ is small and households care more about wait times than applying sincerely, mechanisms that ignore preferences are optimal. When γ is large and households care about allocation, mechanisms involving sincere applications do better.

In the context of public housing, Corollary 1 enables a simple comparison of "take-it-or-leave-it" mechanisms and the BTO mechanism. Referencing Arnosti and Shi (2019), this implies that "take-it-or-leave-it" mechanisms may be optimal. Indeed, more intricate mechanisms feature losses that may not be immediately apparent. Both "take-it-or-leave-it" mechanisms and endogenous supply can be optimal in different parameter regions. Crucially, the social planner's objective as well as the preferences of the recipients need to be considered before determining the appropriate mechanism. For instance, in the US, public housing is primarily used as an alternative to homelessness. In our model, homelessness corresponds to a large value for c, the cost of being homeless an additional period. For contrast, in Singapore, the alternative to receiving an apartment is generally renting or living with family for an additional period. Then, our model suggests that the differences in housing policy between the US and Singapore could be optimal, in contrast to the findings of previous work. Changing the design of US public housing policy to incorporate household preferences may also leave more apartments vacant, increasing waste.

2.5 Competition and Market Thickness

Having characterized the constrained optimal government strategy, we consider the impact of competition on gains from endogenous supply. To begin this section and fix ideas, we informally define our notions of competition and thickness. A household considers a queue to be "competitive" if the household believes there is a high probability another household will later enter that queue. Competition implies that a household expects there are or will be several other households in the same queue it is in. Thickness, while related, refers to the number of households applying for queues simultaneously. Thickness implies competition, but competition may not imply thickness.

We will proceed by showing that competition is undesirable when supply is exogenous. To see why, note that competition has a multiplicative effect on expected wait times. Importantly, when supply is exogenous, competition increases the expected difference in wait times. Then, high levels of competition and exogenous supply dissuade households from applying sincerely. On the other hand, when supply is endogenous, competition gives the government a greater deal of flexibility allowing the government to equalize expected wait times across queues. Additional policy flexibility dominates its multiplicative effect, causing competition to actually improve efficiency in certain parameter regions, though overall welfare still decreases. Thickness, in particular, is highly desirable for the government. We show that the government can artificially generate thickness through batching several applications together. Furthermore, when the government prioritizes sincere applications, i.e., α is large, it is optimal for the government to batch.

2.5.1 Persistence

To begin, we consider the impact of persistent household types. Recall that in the original model, household types were independently distributed with uniform probability. We modify

that assumption here, $\theta_{t+1} = \theta_t$ with probability $p \geq \frac{1}{2}$. The period-0 household still has its type drawn with probability $\frac{1}{2}$ from $\{A, B\}$.

We then find conditions under which the government can implement the natural analog of the first best mechanism, μ_{fb}. In order for it to do so, it must utilize μ_{fb} as in Section 2.4.1. Households must prefer to apply sincerely, and the government must build apartments matching the queue of the old household. Namely, $d(\theta, s) = \theta$ and $\Phi^A(s) = \mathbb{1}_{s=(1,0)}$. We emphasize that the government's hands are equally tied in the setting with persistent types and the previous implementation of the first-best in Section 2.4.2. There is increased competition, but the level of market thickness remains the same. In a given period, the same number of households are present relative to before, but if a household applies sincerely, it expects an increased level of competition in the following period. We will show that persistence hinders the government's ability to implement the first-best outcome. To be exact, the parameter region where households always apply sincerely in μ_{fb} shrinks in p.

We proceed in a manner similar to Section 2.4.2. Computing the wait times conditional upon entering a queue implies the difference in ex-ante expected wait is $\frac{1+2p}{2(2-p)}$. Lemma 15 then implies that the difference in expected wait times is the lower bound on γ.

Proposition 10. *Under persistence p, there exists an optimal mechanism with 0 inefficiency when:*

$$\frac{1+2p}{2(2-p)} \leq \gamma. \tag{2.4}$$

We consider the welfare impact of increasing persistence. To do so, we take the derivative of the difference in wait times with respect to p. The result is positive, as the level of persistence increases, it becomes more difficult to implement the first-best.

Lemma 20. *When demand is persistent, the threshold under which there exists an optimal mechanism with 0 inefficiency is increasing in p.*

Lemma 20 follows due to households' expectation that future applicants are more likely to compete for the same queue. A household's incentive to not apply sincerely increases in p. If a household applies sincerely, and is not matched in the current period, the household expects a longer overall wait time relative to settings with lower values of p. By not applying sincerely, households significantly decrease their expected wait times, due to decreased future competition. Then, γ must be higher to motivate households to apply sincerely under μ_{fb}.

2.5.2 Oversubscription

Going forward, we return to the no persistence case, $p = \frac{1}{2}$. We proceed by considering a natural form of competition, oversubscription. We call a housing market oversubscribed when the number of households balloting is larger than the number of apartments available. For reference, under the BTO scheme, apartments of all sizes are oversubscribed, generally at a minimum of $2 - 3\times$. In Section 2.4.4, we focused on oversubscription at a rate of $2\times$. Larger rates of oversubscription feature an increase in the market thickness, while also directly increasing wait times. In this section, we will show that the increase in thickness dominates, expanding the region within which the government can implement the first-best, thereby reducing inefficiency.

As a simple method of varying the level of oversubscription, we change the number of households that arrive in period 0, without changing the supply of apartments. In every subsequent period, one household appears as before. Let N denote the number of households that arrive in period 0, i.e., the surplus of households. Households have the same information as in previous sections. Household types are private, but households observe the queues other households have entered. In particular, at $t = 0$, all households make their application simultaneously, and do so without information about the other households that are present.

First, we change the number of households appearing at $t = 0$ from 1 to 2. We then derive the optimal mechanism that implements the first-best. In state $(2, 0)$ the government

must build a type-A apartment with probability 1, to avoid vacancies. In state $(1,1)$ the government builds a type-A apartment with probability $\frac{1}{2}$. Without loss of generality, suppose this probability was less than $\frac{1}{2}$. Then, in state $(0,2)$ type-A households wish to apply sincerely only if type-B households wish to apply sincerely in state $(2,0)$. The probability of building a type-A apartment could be increased, strictly improving type-A household's willingness to apply sincerely.

Since when attempting to achieve the first-best there are no vacancies, there are only three possible resulting states $(2,0),(1,1),(0,2)$. We treat the first and last state symmetrically.

Solving for wait times under this mechanism, we find that $w_A(2,0) = \frac{25}{17}, w_A(1,1) = \frac{27}{17}$. Using these values, we can then compute the incentive for new arrivals to apply sincerely in all states. First, in state $(1,1)$, the expected waiting time is independent of the queue entered, and so households always prefer to apply sincerely. Next, we ensure both types wish to apply sincerely when $s = (2,0)$, in order to do so, we compute the difference in expected waiting times conditional on entering queue A as opposed to queue B. The expected wait time from entering queue A is $\frac{2}{3}(1 + w_A(2,0))$, while entering queue B gives a wait time of $1 + w_A(1,1)$. Taking the difference and simplifying generates a value of $\frac{16}{17}$.

Then, a type-A household never has an incentive to not apply sincerely in state $(2,0)$ since doing so entails increasing their expected waiting time. By comparison, type-B households have a strong incentive to not apply sincerely. We do not consider the incentive to apply sincerely in state $(1,1)$, because the mechanism implies waiting times are equal between queues. We show in the appendix that when $N = 2$, the first-best can be implemented iff $\gamma \geq \frac{16}{17}$.

In particular, comparing this value to the constraint when $N = 1$ implies that the immediate effect of oversubscription was to decrease the ability of the government to implement the first-best outcome. It is instructive to consider why this occurs. While the level of competition is higher, the government's hands are tied when it comes to designing the mech-

anism. The allocation of households to queues is fixed to avoid allocation inefficiency. In order to avoid unassignment, it must choose $\Phi^A(2,0) = 1$ and $\Phi^A(0,2) = 0$ to avoid unused apartments. Last, in state $(1,1)$ it must choose $\Phi^A(1,1) = \frac{1}{2}$ as any other value either unbalances the wait times from entering queue A or queue B, and thereby fails to optimize. Then, $N = 2$ maintains the pernicious effects of competition that were present in the persistence extension, while not providing new tools to deal with the increase in wait times.

Corollary 2. *If γ is high enough to incentivize households to apply sincerely under the first-best mechanism when $N = 2$, then γ is also high enough to incentivize households to apply sincerely under the first-best mechanism when $N = 1$.*

The Markovian assumption has some bite. The previous result arises because we restrict consideration to mechanisms that only condition upon the current values of payoff relevant variables. When the state is $(1,1)$, both types of households are more than willing to apply sincerely. Then, the apartment construction probabilities can be altered in order to incentivize households to apply sincerely in extreme states. Since the binding constraint comes from a type-B household in state $(2,0)$, we increase the probability that the government builds a type-B apartment when the current state is $(1,1)$ and the previous state was $(2,0)$. There also exists an upper bound: when type-B apartments are built too often, new type-A households will not apply sincerely. In Appendix Section 2.C, we show that when the Markovian assumption is weakened, this increase in competition is helpful and the first-best can be implemented for $\gamma \geq .48$.

Next, we consider the impact of further increasing N. We show that the negative impact of oversubscription on efficiency when there are two excess households is an anomaly. A mechanism that implements the first-best in this setting cannot allow vacancies, but also must incentivize households to apply sincerely. As such, when the state is $(N,0)$, the government must build a type-A apartment. However, for any intermediate state $(k, N-k)$, where households of both types are present, the government can freely choose any probability

for $\Phi^A(s)$. Of course, as discussed above, when $N = 2$, the government has no additional flexibility in choosing $\Phi^A(s)$. For $N > 2$, increasing the level of oversubscription always improves the ability of the government to implement the first-best. This is a direct result of the government's increased ability to equalize wait times between the two queues.

We proceed by solving for the optimal supply probabilities, as well as the associated restrictions on γ. It is worth focusing on the shape of the optimal mechanism under oversubscription. Thus, the optimal mechanism randomizes; it sometimes builds an apartment in lower demand. In turn, the state is pushed towards a more extreme level: sometimes the less demanded apartment is built, and the incoming household wants the more demanded apartment. In order to achieve the first best, the mechanism cannot build an apartment of the less-demanded type when there are no households in the corresponding queue. This places a hard constraint on the value γ under which the first-best can be implemented. A naïve solution would be to default to building the apartment type that is in higher demand. However, doing so strongly disincentivizes incoming households of the under demanded type from applying sincerely.

Except in this most extreme state $s = (N, 0)$, both apartment types are always built with positive probability by the government. Figure 2.5.1 displays the optimal mechanism for varying numbers of excess households. In general, the probability that a less desired apartment is built is larger than the fraction of households in the corresponding queue.[30]

Based on our previous analysis, two forces are at play. On the one hand, increasing the level of competition exacerbates the loss from missing a match in the current period, because waiting times increase across the board. On the other hand, increasing oversubscription increases the number of free variables the government can use in order to normalize wait times between reports. This outcome relates to the thickness of the market, and how it better enables the government to match households correctly.

[30] We solve explicitly for the optimal mechanism when $N < 5$, and numerically compute it when $N \geq 5$.

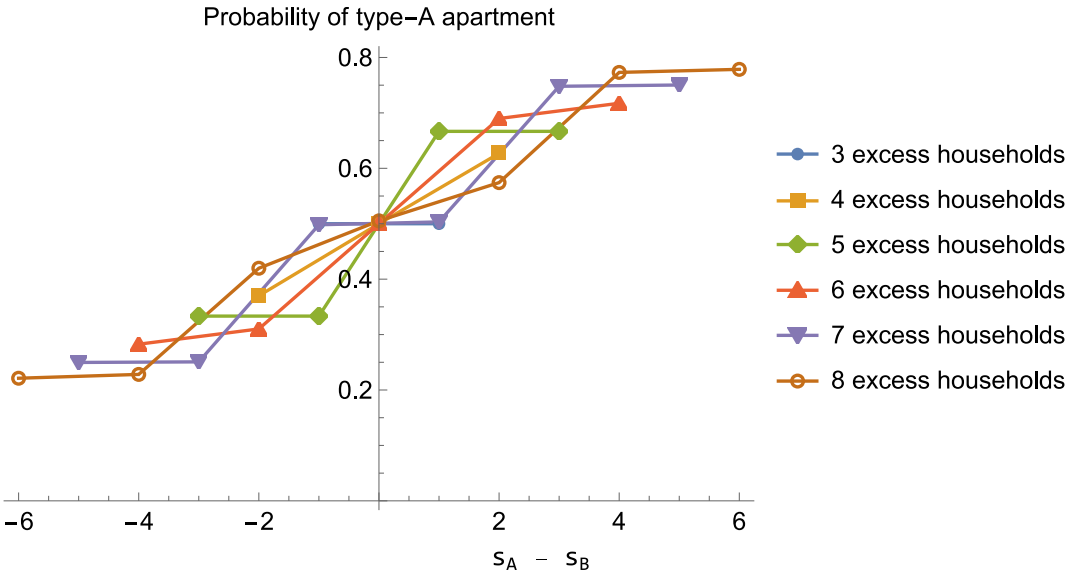


Figure 2.5.1: The probability that a type-A apartment is built under the optimal mechanism with over subscription. The x-axis depicts the number of households in queue A minus the number in queue B. The y-axis is the probability that the government builds a type-A apartment in the corresponding state. States $(0, N)$ and $(N, 0)$ are omitted, the corresponding y-values are 0 and 1 respectively.

To better distinguish between the two forces, we return to the previous baseline where the government randomly builds either apartment type with probability $\frac{1}{2}$, $\Phi^A(s) = \frac{1}{2}$. Households freely choose the queue they wish to enter as before. Such a mechanism can never implement the first-best for any value of γ. There always exists a sufficiently extreme state $(x, N - x)$ such that type-A households prefer to not apply sincerely for x large. Instead, we find a weaker condition under which households apply sincerely until there exists a vacant apartment. That is, households apply sincerely except in state $(N + 1, -1)$, where all households enter the queue for type-B apartments. We then determine conditions on γ under which households are incentivized to follow this strategy profile. Such a measure underestimates the direct effect of competition while supply is exogenous, yet this further serves our point to show that the government's flexibility in choosing the apartment allocation is crucial.

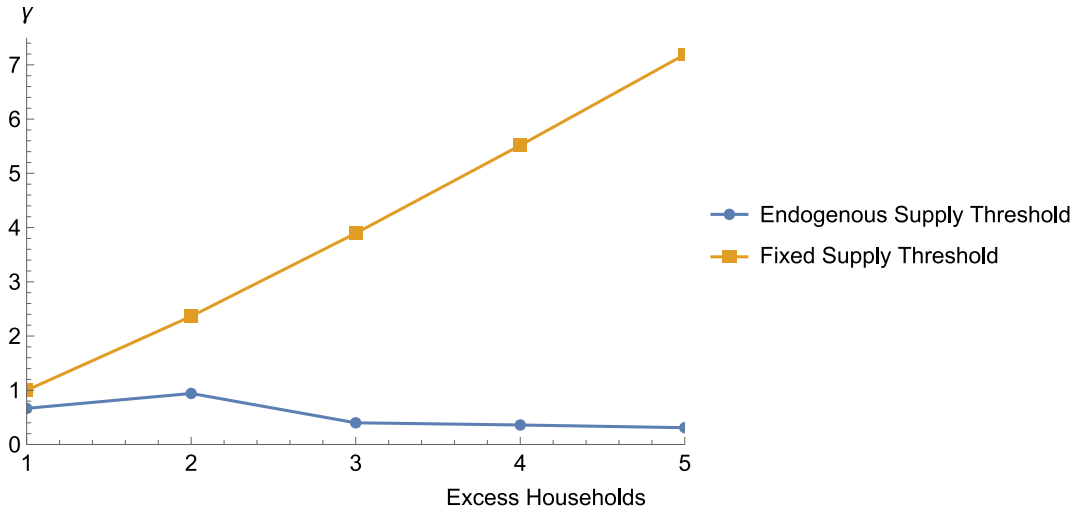


Figure 2.5.2: Lower bound for γ under endogenous supply and exogenous supply. The x-axis displays the number of excess households (in the original model $x = 1$), while the y-axis displays the lower bound on γ. Endogenous Supply Threshold refers to the lower bound on γ under which the first-best can be implemented. Fixed Supply Threshold refers to the lower bound on γ under which individuals are willing to apply sincerely while no apartments are vacant.

Figure 2.5.2 displays the minimal values of γ as the level of competition increases under this mechanism. Despite the increase in thickness generating an increase in expected wait times, the added government control reduces the difference in wait times. Therefore, when competition increases, the necessary value of γ also rises under exogenous supply, while the opposite holds under endogenous supply.

2.5.3 Batching

One natural concern is that by directly increasing oversubscription, all households are made worse off due to increased levels of unassignment. Furthermore, in practice, the government does not control household demand. Nonetheless, the government can manipulate the timing of apartment applications. For instance, the government could opt to have households apply quarterly or annually.[31] Through delaying the timing of applications, the government increases the level of unassignment in the short run, as incoming households must wait until

[31] In Singapore, the government batches applications on the quarterly level.

the next application cycle to enter the market. However, we show that the corresponding increase in market thickness improves the government's ability to incentivize sincere applications.

We adjust the timing of the model to allow the government to choose the amount of time that elapses between application cycles. At the beginning of the game, before $t = 0$, the government declares the length of application cycles, T. Households that arrive during periods that are not a multiple of T, must wait for the period to reach a multiple of T before entering a queue. While a household waits, the household's type remains hidden, and the household continually pays the flow cost of waiting each period while it remains unmatched. When the period is a multiple of T, all households not in a queue, simultaneously choose a queue to enter. At the same time, except when $t = 0$, the government chooses the types of T different apartments to be built. In effect, the government stockpiles its supply of apartments, then builds all of them simultaneously when a cycle begins.

The government must still choose the apartment supply, $\Phi^A(s)$, before observing agent reports. Since the government can now build T apartments simultaneously, we update our notation. Formally, $\Phi^A : S \to \Delta\{A, B\}^T$. That is, the government declares a probability distribution over the types of T apartments. Then, the government's objective is:

$$V(\mu) = \min_{(\Phi^A(s),T)} \alpha m(\mu) + (1 - \alpha)v(\mu),$$

where m and v are the values of allocation and unassignment inefficiency in any steady state of μ as before. The government faces the same objective as before: to minimize the weighted sum of inefficiencies. In this section, unassignment and vacancies are no longer equivalent. This is due to the fact that new households continue to arrive while the government delays the building of apartments. Here, we focus on the original definition of unassignment. Notably, unassignment penalizes high values of T, as the minimal unassign-

ment inefficiency for a given value of T is $\frac{T-1}{2}$. In a sense, batching procedures are overly penalized, implying that the government places a greater weight on the welfare of its citizens relative to the cost of vacant apartments. Nonetheless, we show that batching is still optimal when the government cares deeply about the quality of matches.

We proceed by describing a mechanism with $T = 2$ and finding conditions under which it is optimal. In the appendix, we consider the general setting and show formally the following mechanism is optimal under these conditions. We will abuse notation and refer to it as the $T = 2$ mechanism where appropriate.

Suppose the government aimed to ensure all households applied sincerely, $m = 0$. In state $(1, 0)$, in order to properly incentivize households, the government randomizes between building one apartment of each type or building two type-A apartments. Let $\Phi^A(1, 0) = [(q, (1, 1)), (1 - q, (2, 0))]$. In state $(2, -1)$, the government always builds two type-A apartments, $\Phi^A(2, -1) = [(1, (2, 0))]$. We can then compute expected wait times conditional upon the state. These in turn allow the expected wait times conditional on reports to be computed. By Lemma 15 the difference must be bounded by γ in order for households to apply sincerely.

In the appendix, we compute the difference in wait times with respect to q. Then, we can solve for the optimal level of q that minimizes the difference in wait times. We find the minimal difference is given by $\gamma \approx .294$. Hence, the batching mechanism with $T = 2$ induces sincere applications for $\gamma \geq .294$. Notably, batching attains a substantive improvement for the γ requirement under $T = 1$, namely $\gamma > \frac{2}{3}$.

Since the mechanism achieves 0 allocation inefficiency, the only inefficiency is unassignment of which there are two sources. The two sources are the default $\frac{1}{2}$ unassignment inefficiency from batching with $T = 2$ and the unassignment inefficiency in state $(2, -1)$. State $(2, -1)$ is only entered from state $(1, 0)$ when two type-A households arrive and with $1 - q$ probability the government builds two type-A apartments, or when two type-B households

arrive and with q probability the government builds one apartment of each type. Notably, since a household has $\frac{1}{2}$ probability of being either type, this implies that in the steady state the probability of state $(2,-1)$ is independent of q. Then, so long as γ is high enough such that households apply sincerely and q is accordingly chosen, the level of unassignment inefficiency is independent of q.

As in the previous welfare analysis, we combine states that are symmetric in A and B. The transition probabilities are given by:

	$(1,0)$	$(2,-1)$
$(1,0)$	$\frac{3}{4}$	$\frac{1}{4}$
$(2,-1)$	$\frac{3}{4}$	$\frac{1}{4}$

It is then immediate to observe that the expected time spent in state $(2,-1)$ is $P(2,-1) = \frac{1}{4}$. Therefore, the total level of unassignment inefficiency generated by this mechanism is the sum of unassignment from $T = 2$ and state $(2,-1)$ or $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. We observe that any mechanism with $T > 2$ occurs a minimum unassignment inefficiency of 1 and therefore is dominated by the optimal $T = 2$ mechanism when $\gamma > .294$. It remains to determine if there exists a superior mechanism for $T = 2$. We note that a superior mechanism also needs to overcome the pooling mechanism, which generates 0 unassignment inefficiency but $\frac{1}{2}$ allocation inefficiency.

A superior mechanism with $T = 2$ must generate a lower level of unassignment inefficiency. At the same time, it must improve upon the allocation inefficiency generated by the pooling mechanism. However, in order to improve upon the inefficiency of the pooling mechanism, households must not be matched uniformly. In the appendix, we show that no mechanism can do both. The key tension is that mechanisms that improve the level of unassignment inefficiency do so at the cost of increasing allocation inefficiency. However, due to the inherent $\frac{1}{2}$ unassignment inefficiency due to setting $T = 2$, such mechanisms are

dominated either by the previously defined $T = 2$ batching mechanism or by the pooling mechanism for all α.

As an aside, we note that for $\gamma \in (.294, \frac{2}{3})$, the optimal mechanism depends on α. When α is small, $\alpha \approx .17$ the pooling mechanism is optimal. As α increases, μ_q becomes optimal. Finally, when α is large $\approx .94$, the batching mechanism is optimal.

Theorem 4. *When $\gamma \in (.294, \frac{2}{3})$, the optimal mechanism for $\alpha < \frac{2-3\gamma}{8-5\gamma}$ is the pooling mechanism. For $\alpha \in \left[\frac{2-3\gamma}{8-5\gamma}, \frac{34-9\gamma}{38-15\gamma}\right]$ it is μ_q. Last, for $\alpha > \frac{34-9\gamma}{38-15\gamma}$ the optimal mechanism is the T=2 batching mechanism.*

The key implication of Theorem 4, is that batching is only a useful tool when allocation is a greater concern than unassignment. If this is the case, then through increasing the thickness of the market at the cost of increasing the temporary level of unassignment, batching can drastically improve the quality of matches.

2.6 Descriptive Analysis

In this section, we present several descriptive findings to show that, 1) the Singaporean government does indeed take household demand into consideration when determining supply and 2) the BTO mechanism matches households to apartments below their top preference a significant fraction of the time.

2.6.1 Data

Our data are taken from the HDB. The data set comprises applications to all BTO developments between 2012 and 2020. We refer to each period with applications and matching as a "cycle."[32] These data were constructed through liberal use of the Wayback Machine to

[32]The number of cycles per year varies during this period: there are 6 cycles each year in 2012-2014; 3 cycles in 2015 and 2020; and 4 cycles in the remaining 4 years, resulting in a total of 40 cycles.

scrape historical BTO results from the HDB website.[33] We split the apartment data into types based on the number of rooms, i.e., 3-, 4-, or 5- room apartments.[34] The types of apartments in the previous theoretical sections corresponds to the number of rooms provided. We use "supply" to refer to the number of apartments of a certain type available in a given period. Similarly, "demand" will refer to the number of applications in a period for those apartments. We aggregate the data across mature and non-mature neighborhoods.

2.6.2 Lagged Effects of Supply and Demand

To mimic the model, we use proportional demand and supply. Rather than using the total number of k-room apartments built, we transform the data to calculate the per-period proportion of k-room apartments built:

$$dd4_t = \beta_0 + dd4_{t-3}\beta_1 + dd4_{t-4}\beta_2 + \sum_{l=1}^{4} ss4_{t-l}\gamma_l + u_t.$$

We regress proportional supply on proportional demand from 3-4 cycles ago and lags of supply from 1-4 cycles ago. This specification is chosen because the Public Housing Authority announces its developments approximately two cycles before the cycle takes place, so it can only condition on the observed demand from 3-4 cycles before. From the results of the regression, displayed in Table I, the third lag of demand is significant at the 5% level in predicting supply. The coefficient is also positive, suggesting that when there is higher demand for 4-room apartments in previous periods, the corresponding supply increases in subsequent periods. Lags of supply are not significant when regressing demand on lags of demand and supply. Hence, demand likely does not depend on past supply. We have

[33] http://www.archive.org/

[34] Applicants apply to a (location, size) pair in every cycle. The majority of the population chooses among 3-, 4-, and 5- room apartments. The Singaporean Public Housing Authority chooses the proportion of each type of apartments (characterized by size) in each period, but often does not have a choice of location.

Table I: Lagged Demand Predicts Supply

Regressor	Estimate	Std. Error	t-stat	p-value
(Intercept)	0.495	0.192	2.581	0.015
$dd4_{t-3}$	0.435	0.167	2.595	0.015
$dd4_{t-4}$	0.159	0.176	0.904	0.374
$ss4_{t-1}$	0.183	0.183	1.004	0.324
$ss4_{t-2}$	0.295	0.161	1.829	0.078
$ss4_{t-3}$	-1.032	0.304	-3.395	0.002
$ss4_{t-4}$	0.010	0.336	0.029	0.977

Regression of proportional supply of 4-room apartments ($ss4_t$) in period t on proportional supply of 4-room apartments in period $t-1$ to $t-4$, and proportional demand of 4-room apartments in periods $t-3$ and $t-4$ (i.e., $dd4_{t-3}$ and $dd4_{t-4}$). There are $T = 40$ cycles.

suggestive evidence that the government takes previous household requests into consideration when determining what types of new apartments to build, which justifies the modelling focus on the government's problem.

2.6.3 Estimating the Upper Bound on Sincere Applications

We only observe the total number of applicants in each queue. Hence, we cannot observe true household preferences, or even track households at the individual level to observe if their applications are changing over time. This makes it difficult to measure match inefficiency, as we do not know whether a match corresponds to a household receiving its top choice.

Nonetheless, we can estimate an upper bound on truth-telling with aggregate data. We assume individual-level preferences do not change over time and no exiting. Consider the number of households that apply for an apartment type in a given period minus the number of households that actually receive that apartment. Call this number the *oversubscription* (os) for that apartment type. In the next period, the *number of households applying to that apartment type* (dd) should be at least as high as the current period's oversubscription. If it is not, this means that households must have applied for different apartment types (i.e., they have switched). Then, in period t, the statistic $os_{t-1} - dd_t$ is a *lower bound on the number*

of period-t switchers (lsw_t). If $os_{t-1} < dd_t$, we take the conservative view that there are no switchers. Hence,

$$lsw_t := \max\{0, os_{t-1} - dd_t\}.$$

Switching can occur at the individual level, but be indiscernible in the aggregate. As such, lsw is a lower bound for the number of switchers. For instance, if one household switches its application from 3-room apartments to 4-room apartments and another switches its application from 4-room apartments to 3-room apartments, aggregate applications would look identical for both periods. We would be unable to tell that two switches had occurred. The number of switchers is a lower bound for the extent of misreporting. With persistence in household preferences, switchers must have misreported their type at least once. Further, some households may consistently misreport their type, and would not be captured by the switching statistic. Hence, our lsw statistic is a conservative lower bound for the total extent of misreporting in the system.

Figure 2.6.1 shows there are several instances where the demand in a given period is below the oversubscription rate from the previous period, indicating the existence of switchers.

With a lower bound on the extent of misreporting, we can approximate a lower bound for mismatch present in the economy. Due to the uniform allocation lottery, the probability that any household receives an apartment in a given period for a given type is simply $P(match_t) = ss_t/dd_t$, where ss_t is the period t supply. Then, the expected number of switchers that are matched to an apartment in period t is:

$$lmis_t := P(match_t)lsw_t.$$

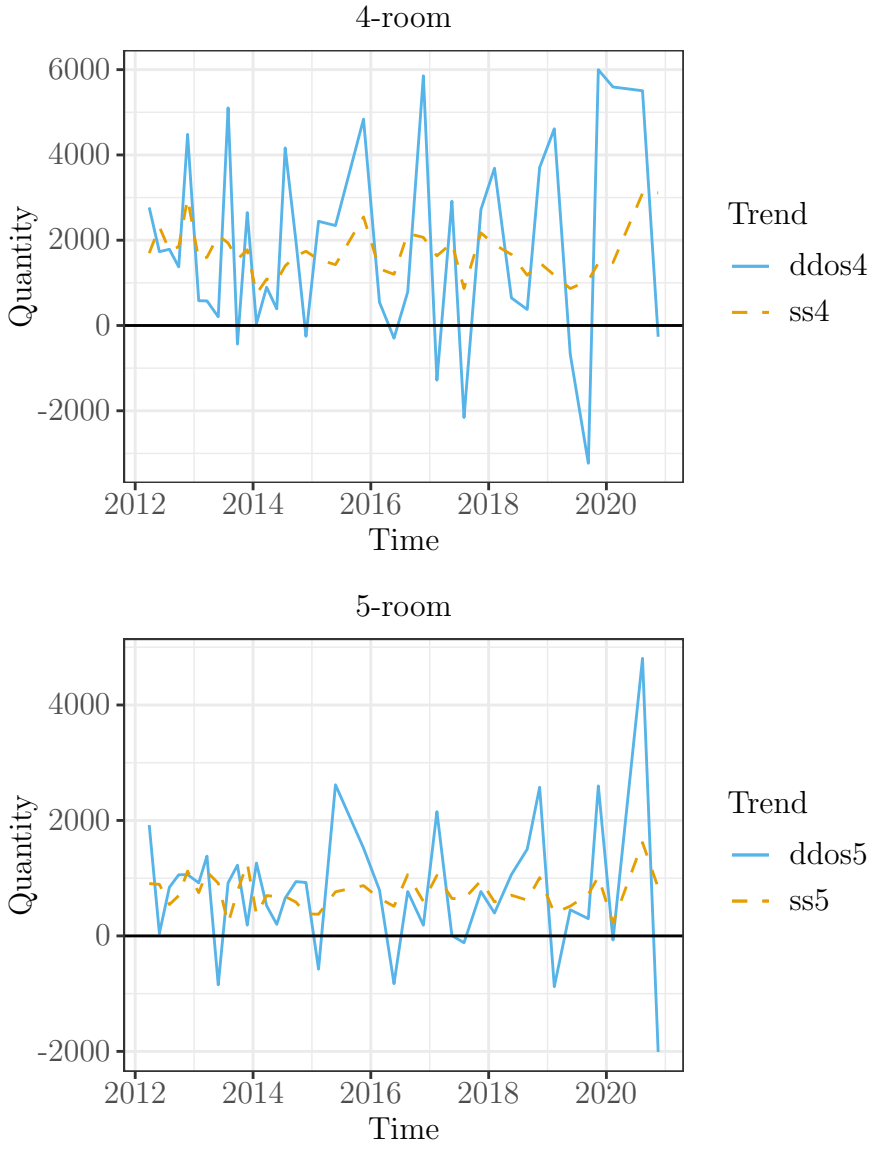


Figure 2.6.1: 4-room apartments are represented in the top figure, while 5-room apartments are represented in the bottom figure. Solid lines indicate the difference between previous oversubscription and current demand, $dd_t - os_{t-1}$. Dashed lines indicate the supply of apartments.

Since lsw_t is a lower bound for the total number of people misreporting in period t, the expected number of mismatches in period t must be at least as high as $lmis_t$.[35] We then normalize $lmis_t$ by dividing by the total supply of apartments of that type, and use $lmis_t/ss_t$ as our measure of mismatch. By taking the mean of $lmis_t$ for 4-room and 5-room developments in our time period, we find that at least 6.4% of 4-room apartments and 10.2% of 5-room apartments are allocated to households that would have preferred a different apartment under this measure. Given that this measure underestimates actual levels of mismatch, the potential to improve allocative efficiency is large.

2.7 Discussion

Beyond the public-housing setting, our study holds broader implications for any market where the mechanism designer controls the supply of goods available. In economics, it is generally taken for granted that markets can achieve an "efficient" solution.[36] However, this notion of efficiency inherently fails to address certain societal objectives, such as avoiding inequality or racial segregation. That is, an environment where one "dictator" receives everything is considered efficient by this notion. If a government wishes to address those objectives, it runs the risk of distorting the market. This can potentially lead to a worse outcome than if the government had done nothing. In particular, when market forces cannot ensure demand is met by supply, the government struggles to accurately measure household preferences. Indeed, many attempts to create centralized markets in the past have failed.

Through the BTO program, the Singaporean government has treated concerns regarding racial and socioeconomic inequality, while also incentivizing truthful reporting. These concerns have often failed to be addressed by private markets in other cities. For instance, the

[35]Since people are matched to their reports, the number of people who misreport and are matched is equal to the number of mismatches in a given period.

[36]For instance, the First Welfare Theorem states that in a competitive market with minor regularity conditions, any equilibrium is Pareto efficient.

Singaporean government wishes to ensure that adequate housing is affordable for all families. In a city-state like Singapore, where land is a scarce commodity, leaving housing to market forces has historically failed to ensure affordability. Indeed, in many large modern cities such as New York City or Hong Kong, housing costs have skyrocketed in recent decades. In practice, this exacerbates the effects of inequality, dividing the rich and the poor. Those who seek jobs or amenities provided by the city are often forced to commute long hours or settle for cramped living quarters. While we will abstract from these concerns in our model, they provide a motivation for centralization of the market.

We developed a model of allocation with endogenous supply. The model predicts that extreme shifts in preferences are underestimated by simple counts of applications. For instance, if a given apartment type experiences a commonly known surge in popularity, a portion of households will strategically apply for less desirable housing to avoid extended wait times. The model shows that market thickness improves the government's ability to match households to apartments correctly. One policy implication is that the government should delay the timing of housing developments to increase market thickness artificially.

We also provide a normative statement regarding the added benefit from endogenous supply as opposed to exogenous supply. In many situations, the mechanism designer has the ability to change the flow of incoming goods, potentially at cost. This model suggests that the gains from doing so can be quite large. Indeed, the gains are likely more than a naïve estimate of household preferences would suggest. This follows from a household incentive to not apply sincerely when supply is exogenous. Current mechanism design setups generally focus on allocating objects that arrive exogenously. Where supply can be adjusted—such as housing, transportation, and foodstuffs—the conclusions for optimal design differ starkly from settings where goods arrive randomly.

Appendix

2.A Appendix—Proofs

Proof of Lemma 15. As noted in the text, in state $s = (s_A, s_B)$, a type-A household prefers to apply sincerely if and only if $u(d_t = A, \theta_t = A, s) \geq u(d_t = B, \theta_t = A, s)$. Rearranging the utility function yields:

$$\begin{aligned} u(d_t = A, \theta_t = A, s) &\geq u(d_t = B, \theta_t = A, s) \\ h - cW_A(s) &\geq l - cW_B(s) \\ h - l \geq c[W_A(s) - W_B(s)] \implies \frac{h-l}{c} &\geq W_A(s) - W_B(s). \end{aligned}$$

An identical computation holds for type-B households. The claim follows. □

Proof of Proposition 7. As shown in the text, the first-best mechanism is the only mechanism that achieves 0 inefficiency. Under the first-best mechanism, the expected wait times for incoming households in state $(1, 0)$ are:

$$
\begin{aligned}
W_B(1,0) &= \sum_{i=0}^{\infty} (\frac{1}{2} \cdot \frac{1}{2})^i \\
&= \frac{4}{3}, \\
W_A(1,0) &= \sum_{i=0}^{\infty} \frac{1}{2} (\frac{1}{4})^i \\
&= \frac{2}{3}.
\end{aligned}
$$

Because $W_B(1,0) > W_A(1,0)$, type-A households will always be willing to apply sincerely in state $(1,0)$. Lemma 15 implies that type-B households will be willing to apply sincerely if $\gamma \geq W_B(1,0) - W_A(1,0) = \frac{2}{3}$. By design, the mechanism is symmetric in states $(1,0)$ and $(0,1)$. Therefore, in state $(0,1)$, type-B households are always willing to apply sincerely, and type-A households are willing to apply sincerely if $\gamma \geq \frac{2}{3}$. The mechanism never leaves the set of states $\{(1,0),(0,1)\}$, which implies that no other constraints are relevant. Then, when $\gamma \geq \frac{2}{3}$, the first-best mechanism is implementable, and if $\gamma < \frac{2}{3}$, no mechanism achieves 0 inefficiency. □

Proof of Lemma 16. Suppose not. Then, there exist two states s, s', both of which occur with positive probability, such that $s_\theta > s'_\theta$, but type-θ households enter queue-θ in state s and not state s'. However, $W_\theta(s) > W_\theta(s')$ implies that type-θ households must either strictly prefer to enter queue-θ in state s' or the other queue in state s. Therefore, in at least one of these states, households must have a profitable deviation. Then, the original strategy profile cannot be an equilibrium. □

Proof of Lemma 17. The wait time in a given state and queue depends on the probability of receiving an apartment immediately, as well as the transition probabilities. As such, we can recursively define the wait times $w_A^m(k, -(k-1))$, when $0 < k < m$, as the solution to the following system of equations using the transition matrix:

$$w_A^m(k,-(k-1)) = \frac{1}{2}\left[\frac{1}{2}\cdot\frac{k}{k+1}(1+w_A^m(k,-(k-1))) + \frac{1}{2}\cdot\frac{k-1}{k}(1+w_A^m(k,-(k-1)))\right] + \frac{1}{2}\left[\frac{1}{2}(1+w_A^m(k+1,-k)) + \frac{1}{2}(1+w_A^m(k,-(k-1)))\right],$$

$$w_A^m(m,-(m-1)) = \frac{1}{2}\frac{m-1}{m}(1+w_A^m(m-1,-(m-2))) + \frac{1}{2}(1+w_A^m(m,-(m-1))) = \frac{2m-1}{m} + \frac{m-1}{m}w_A^m(m-1,-(m-2)).$$

Next, note $w_A^m(m+1,-m)$ only occurs when a household has deviated and entered a queue that is already at capacity while the government simultaneously fails to build a type-A apartment. Its value comes directly from the previous equations:

$$w_A^m(m+1,-m) = \frac{1}{2}\frac{m}{m+1}(1+w_A^m(m,-(m-1))) + \frac{1}{2}(1+w_A^m(m+1,-m)),$$
$$= \frac{m}{m+1}(1+w_A^m(m,-(m-1))) + 1.$$

When $m=2$ this process generates the following system of equations:

$$w_A^2(2,-1) = \frac{1}{2}\cdot\frac{1}{2}(1+w_A^2(1,0)) + \frac{1}{2}(1+w_A^2(2,-1)),$$
$$w_A^2(1,0) = \frac{1}{2}[\frac{1}{2}\cdot\frac{1}{2}(1+w_A^2(1,0))] + \frac{1}{2}[\frac{1}{2}(1+w_A^2(2,-1)) + \frac{1}{2}(1+w_A^2(1,0))].$$

The solution to the system of equations is given by $w_A^2(2,-1) = \frac{5}{2}$ and $w_A^2(1,0) = 2$. Given the previous two values, $w_A^2(3,-2)$ can be computed and is equal to $\frac{10}{3}$. Since $\kappa = 2$ is an equilibrium only if both $IC_\kappa(\kappa)$ and $IC_\kappa(\kappa-1)$ are satisfied, this implies the threshold equilibrium $\kappa = 2$ requires $\gamma \in [\frac{5}{2}, \frac{10}{3}]$. □

Proof of Proposition 8. To begin, note that as the threshold κ increases, the average wait time does as well. To see why, consider the exact difference between a threshold κ strategy

and a threshold $\kappa + 1$ strategy. In particular, the different arises when there are $\kappa - 1$ households in a given queue and the incoming household is of that type. Under the threshold $\kappa - 1$ strategy, the incoming household enters the empty queue and is immediately allotted an apartment. Under the threshold strategy, the incoming household enters the full queue, further increasing all present households wait times. Therefore, the ex-ante expected wait time is greater under a threshold equilibrium in any state.

Then, consider the $IC(\kappa)$ constraint. Since each $w_A^\kappa(s) > w_A^{\kappa-1}(s)$ and $w_A^\kappa(s+1) > w_A^\kappa(s)$, it follows that both $w_A^\kappa(\kappa, -(\kappa-1)) > w_A^{\kappa-1}(\kappa-1, -(\kappa-2))$ and $w_A^\kappa(\kappa+1, -\kappa) > w_A^{\kappa-1}(\kappa, -(\kappa-1))$. Last, a direct comparison of $IC_\kappa(\kappa-1)$ and $IC_{\kappa-1}(\kappa-2)$ then implies that the solution to the first must be larger than the solution to the section. □

Proof of Lemma 19.

1. To begin, Lemma 18 implies that in state $(2, -1)$ all incoming households strictly prefer to enter queue B. Then, the state space is bounded by $(2, -1)$ and $(-1, 2)$, and a finite number of states are recurrent. Therefore, at least one steady state exists by standard results in dynamics.

2. For the steady state to fail to be unique, there must be an absorbing state separating either $(2, -1)$ from $(1, 0)$ or $(1, 0)$ from $(0, 1)$. A mechanism that remains in $(2, -1)$ indefinitely cannot be optimal, since such a mechanism would have both allocation and unassignment inefficiency higher than the pooling mechanism. On the other hand, a mechanism that fails to transition between $(1, 0)$ and $(0, 1)$ implies uniqueness up to queue symmetric steady states unless the mechanism has different steady states when beginning with a type-A household as opposed to a type-B household. However, this cannot be optimal due to the symmetry of the problem. Suppose a mechanism generated two steady states that were not queue symmetric, which yielded different efficiencies. Then, because both must involve equilibrium behavior of the part of the

household, the government could simply use a strategy equivalent to that of the steady state with lower inefficiency everywhere. Through doing so, household behavior must still be equilibrium behavior, and inefficiency would have been lowered, proving that the original mechanism was not optimal.

□

Remark 2. *When the government utilizes μ_q with q^* an incoming type-A household in state $(1,0)$ does not have incentive to deviate.*

The following simple result will prove useful for the proof of Proposition 9.

Lemma 21. *The maximal equilibrium allocation inefficiency is $\frac{1}{2}$.*

Proof of Lemma 21. Let $W_A(s) - W_B(s) < \gamma$ for some state s, then each type-A household strictly prefers to enter queue A. As such, allocation inefficiency in s is at most $\frac{1}{2}$ since all type-A households apply sincerely. When $W_B(s) - W_A(s) < \gamma$ a similar argument holds for type-B households. Then, in any state, allocation inefficiency is at most $\frac{1}{2}$. Last, since overall allocation inefficiency is a weighted average of the allocation inefficiency in each state, the overall allocation inefficiency must be at most $\frac{1}{2}$. □

Proof of Proposition 9. Lemma 18 allows us to focus on mechanisms with state spaces restricted to states $(2,-1)$ through $(-1,2)$. Since incoming households always enter the short queue in $(2,-1)$ and $(-1,2)$, the state can never exceed $(2,-1)$ or $(-1,2)$. In state $(2,-1)$, allocation inefficiency is $\frac{1}{2}$ and vacancy inefficiency is 1. Note that both inefficiency values are weakly larger than inefficiency in state $(1,0)$ by Lemmas 18 and 21. It follows that the government aims to minimize the proportion of time spend in states $(2,-1)$ and $(-1,2)$.

Importantly, if allocation inefficiency is lower than $\frac{1}{2}$, the difference in expected wait times in states $(1,0)$ and $(0,1)$ can be at most γ. That is, $|W_A(1,0) - W_B(1,0)| \leq \gamma$, otherwise type-$A$ households or type-B households strictly prefer to not apply sincerely.

We then optimize over all such possible mechanisms, and show that μ_q minimizes inefficiency. Firstly, note that if a type-B household in state $(1,0)$ strictly prefers to apply sincerely, $\Phi^A(1,0)$ can be increased while ensuring type-B still has incentive to continue applying sincerely. Furthermore, increasing $\Phi^A(1,0)$ reduces the probability that the mechanism enters state $(2,-1)$. Then, the optimal mechanism must set $\gamma = W_B(1,0) - W_A(1,0)$.

We proceed by considering a mechanism more general than μ_q. There are two primary differences. First, the government sometimes builds the wrong apartment in state $(2,-1)$, that is $\Phi^B(2,-1)$ is not necessarily 0. Second, type-B households are incentivized to apply sincerely with probability below 1 in state $(1,0)$. Let $x^B(1,0)$ be the probability with which a type-B household enters queue A in state $(1,0)$.

The level of inefficiency generated by this mechanism:

$$\frac{2(1+x^B(1,0))q - \alpha(q + x^B(1,0)(-2+q+2\Phi^B(2,-1)))}{2(2+q+x^B(1,0)q - 2\Phi^B(2,-1))}.$$

Taking the derivative of the above equation with respect to $\Phi^B(2,-1)$ yields the following, where A is constant with respect to $\Phi^B(2,-1)$:

$$-\frac{4x^B(1,0)(1+x^B(1,0))(-2+\alpha+\alpha x^B(1,0))(-3+x^B(1,0)+\gamma(-1+x^B(1,0))(-1+\Phi^B(2,-1))+2\Phi^B(2,-1))}{A^2} + \frac{\sqrt{-4(2+\gamma(-3+x^B(1,0)))x^B(1,0)(-1+\Phi^B(2,-1))+(-3+x^B(1,0)+\gamma(-1+x^B(1,0))(-1+\Phi^B(2,-1))+2\Phi^B(2,-1))^2}}{A^2}.$$

where the exact value of A is irrelevant since A^2 must be positive. The numerator is always negative, note that it can be rewritten as $B + \sqrt{4(2+\gamma(x^B(1,0)-3))x^B(1,0)(1-\Phi^B(2,-1))+B^2}$, where $B = 4x^B(1,0)(1 + x^B(1,0))(-2+\alpha+\alpha x^B(1,0))(-3+x^B(1,0)+\gamma(-1+x^B(1,0))(-1+\Phi^B(2,-1))+2\Phi^B(2,-1)$. However, $\gamma \leq \frac{2}{3}$ implies that $4(2+\gamma(x^B(1,0)-3))x^B(1,0)(1-\Phi^B(2,-1)) \geq 0$ and therefore $\sqrt{4(2+\gamma(x^B(1,0)-3))x^B(1,0)(1-\Phi^B(2,-1))+B^2} \geq -|B|$. Then the derivative of

inefficiency with respect to $\Phi^B(2,-1)$ is always positive, and it is optimal to minimize $\Phi^B(2,-1)$ by setting it equal to 0.

We then proceed by taking the derivative of inefficiency with respect to $x^B(1,0)$. The derivative is:

$$\frac{8x^B(1,0)(-3+\gamma+x^B(1,0)+4\alpha x^B(1,0)-\gamma x^B(1,0)}{A^2}$$
$$+\frac{\sqrt{4(2+\gamma(-3+x^B(1,0)))x^B(1,0)+(-3+\gamma+x^B(1,0)-\gamma x^B(1,0))^2)}}{A^2}.$$

Where again A is an constant we omit because A^2 must be positive. Furthermore, a similar line of reasoning works here to imply that this derivative is positive and therefore $x^B(1,0)$ should be set to 0 in an optimal mechanism. We summarize these two findings in the following Lemma.

Lemma 22. *Under Assumption 2, if the pooling mechanism is not optimal, then the optimal mechanism sets* $x^B(1,0)=0$ *and* $\Phi^B(2,-1)=0$.

Then, Lemma 22 implies that the last variable to consider is the value of q. As shown above, q must be minimized subject to the constraint that type-B households in state $(1,0)$ apply sincerely. Therefore, μ_q is optimal anytime type-B households in state $(1,0)$ have incentive to apply sincerely. Then, given the previous computation of q^*, μ_q is optimal. □

We proceed by determining the wait times for μ_q. These will prove useful for determining the region in which μ_q is optimal.

Wait time computations for two-state mechanism:

$$w_A(1,0)=\frac{1}{2}\cdot q[1+w_A(1,0)]+\frac{1}{2}\cdot q[1+w_A(2,-1)]+\frac{1-q}{4}\cdot[1+w_A(1,0)],$$
$$w_A(2,-1)=\frac{1}{2}[1+w_A(1,0)].$$

Solving yields $w_A(1,0) = \frac{4q+1}{3-2q}$ and $w_A(2,-1) = \frac{2+q}{3-2q}$.

Then, in order for households to apply sincerely, Lemma 15 implies the following constraints on γ:

$$\gamma \geq (1-q)[1+w_A(1,0)] - [q(1+w_A(2,-1)) + \frac{1-q}{2}(1+w_A(1,0))],$$
$$\gamma \geq \frac{1-q}{2}(1+w_A(1,0)) - q(1+w_A(2,-1)).$$

□

Lemma 23. *Under μ_q, the level of allocation inefficiency is $\frac{2-3\gamma}{14(2-\gamma)}$ and the level of unassignment inefficiency is $\frac{2-3\gamma}{7(2-\gamma)}$.*

Proof of Lemma 23. Since μ_q only generates inefficiency in states $(2,-1)$ and $(-1,2)$, the total inefficiency is directly proportional to the proportion of time spent in those two states. Under the optimal mechanism, the proportion of time in states $(2,-1)$ and $(-1,2)$ is $\frac{2-3\gamma}{7(2-\gamma)}$. With $\frac{1}{2}$ probability, a household fails to apply sincerely, and there is always a vacant apartment in $(2,-1)$ or $(-1,2)$. Therefore, the allocation and vacancy inefficiencies are given by $\frac{2-3\gamma}{14(2-\gamma)}$ and $\frac{2-3\gamma}{7(2-\gamma)}$. □

Proof of Theorem 3. Proposition 9 shows that either μ_q or μ_p is optimal. We proceed by comparing the inefficiencies generated by the two mechanisms. Lemma 23 directly provides the individual inefficiencies for μ_q. Summing them with weights from the government's objective implies that the total level of inefficiency is:

$$\frac{\alpha}{2}\frac{2-3\gamma}{7(2-\gamma)}(1-\alpha)\frac{2-3\gamma}{7(2-\gamma)} = (1-\frac{\alpha}{2})\frac{2-3\gamma}{7(2-\gamma)}.$$

The level of inefficiency under the pooling mechanism is $\alpha\frac{1}{2}$; there is no vacancy inefficiency and half of the households fail to apply sincerely. Comparing the two and solving for α yields the threshold $\frac{2-3\gamma}{8-5\gamma}$. □

Corollary 3. *The threshold which dictates whether μ_p or μ_q is optimal, $\frac{2-3\gamma}{8-5\gamma}$, is decreasing in γ.*

Proof of Corollary 3. We take the derivative of the threshold in Theorem 3 with respect to γ.

$$\frac{\partial \frac{2-3\gamma}{8-5\gamma}}{\partial \gamma} = \frac{-14}{(8-5\gamma)^2} < 0.$$

□

Proof of Corollary 1. A computation similar to the above implies that the government prefers μ_q when $2(\gamma)^2 - 9\gamma + 2 < 0$ or when $\gamma > \frac{9-\sqrt{65}}{4} \approx .234$. □

Proof of Proposition 10. We begin by computing the difference in the expected wait times under the implementation of the first-best mechanism.

$$\begin{aligned}
W_B(1,0) &= p \cdot \frac{1}{2} + (p \cdot \frac{1}{2})^2 + \dots \\
&= \frac{1}{1-\frac{p}{2}}, \\
W_A(1,0) &= \frac{1}{2} + \frac{1}{2}\frac{1-p}{2} + \frac{1}{2}\frac{1-p}{2}\frac{p}{2} + \dots \\
&= \frac{1}{2}\frac{3-2p}{2-p}, \\
W_B(1,0) - W_A(1,0) &= \frac{1+2p}{2(2-p)}.
\end{aligned}$$

Lemma 15 then implies that this mechanism is an equilibrium only if γ bounds the differences in expected wait times. □

Proof of Lemma 20. The derivative of equation 2.4 with respect to p is given by:

$$\begin{aligned}\frac{\partial[w_B(1,0) - w_A(1,0)]}{\partial p} &= \frac{2(2(2-p)) - (1+2p)2(-1)}{4(2-p)^2}, \\ &= \frac{5}{2(2-p)^2} > 0.\end{aligned}$$

Then, since the derivative is positive, as p increases, so must γ in order to incentivize households to apply sincerely. □

Lemma 24. *When $N = 2$, the first-best can be implemented iff $\gamma \geq \frac{16}{17}$.*

Proof of Lemma 24. Wait times under the first-best mechanism, when $N = 2$ can be described according to the following system of equations:

$$\begin{aligned}w_A(2,0) &= \frac{1}{2} \cdot \frac{2}{3}(1 + w_A(2,0)) + \frac{1}{2} \cdot \frac{1}{2}(1 + w_A(1,1)), \\ w_A(1,1) &= \frac{1}{2}\left[\frac{1}{2} \cdot \frac{1}{2}(1 + w_A(1,1)) + \frac{1}{2}(1 + w_A(2,0))\right] + \frac{1}{2} \cdot \frac{1}{2}(1 + w_A(1,1)).\end{aligned}$$

The solution to the above system of equations is $w_A(2,0) = \frac{25}{17}, w_A(1,1) = \frac{27}{17}$, the remainder of the proof follows directly from surrounding arguments. □

Proof of Corollary 2. Referring back to Proposition 7, under $N = 1$ the first-best could only be implemented when $\gamma \geq \frac{2}{3}$. The threshold, $\frac{16}{17}$ from Lemma 24 is higher, and therefore imposes a stricter condition on γ. □

Proof of Theorem 4. In order to show that the $T = 2$ batching mechanism is optimal, we proceed in a similar manner to the proof of the optimality of the q^* mechanism. As shown in the main body of the chapter, whenever the $T = 2$ mechanism encourages households to

apply sincerely, the $T = 2$ mechanism dominates any mechanism with $T > 2$. This follows from the fact that if $T > 2$, then unassignment inefficiency is 1 at a minimum.

Since we have already characterized the optimal mechanisms with $T = 1$, we proceed by proving the $T = 2$ mechanism is optimal among all mechanisms with $T = 2$. Note that for reasons similar to that in the q^* mechanism, it is never optimal to build type-B apartments in state $(2, -1)$. Formally, the problem is the following:

$$V(\mu) = \min_{\Phi^A(1,0) \in \Delta\{0,1\}^2} \alpha m + (1 - \alpha) v(\mu).$$

Standard optimization techniques imply that $\Phi^A(1, 0)[0, 2] = 0$. Last, we determine the optimal value for q under the $T = 2$ mechanism. We can compute expected wait times for a household already in queue A as the solution to the following pair of equations:

$$\begin{aligned} w_A(1,0) &= \frac{1}{4}\left[\frac{2q}{3}(2 + w_A(2,-1)) + (1-q)\frac{1}{3}(2 + w_A(1,0))\right] + \frac{1}{2} \cdot \frac{q}{2}(2 + w_A(1,0)), \\ w_A(2,-1) &= \frac{1}{4} \cdot \frac{1}{2}(2 + w_A(2,-1)) + \frac{1}{2} \cdot \frac{1}{3}(2 + w_A(1,0)). \end{aligned}$$

Algebra yields the following solutions with respect to q:

$$\begin{aligned} w_A(1,0) &= \frac{42 + 196q}{231 - 50q}, \\ w_A(2,-1) &= \frac{162 + 4q}{231 - 50q}. \end{aligned}$$

Conditional on the state, the difference in wait times is:

$$w_A(1,0) - w_B(1,0) = \frac{1}{6}\frac{-1725 + 2357q + 62q^2}{-231 + 50q}$$
$$w_A(2,-1) - w_B(2,-1) = \frac{453 - 142q}{1386 - 300q}$$

This batching mechanism minimizes the maximum of the two differences when $q = \frac{3}{124}(-833 + \sqrt{753905})$, implying that households apply sincerely for γ above $\frac{-453+\frac{213}{62}(-833+\sqrt{753905})}{6(-231+\frac{75}{62}(-833+\sqrt{753905}))} \approx .294$. □

2.B Queue Hopping

Under the real BTO mechanism, households can freely switch queues between application cycles. In this section, we show that the ability to do so does not change the optimality of the mechanisms presented in the main body of the chapter. Under the mechanisms presented in the main body of our chapter, no household wishes to change its queue at the beginning of a period. We will formally prove this below.

To provide intuition for the results that follow, note that the incoming household in a given period always has more information than households currently in a queue. In particular, if a household in a given queue is willing to change queues, then all incoming households will strictly prefer to enter the queue that that household swapped to. As a result, correct allocation is impossible when households switch, causing inefficiency in mechanisms that take advantage of swapping to be high. This insight is specific to stationary markets, wherein households continually arrive to be matched. In a static market, where no new households arrive, swapping could very well be part of an optimal mechanism.[37]

[37] Consider a simple static setting with two households and one apartment of each type. If both households initially apply to the same queue, the household that loses the resulting lottery would prefer to switch queues.

The timing is as follows. In every period when the incoming household would enter a queue, all present households simultaneously also choose a queue to enter. That is, households choose the queue they wish to enter without knowledge of the queue other households are about to enter. We denote the period τ choice of the household that arrived in period t by d_t^τ.

In this setting, we need to define the state variables with more care. Previously, once a household had entered a given queue, its actual type became irrelevant for both the household and the government. Since it could not switch, incoming households only cared about its selected queue, and the government could no longer influence that household's match. However, now household's can switch queues between periods, implying that tracking their type is important. Furthermore, households can carry beliefs regarding other household's types. This is important because their type informs their switching probabilities. In practice, this behavior seems unrealistic. While households might observe application rates, they will not track applications on a household level. We then make the following assumption of naïvete, households only observe the length of each queue, they do not observe the types of other households or the history of household level applications.

Assumption 3. *Households are Markovian—their strategies are a function of the state.*

It is trivial to show that whenever the first-best mechanism was implementable in the original model it remains optimal in the new setting, namely when $\gamma \geq \frac{2}{3}$. To see why, recall that the first-best mechanism instructed households to report truthfully, and always built an apartment matching the type of the household currently present. Then, a household in the queue that matches its type never wishes to change its queue. If it were to do so, that household could not receive an apartment this period, and furthermore ensures the government will build the "wrong" apartment next period. We then focus our attention on the case where the first-best is not implementable, namely when Assumption 2 holds.

The new incentive constraints implied by the ability to switch are never violated by the pooling mechanism. Switching queues merely ensures that the household cannot receive an apartment in the given period, and will not change the types of apartments the government builds in the future.

We will show that the natural translation of Section 2.4.2's two-state mechanism continues to be an equilibrium in household strategies. As before, μ_q is optimal whenever the pooling mechanism or first-best mechanism are not optimal.

Proposition 11. *Households never switch their queues under μ_q.*

Furthermore, if neither the pooling nor the first-best mechanisms are optimal, then μ_q is optimal.

Proof. We first show formally that no household wishes to change its queue under μ_q on the equilibrium path. We translate μ_q to the current setting by fixing the government's strategy and incoming household's strategies. Old households reenter their queue in every period. Notably, in state $(2, -1)$ incoming households report type B independently of their type. In addition, we require that households do not swap the queue they have entered in later periods. This generates two new constraints, one for each possible state.

In state $(1, 0)$, it is easy to see that the type-A household does not wish to switch queues. q, the probability with which an apartment of type B is built, was selected to render incoming type-B households indifferent between the two queues. Relative to incoming type-B households, current type-A households expect less competition in queue A and prefer to match correctly. Then, if an incoming type-B household is indifferent, current type-A households strictly prefer to remain in queue A.

Similarly, in state $(2, -1)$ current type-A households expect the same wait time independent of the queue they select. All incoming households select queue B and the government always builds a type-A apartment. By remaining in queue A, they compete with one other household for one apartment, the other household from the previous period. By switching

to queue B, they compete with one other household for one apartment as well, the incoming household in this case. Furthermore, in the event the household does not receive an apartment in the current period, it prefers state $(1,0)$ to state $(0,1)$.

It remains to prove that μ_q is optimal in the current environment. Proposition 9 implies that μ_q is optimal among mechanisms that do not utilize swapping on the equilibrium path. Next, suppose a mechanism involved households swapping queues with probability k, in state $(2,-1)$, where $0 < k < 1$. The willingness to randomize would imply that present households are indifferent between the two queues. Such a mechanism must fail to improve upon μ_q with respect to allocative inefficiency in state $(2,-1)$. To see why, note that under the two state mechanism, households always sincerely apply except when in state $(2,-1)$ in which case they are immediately matched and exit the market. It remains to show that such a mechanism cannot reduce the proportion of time spent in state $(2,-1)$ through swapping in either state.

Suppose a mechanism involved households swapping their queue in state $(1,0)$ with probability $0 < k < \frac{1}{2}$, while incoming households sincerely apply. The willingness of the present household to randomize implies they are indifferent between the two queues. This is despite the fact that the present household knows there is a $\frac{1}{2}$ chance that the incoming household is of type-B and enters queue B. However, the incoming household of type A then must strictly prefer to enter queue B. To see why incoming households prefer to enter queue B, note that there is a $k < \frac{1}{2}$ chance that the present household enters queue B. If the present household was indifferent between the two queues, then the incoming household must have a strict preference for queue B. Then, allocative inefficiency under such a mechanism is equal to that of the pooling mechanism.

Last, suppose instead that $\frac{1}{2} \leq k < 1$. For households to be willing to switch queues, either $\Phi^A(1,0) < \Phi^B(1,0)$ or incoming households must not be applying sincerely. In both

cases, allocative inefficiency is comparable to that of the pooling mechanism, which by assumption is suboptimal. □

2.C Suboptimality of Markovian Mechanisms

While throughout this chapter we focus on Markovian mechanisms, we note that doing so is with potential loss of generality. To provide intuition for why the optimal mechanism is not necessarily Markovian, we construct a mechanism that improves upon the oversubscription mechanism developed in Section 2.5.2 when $N = 2$. However, while the optimal probabilities change, the structure of the optimal mechanism is similar to the mechanisms utilized throughout the main body of the chapter.

We weaken the Markovian assumption and allow the government to condition Φ_t^A not only on s_t but also on s_{t-1}. That is, the government can "reward" households from the previous period that have waited. We focus on allocation in state $(1, 1)$, where incentive constraints are lax for incoming households of either type. The government can then modify the allocation probabilities in order to better incentivize sincere applications in states $(2, 0)$ and $(0, 2)$. Altering notation slightly, we use $\Phi_n^A(s)$ to indicate the probability with which the government builds a type-A apartment in state (s) when the previous state was $(n, 2-n)$.

Again, previous reasoning implies that in order to implement the first-best, if the state is $(2, 0)$ or $(0, 2)$ the government must always build a type-A or type-B apartment respectively. However, in state $(1, 1)$, the government may build either apartment type freely. We will suppress the state when indicating $\Phi_n^A(1, 1)$, since $\Phi_n^A(2, 0) = 1$ and $\Phi_n^A(0, 2) = 0$. Utilizing our above methodology, we proceed by computing the expected waiting times for a household in each queue. We will slightly alter our notation to accommodate the new conditioning of w. Let $w_{T,n}(s)$ indicate the expected waiting time for a household in queue T when the previous state was $(n, 2-n)$ and the current state is s. Similarly, $W_{T,n}(s)$ indicates waiting times for the incoming household.

In state $(2,-1)$ all incoming households are allocated to queue B. If there are different wait times depending upon their reported type, they are incentivized to misreport their type in state $(2,-1)$. In the following equations, we drop unnecessary notation. Then, the following system of equations defines expected wait time for an individual already in queue A:

$$\begin{aligned}
w_{A,n}(2,0) &= \frac{1}{2}\cdot\frac{2}{3}(1+w(2,0)) + \frac{1}{2}\cdot\frac{1}{2}(1+w^2(1,1)),\\
w_A^n(1,1) &= \frac{1}{2}[\Phi_n^A(1,1)\frac{1}{2}\cdot(1+w^1_{(a,A)}(1,1)) + (1-\Phi_n^A)(1+w(2,0))]\\
&\quad + \frac{1}{2}\cdot(1-\Phi_n^A)(1+w^1_{(b,A)}(1,1)).
\end{aligned}$$

We compute the the following expected wait times for each state, and use the symmetry of the problem to deduce that $\Phi_1^A = 1-\Phi_1^B$ and $\Phi_2^B = \Phi_A^0$:

$$\begin{aligned}
w(2,0) &= \frac{76-37\Phi_1^A-15\Phi_2^B}{44-29\Phi_1^A+9\Phi_2^B},\\
w^2(1,1) &= \frac{100-31\Phi_1^A-61\Phi_2^B}{44-29\Phi_1^A+9\Phi_2^B},\\
w_b^1(1,1) &= \frac{3(12+11\Phi_1^A+\Phi_2^B)}{44-29\Phi_1^A+9\Phi_2^B},\\
w_a^1(1,1) &= \frac{100-95\Phi_1^A+3\Phi_2^B}{44-29\Phi_1^A+9\Phi_2^B},\\
w^0(1,1) &= \frac{36-31\Phi_1^A+67\Phi_2^B}{44-29\Phi_1^A+9\Phi_2^B}.
\end{aligned}$$

The next step is to then compare expected wait times for individuals in different settings, we begin with a type-B household in state $(2,0)$. The household knows that the supplied apartment will always be type-A, and must decide if waiting is worthwhile:

$$
\begin{aligned}
W_B(2,0) &= 1 + w_b^2(1,1), \\
W_A(2,0) &= \frac{2}{3}(1 + w(2,0)), \\
W_B(2,0) - W_A(2,0) &= \frac{1}{3} + w_b^2(1,1) - w(2,0), \\
\implies \gamma &\geq \frac{16(\Phi_1^A - 5\Phi_2^B)}{-44 + 29\Phi_1^A - 9\Phi_2^B}.
\end{aligned}
$$

An identical restriction can be computed in state $(0,2)$ by the symmetry of the problem.

Last, we compute the differences in wait times in state $(1,1)$ dependent upon the previous period's state and report.

When the previous state and report are s and r:

$$
\begin{aligned}
W_{(r,B),n}(1,1) &= \Phi_n^A(1 + w(2,0)) + (1 - \Phi_n^A)\frac{1}{2}(1 + w_A^1(1,1)), \\
W_{(r,A),n}(1,1) &= \frac{\Phi_n^A}{2}(1 + w_A^1(1,1)) + (1 - \Phi_n^A)(1 + w(2,0)).
\end{aligned}
$$

In effect, the wait times are the same except Φ_n^A is replaced with $(1 - \Phi_n^A)$, hence why setting both equal to $\frac{1}{2}$ was previously optimal. In this instance, since this constraint in state $(1,1)$ is lax, changing these values may weaken constraints in states $(2,0)$ and $(0,2)$.

This generates the following differences in wait times:

$$
\begin{aligned}
W_{(r,B)}^2(1,1) - W_{(r,A)}^2(1,1) &= \frac{4(\Phi_1^A + 3(-4 + \Phi_2^B))(-1 + 2\Phi_2^B)}{44 - 29\Phi_1^A + 9\Phi_2^B}, \\
W_{(r,B)}^1(1,1) - W_{(r,A)}^1(1,1) &= \frac{(4(-1 + 2\Phi_1^A)(\Phi_1^A + 3(-4 + \Phi_2^B))}{44 - 29\Phi_1^A + 9\Phi_2^B)}.
\end{aligned}
$$

Then, we proceed by minimizing these differences in wait times through selection of C. Rudimentary optimization yields $\Phi_1^A = .28 = \Phi_2^B$ and $\gamma \geq .48$. This is a substantial improvement upon the minimal γ for $N = 2$.

2.D Labelling

In this section, we consider the impact of relabelling one of the apartment types on the set of achievable outcomes. In particular, suppose that apartment type-B was relabelled as two separate types $B1$ and $B2$. All preferences are maintained, that is if a household is of type B, then it gains h in utility from receiving either $B1$ or $B2$. Similarly, a type-A household gain l in utility from receiving either $B1$ or $B2$. The government can choose any of the three housing types when building an apartment.

Consider what happens if the government implements the previous first-best mechanism, replacing any instance of a type-B apartment with apartment type $B1$. Similarly, any household that would have previously applied for a type-B apartment instead applies for apartment type $B1$. Households still always wish to apply sincerely, the existence of an extra housing type that will never be constructed provides no incentive to deviate. Then, the previous strategy profile remains an equilibrium. Furthermore, suppose the mechanism involved some form of randomization between the two different housing types. Then, with positive probability, an apartment will lie vacant, implying that the first-best cannot be achieved.

Lemma 25. *The first-best in the standard two type case can be achieved if and only if it can also be achieved under relabelling.*

2.E M Apartment types

Here we consider the impact of actually increasing the number of apartment types. Suppose there are now $|\Theta| = 3$ different types. Households are still born with a type in Θ. If they receive an apartment of their type they gain h in utility, if they receive a different apartment type they receive l.

Consider the first-best outcome. Begin with μ_{fb} from the $|\Theta| = 2$ case. It is simple to see that households wish to apply sincerely here. A household that did not receive an apartment in the previous period never has incentive to misreport because it knows an apartment of its type will be built in the current period. Incoming households of a different type face the exact same incentive constraint as in the original model, and so face no incentives to switch. Last, incoming households that match the current type never have incentive to switch under this mechanism, and so it remains an equilibrium.

Furthermore, no other mechanism can achieve the first-best, in doing so they must risk positive probability of vacancies. Last, this argument holds for all $|\Theta| > 2$, the above arguments did not utilize the fact that $|\Theta| = 3$.

Lemma 26. *The first-best in the standard two type case can be implemented if and only if it can also be achieved in the m type case.*

Chapter 3

The Dynamic Allocation of Public Housing: Policy and Spillovers

3.1 Introduction

This chapter considers the design of public housing policy. Governments build public housing, then rent or sell these apartments below market prices.[1] In these programs, governments seek to minimize apartment vacancies, lower wait times for needy households, ensure price stability in related housing markets, among other objectives. Insights from market design suggest that tweaking allocation rules may help achieve these outcomes.[2] Meanwhile, popular discourse emphasizes building more housing. How do these two policies impact our desiderata, and how does adding supply interact with changing the rules? To answer these

[1]Public housing is a common form of housing assistance worldwide. For instance, 45% of residents in Hong Kong live in public housing (Hong Kong Housing Authority, 2017), as do 80% in Singapore (Yuen, 2021). Public housing allocation has been studied carefully in the contexts of Cambridge, MA (Waldinger, 2021) and in Amsterdam (Van Dijk, 2019). Unlike in the latter paper, we do not speak to the effect of selection into/out of public housing on life outcomes, like labor market participation, educational attainment, and criminal behavior (see, e.g., Chyn (2018)).

[2]For instance, better rules to allocate kidneys can improve patient welfare by an amount "equivalent to an 18.2% increase in donor supply" (Agarwal et al., 2021a). For a review of why governments disburse transfers in kind, see Currie and Gahvari (2008). For a theoretical analysis of rationing housing assistance via lotteries and waitlists, see Arnosti and Shi (2020). For why the Singapore government provides public housing, see Subsection 3.3.2.

questions, we form a dynamic choice model and estimate it on novel data from Singapore. In counterfactual simulations, we find that better rules and adding supply complement each other. Our proposed mechanism reduces vacancies and wait times, but raises prices on the aftermarket for apartments. To relieve this upward pricing pressure, building more apartments can help, while also keeping wait times low. However, in the existing mechanism, at the margin, adding supply is ineffective at reducing wait times.

We study a large-scale public housing program, the Singapore Build-to-Order mechanism (BTO), described in Section 3.3. In Singapore, the government supplies 80% of the housing stock, seeking to boost fertility, ensure racial and socioeconomic mixing, and, most importantly, to encourage homeownership. In each quarter, different developments are offered within a period and across time; households may express a preference for at most one such development. These apartments are rationed by lottery and can be resold by their occupants on an aftermarket. The government subsidizes matched applicants in two ways: implicitly, by pricing apartments below their counterparts on the secondary market; and explicitly, by giving the poor a further discount. Ultimately, BTO is a clean setting for us to capture how young households behave when they apply for public housing.

In Section 3.6, we provide a theoretical example showing that injecting supply may exacerbate the problem of rational delay, harming applicants in equilibrium. In the current mechanism, young households are incentivized to wait for better apartments arriving next period, leaving today's apartments empty.[3] This example suggests the following questions: We can precisely target rational delay using a mechanism that eliminates intertemporal risk. How does our proposed mechanism impact vacancies, wait times, and prices on the secondary market? Moreover, what are the effects of a small supply increase on these outcomes, both in the actual mechanism and our proposed one?

[3]In the February 2023 BTO exercise, a 23-year-old university student stated that she "chose to sit out the February exercise because they preferred to save their application for May to try for a unit in Serangoon," a popular neighborhood (Ong, 2023).

To answer these questions, we consider two policy interventions. To shut down intertemporal risk, we simulate a dynamic version of Random Serial Dictatorship, in which households express preferences over housing units that may be available in the future. Meanwhile, to implement a small supply increase, we build 10% more apartments in oversubscribed developments.

To evaluate the effects of these interventions, in Section 3.7, we set up a dynamic choice model over housing lotteries, which we take to the data. In this model, new housing developments arrive exogenously. Young households are born rich or poor. In each BTO cycle, given the developments on offer, as well as the prices of apartments on the secondary market, these households can select one development, take an apartment on the secondary market, or wait till the next period. Waiting can be rational because, in the mechanism, reneging on an offered apartment is costly. These potential applicants trade off applying to a safe development today, trying for a risky but attractive development, and waiting for "better" developments in the next cycle. Finally, to close the model, on the secondary market, existing owner-occupiers observe the price of their apartment and decide whether to continue living in it or to sell it off and exit the market. These owner-occupiers trade off selling their apartment today against selling for a higher price tomorrow. In equilibrium, the BTO mechanism clears by lottery odds, while the secondary market clears by price.

We take this model to the rich data described in Section 3.4. In the primary market, we see every offered development, as well as its characteristics and the number of young households who applied to it. Moreover, on the secondary market, we see all transactions that have transpired.

With these data, we undertake descriptive analysis in Section 3.5, exhibiting development characteristics and dynamics that influence BTO demand. More young households apply to developments that are closer downtown and to the nearest metro stop, among other amenities. Moreover, dynamics matter. In any given BTO cycle, average distances to nearby

amenities do not predict total applications. Yet, when comparing pairs of cycles adjacent in time, more attractive choice sets do predict more applications. These observations bridge a gap between the data and our analysis of new rules and larger supply, but do not close it: Equilibria in our counterfactuals not only feature dynamics, but also endogenous lottery odds and aftermarket prices.

Cognizant of the limitations of descriptive analysis, we estimate our proposed model for the 2012-2015 BTO cycles in Section 3.8. On the supply side, we exploit the fact that selling off one's apartment is a terminal action (Arcidiacono and Miller, 2011), so the price elasticity of supply can be recovered using a static instrumental variables (IV) regression. This elasticity is identified by variation across mover shares and prices by apartment size by neighborhood. To deal with unobserved supply shifters, we instrument for price using demand shocks, including the timing of changes in the government's subsidy scheme and eligibility for the BTO mechanism.

We also estimate a dynamic random coefficients demand model over new developments and apartments currently in circulation (see e.g., Gowrisankaran and Rysman (2012)). Variation in development characteristics within a period, as well as choice sets across periods, identifies each applicant's preference parameters. Additionally, variation in subsidies disbursed by income type allows us to disentangle how sensitive the poor and rich are to prices. To deal with unobserved development quality, we instrument for price using cost shifters and the timing of a policy change affecting households buying multiple properties. We combine these instruments in a Differentiation IV framework (Gandhi and Houde, 2021).

We show, in Section 3.9, that our parameter estimates are plausible and our model fits the data well. Our demand model successfully recovers that unobserved development amenities are orthogonal to various observed amenities. The demand model also does well on untargeted moments: the recovered residuals do not capture a time trend or amenity from having a (non-prestigious) school nearby. Moreover, in our simulations, we find that

most applicants do make less than four applications, corroborating a ministerial response to a question posed in the Singapore parliament.

In our counterfactuals, described in Section 3.10, we find that better rules complement higher supply. Suppose the Singapore government (1) eliminated intertemporal risk; and (2) built 10% more apartments in oversubscribed developments. Then, vacancies fall from 12% to 10%; meanwhile, wait times fall by 8% for the poor and 23% for the rich. Young households gain from being allowed to list two developments in the same period and possibly exiting early. Exiting early pushes up prices on the secondary market by 9%, or 12% in the case without additional supply. As in our theoretical example, it is ineffective to add supply without changing the rules. Under the current mechanism, the demand response eclipses the supply increase, raising vacancies from 12% to 16% while leaving wait times unchanged. This behavior arises because marginal agents switch from less desirable developments today and apply to better developments next period, leaving today's developments empty. Thus, in a congested market, when dynamic incentives are influenced by supply, increasing it is not a panacea. Policymakers should also pay attention to the rules under which goods are allocated, so that their constituents can best utilize the additional supply.

3.2 Recent Literature

This chapter speaks to four strands of literature in Economics. Our core contribution speaks to the literature on centralized assignment mechanisms (Abdulkadiroğlu et al. 2017; Agarwal and Somaini 2018; Fack et al. 2019; Hastings et al. 2009; Narita 2018). Our chapter is most closely related to four papers. The first, Verdier and Reeling (2021), focuses on the dynamic assignment of bear hunting licenses with accumulated "preference points" in Michigan.[4] The second, Agarwal et al. (2021a), focuses on the allocation of deceased donor kidneys as an

[4]In follow-up work, Reeling et al. (2020) estimate applicant willingness to pay for these bear hunting permits in a dynamic discrete choice framework.

optimal stopping problem. The third and fourth, Waldinger (2021) and Naik and Thakral (2021), study alternative mechanisms to allocate public housing in various cities in the United States and the United Kingdom. We build on this literature in two ways.

We are the first to estimate a model of public housing with a secondary market and nontrivial dynamics. Unlike in rented social/council housing, owner-occupants hold equity in their homes, necessitating dynamic considerations. When deciding which apartment to purchase, putative homebuyers need to account for how their property prices will evolve. Hence, after 5 years, these owner-occupants continually trade off the "option value" of keeping their apartment against the cash they would obtain from selling it off on the secondary market. Since this secondary market exists, in each BTO cycle, an unhoused young couple not only has to decide to which development to apply, but also whether to apply at all. "Used" apartments remain attractive to richer households who are impatient or nearing the end of their search. Thus, any policy change in the BTO mechanism may spill over to the secondary market for apartments. Unlike our predecessors, we model this market and show that BTO rule changes can affect owner-occupier home equity by an economically important margin.

Moreover, we find that rules matter when considering equivalent changes in supply. Our predecessors find that, in the setting of auctions, the value of negotiating skill may be small relative to inviting one additional bidder (Bulow and Klemperer, 1996). Often, design changes are said to improve welfare by an equivalent increase in the supply, for example, of donor kidneys (Agarwal et al., 2021a). In our chapter, we qualify this sentiment. How effective is increasing supply? The answer depends crucially on the rules under which this supply is being allocated, as well as the characteristics of the supplied goods.

Our chapter is also related to a burgeoning literature on neighborhood choice in Empirical Industrial Organization (see, e.g., Bayer et al. 2007; Geyer and Sieg 2013; Galiani et al. 2015; Bayer et al. 2016; Almagro and Domínguez-Iino 2019; Van Dijk 2019). We account for

geography and study a centralized mechanism, differing from Sieg and Yoon (2020), which models search, queueing and mobility in the market for affordable housing in Manhattan. Unlike in their setting, at the margin, expanding the inflow of affordable housing in Singapore does not seem to impose large costs on existing owner-occupiers.

Furthermore, our chapter speaks to demand estimation in general (Berry 1994; Berry et al. 1995; Greene and Hensher 2003) and the estimation of demand for durable goods in particular (Erdem et al. 2003; Hendel and Nevo 2006; Doraszelski and Pakes 2007; Gowrisankaran and Rysman 2012; Aguirregabiria and Nevo 2013; Gowrisankaran and Rysman 2020). We build on this literature by considering demand over apartment lotteries in the style of Agarwal and Somaini (2018), resulting in an additional nested fixed point (apartment success rates) that we solve for in our counterfactuals. Moreover, relative to Gillingham et al. (2022), who study the market for new and used automobiles, their "primary market" for cars clears by price while ours clears by lottery and wait times. To obviate repeatedly iterating value functions when estimating supply, we exploit finite dependence (Arcidiacono and Miller, 2011).

Finally, our chapter is directly related to various studies of the Singaporean housing market in the Urban and Transportation Economics literature (Wong 2013; Wong 2014; Diao et al. 2017; Agarwal et al. 2021b; Lee and Tan 2023). We build on this literature on two fronts: not only do we showcase new data from the Housing and Development Board of Singapore, but we are also the first to conduct a structural estimation of a matching model in this rich setting.

3.3 Policy Background

3.3.1 Housing stock in Singapore

In Singapore, 80% of the housing stock is government housing, administered and maintained by the Housing and Development Board (HDB).[5] Over 80% of resident households in Singapore live in these apartments. During our sample period (2012-2015), the average apartment sold for S$460,000 on the secondary market, or about 4.6 times the median yearly income of a Singaporean household in that period.

The remaining housing stock is private, with the vast majority (19% of the housing stock) being private apartment complexes known as "condominiums." These complexes are gated communities with their own facilities, like swimming pools, gyms, barbecue pits, and so on. In July 2012, the median price per square foot of new private apartment sales was S$1,500; given a 2-bedroom apartment size of 900 square feet, said apartment would go for S$1,350,000, or 13.5 times a household's median yearly income.[6] Meanwhile, non-"condominium" private housing, known as "landed property," are even more expensive.

In this chapter, given that private housing in Singapore is unaffordable to the young households the government wishes to target, we will assume that the private housing market is separate from the market for new and "used" government housing.

3.3.2 The Housing and Development Board and its objective

Shortly after Singapore took over full internal self-government from the British in 1959, the HDB was established in 1960 to quickly build apartments to house all Singaporeans, the majority of whom lived in slums. The agency successfully built 21,000 apartments within 3 years of its formation, and 54,000 by 1965. Today, the HDB manages "1 million

[5]For more details on the HDB, consult https://www.hdb.gov.sg/cs/infoweb/about-us.

[6]At the time of writing, one US dollar is worth 1.39 Singapore dollars. For the rest of this chapter, we will take the Singapore dollar as our default currency.

[apartments]... in 23 towns and 3 estates" islandwide (Housing and Development Board, 2021a).

The HDB's stated objective is to provide "affordable, quality public housing and a great living environment where communities thrive."[7] The Singapore government views homeownership as a valuable endeavor in its own right. Owning one's home dovetails well with several of the government's social goals: fostering a sense of belonging and responsibility to one's neighborhood, socioeconomic mixing, "reinforcing the family institution" (Huat, 2017, 85), and so on.

In service of this objective, the Singapore government has to balance several competing concerns (Huat, 2017, 96). Chief among its concerns is that it must build enough new, "affordable" apartments to house putative homeowners and the poor. By "affordable," the government means that prices for new apartments in neighborhoods farther away from the city center should not be higher than four times an applicant's median income (Chin and Chang, 2013). On the other hand, the HDB refuses to build "too many" new apartments for two reasons. First, vacant apartments are very costly for the government to hold and allocate (Mah, 2011, 29). Second, the prices of apartments on the secondary market may fall, hurting the home equity of existing owner-occupiers.[8] Since Singaporeans hold much of their wealth in real estate, volatility in apartment prices would cast a pall over many households' retirement plans, a scenario the Singapore government seeks to avoid.

In contrast with public housing in the United States (US), where households rent the public housing apartments they are assigned to, over 90% of Singaporean households in HDB flats "own their home." That is, these households have signed a deed giving them the

[7] In related work, Ferdowsian et al. (2023) investigate optimal supply choices in public housing through the lens of dynamic mechanism design and queuing theory. In their setting, the government prefers to minimize vacancies and wait times, but may weight each outcome differently. Given these different social welfare functions, they characterize conditions under which it is possible to achieve a constrained efficient allocation of apartments, and discuss mechanisms implementing this allocation.

[8] In Singapore, a household can finance its government apartment through its mandatory retirement savings account in the Central Provident Fund (Huat, 2017, 86).

right to live in their assigned apartment for 99 years, and are allowed to sell the apartment on a secondary market after living in it for 5 years.[9]

3.3.3 The Build-to-Order Scheme

In Singapore, many government apartments are first introduced into the housing stock via a process known as the Build-to-Order Scheme (BTO). Introduced in April 2001 and taking place every 2-3 months, the BTO exercise allows potential homeowners to ballot for their preferred neighborhood and apartment size. Applicants apply for BTO as a family unit, typically a married couple. One person in the family unit must be a Singapore citizen, while the other must be at least a permanent resident. Both people must be at least 21 years old and have a combined annual income below a threshold (S$168,000 as of the time of writing). This threshold is set at roughly the 60th percentile of incomes for young households. Finally, no one in the family unit is allowed to own real estate in Singapore or overseas.

In each BTO cycle, putative applicants observe the developments on offer. Then, they decide whether to apply to one development; take an apartment on the secondary market; or wait till the next BTO cycle. Applicants may choose to wait because reneging on an offered apartment could cause them to lose priority in future application cycles. When young households remain unmatched, they typically live with their parents, for cultural and financial reasons.[10]

After the application phase, the HDB assigns each applicant a queue number, indicating the sequence in which she can select her preferred apartment.[11] At any point, an applicant

[9] Our chapter focuses on the primary market for HDB apartments. For a paper that studies the broader spatial organization of economic activity in Singapore, see Lee and Tan (2023).

[10] The three dominant racial groups in Singapore are the Chinese, Malay, and Indians. The typical household structure for these groups is the extended family. Moreover, renting is generally unaffordable and "avoided by citizens": a 2-bedroom apartment outside the city center costs "more than half of the median salary" (Choo, 2021). In 2015, only 23,700 Singapore residents under 35 lived alone or away from their parents. In this age group, at most 30% are married. Thus, the fraction of putative BTO applicants who rent is likely to be miniscule.

[11] To enforce social mixing, there are ethnic quotas for each housing development, i.e. a maximum number of Chinese, Malay and Indian households in each government apartment building. These quotas are enforced

may withdraw her application and participate in a future cycle. However, as mentioned earlier, this withdrawal is not costless in general. If an applicant is twice chosen to select an apartment but never does so, she loses priority in the BTO process for a year, drastically lowering her chances of "winning" future BTO lotteries.

After all applicants have either selected an apartment or withdrawn their application, HDB will commence construction if 70% of all properties in a development have been allocated. In our sample period, despite some vacancies in developments farther from the city center, all BTO developments have crossed this 70% threshold, and thus have been constructed. In practice, most developments of all sizes are oversubscribed, some by at least 2 to 3 times. These apartments are typically ready for homebuyers to move in within 3 years of the corresponding BTO exercise. To deter arbitrage, the government imposes two restrictions. First, an owner may only sell her apartment after she has lived in it for 5 years. Because their units are small, few owner(-occupier)s rent out their homes.[12] Second, quotas for second-time applicants to BTO are much smaller than those for first-timers.[13] Thus, once an applicant is housed, she is deterred from selling off her apartment, living in temporary accommodation, and applying for a new BTO apartment.

These apartments are often oversubscribed because they are sold at highly subsidized prices, with the amount of the subsidy decreasing in apartment size. The modal grant awarded is over S$66,000 of a sticker price of S$200,000 for a 2-bedroom apartment in a "non-mature town," an area typically farther away from the city center and with fewer

at both the BTO stage and on subsequent resale. This aspect of the housing system has been extensively studied in previous work (Wong, 2013, 2014). The data from our mechanism do not contain information about each applicant's race, so we will not speak to equity concerns between ethnicities. Instead, we focus on the government's tension between encouraging homeownership for young households and maintaining the home equity of existing owner-occupiers.

[12]In 2012, about 2.7% of apartments in Singapore were rented out (Housing and Development Board, 2022). These rentals include rentals of whole apartments, as well as that of a bedroom within an apartment. The vast majority of owners do not rent out their homes because their apartments are small. The largest apartment type has three bedrooms in total.

[13]For instance, 95% of the apartments in developments in "mature" estates are reserved for "first-timers." These "mature" estates tend to be more expensive and desirable to live in because they are closer to the city center.

completed apartments; and over S$50,000 for a similar apartment costing about S$300,000 in a "mature town."[14]

During our sample period of January 2012 to May 2015, about 69,500 apartments were built and offered at BTO, of which about 60,000 were reserved for "first-timers," or young households who have never been assigned a BTO apartment. To these developments about 183,000 applications were made, of which about 110,000 were made by first-timers. These first-time applicants faced an average oversubscription ratio of 1.91:1, with 5th and 95th percentiles of 0.47:1 and 4.73:1 respectively. At the same time, on the secondary market, about 70,000 transactions were made.

3.3.4 Information available to household at application

When a putative homebuyer is deciding to ballot for a BTO sales launch, she accesses the sales launch subsection on the HDB website (see Figure 3.I.8 in the Appendix for this and all subsequent images). After clicking through to that subsection, she sees the projects on offer, so she learns the price for each apartment size-development pair (top right). Additionally, she can also view a map of the neighborhood for her desired development (bottom left) and the associated site plan (bottom right). The applicant thus has a relatively complete picture of the characteristics and surrounding amenities of the apartment in question.

When the application portal is launched, a putative homebuyer has a one-week window to decide which apartment to apply for. Throughout the process, she can view the "flat supply and applications received" for each apartment option, and can condition her choice on the number of applications received so far. However, she will not know the characteristics of the applicants competing with her for these apartments.

In theory, a rational homebuyer should always wait till the last possible moment to submit her application, since she obtains the most amount of information about the application

[14]We describe the subsidy scheme in detail in Section 3.A.3 in the Appendix.

rates to each apartment option. Moreover, once she has submitted her application, rescinding it is somewhat troublesome: she has to either submit a written request or fill out another application in full before canceling the first. In practice, we see many homebuyers submitting their applications well in advance of the deadline. For instance, in the August 2021 application cycle, 47.5% of all applications to 3-room and larger apartments are made on August 13, despite the application window only closing on August 17.

In this chapter, because of data limitations, we will assume all homebuyers submit their applications simultaneously, consistent with the principle of strict dominance outlined above. We note that if one observes the number of applicants on each day of the application window, one may be able to back out applicant types by when they submit their applications, thus allowing inferences on homebuyer attrition and entry. This extension is left for future work.

3.4 Data

Our data cover the universe of applications in the BTO mechanism, as well as all transactions on the secondary market for HDB housing. To our knowledge, we are the first to observe the universe of applications to and subsequent trade of public housing.

Our main data set contains aggregate applications to all BTO developments between 2010 to 2020. These data were constructed partly through a vendor, and partly through liberal use of the Wayback Machine to scrape historical BTO results from the HDB website.[15] In this data set, each observation is a development-apartment size pair (i.e. "product-level data"). We observe sticker prices, the number of units offered ("quantity") and the number of applications to that product. These data are used to infer applicant demand over developments in each period. In our sample (January 2012 to May 2015), we observe

[15]Our data provider is Teoalida, an independent researcher collating information on past BTO cycles. These data are validated against historical BTO listings and results using the Wayback Machine (http://www.archive.org/). See the Online Appendix for more details on data construction.

about 110,000 applications to 195 developments over 20 BTO cycles. Table 3.I.1 contains the summary statistics corresponding to applications to the BTO mechanism.

As a first step to compute the distances of each development to various amenities, we combine these application data with latitudes and longitudes of each BTO development by calling the appropriate APIs on Onemap.sg, a government mapping portal.[16] The algorithm we used selects the first search result corresponding to each development.

To recover applicant preferences for the resale potential of each apartment, we augment our data with the transaction history of all HDB apartments on the secondary market from 2000 to 2021. Here, each observation is an apartment transaction. We observe the street address of the apartment being transacted; its sale price; its approximate floor level; the number of years left on its lease; the number of rooms it has; as well as the HDB town in which it is located.[17]

Finally, to recover applicant preferences for various amenities, we use public data from http://data.gov.sg. These data include the locations of all elementary schools, and an indicator of whether each school receives additional funding under the Special Assistance Plan (SAP); the locations of all metro stations; the locations of all supermarkets; and the locations of all "hawker centers," or open-air food courts selling street food.[18] We compute the distance between each BTO and secondary market apartment and the nearest amenity of each type. More details on data acquisition and processing can be found on the Online Appendix.

For dynamic supply estimation, we collapse the apartment transactions on the secondary market by neighborhood, apartment type and corresponding BTO cycle, then merge in the

[16] http://onemap.sg/

[17] We associate each resale transaction with a (latitude, longitude) pair via programmatically calling the Onemap.sg API with the street address of the apartment, falling back on Google Maps if no results are found on the former portal.

[18] School data are from the Singapore Ministry of Education. Transit data are from the Singapore Land Transport Authority. Supermarket and hawker center data are from the Singapore National Environment Agency.

corresponding apartment stock obtained from the Housing and Development Board. Table 3.I.2 displays the summary statistics for the dynamic supply estimation data set.

3.5 Descriptive Analysis

Before we exposit our model and estimation procedure, we describe the apartments chosen by households and provide evidence for several determinants of apartment demand.[19]

3.5.1 Apartments in BTO developments are homogeneous

Apart from their location, apartments built in BTO developments are relatively standardized. This is to ensure that they can be built relatively cheaply. Apartments within a development are primarily differentiated along two dimensions. First, we observe the number of rooms per apartment in each development; see Table 3.I.3 for a list of indicative floor areas for an apartment with a given number of rooms.[20] For apartments with two to four rooms, one room is a living room while the remaining are bedrooms. The notable exception is the five-room apartment, which has a living room, three bedrooms and a dining room. Fittings are standard across apartments.[21]

Within a (neighborhood, apartment size, BTO cycle) bin, apartments may differ along some dimensions that we do not observe. For instance, an apartment on a high floor facing a hill is likely more expensive and more desirable than a similar unit on a low floor facing a parking garage. At the application stage, a household only knows the distribution of these

[19] Additional descriptive analysis can be found in Section 3.A in the Appendix.

[20] We omit one-room apartments ("rental flats") as they are meant for the poorest households; to our knowledge, these apartments are allocated via a waitlist system akin to those common in the United States (see e.g. Waldinger (2021)). There also exist 3-Gen (4-bedroom) and Executive units which do not fall neatly into the framework above. We classify these larger units as 5-room units because (1) they are few in number and (2) they are targeted for larger families, most of whom would consider standard 5-room flats.

[21] These fittings include: lighting, window grilles, built-in kitchen cabinets with an induction hob and cooker hood, a kitchen sink, tap, and dish drying rack, a built-in wardrobe, a water heater and a mirror and toilet roll holder in each bathroom. See for example the November 2020 BTO offerings for a development in Sembawang.

unobservable characteristics for each development; she only learns the characteristics of the (remaining) apartments in her choice set after HDB calls her to select one.

The remaining variation between developments boils down to observables: agents observe the distance of each development to downtown, to the nearest metro stop, and so on.

In our main analysis, we will restrict our market to all households choosing between 3-room, 4-room and 5-room units. This is because 2-room units attract a large number of "joint single" households, or are shorter-lease units targeted to the elderly. We focus on the largest demographic group applying to BTO apartments: newly-married couples applying for their first apartment.

3.5.2 Construction costs shift BTO prices

Inputs into construction shift the prices of BTO apartments. In recent years, "higher building costs" have pushed the highest BTO prices "closer to [the] million-dollar mark" (Elangovan, 2022).

One key input into apartment construction is concreting sand, which we will use as a cost shifter in demand estimation. Concreting sand is an essential ingredient in mixtures of cement and asphalt. In Figure 3.I.1, we display a binned scatter plot of BTO prices against a standardized price index of concreting sand. These BTO prices have been residualized on controls including distances to downtown and to the nearest metro stop. As expected, higher concreting sand prices predict higher BTO prices, all else equal.

In demand estimation, we will instrument for endogenous BTO prices by interacting such cost shifters with the number of rooms of each apartment in the BTO development. Since larger apartments require more concreting sand to construct, when prices of concreting sand rise, prices of larger apartments should rise by a greater amount, all else equal. This specification is validated by regressions in Table 3.I.17 in the Appendix.

3.5.3 Access to (good) schools and public transit

Preferences

For a credible model of BTO applicant demand, we need to control for observable measures of development quality. Such measures include access to downtown, to transit, to good schools, as well as other amenities.[22] In Figure 3.I.2, we display binned scatter plots of how oversubscribed BTO developments are, relative to their distance from various amenities. As expected, the closer a development is to downtown, the nearest metro stop or the nearest Special Assistance Plan (SAP) school, the more oversubscribed it is, all else equal. However, there appears to be no relationship between how popular a development is and how far it is from the nearest (ordinary) school. We distinguish SAP from non-SAP schools to proxy for school quality: SAP schools specialize in bilingual education in English and Chinese and are academically competitive.

Residential sorting and congestion effects of BTO are small

In our modeling below, we will take the evolution of neighborhood compositions and that of amenities as invariant to our counterfactual simulations. Throughout, we maintain that our counterfactuals do not change the paths of how many singles marry and whether these newly-married couples demand housing at all. Then, because of the income eligibility limits and aforementioned racial quotas, government housing neighborhoods are relatively homogeneous by race and income. Thus, endogenous neighborhood composition is not of first-order concern for our purposes.

However, as more residents live in a particular neighborhood, local amenities could become more congested. First, we rule out traffic congestion. Since cars can cost as much as a

[22] Applicants have cited poor amenities as a reason to delay application to the next BTO cycle. In February 2023, Ms. Low Wei Xin [gave] "the current sales exercise a miss" because most of the developments "felt too ulu (Singaporean slang for remote)" (Ong, 2023).

house, and the transit system covers much of the city, most workers commute by bus, metro, or on foot.[23] Second, we argue that schooling amenities are unlikely to be more congested for small perturbations in apartment quantity. In Figure 3.I.3 below, we plot summary statistics, from 2009 to 2021, of the percentage of vacancies in first grade by school. Our two comparison neighborhoods, Ang Mo Kio and Punggol, are representative of mature and non-mature estates respectively. In both neighborhoods, we find mean vacancies per school of about 20%, with interquartile ranges between 0% and 50%. These percentage vacancies have held steady despite school mergers in the past decade. Furthermore, since the Singaporean school system is funded by the national government and cohort sizes were larger pre-2009, schools remain capable of accommodating future rises in local enrollment. Therefore, we do not regard congestion in elementary schools as of first-order concern.

3.5.4 Households are forward-looking

How do dynamics matter in our setting? Each young household knows the developments on offer both today and in the next cycle; if the developments next quarter are "better" than today's, a putative applicant may elect to wait and only apply to a development next cycle. Waiting can be rational because reneging on an offer is costly. If a young household applies to a development today, is invited to select an apartment but turns down the offer twice, she loses priority in the BTO cycles in the coming year.

As we saw in Section 3.5.3, young households seem to like developments that are closer to the nearest metro stop. If the developments next cycle are closer to the nearest metro stop on average than today's, more applicants should apply to those developments relative to today's, all else equal. This stylized fact is displayed in Figure 3.I.4. In the left panel, we display a binned scatter plot of oversubscription rates in a BTO cycle against the mean

[23] Commuters rely on taxis and ride-sharing services for occasional late night trips. As of this writing, a Toyota Corolla costs about S$150,000, or about 1.5 times the yearly income of the median household. On the other hand, transit fares range from S$0.95 to S$3.30 per trip, depending on travel distance. See Lee and Tan (2023) for details.

distance to the nearest metro stop of all developments offered in that cycle. In this plot, there appears to be a weak negative relationship between the oversubscription rate in a cycle against the aforementioned mean distance. In contrast, in the right panel, we plot first differences over time in both oversubscription rates and mean distances. In this second plot, if the developments next cycle are farther from a metro stop relative to today's, we predict that developments next cycle will be less oversubscribed relative to today's, all else equal. Since the linear fit improves significantly in the second plot relative to the first, we have suggestive evidence that households are forward-looking.

3.6 Theoretical Example

Our main design intervention targets inefficiencies from limited choice expression and rational delay. Under the current mechanism, in any one BTO cycle, households are not allowed to list a second development as acceptable to them. Moreover, if next cycle's developments are much more attractive than this cycle's, putative applicants may wait till next cycle and try for a competitive development. Some of these applicants will fail to get an apartment in said development, and experience ex post regret. Rather than remaining unmatched at all, they would have been willing to "settle for" an apartment in the previous cycle.

In Section 3.7, we will describe a richer model that we can take to the data and from which we can simulate our proposed interventions. For the sake of exposition, in this section, we illustrate key tradeoffs with a simple theoretical example. We abstract from equilibrium in the secondary market and limit our attention to a setup with three agents and two developments of capacity 1. First, we characterize the BTO equilibrium. Next, we demonstrate that, under BTO, building more apartments may be ineffective at reducing vacancies precisely because of rational delay. Finally, we show that if the government adopts a strategyproof mechanism, adding supply achieves its intended effect, matching all agents to an apartment.

3.6.1 Setup

There are two periods, $t = 1, 2$. In Period 1 (2), one house A (B) arrives. At the end of each period, vacant houses vanish. Three identical players $i = 1, 2, 3$ are born at the beginning of time and are perfectly patient. Every period, each active player i simultaneously takes one of three actions. She can apply for the available house; wait one period; or exit the market. If house j is oversubscribed, the winning agent is chosen via uniform random lottery. Agents who are matched to a house or exit the market stop making further decisions. If an agent waits or otherwise remains unmatched in Period 2, she exits the market.

Agents value house A at 2 utils and B at 5 utils. Unmatched agents between Periods 1 and 2 incur a small waiting cost of $\epsilon \in (0, 1)$. Agents exiting the market in Period 1 (2) receive a payoff of -2 (-1) utils, representing higher housing congestion in earlier periods from a smaller stock of available developments.

3.6.2 BTO Equilibrium

We seek a Subgame Perfect Nash Equilibrium. We claim that in Period 1, everyone applies to house A; then, in Period 2, the remaining agents apply to house B. Proceeding by backward induction, consider any subgame in which two agents are unmatched at the start of Period 2. It is strictly dominant for each agent to apply to house B, since remaining unmatched yields a negative payoff for sure. Thus, each agent's expected utility at the start of Period 2 is

$$V_2 = \frac{1}{2} \times u(B) + \frac{1}{2} \times u(\text{exit}_2) = 2.5 - 0.5 = 2.$$

Then, any agent left unmatched at the end of Period 1 incurs the waiting cost ϵ, so their expected payoff at the end of Period 1 is $2 - \epsilon$.

At the start of Period 1, we conjecture that everyone applies to house A. As in Period 2, exiting the market yields a negative payoff and is thus strictly dominated by applying to

house A. Applying to house A yields a payoff of

$$v_1(a; -a) = \frac{1}{3} \times u(A) + \frac{2}{3} \times (2 - \epsilon) = 2 - \frac{2}{3}\epsilon,$$

whereas waiting yields a payoff of

$$v_1(w; -a) = 2 - \epsilon < v_1(a; -a).$$

Therefore, in Period 1, all agents apply to house A; then, in Period 2, the remaining agents apply to house B.

The resulting allocation of houses is a permutation of $x = (A, B, \emptyset)$. No houses are vacant.

3.6.3 Under the current rules, building more "B" is ineffective

We show that, under the current rules, added supply can be ineffective. Suppose the government built one more house B, arriving in Period 2. We verify that in the new equilibrium, in Period 1, everyone waits; then, in Period 2, everyone applies for B. Proceeding by backward induction, consider any subgame in which all three agents are unmatched at the start of Period 2. It is strictly dominant for them to apply to B, since remaining unmatched yields a negative payoff for sure. Thus, each agent's expected utility at the start of Period 2 is

$$\tilde{V_2} = \frac{2}{3} \times u(B) + \frac{1}{3} \times u(\text{exit}_2) = \frac{10}{3} - \frac{1}{3} = 3.$$

Then, any agent left unmatched at the end of Period 1 incurs the waiting cost ϵ, so their expected payoff at the end of Period 1 is $3 - \epsilon$.

At the start of Period 1, we conjecture that everyone waits. As in Period 2, exiting the market yields a negative payoff and is thus strictly dominated by applying to house A.

Applying to house A yields a payoff of

$$\tilde{v}_1(a; -w) = 2,$$

whereas waiting yields a payoff of

$$\tilde{v}_1(w; -w) = 3 - \epsilon > \tilde{v}_1(a; -w).$$

Therefore, in Period 1, all agents wait; then, in Period 2, everyone applies to B.

The resulting allocation of houses is a permutation of $\tilde{x} = (B, B, \emptyset)$. Vacancies rise (house A is empty). Average wait times rise because everyone waits till Period 2.

Two insights can be gleaned from this exercise. The first is the potential for ex post regret. The unmatched agent prefers to be matched with house A than remain unmatched, even though it was rational for them to try for house(s) B. The second is that vertical quality and intertemporal substitution matter. If houses A and B were of identical quality, no one would have an incentive to wait. Nor would anyone wait if they were sufficiently myopic, e.g. if each agent discounted her payoffs in period 2 by a factor of $\delta = 0.5$. Thus, when we amend the BTO program, dynamics and heterogeneity in developments play a crucial role in driving our results.

3.6.4 Once we shut down intertemporal risk, adding supply can be effective

Once we shut down intertemporal risk, adding supply succeeds in matching all the applicants. To target this source of inefficiency, we implement a version of Random Serial Dictatorship (RSD) closely resembling the mechanism we simulate in Section 3.10. The stock of available housing is 1 unit of A in Period 1, and 2 units of B in Period 2. Under the new rules, at the beginning of time, all three agents submit a rank-ordered list over house A and house B.

Then, at random, each agent is assigned her highest-ranked option that is still available to her. If an agent is unassigned, she exits either in Period 1 or Period 2, each with probability 0.5.

Since RSD is strategyproof, all agents truthfully specify that they prefer house B to house A to remaining unmatched. The resulting allocation of houses is a permutation of $x' = (A, B, B)$. No houses are vacant. Thus, under the new rules, building more apartments can ease BTO congestion.

Our illustrative model abstracts from equilibrium in the secondary market for apartments. Moreover, it does not capture the rich substitution patterns between developments in the same BTO cycle, as well as between one cycle and the next. To fully capture the equilibria and spillovers from our proposed interventions, we need a model we can take to the data.

3.7 Model

We expand on our illustrative model to recover the preferences of young households and owner-occupiers of existing apartments. Our demand model closely resembles the decision-making process of each putative BTO applicant, while our supply model captures the tendency for each owner-occupier on the secondary market to "cash in" on higher home values.

Each period, new BTO developments arrive exogenously, each with a fixed number of units. On the secondary market, each owner-occupier decides whether or not to sell their apartment in the time frame corresponding to the prevailing BTO cycle. The trade-off for them is between selling today and selling at a higher price tomorrow. Once she sells her apartment, an owner-occupier exits the market. We interpret "exit" to mean a resident upgrading to private property, or moving in with their children in their old age.[24]

On the demand side, every BTO cycle, new young households become active for a maximum of $\bar{\tau}$ periods. All young households choose one BTO development to which to apply;

[24] See Subsection 3.7.6 for details.

wait; or exit to the secondary market. Their choice depends on their preferences, lottery odds and prices on the secondary market. Matched households obtain their apartment and exit the market.[25] Unmatched households of age $\bar{\tau}$ exit to the secondary market at the beginning of the $\bar{\tau}+1$-th period in which they become active.[26]

3.7.1 Developments offered at BTO

Time is discrete and indexed by $t = 0, 1, 2, \ldots, T$. Each time period corresponds to a BTO application cycle. In the BTO mechanism, every period, developments $a = 1, 2, \ldots, A_t$ arrive exogenously, each with its respective number of units $Q_{1t}, Q_{2t}, \ldots, Q_{A_t t}$. Let $\mathcal{A}_t$ represent the set of developments available at time t. The prices of BTO apartments are set exogenously by the government.

3.7.2 Model timing

Fix a time period t. Putative BTO applicants have perfect foresight and are finitely-lived. They age deterministically across periods. At the posted secondary market prices p_t^S and BTO success probabilities q_t, on the demand side,

1. Putative applicants in cohort t are born.

[25] Recall that a young household that purchases a government apartment cannot sell it off until 5 years have elapsed. Since our sample period is shorter than 5 years in length, we assume these (new) owner-occupiers do not take any further action.

[26] Though we specialize our model to our empirical setting, our approach can be adapted to other instances of dynamic demand over lotteries. For example, our model can be adapted to school and course choice settings that do not utilize First-In-First-Out (FIFO) mechanisms.

We give an example of a non-FIFO mechanism, the "Primary One Registration" exercise governing admission into first grade in Singapore. Every admissions cycle, there are several priority phases for children with siblings in the same school, children with alumni parents, and so on. Priority phases take place sequentially. If a child applies to a school under the first priority phase, and the school is oversubscribed in that phase, school places are allocated by uniform random lottery. If the child fails to secure a place in the first phase, she can apply to a different school in the second phase (or the same school), and this process continues until all children are matched. The efficiency of such a mechanism, vis-à-vis deferred acceptance or other centralized assignment systems, remains an open question.

2. All agents observe developments $a = 1, 2, \ldots, A$ on offer via BTO in that period, as well as the price p_t^S of taking an apartment on the secondary market. For each development a offered at time t, let the quantity of apartments available in that development be Q_{at}.

3. Each household either applies for an apartment in BTO ($a \in \mathcal{A}_t$), exits the mechanism ($a = E$), or opts to wait: ($a = W$).

4. If a household exits the mechanism (i.e. $a = E$), it purchases an apartment on the secondary market and takes no further action. Else if a household has applied to an apartment within the mechanism, it learns if it is matched to its preferred development. Matched households exit the mechanism and collect utility from purchasing their BTO flat. Unmatched and waiting households pay a waiting cost and remain in the market.

5. Time advances to period $t + 1$.

6. Unmatched agents aged $\bar{\tau} + 1$ in period $t + 1$ exit the mechanism, buy an apartment on the secondary market and take no further action.

Each putative applicant is identified by its (permanent) income type y_i and age $\tau = 1, 2, \ldots, \bar{\tau}$. Household income types are drawn from an exogenous distribution $\mathcal{F}_y$. The aggregate endogenous state of our economy is μ_t, the distribution of alive households by income level and age. The exogenous states are the characteristics of all BTO developments offered at time t.

To simplify estimation, recall that each BTO applicant has perfect foresight over the set of apartments offered throughout the entire history of the mechanism. Hence, the only uncertainty each household faces is over its idiosyncratic preference shocks in each period, as well as whether it is offered the apartment to which it applied.

Simultaneously, on the supply side, owner-occupiers have perfect foresight (through at least 2022). At the posted secondary market prices p_t^S,

1. She decides whether or not to sell off her apartment.

2. If she sells off her apartment, she exits the market. If she does not, she retains her apartment and time advances to period $t+1$.

3.7.3 Supply in the secondary market

On the secondary market, each owner-occupier i lives in an apartment of type l at time t. Every period, she decides whether to sell off her apartment. If she sells, she exits the market. In this case, her indirect utility is

$$w_{i1t|l} = \rho_0 \bar{p}^S_{t|l} + x'_{1t|l}\rho_1 + \xi_{1t|l} + \epsilon_{i1t|l}. \tag{3.1}$$

That is, her indirect utility depends on the sale price of her apartment $\bar{p}^S_{t|l}$, other covariates $x_{1t|l}$, the unobservable attractiveness of selling $\xi_{1t|l}$ and an idiosyncratic shock to the attractiveness of selling, $\epsilon_{i1t|l}$. We assume that this shock, $\epsilon_{i1t|l}$, as well as its counterpart for not selling, $\epsilon_{i0t|l}$, are independently and identically distributed Type-1 Extreme Valued (T1EV).

On the other hand, if she opts not to sell, her indirect utility is

$$w_{i0t|l} = \delta^o \mathbb{E}V^o_{it+1|l} + \xi_{0t|l} + \epsilon_{i0t|l}. \tag{3.2}$$

Let $\bar{w}_{dt|l}$ be the mean utility, aggregated across idiosyncratic shocks $\epsilon_{idt|l}$, of taking action d. Then, the agent's indirect utility depends on her continuation value

$$\mathbb{E}V^o_{it+1|l} = \gamma^E + \ln\left[\exp(\bar{w}_{0t+1|l}) + \exp(\bar{w}_{1t+1|l})\right], \tag{3.3}$$

with γ^E being the Euler's gamma constant; the unobserved attractiveness of not selling $\xi_{0t|l}$; as well as her idiosyncratic shock $\epsilon_{i0t|l}$. Therefore, the market share of sellers of apartments

of type l at time t can be written

$$s_{1t|l} = \frac{\exp(\bar{w}_{1t|l})}{\exp(\bar{w}_{0t+1|l}) + \exp(\bar{w}_{1t+1|l})},$$

and the supply of apartments on the secondary market at time t can be written as

$$Q_t^S(\vec{p}_t^S) = \int s_{1t|l}(p_{t|l}^S)\, d\lambda_t(l), \tag{3.4}$$

where λ_t is the measure of housing stock over all secondary market apartment types l at time t.

3.7.4 Decision problem for eligible BTO applicants

Each BTO applicant faces a dynamic discrete choice problem in her lifetime, specifically in which apartment to apply for each period. We specify household preferences backwards: starting from its decision to accept an offered apartment, then which apartment to apply for. Throughout, fix a time period t and household i of age τ and with income type y.[27]

Apartment application choice

Consider a household of income level y recently matched to a BTO development a at time t. Then the utility from the apartment she is matched to is

$$U_t(a; y) = x_{at}'\beta^y - \alpha^y p_{at}^y + \alpha^{Ry} p_{at}^R + \xi_{at}. \tag{3.5}$$

The household likes better apartment characteristics x_{at} but dislikes paying a higher price p_{at}^y. We allow prices to depend on a household's type because poor households receive explicit subsidies that depend on the size and location of the apartment they apply to.

[27] Following recommendations in Andrews et al. (2020), we discuss alternate model specifications and key channels driving our results in Section 3.H in the Appendix.

Meanwhile, households may also value a higher five-year-ahead resale price p_{at}^{Ry}, as well as the development's unobservable quality ξ_{at}.[28]

If a household is unsuccessful at obtaining the development she applied for, she pays a waiting cost and continues to the next period, i.e.

$$U_t(0; y, \tau) = -c^y + \delta V_{t+1}(y, \tau + 1), \tag{3.6}$$

where $-c^y$ is her income-specific cost of waiting and δ is her discount factor.

If a household exits to the secondary market, she dislikes a higher mean apartment price[29] in the time period she is searching over:

$$U_t(E; y) = -\alpha^y \bar{p}_t^S + \xi_t^{Sy}.$$

The term ξ_t^{Sy} is the income-specific unobserved attractiveness of exiting to the secondary market in period t. Meanwhile, waiting households pay a waiting cost and advance to the next period, so their payoff from waiting is is

$$U_t(W; y, \tau) = \xi_{Wt}^y + U_t(0; y, \tau).$$

[28]In our model, we assume five-year-ahead resale prices are invariant in our counterfactuals. In our counterfactuals, we implicitly assume that our rule changes and supply increases do not persist beyond the end of our sample period, May 2015. Since our interventions are temporary, given the government's racial quotas by neighborhood, as well as income restrictions, we do not expect significant differences in long-run neighborhood sorting by race and income. Moreover, the supply increases we consider are miniscule relative to the stock of existing apartments. Hence, in our alternate scenarios, long-run resale prices are unlikely to differ from the realized path.

[29]For tractability, we capture the indirect utility of exiting to the secondary market by the mean price of an apartment in that period, as well as an unobserved shifter common to all applicants. Implicit in this modeling choice is the assumption that the value of exiting is well captured by the mean price on the secondary market.

An alternative approach is to fully model household preferences over each (neighborhood, apartment size) pair offered on the secondary market. For each period t, given a guess of preferences, we can obtain the value of her choice set, then insert this value into the household's choice problem. Unfortunately, this approach suffers from the curse of dimensionality, with the key computational bottleneck being the "middle loop" setting model market shares equal to observed market shares.

Similar to exiting, ξ^y_{Wt} represents the income-specific unobserved attractiveness of waiting a period.

We now step backwards to Step 3 and describe the household's decision problem. In each period, she selects a BTO development ($a \in \mathcal{A}_t$), exits ($a = E$) or waits ($a = W$) to maximize expected utility:

$$V_t(y, \tau) = \mathbb{E}\left\{ \max_{d \in \mathcal{A}_t \cup \{E,W\}} \mathbb{E}_t[v_{it}(d; y, \tau)|d, \mu_t] \right\}. \tag{3.7}$$

The agent makes her choice according to $v_{it}(\cdot; y, \tau)$, her choice-specific value function:

$$\begin{aligned} v_{it}(a; y, \tau) &= q_{at} U_t(a; y) + (1 - q_{at}) U_t(0; y, \tau) + \epsilon_{iat}; \\ v_{it}(W; y, \tau) &= U_t(W; y, \tau) + \epsilon_{iWt}; \\ v_{it}(E; y) &= U_t(E; y, \tau) + \epsilon_{iEt}. \end{aligned}$$

When a household applies to development a at time t, she chooses a lottery where she gets her apartment with probability q_{at}, and is unsuccessful otherwise. Finally, the agent's idiosyncratic shocks $\epsilon_{i \cdot t}$ are drawn iid from a Type-1 Extreme Valued (T1EV) distribution. Let N_{at} be the number of applicants to apartment a at time t.

Applicant state transitions

- If a household chooses to wait, she pays a waiting cost and remains unmatched next period (or, if aged $\bar{\tau}$, is compelled to exit the mechanism).
- If a household takes an apartment on the secondary market, she exits the mechanism.
- If a household applies for BTO development a, and if $N_{at} \leq Q_{at}$, she is matched to development a with probability 1. Else option a is oversubscribed, so the government selects successful households via a uniform random lottery. With probability q_{at}, the

household obtains an apartment in development a. Otherwise, she pays a waiting cost and remains unmatched next period (or, if aged $\bar{\tau}$, is compelled to exit the mechanism).

- If an agent remains unmatched at age $\bar{\tau}$ at time t, she is compelled to exit at time $t+1$, so her expected utility is

$$V_{t+1}(y, \bar{\tau}+1) = \gamma^E + U_t(E; y).$$

Market shares

Let $\bar{v}_t(\cdot; y, \tau)$ be the mean utility over all applicants of type (y, τ) taking a particular action. The market share for development a is obtained by integrating over the conditional choice probabilities of each type, i.e.

$$s_{at} = M_t^{-1} \int \frac{\exp(\bar{v}_t(a; y, \tau))}{\sum_{a' \in \mathcal{A}_t} \exp(\bar{v}_t(a'; y, \tau)) + \exp(\bar{v}_t(E; y)) + \exp(\bar{v}_t(W, y))} \, d\mu_t(y, \tau), \tag{3.8}$$

where M_t is the number of households active at time t:

$$M_t = \int d\mu_t(y, \tau).$$

3.7.5 Solution concept

To facilitate estimation, we will make an assumption of large markets. That is, we assume households take success probabilities q_t and secondary market prices $\bar{p}_t^S$ as given. If a household i unilaterally deviates from her equilibrium report σ_{it}, she does not significantly change other agents' success probabilities. Nevertheless, in our counterfactuals, it is important to let application success probabilities q adjust with changing policy regimes.

Definition. *Let $\mathcal{F}_t$ represent all information available at time t. A* ***BTO equilibrium*** *comprises a profile of type-specific strategies $\sigma_t^y = \sigma^y(\mathcal{F}_t)$, secondary market prices $\bar{p}_t^S$ and application success probabilities q_t such that*

1. *Households behave optimally. Conditional on the realization of household-specific shocks ϵ_{iat}^y, secondary market prices p_t^S and success probabilities q_t, the strategy σ_t^y solves the household optimization problem (3.7).*

2. *Application success probabilities are consistent with household choices. Given the strategy profile σ_t^y, the application success probabilities satisfy, for all t,*

$$q_{at} = \min\left\{\frac{Q_{at}}{s_{at}(\sigma_t^y; q_{at}, p_t^S)M_t}, 1\right\}. \tag{3.9}$$

3. *Given household behavior and application success probabilities, secondary market prices clear the market for owner-occupied apartments:*

$$Q_t^S(\vec{p}_t^S) = s_{Et}(\vec{p}_t^S)M_t.$$

Proposition 12. *A BTO equilibrium exists.*

We observe that success probabilities q lie in $[0,1]^{|\mathcal{A}|}$, a compact and convex set. Moreover, mean secondary market prices $\bar{p}^S$ can be constrained to lie in $[0, \bar{p}^{S,\max}]^T$, also a compact, convex set. Furthermore, given preference parameters (β, α, c, ρ), the mapping from $\omega = (q, p^S)$ to implied success probabilities and excess demand is a continuous function, by the aforementioned large-market assumption and an interpolation procedure. By Brouwer's Fixed Point Theorem (Brouwer, 1911), this mapping has a fixed point ω^*.

However, given preference parameters (β, α, c) and the initial conditions on the market size and the distribution of incomes and ages, the resulting lottery odds and secondary

market prices $\omega = (q, p^S)$ may not be unique.[30] This observation does not pose a problem in estimation, since we will recover applicant parameters in the equilibrium played in the data. In counterfactual simulations, we will impose the following equilibrium selection rule: we will start our optimizer at the ω observed in our data. In practice, starting our optimizer at other feasible points does not change our qualitative results.[31]

3.7.6 Model discussion

When the government builds more apartments, two economic forces are at play: Higher supply relieves congestion but the demand response exacerbates it. If apartments were all homogeneous, fixing the flow into the mechanism, building more apartments reduces congestion mechanically. However, in equilibrium, building more apartments in a desirable area may lead to agents switching from their original choice to the "better" development, partially undoing the supply expansion. Since agents are forward-looking, the aforementioned substitution not only happens across space, but also across time. Since eligible agents are always free to wait, they may try for a "better apartment" next period instead of "settling" for an apartment today. This intertemporal substitution is bad for the government in two ways: agents remain unmatched for longer and more apartments remain vacant.

Therefore, which of these forces dominates is an empirical question. Below, we discuss other model assumptions.

Large markets. In our model, each agent maximizes her utility, while believing that she cannot influence other agents' success probabilities. We believe this assumption is

[30] We display a simple example of multiple equilibria in our setting. Fix an equilibrium E and a period t. Suppose, in that period, developments A and B confer identical match utilities and are not congested (i.e. have fewer than 1 applicant to 1 apartment). We construct a second equilibrium E'. Let all of the applicants, except for one, behave as in equilibrium E. Let the "deviating" applicant switch from development A to development B (or vice versa). Then equilibrium success probabilities and secondary market prices in E' would remain the same in E, but we obtain a different allocation of apartments to applicants.

[31] Other starting points include setting success probabilities equal to 1 for all BTO developments; starting secondary market prices at the mean value during the sample period; and so on.

innocuous because an average of 322 apartments were offered per period per apartment type, with 95% range (98, 888). Similarly, the number of applicants per period per apartment type averaged 978, with 95% range (93, 3239). Therefore a deviation by a household shifts success probabilities by at most 1/93, or less than 1.1%. Nonetheless, this assumption is likely to be more valid for larger developments than smaller ones.

No peer effects other than through ξ. In our model, peer effects cannot enter the applicant's decision calculus other than through a development's unobserved quality ξ. This assumption rules out cases where two households apply to the same development because they want to live near each other's families. From speaking with BTO applicants, we understand that they apply to developments closer to their workplaces and to the nearest metro stop, rather than near where their friends live. This sentiment could be because Singapore is sufficiently small and well connected by transit.

Perfect foresight. Young households are assumed to have perfect foresight over the approximately 2.5 years they remain active. This assumption is valid in our context because of the copious development, neighborhood and pricing information provided by the government.[32] Moreover, the government routinely announces the next year's government apartment supply and neighborhoods well in advance of each BTO cycle. Thus, in practice, agents can foresee the quality of developments at least 4-5 cycles out. We expect a model of agents with limited foresight to deliver the same qualitative predictions as ours.

Owner-occupiers exit after selling their apartment. After an owner-occupier sells off her apartment on the secondary market, we assume that they exit the market and take no further action. Our assumption covers two major cases of movers. First, while some

[32] We assume households can observe all amenities near each (development, apartment size) pair because of the official maps provided by the HDB (see Figure 3.I.8 in the Appendix). The HDB also furnishes applicants with approximate prices for each apartment, as well as resale prices of comparable nearby apartments. See, for example, the HDB news release for the May 2021 BTO cycle.

elderly households age in place, many others dispose of their apartments and move in with their children, financing their retirement using the home equity they cashed in. Second, several owner-occupiers experience a high income or wealth shock and elect to move into private housing. However, the structure of our model rules out a third class of movers: residents disposing of their government apartment and purchasing another. Since we only observe apartment transactions, but not who sold to whom, we do not know the precise level of "residual demand" from agents eligible for government housing but not BTO.[33] As family sizes fall and transit becomes more efficient, we expect this last group to be small relative to the rest of the market.

3.8 Estimation

With our model, we intend to simulate our policy interventions: changing rules and adding supply. In this section, we recover the preferences of putative BTO applicants and to what extent owner-occupiers move when their homes appreciate. Since (secondary market) supply only relies on data from the secondary market, we estimate supply offline before tackling demand:

1. (Supply.) Use finite dependence to eliminate the value function in an owner-occupier's moving decision. Then, regress adjusted log differences in moving shares on adjusted differences in prices between today and the next BTO cycle.

2. (Demand.) Recover BTO applicant preferences via the Generalized Method of Moments (GMM), conditioning on the observed success probabilities and prices. The

[33]There are no statistics on movers from one government apartment to another. However, in the 2018 HDB Sample Housing Survey, just under 14% of current residents made "lateral moves," from one government apartment to another of the same size (Housing and Development Board, 2021b). Since the time they were married, 63.6% of residents in government housing made one or fewer move, i.e. into their first government apartment. When asked if they were "content with their [apartment]," between 60% and 75% of residents answered that they were. Hence, lateral movers are not of first-order concern in our setting.

shares of each development, as well as exiting, are matched exactly. Then, we minimize an IV-GMM criterion.

3.8.1 Secondary market supply

We estimate supply on the secondary market by exploiting finite dependence (Arcidiacono and Miller, 2011). Specifically, an owner-occupier exits the market once she has sold her apartment. Thus, no matter whether she sells off her apartment today or tomorrow, her continuation value "cancels out," allowing the econometrician to collapse a dynamic problem to a static one.

Our model is identified off variation in mean sale prices of each apartment type over time and in space, vis-à-vis variation in the shares of movers and stayers in each (neighborhood, apartment size) in each quarter. The model "corrects" for forward-looking behavior by leveraging one-period-ahead mover shares of the corresponding apartment type.

Fix an apartment type l and a time period t. Using the Hotz-Miller (1993) inversion, we take log differences between the share of movers and that of stayers. Then, the relative attractiveness of selling, against not selling, is

$$\ln s_{1t|l} - \ln s_{0t|l} = \bar{w}_{1t|l} - \bar{w}_{0t|l} = \rho_0 \bar{p}^S_{t|l} + x'_{1t|l}\rho_1 - \delta^o \mathbb{E}V^o_{t+1|l} + \xi_{1t|l} - \xi_{0t|l}. \tag{3.10}$$

In our model, selling one's apartment is a terminal action, so the inclusive value at time $t+1$ can be written as

$$\mathbb{E}V^o_{t+1|l} = \gamma^E + \rho_0 \bar{p}^S_{t+1|l} + x'_{1t+1|l}\rho_1 + \xi_{1t+1|l} - \ln s_{1t+1|l}.$$

Substituting this inclusive value term into Equation (3.10), we obtain

$$\begin{aligned}\ln s_{1t|l} - \ln s_{0t|l} - \delta^o \ln s_{1t+1|l} = &- \delta^o \gamma^E + \rho_0(\bar{p}^S_{t|l} - \delta^o \bar{p}^S_{t+1|l}) \\ &+ (x_{1t|l} - \delta^o x_{1t+1|l})' \rho_1 + (\xi_{1t|l} - \xi_{0t|l} - \delta^o \xi_{1t+1|l}).\end{aligned} \tag{3.11}$$

We are interested in ρ_0, which governs the responsiveness of an owner-occupier's moving decision to changes in their apartment's price. This parameter can be recovered by forming the composite regressors on the left- and right-hand sides of Equation (3.11), then running a regression with fixed effects for each apartment's neighborhood and its number of rooms.

A concern in our setting is omitted variable bias (OVB) in ρ_0. The composite unobservable term

$$\tilde{\xi}_{t|l} = \xi_{1t|l} - \xi_{0t|l} - \delta^o \xi_{1t+1|l}$$

may contain terms jointly affecting sale prices and the propensity of each owner-occupier to move. For instance, during the COVID-19 pandemic, owner-occupiers are less able to search for new homes and value their existing living arrangements more. In this example, these homeowners will demand a higher price before they are willing to move, and the share of movers will fall.

To mitigate the effect of OVB, we will instrument for prices with (demand-side) policy shocks in August 2015, July 2018 and August 2019. In 2015 and 2019, the government unexpectedly raised subsidies and expanded eligibility for applicants to the BTO mechanism. Unmatched households would thus value applying to BTO over taking an apartment on the secondary market, resulting in a negative demand shock to the latter. In 2018, the government unexpectedly tightened borrowing limits on residential property by 5%, resulting in a negative demand shock over all housing.[34] For our instruments to be valid, the aforemen-

[34]To curb speculation on the market for private housing, the Monetary Authority of Singapore (MAS) mandated that the maximum Loan-to-Value (LTV) limits on individual borrowers be lowered from 80% to 75% on first housing loans in general. If the tenure of the loan exceeded 30 years or extended past the borrower's 65th birthday, the corresponding LTV limits also fell from 60% to 55%. See the corresponding MAS news release for more details.

tioned shocks must be correlated with sale prices but not with the composite unobservable $\tilde{\xi}_{t|l}$. We believe this exclusion restriction is valid because the timing and magnitude of these policy changes could not have been anticipated by the owner-occupiers.

3.8.2 Demand

We estimate demand for apartments via the Generalized Method of Moments (GMM, see e.g. Hansen and Singleton (1982)). We recover preferences and waiting costs associated with the two latent types in our model, corresponding to low- and high-income applicants respectively.

Our estimation procedure is a doubly nested fixed point procedure, comprising three loops. This procedure is akin to that of Berry et al. (1995), extended to deal with dynamics and choice over lotteries. In estimation, we condition on prices and success probabilities observed in the data. In the innermost loop, given these observed data, a guess of parameters, and a guess of the mean utility of each development, we solve the dynamic program for each type of applicant. From this loop, we obtain the model-predicted choice probabilities over each development, waiting, and exiting. In the middle loop, we find mean utility parameters that set model-predicted market shares equal to the observed shares. Having found the "correct" mean utility parameters that match observed and predicted shares, we form an IV-GMM criterion. This criterion function imposes orthogonality between the mean utilities of each development and the instruments. Finally, in the outermost loop, we search over nonlinear parameters to minimize this criterion function.

In this section, we first outline an informal identification argument. Next, we describe the GMM estimation problem. Then, we describe the instruments used to form the associated moment conditions. Finally, we specify other calibrated quantities necessary for estimation.

Identification of demand in practice

Identification for the demand model follows standard arguments in Berry and Haile (2014). The key threat to identification is that developments are attractive in ways that we do not see. Hence, in the GMM criterion, we need instruments that shift demand but not through the unobserved attractiveness of each development. To instrument for endogenous prices and success probabilities, we employ cost shocks, a policy change affecting only secondary market supply, and differentiation IVs.[35]

Given mean utilities $U_t(a)$, variation in each regressor identifies its corresponding coefficient in the indirect utility expression. Moreover, there are several developments at which agents will be successful with probability 1, so wait costs do not enter the indirect utility of applying to these developments. Thus, since the mix of developments offered varies from period to period, we can separately identify the level of indirect utilities and wait costs. Finally, variation in one-period-ahead values affects parameter estimates through contemporaneous market shares. A higher inclusive value tomorrow tilts today's applications towards apartments with lower success probabilities but higher indirect utilities.

How are random coefficients on price and wait costs identified? We leverage variation in the BTO subsidy scheme: Poor households are subsidized the most for smaller apartments in "non-mature" estates far from the city center, and these subsidies decline if applicants select larger apartments or developments in "mature" estates closer downtown.

Suppose for now that poor households were unsubsidized, and consider a comparative static in which these subsidies gradually "rose" to their true levels. At the outset, our model would predict that poor and rich households have the same coefficients on price and wait costs. As poor households become more and more subsidized, they tilt their applications towards developments with smaller apartments in "non-mature" estates. If all such developments are unexpectedly unpopular, the model rationalizes this phenomenon by lowering

[35] Further details on our instruments are available in Subsection 3.8.2.

how sensitive the poor are to sticker prices. Furthermore, suppose that in three consecutive BTO cycles, $(t, t+1, t+2)$, we see many applications to the aforementioned developments in periods t and $t+2$, but not in period $t+1$. Then the model predicts that many poor households must have waited in period $t+1$, implying lower waiting costs for the poor relative to the rich.

Estimation procedure

Our estimation procedure is two-step optimal GMM. Formally, the estimation problem is:

$$\begin{aligned} \min_{\theta} \quad & g(\xi;\theta)' W g(\xi;\theta) \quad \text{s.t.} \\ g(\xi;\theta) &= Z\xi(\theta) \\ \xi(\theta), V(\theta) &\in \{\tilde{\xi}, \tilde{V} : s(\tilde{\xi}, \tilde{V}; \theta) = \hat{s} \text{ and } \tilde{V} = \mathbb{T}(\xi;\theta)\tilde{V}\}. \end{aligned} \tag{3.12}$$

In each GMM step, we minimize the GMM quadratic form with associated weight matrix W. The moments $g(\xi;\theta)$ are the inner product of the instruments Z and the recovered unobserved development qualities $\xi(\theta)$. These unobservables $\xi(\theta)$ are chosen, via a Newton root-finding procedure, to set predicted market shares $s(\xi, \cdot; \theta)$ equal to actual market shares $\hat{s}$. Finally, in the "innermost" loop, the value function V is iterated (via the Bellman operator $\mathbb{T}$) until convergence. See Appendix 3.D for computational details.

In the next subsection, we describe the instruments Z with which we formed the aforementioned moment conditions. Then, we specify the income distribution, initial market size and age distribution, and other calibrated parameters.

Instruments

Our moment conditions rely on unobserved development quality ξ being uncorrelated with our instruments Z. Moreover, to identify the effect of price on demand, these instruments must affect prices and shift demand (but not through ξ).

We use three sets of instruments:

1. (Cost shifters.) For each development, we interact the contemporaneous prices of granite and concreting sand with the number of rooms in each apartment in that development. These cost shifters are relevant because granite and concreting sand are inputs into production. More granite and concreting sand are used to construct larger apartments; hence, when granite becomes more expensive, the cost to build a larger apartment should rise by more than that of a smaller apartment, all else equal. The exclusion restriction requires that prices of raw materials be plausibly uncorrelated with unobserved development amenities.

2. (Unanticipated stamp duty shock in January 2013.) We include a dummy for BTO periods on and after January 2013, in which the Singapore government raised Additional Buyer Stamp Duty (ABSD) rates for Singaporeans purchasing their second and subsequent residential properties. Since BTO applicants must have no other property to participate in the mechanism, they are not directly affected by this policy change. However, at the margin, the owner-occupiers "upgrading" to private property are more likely to dispose of their existing government apartments before purchasing private property. Thus, the ABSD rate changes constitute an unexpected supply shock that is plausibly uncorrelated with unobserved development amenities.

3. (Differentiation IVs.) Finally, we construct differentiation IVs (Gandhi and Houde, 2021) using the following procedure. First, we predict contemporaneous prices using a flexible function of the aforementioned instruments and other observables. Then, we

construct quadratic terms exploiting the isolation of each development j in predicted prices and exogenous observables k, as well as cross terms over different dimensions k and k'.[36]

Instrument exclusion requires that, with respect to its contemporaneous alternatives, the isolation of each development in characteristics space is uncorrelated with its unobserved quality. This exclusion restriction is plausible because, each period, the government offers a mix of developments across neighborhoods to cater to a wide range of applicants. The main constraints limiting which developments can be offered each period include site readiness, town planning, and bureaucratic approvals that are beyond the purview of the Housing and Development Board.[37] Thus, the isolation of each development in its distance to downtown, to the nearest metro stop, and other such characteristics is unlikely to systematically vary with the unobserved quality of said development.

Income distribution

For our two income types, using the 2013 Household Expenditure Survey, we discretize the income distribution for 25 to 40-year olds who are eligible to apply for BTO housing. We impute that 20% of households are poor, with a mean income level of S$36,000 a year; the remaining 80% are rich, with a mean income level of S$72,000 a year.[38] We assume two income types for tractability; with continuous distributions for random coefficients over price and wait times, solving the nested fixed point problem is computationally infeasible.

[36] We describe the construction of these instruments in Appendix 3.E.

[37] In a recent parliamentary question, the Singaporean Minister for National Development wrote that "Since May 2011, HDB has called for and awarded the construction tenders for the vast majority of BTO projects ahead of the outcome of the flat selection... However, how early the construction tenders can be launched depends on a variety of factors, such as readiness of sites, status of the planning approval and design development process, among others" (Lee, 2022).

[38] The income level of the rich households is not important to our analysis, since these households earn too much to be given any explicit subsidies under the 2011-2015 regime. In robustness checks, varying the calibrated income level for the poor does not change our qualitative results.

Market size and an agent's birth-death process

Now, we discuss market sizes and each agent's birth-death process. First, we assume each agent searches for a maximum of 10 periods, or two and a half years. We set $\bar{\tau}$ to 10 to strike a balance between capturing each agent's forward-looking behavior and reducing the complexity of the state space.[39]

Next, we assume that BTO applicants are responsible for 30% of the transactions on the secondary market. We believe that demand on the secondary market is largely driven by households who are ineligible for BTO and cannot afford private apartments or houses; thus, they are likely less sensitive to prices than agents who are eligible for BTO. Hence, we assume that this residual demand is invariant to our counterfactuals.

Then, we need to specify the inflow of new applicants in each period. We assume that about 4,000 new households are born each period, split 20:80, i.e. 800 poor and 3,200 rich. This number arises if the government builds housing at approximately the rate that new households enter.

Finally, we need to specify, at $t = 1$, the market size M_1 as well as the age and income distribution of active applicants μ_1. These values are obtained implicitly by estimating the model twice.[40] First, we assume 16,000 households were active at the beginning of our sample; these households are 20% poor and 80% rich; and their ages decay exponentially.[41] We call these initial conditions `Init`. Next, we estimate the model assuming (M_1, μ_1) are the true initial conditions, obtaining unobserved development qualities $\hat{\xi}_0 \sim \hat{F}_\xi$. Then, we match the development characteristics in the 5 periods preceding our sample with the

[39] Most households who participate in BTO are matched in very few tries: the "mean number of attempts by [successful] first-timer families... is 1.3" (Lee, 2021). Varying the maximum number of search periods $\bar{\tau} \in \{8, 12\}$ does not change our qualitative results.

[40] We augment the data because we see the development characteristics before our sample period, but we do not see the number of applicants with the correct priority level, i.e. "first-timers."

[41] We chose the value of 16,000 because this number of Singaporean couples between the ages of 25 and 40 were first married within a 2.5-year window corresponding to $\bar{\tau} = 10$. In robustness checks, varying the initial market size $M_0 \in \{15000, 17000\}$ does not change our qualitative results.

"closest" sequence of 5 periods within our sample. Finally, we append the matched sequence to the beginning of the sample and redraw the unobserved development qualities $\tilde{\xi}_{jt}$ for the matched sequence from the first-step empirical distribution $\hat{F}_{\xi}$. Holding these draws $\tilde{\xi}_{jt}$ fixed, we re-estimate the model with the augmented sample, now assuming that the initial conditions at the beginning of said sample are `Init`. Further details are available in Appendix 3.F.

Other calibrated parameters

We describe our remaining calibrated parameters. We maintain our assumptions that all households have the same priority, and the applicants cannot affect the success probabilities of other agents significantly by applying to a different apartment. We also calibrate the annual discount factor to the standard value of 0.96.[42]

3.9 Parameter Estimates

Having discussed our estimation procedure, we now report our parameter estimates. First, we report the price sensitivity of the owner-occupiers on the secondary market. Next, we discuss the young households' preference parameters and wait costs we recovered in our GMM procedure. Finally, we present several measures of model fit and overidentification checks.

3.9.1 Parameter Estimates — Supply

We estimate this regression offline. Our data span all secondary market transactions during the 2010-2020 BTO cycles. We drop pairs of observations in which either no sales were observed at times t or $t+1$ for a given apartment type l. As is typical in the literature, we assume an annual discount factor of 0.96.

[42] In robustness checks, varying the discount factor $\delta \in \{0.9, 0.98\}$ does not change our qualitative results.

The regression results can be found in Table 3.I.4. We include fixed effects for each neighborhood and the number of rooms in each apartment. In our preferred specification (4), we find a price coefficient $\hat{\rho}_0$ of 0.712, significantly different from zero at the 5% level. This price coefficient maps to a mean price elasticity of supply of 3.64. Our elasticity is substantially higher than similar measurements in prior work, including the mean across American cities of 1.75 (Saiz, 2010) and the mean census tract-level housing unit provision elasticity of 0.3 (Baum-Snow and Han, 2021). However, we caution that our estimated elasticity is not directly comparable with its counterparts. While our predecessors analyze static housing provision, we model our owner-occupiers as forward-looking and responding to the path of future housing prices.

3.9.2 Parameter Estimates — Demand

After estimating supply offline, we tackle demand. We estimate our model for the January 2012 to May 2015 BTO cycles, restricting random coefficients to be on wait costs and BTO sticker prices.[43] We restricted our sample because later periods may not have available five-year-ahead resale prices; moreover, in August 2015, there was an unexpected policy change.[44] Our results are presented in Table 3.I.5. Mean utilities depend on

- the distance to the Marina Bay Financial Centre, a proxy for proximity to downtown;
- the distance to the nearest Mass Rapid Transit (MRT) station, a proxy for transit connectivity;
- the distance to the nearest SAP school;

[43] We limit the number of random coefficients we estimate because of power concerns — we only have access to "development-level demand." We are currently requesting data from the Central Provident Fund of Singapore to obtain income and family sizes of matched households, allowing us to estimate a richer model.

[44] See Subsection 3.A.3 in the Appendix for more details. It is difficult to incorporate the August 2015 policy change "in sample" because richer agents became eligible for BTO housing, necessitating a third income type we must track, among other computational burdens.

- apartment sticker price; and
- a "forward resale price" term, proxied for by the five-year-ahead resale price of similarly-sized apartments in the same neighborhood of the development in question.[45]

Our results are consistent with our stylized findings. Transit connectivity and distance to downtown are important factors, as are the apartment size and sticker price. We compute the willingness for each income type to pay for various amenities in Table 3.I.6, finding that the rich are more willing to pay for transit access and especially to be closer downtown. To assess the plausibility of our willingness-to-pay (WTP) estimates, we attempt to convert them into Singapore dollars per month. Since each BTO apartment must be held for 5 years before it can be sold off, if we divide the WTP value reported in our table by 60, we obtain an indicative measure of the "monthly user value" of the amenity in question. As an example, the poor appear to value living 100 meters closer to a metro stop at S$35/month, while the rich value the same amenity at about S$142/month. This could be because the rich more are likely to work in knowledge-intensive industries, which are currently clustered downtown. Moreover, the rich are more likely to travel longer distances for work and play, while the poor are less likely to enjoy leisure activities outside of their immediate vicinity.[46]

[45] We included resale prices five years ahead because those prices are realized at the earliest time at which the household is legally allowed to sell its apartment off. Our other covariates were chosen because government officials and participants in the mechanism viewed them as important. See Appendix 3.C for more details.

[46] The fares for a bus and subway trip scale with distance. With a transit farecard, each trip costs a minimum of S$0.92 and rises to S$2.77 (see the Public Transport Council website for a detailed breakdown of Singaporean public transit fare structure). In contrast, taxis and platform ride-hailing services are much more expensive. Thus commuters are likely much more reliant on bus and rail than taxis and private car services.

Moreover, our WTP results corroborate the findings in Lee and Tan (2023), who show that the construction of a new subway line may help the rich but hurt the poor in the long run. They find that rich households choose to shop downtown, so service-intensive sectors relocate away from the suburbs, extending commute times for the poor.

3.9.3 Model fit

To see how well our demand model fits the data, we display several targeted and untargeted moments in Table 3.I.7. Associated binned scatter plots can be found in Figure 3.I.1 in the Appendix. These targeted moments cover amenities valued by households, while the untargeted moments validate that there is unlikely to be a time trend or schooling amenity captured by unobserved development quality; moreover, the application behavior implied by our model is consistent with figures brought up in the Singapore parliament.

First, we inspect the distribution of unobserved development qualities ξ. The GMM algorithm successfully matches (among others) three targeted moments: ξ should be of mean zero and uncorrelated with the distances of each development to downtown and the nearest metro stop.

Turning to untargeted moments, we note that unobserved development qualities ξ are uncorrelated with the BTO cycle in which it is offered, as well as the distance of said development to its nearest school. Thus, ξ is unlikely to capture a time trend or unobserved schooling amenities. Our final untargeted moment is the percentage of agents making three or fewer applications. The Singapore Ministry of National Development (MND) notes that most young households applying for apartments in non-mature estates "are able to book a flat... on their third try" (Wong, 2017b). In simulations detailed in Section 3.10.1, we find that 90% of agents make 3 or fewer applications, which is congruent with the MND report.

3.10 Counterfactuals

3.10.1 Alternative policies

With parameter estimates in hand, we can now simulate alternative policies targeting key weaknesses of the BTO mechanism. Given the government's stated objectives, we will evalu-

ate these policies on apartment vacancies, applicant wait times, and prices on the secondary market.

The first of these alternate policies is to change the rules of the mechanism to eliminate intertemporal risk. Under the current mechanism, if young households try for an oversubscribed development this period and fail, they cannot go back in time and "settle for" a previously offered development. To mitigate this risk, we will implement a strategyproof benchmark, akin to a dynamic version of Random Serial Dictatorship (RSD). At the given prices, let applicants truthfully declare their preferences; then, draw household priorities at random and assign households to their preferred developments/period of exit. To compute this counterfactual, we solve for the new equilibrium. Equilibrium prices must set excess demand for apartments on the secondary market to zero.

There are two main computational challenges in this procedure: computing household preference lists and solving for prices that clear the secondary market. We sketch the mechanism below:

1. Given prices on the secondary market, applicants truthfully declare their preferences over all apartments over which they are eligible, as well as exiting to the secondary market. For an applicant of type (y, t_0), where t_0 is the her birth period, she is eligible for all developments in periods $t_0, t_0+1, \ldots, t_0+\bar{\tau}-1$, and she can exit in periods $t_0, t_0+1, \ldots, t_0+\bar{\tau}$.[47]

2. The PHA draws household priorities at random and assigns households to their preferred developments (or exiting, if applicable).

The simulation algorithm searches over prices on the secondary market until excess demand for those apartments is zero.[48]

[47] Despite wait costs, in our simulations, a young household who fails to get her preferred developments may not want to exit immediately. This behavior is driven by two factors. First, secondary market prices may fall in the near future, so a sufficiently patient agent may wait to exit in a later period. Second, an applicant may have a high idiosyncratic utility draw for waiting today, leading her to exit tomorrow.

[48] Details are available in Appendix 3.G.1.

We showcase this alternate mechanism because it has two attractive properties. First, this policy completely eliminates intertemporal risk. Additionally, this mechanism is strategyproof: It is a dominant strategy for BTO applicants to report their true preferences. Hence, when finding secondary market-clearing prices, it suffices to recompute agent preferences once. Then, we can simulate many realizations of allocations from the mechanism given each draw of agent priorities, simplifying the computation of excess demand on the secondary market. However, this mechanism also has a drawback: the government has to run the assignment "once and for all" at the start of the sample period, accounting for all potential entrants.[49] To avoid this drawback, the planner could consider a version of the "common lottery" (Shi and Yin, 2022). Upon their first entry in the mechanism, each household is assigned a persistent lottery number. With this lottery number, each quarter, agents apply for at most one BTO development, and are assigned apartments based this lottery number.[50] Under perfect foresight over future developments and knowledge of the equilibrium success probabilities at each development, this new mechanism also eliminates intertemporal risk.[51]

The second alternate policy directly targets housing scarcity: building more. To operationalize this policy, we increase the quantity supplied in oversubscribed developments by 10%. We only consider increasing supply by a small amount to abstract from changes in

[49] Given infinite computational power, we would prefer our agents to specify their preferences period-by-period over offered developments, exiting, and waiting. Unfortunately, we are cursed by dimensionality. For us, this alternate approach entails three computationally-intensive loops. Given a guess of secondary market prices and the value of waiting, we must compute preference lists and simulate assignments. From these assignments, we can obtain an implied value of waiting; then we iterate value functions until convergence. Finally, given assignments at the converged values of waiting, we need to check if the secondary market clears. Thus, cursed by dimensionality, we are forced to compute a benchmark that is feasible to us, but infeasible to the planner.

[50] Another feasible alternative would be allowing applicants to specify (1) the development to which they want to apply; and (2) if they fail, which development next period at which they want priority. Then, if today's applicants do not get their preferred development, they will get an additional lottery draw for their specified development next period.

[51] While the "common lottery" places great burdens on what the decision maker needs to know and anticipate, this alternate mechanism can be implemented and can accommodate priorities based on income or household size. We will investigate the "common lottery" mechanism, adapted to our context, in future work.

equilibrium prices and quantities in the forward resale market.[52] Doing this requires two substantive assumptions. First, changing apartment supply must not significantly shift patterns of demand and supply for local amenities. Second, patterns of residential sorting must remain invariant under the counterfactual scenario. We argue that, in our setting, both assumptions are reasonable. Residential sorting is plausibly invariant because the government imposes (binding) racial quotas in each block of public housing, and balloting for new apartments is limited to households below an income threshold. Furthermore, most neighborhoods in Singapore see their housing stock grow by under 5% per year. If we limit our counterfactual supply expansion to about 10% per development, then each neighborhood sees a further increase of 0.5% relative to the baseline, which we argue is not of first-order consequence.

To find a new equilibrium when more apartments are built, recall that we need to find success probabilities q and secondary market prices p^R such that, at the prevailing prices, the success probabilities q are self-fulfilling; and, at the given success probabilities, the secondary market clears at the prevailing prices. Computational details, as well as notes on equilibrium existence and uniqueness, are available in Appendix 3.G.2.

3.10.2 Outcome comparisons

To assess the efficacy of each counterfactual in terms of efficiency and equity, we compute three sets of summary statistics:

[52] The reader may be concerned that even the "sticker prices" for each BTO development might change in response to the government's own policy changes. Suppose the government expanded supply in a particular development. Even if the effects of this policy "wash out" far in the future, it may be that contemporaneous resale prices fall. If the government had a responsive pricing rule (not disclosed in detail as of this writing), they might then lower the "sticker price" of said development in which they expanded supply.

We argue, given the set of covariates available to us, that contemporaneous resale prices fail to explain the residual variation in the "sticker prices" of each development. We first regress the "sticker price" of each development on our covariates, obtain the residual from that regression, then regress this residual on contemporaneous resale prices and the number of apartments in the offered development. We fail to reject the null hypothesis that contemporaneous resale prices and development quantity do not explain the residual variation in "sticker prices." Therefore, the government likely does not price according to information that we do not have.

1. Simulated wait times

2. Apartment vacancies

3. Prices on the secondary market

These summary statistics map directly into outcomes the Singapore government cares about, as elucidated in Subsection 3.3.2. The primary objective is to assign young households to apartments as soon as possible, subject to budget and incentive constraints. These households benefit by having a home in which to settle down and start a family; the government gains from de-congesting the "queue" of active applicants. Moreover, vacant apartments should be minimized, as they are expensive for the government to hold and (re-)allocate. Finally, existing homeowners should not have their apartments collapse in value, jeopardizing home equity they could draw down for retirement.

Changes in these outcomes are displayed in Table 3.I.9. In the actual mechanism, we directly observe development vacancies and mean prices on the secondary market. First, the Singapore government allocates 88% of the apartments it offers at BTO through the mechanism, so it is mostly successful at avoiding vacancies. By way of comparison, as of 2018, half of the 200,000 partially complete Mehr Project apartments in Iran were "left without prospective buyers" (Radio Farda, 2018). Second, prices on the secondary market over the sample period averaged S$460,000, against S$100,000 for the median household's income from work in 2015. The (secondary market) apartment price to median income ratio is thus 4.6:1, very slightly exceeding the 4:1 "rule of thumb" measure of housing affordability preferred by the Singapore government. We note that our chosen measure is conservative in two ways: we consider the mean price of all secondary market transactions in Singapore, not just those in neighborhoods farther from the city center. Moreover, prices on the secondary market are higher than comparable apartments allocated at BTO. After considering these two mitigating facts, we find that public housing in Singapore was affordable in general between 2012 and 2015.

Third, the poor wait 1.5 years to be assigned to an apartment, against 10 months for the rich. After being assigned, each household has to wait approximately 3 years for their development to be built. Therefore, households can wait upwards of 4 years between starting their search and taking possession of their apartment. For a (distant) comparison, between 2010 and 2014, among the developments administered by the Cambridge Housing Authority in Massachusetts, 327 successful applicants waited between 1.6 to 3.8 years before being housed (Waldinger, 2021). We note that these numbers are not directly comparable because the Singapore and Cambridge programs differ in scale and purpose. The Singapore mechanism targets households from the 5th to the 80th percentile of incomes, i.e. not the destitute. These households are more likely to be able to wait for their preferred development than the agents in Cambridge are.

Building more apartments

Under the current mechanism, at the margin, adding supply is ineffective because demand substitution undoes the supply response. If the designer builds 10% more apartments in oversubscribed developments, vacancies rise from 12% to 16%. Wait times fall slightly for the rich, by about 3%. However, the poor do not wait less. These findings contrast with the case where developments are homogeneous: in that scenario, wait times would unambiguously fall. In our setting, when supply expands in competitive developments, agents may substitute from a "safe" development today to a "risky" development tomorrow, i.e. choosing to wait longer to gamble for a better apartment. Though substitution attenuates the effect of expanded supply on wait times, the rich still wait about 1.3 weeks less on average.

More notably, prices on the secondary market remain unchanged. This effect is unsurprising to us: a 10% expansion of supply in oversubscribed developments represents a 7.8% increase in the flow of apartments between January 2012 and May 2015. Since the flow of apartments is miniscule relative to the stock, prices on the secondary market do not change.

In summary, because of substitution within a period and across periods, the demand response almost completely undoes the effect of added supply.

Changing rules

Our proposed strategyproof mechanism successfully reduces vacancies and wait times, but at the cost of higher (short-term) congestion on the secondary market. Vacancies fall from 12% to 7%, and wait times decrease by 8% for the poor and 23% for the rich. Two economic forces are at play. First, agents benefit from being allowed to list more than two developments (and/or exiting) in the same period. Since more agents are matched to developments within the mechanism, vacancies fall. Next, instead of failing at a development late in their search, an agent may prefer exiting early. Therefore, on average, agents wait less. However, since more agents exit, prices on the secondary market must rise to incentivize owner-occupiers to sell their units. These prices rise by 12%, a substantial amount.

These findings demonstrate that rule changes are not free: spillovers on the secondary market benefit owner-occupiers at the expense of young households. Is there a way to mitigate the increase in prices on the secondary market?

Changing rules and adding apartments

In our final counterfactual, we adopt the aforementioned mechanism and also have the designer build 10% more apartments in oversubscribed developments. In this intervention, we find that adding supply partially alleviates the price increase in the secondary market. Relative to the true mechanism, vacancies fall from 12% to 10%. Wait times fall by similar magnitudes as in the scenario where only the rules were changed. The added supply relieves upward pressure on prices in the secondary market, leading them to rise by only 9% (as opposed to 12% when only DA was implemented).

Given the supply increase, young households still substitute away from undesirable developments to more popular ones. Thus, 10% of apartments remain vacant, as opposed to 7% when only the rules were changed. However, at the margin, young households benefit from the added capacity in desirable developments, so fewer of them exit to the secondary market.

Thus, our main takeaway from this section is that adding supply alone is not a panacea for congestion. Policymakers must pay attention to the rules through which this supply is allocated.

3.11 Conclusion

We study how public housing can be allocated more efficiently and equitably. In particular, we evaluate what happens when the rules are changed, when more apartments are built in overdemanded developments, and when the two interventions are combined. To this end, we combine tools from Urban Economics and Industrial Organization to estimate a dynamic choice model over housing lotteries.

In our setting, a key source of inefficiency is intertemporal risk: if an agent applies for a competitive development today and fails, she cannot go back in time and "settle for" a safe development in a previous period. To target this source of inefficiency, we consider a strategyproof mechanism eliminating intertemporal risk.

We also consider what happens when the government builds slightly more apartments in oversubscribed developments. We find that better rules complement building more apartments. Expanding supply alone fails to reduce wait times because of substitution within and across periods. However, under our proposed mechanism, wait times and vacancies fall, but at the cost of higher prices on the secondary market. In the new mechanism, building more apartments not only preserves the lower wait times, but also relieves some of the upward pricing pressure on the secondary market.

Our framework and insights on market design and supply control are applicable beyond public housing. Previous work investigates environments in which the designer cannot directly supply the good, like kidney donation (Agarwal et al., 2021a) and school choice (Abdulkadiroğlu and Sönmez, 2003). We bridge a gap between the former settings and that of platform design on two-sided marketplaces (see e.g. Dinerstein et al. (2018)). Typically, platforms intermediate between customers and sellers, needing to keep prices low to attract customers while sharing enough surplus with the sellers so that they do not exit (Lee and Musolff, 2021). A similar dynamic is present in our setting: the government must conduct policy that balances housing access for young agents against preserving home equity for existing owner-occupiers. In Singapore, the effects of higher supply on the housing market depend on the rules under which the supply is allocated. Similarly, a platform could control the supply of goods on the marketplace it hosts (e.g. AmazonBasics). How does control over supply interact with the rules by which products and merchants are recommended?

Our approach comes with several caveats. In our model, we do not permit agents to learn from past play. In our setting, we believe this restriction is not first order, since households only apply for one apartment each period and are given ample information about current and future developments. Moreover, we abstract from state transitions in household financial status and family size, since we do not observe these variables at the individual level. Because most households are matched to apartments in less than 2 years, we believe these concerns do not detract from our mainline results. More can be done with better household-level data: this way, one can investigate differences in apartment choice probabilities when applicants expect a baby in the near future.

Finally, little is known about where the government should optimally locate public housing, given that local amenities respond to large changes in the housing stock. In future work, we intend to embed our allocation problem in a general equilibrium setting to answer these questions.

Appendix

3.A Additional descriptive statistics

3.A.1 Additional correlation plots

Our main data set used in estimation has 195 rows and 36 columns. We scraped the HDB website for data on BTO application cycles from August 2010 to November 2015 and subsetted it to January 2012 to May 2015. From each news release on the HDB website, we obtained the town, size, quantity and price details of each BTO development. These details were merged with aggregate application data on the HDB website after each BTO cycle has concluded. Finally, we corroborated the HDB data with resale flat prices from http://data.gov.sg, as well as amenity data from http://onemap.sg/. Our amenity data take the form of, for instance, "distance in kilometers to nearest MRT (metro) station."

In Figure 3.I.5, we correlate the main dependent variable in our analysis, application rates, with our main covariates. In this analysis, the key amenities are the distances to downtown, the nearest metro stop and the nearest Special Assistance Plan (SAP) school. As in the previous paragraph, the closer a development is to any of these amenities, the higher application rates are, all else equal. These amenities are correlated in the expected direction with the average price of an apartment within each development.

3.A.2 BTO apartments are oversubscribed

BTO apartments are typically oversubscribed by a factor of at least 2:1. Below, in Table 3.I.10, we display selected BTO application rates in the May 2015 cycle. Apartments in non-mature estates generally see lower oversubscription than in mature estates, owing to their being farther away from the city center and other amenities.

From our conversations with BTO applicants and government officials, we learned that when a household is offered an apartment within the mechanism, they almost always take it up. This observation arises because if an applicant forfeits her spot, she loses priority and becomes exceedingly unlikely to be matched in a future cycle. What few unmatched apartments exist are then offered to households applying to irregular clearinghouse cycles.[53] Motivated by these observations and to improve model tractability, we do not allow households to reject apartments they have been matched to, and assume that the PHA only shortlists applicants up to 100% the housing supply.[54]

3.A.3 Implicit subsidies differ by apartment size and neighborhood

We define implicit subsidies given by the government as the difference between an apartment's sticker price at BTO and its five-year-ahead resale price. The government prices BTO flats below market rates "to ensure that new flats are affordable to buyers" (Wong, 2017a). Though not much is known about how the government sets BTO prices, we believe it considers "the prices of comparable resale flats in the vicinity, as well as the specific attributes of the flats, such as storey height and design" (Wong, 2017a).

[53] In Table 3.I.12 in the Appendix, we display application rates for one such clearinghouse cycle, the May 2018 Sale of Balance Flats exercise.

[54] In practice, the PHA shortlists households up to three times the housing supply, resulting in almost all apartments in a development being matched to some household.

Such an implicit subsidy may be conducive to other policy objectives; for example, by making homes more affordable, the government is able to encourage family formation and fertility. However, such subsidies may not be allocatively efficient. Given that BTO apartments are oversubscribed, in a hypothetical decentralized market, there exist prices that exactly match demand with supply. Instead, in the BTO mechanism, the sticker prices set by HDB do not clear the market for public housing, and provide an incentive for household arbitrage.[55] Figure 3.I.6 displays the implicit discount conferred upon households applying to each apartment type. We see that most apartment offerings are priced far below the five-year-ahead resale prices of similar apartments, with a mean of S$88,000; yet there is a fair amount of dispersion, ranging from the BTO sticker price being S$200,000 below to S$50,000 above comparable resale prices.

To be clear, not every household who applies for a BTO flat solely values its resale potential; there is also "consumption value" in living in an apartment that is close to amenities. Yet the dispersion in implicit subsidies can be a powerful force spurring BTO applicants to prioritize apartments with higher resale value net of the sticker price.

3.A.4 Hedonic regressions

We display additional hedonic regressions exploring how distances to various amenities predict oversubscription rates. We specify our first regression as:

$$\text{rate}_{jlt} = x'_{jlt}\beta + \gamma d^{\text{SAP}}_{jlt} + \xi^{\text{mature}} \times \mathbf{1}\{l \in \text{mature}\} + \xi_j + \xi_t + \epsilon_{jlt}. \tag{3.13}$$

That is, we regress the number of BTO applications per apartment in each development (j, l, t), "rate", on various amenity covariates and an indicator of whether the estate is ma-

[55] One can view public housing apartments in our setting as being rationed by a combination of prices and wait times.

ture. To control for persistent differences in application rates across apartments of different sizes and in different BTO cycles, we include apartment size and time fixed effects.

Our results are in Specification (1) of Table 3.I.11. Our amenity covariates include the distance to the Marina Bay Financial Centre (a proxy for distance to downtown), to the nearest metro station, to the nearest hawker center, to the nearest supermarket, to the nearest non-Special Assistance Plan (SAP) school and to the nearest SAP school. We distinguish SAP from non-SAP schools to proxy for school quality: SAP schools specialize in bilingual education in English and Chinese and are academically competitive.[56] The regression coefficients we obtain are mostly as expected, though application rates are higher for housing estates farther from downtown. This could be because many households work in areas outside the city center.[57] Our other coefficients behave as expected. For example, an apartment 100 meters farther away from a metro stop is associated with a 0.104 lower application rate, all else equal.

A particularly noteworthy amenity is the proximity of an apartment to the nearest school. In Specification (1), we note that a 100m reduction in distance to a SAP school is associated with a 0.0292 higher application rate, all else equal. But we know more about the dependence between school distance and application rates: the distance to a school factors into admission priorities for first grade in Singapore.[58] Specifically, households living within 1 kilometer of a school have the highest priority for admission to that school, followed by households living between 1 and 2 kilometers, and finally households living outside 2 kilometers. In our sample,

[56]By some measures, students in SAP schools receive additional funding relative to students in non-SAP schools; for instance, S$300 per student is earmarked for "develop[ing] their proficiency and interest in Chinese language-related studies." See, for example, this statement by then-Minister for Education Ong Ye Kung.

[57]Many neighborhoods in Singapore are incompletely specialized between residences and workplaces. See Lee and Tan (2023) for more details.

[58]The admission exercise for first graders ("Primary One Registration Exercise") has a complex system of priorities which is beyond the scope of this chapter. The first phase, for rising first-graders who have a sibling currently enrolled in the school, does not employ distance priorities. A minority of students enroll through this phase. Within each of the remaining phases, if there are more applicants than school places, distance priorities confer an absolute advantage in the subsequent lottery. See the Singapore Ministry of Education website for more details.

very few BTO developments are located within 1 kilometer of a SAP school. However, several developments are located within 2 kilometers.

A first pass is to regress application rates on amenity covariates, as well as a dummy for whether the BTO development is within 2 kilometers of a SAP school. The results can be found in Specification (2) of Table 3.I.11. We find that apartments within 2 kilometers of a SAP school are about 1.9 times more oversubscribed than apartments outside 2 kilometers, all else equal.

To further clarify the dependence of apartment application rates on the distance to a SAP school, we collect BTO developments into 500-meter bins by distance, then regress application rates on dummies for each bin. The results are displayed in the left plot in Figure 3.I.7. We note that the mean application rate within the 2-kilometer boundary is about 2.5, which is about double that of the mean outside the boundary. To control for systematic differences between apartments within 2 kilometers vis-a-vis those outside the boundary, we control for the other amenity covariates and display the results in the right plot in Figure 3.I.7. Our results remain qualitatively unchanged. For robustness, we run falsification tests on average BTO price and distance to metro stop by distance to SAP elementary school, and find essentially no difference. These falsification plots can be found in Figure 3.I.9 in the Appendix.

We hesitate to interpret the results in Figure 3.I.7 as causal because of two possible confounders. The first is that apartments within 2 kilometers of a SAP school may have higher unobserved quality than apartments outside 2 kilometers. For example, SAP schools may be located adjacent to green or water features (by virtue of zoning requirements), so any "average treatment effect at 2 kilometers" would be confounded by the effect of proximity to natural features. The second is that BTO application rates are equilibrium objects. These objects depend on the choice sets of households in that particular BTO cycle, as well as

which households choose to participate then. Therefore, a full structural model is required to disentangle substitution patterns between offered BTO apartments.

3.B Under new rules, do BTO applicants pay more for aftermarket apartments?

In our main analysis, we find that our proposed rules lower vacancies and wait times, but raise prices that young households pay for government apartments on the aftermarket. Could young households pay less for government apartments under the new rules? In two theoretical examples, we show that young households could pay less (or more) for aftermarket apartments.

If the secondary market were stationary, then, when more households are matched, prices should go down. However, in our sample period, aftermarket congestion was particularly severe in earlier periods. Under the new mechanism, households incur waiting costs by delaying their applications. If they know they will not get their preferred apartments late in their search, they might prefer exiting early, even though they face higher prices in the short run.

3.B.1 Our proposed rules could lower average aftermarket prices

We first illustrate how, under our proposed mechanism, aftermarket prices may fall on average. As in the theoretical example section, There are two periods, $t \in \{1, 2\}$. In Period 1 (2), one house A (B) arrives. At the end of each period, vacant houses vanish. Three identical players $i \in \{1, 2, 3\}$ are born at the beginning of time and are perfectly patient. Every period, each active player i simultaneously takes one of three actions. She can apply for the available house; wait one period; or exit the market. If house j is oversubscribed, the winning agent is chosen via uniform random lottery. Agents who are matched to a house

or exit the market stop making further decisions. If an agent waits or otherwise remains unmatched in Period 2, she exits the market.

Agents value house A at 0 utils and B at 10 utils. In any period, agents exiting the market receive a payoff of $-1 - P(n_E)$, where $P(n_E)$ is the aftermarket price. In this section, the aftermarket price only depends on the total number of households who exit. Assume agents face zero waiting costs, and set $P(x) = x$ for $x > 0$ and $P(x) = 1$ otherwise.

We seek a Subgame Perfect Nash Equilibrium in ex ante symmetric strategies. For the BTO mechanism, we conjecture that all households wait for and apply to house B. We solve for this equilibrium by backwards induction. Consider any subgame in which all three households are still unmatched at the start of the second period. Then, it is strictly dominant for them to apply to house B over exiting. Their expected utility is thus

$$V_2(3) = \frac{1}{3} \times 10 + \frac{2}{3} \times (-1 - P(2)) = \frac{10}{3} - 2 = \frac{4}{3}.$$

In the first period, suppose all other agents passed on house A. Then the remaining agent should also pass on house A, because waiting gives higher expected utility than getting house A for sure:

$$v_1(w, -w) = \frac{4}{3} > 0 = v_1(a, -w).$$

Under this equilibrium, one agent is matched to house B and the other two agents remain unmatched. The price on the aftermarket is $P(2) = 2$.

Suppose we implemented our version of Random Serial Dictatorship. Since the mechanism is strategyproof, all agents truthfully express that they prefer house B to house A to remaining unmatched. The planner assigns one agent to house B, one to house A, and one to the aftermarket. The price on the aftermarket is then $P(1) = 1 < 2 = P(2)$. Therefore, under our proposed rules, aftermarket prices could fall.

3.B.2 Our proposed rules could raise average aftermarket prices

However, our proposed rules could also raise average aftermarket prices, especially if aftermarket congestion is severe in earlier periods.[59] The key intuition is that, in lieu of paying waiting costs to "try for" developments late in their search, agents are willing to exit early, even if they pay higher prices on the aftermarket.

Suppose agents value house A at 1 utils and B at 3 utils. Their payoff from exiting is $1 - P_t$ utils, where P_t is the price they pay for an apartment on the aftermarket in period t. Between periods 1 and 2, agents incur a waiting cost of 1 util. In each period, housing supply is perfectly elastic, with prices $P_1 = 3/2$ utils and $P_2 = 1$ util respectively.

In the BTO equilibrium, we conjecture that agents apply for house A, then the remaining unmatched agents apply for house B. Consider any subgame in which two households remain unmatched at the start of period 2. Then, it is strictly dominant for them to apply to house B over exiting. Their expected utility is

$$V_2(2) = \frac{1}{2} \times 3 + \frac{1}{2} \times (1 - 1) = \frac{3}{2}.$$

In the first period, suppose all other agents apply to house A. Then it suffices to show that the last agent has no profitable deviation by waiting or exiting in period 1:

$$v_1(a, -a) = \frac{1}{3} \times 1 + \frac{2}{3}(-1 + \frac{3}{2}) = \frac{2}{3} > \frac{1}{2} = v_1(w, -a) > 1 - P_1 = -\frac{1}{2} = v_1(e, -a).$$

[59] In our sample period, aftermarket congestion in Singapore was much more severe in earlier years because of rapid population growth; furthermore, the government was slow to expand the supply of new apartments. Housing congestion and affordability played a role in the ruling party's historically poor performance in the 2011 general election (Huat, 2017, 88). After 2011, the Housing and Development Board ramped up the supply of new BTO apartments. Combined with borrowing limits and transaction taxes on second homes, housing congestion was relieved substantially by the end of our sample period, contributing to the ruling party's landslide victory in the 2015 general election (Huat, 2017, 94).

Under this equilibrium, one agent is matched to house A, one to house B and the last exits in period 2. The average price paid by households on the aftermarket is $P_2 = 1$.

Under our version of Random Serial Dictatorship, all agents state that they prefer house B to A to exiting in period 1 to exiting in period 2. Under this mechanism, one agent is matched to house A, one to house B, and the last exits in period 1. The average price paid on the aftermarket is $P_1 = \frac{3}{2} > 1 = P_2$. Thus, under our proposed rules, it is possible for average aftermarket prices to rise.

In sum, whether aftermarket prices rise or fall depends on which economic forces dominate. If the aftermarket is stationary and the new mechanism successfully matches more applicants, then prices should fall. On the other hand, if agents are willing to pay to avoid incurring waiting costs, these agents may exit early, driving up aftermarket prices. Therefore, to evaluate the relative importance of these economic channels, we need to take our full model in Section 3.7 to the data.

3.C Motivating our specification of household preferences

In informal conversations with officials at the Housing and Development Board, we understood that BTO applicants tended to prioritize apartment size over development location, because "the difference in [apartment] sizes [is] rather substantial," and

> "And even for resale value, while in terms of [percentage] gains, smaller flat types might make sense, Singaporeans generally still lean to[wards] nominal profits."
> — Mr. XY[60]

[60] Names have been changed to preserve confidentiality.

Despite the hefty subsidies given to successful BTO applicants, the official noted that only about 10% of BTO buyers eventually sold their apartments within 5 to 10 years of purchase, with about 5% of all HDB owners buying/selling an apartment in any given year.

We also spoke to participants in the BTO mechanism. They generally prioritized a development's proximity to downtown, its distance to the nearest metro stop and its distance to the nearest mall.

3.D Estimation algorithm for demand

The computational procedure for demand proceeds as follows:

1. Set the GMM weight matrix W as the identity matrix.
2. Guess a value of the parameters θ.
3. Given θ, compute structural residuals ξ and value functions V using a nested fixed point procedure:
   (a) Given θ and a guess of ξ, iterate the value function V until convergence.
   (b) Compute the path of market sizes M_t, type distributions $\mu_t(y, \tau)$ and model-implied market shares $s(\xi, V; \theta)$.
   (c) If the model-implied shares $s(\xi, V; \theta)$ exactly equal the observed market shares $\hat{s}$, stop and return ξ. Else update $\xi \mapsto \xi'$ by taking a Newton step, then repeat steps (a) to (c) till convergence.
4. Form the GMM objective function specified in Problem 3.12.
5. If the GMM objective function is below a specified tolerance, stop and return the parameter estimate $\check{\theta}$. Else repeat steps 2 to 5 until the optimization routine has converged.

6. Given the estimated parameters, compute the optimal weight matrix $W(\check{\theta})$.

7. Repeat steps 1-5, replacing the identity matrix in Step 1 with the optimal weight matrix computed in step 6.

3.E Construction of Differentiation IVs

This section describes the construction of Differentiation IVs, previously sketched in Section 3.8.2. In the first stage, we regress the sticker price and forward resale price on cost shifters z_{at}^{cost} and exogenous amenities,

$$p_{at} = z_{at}^{\text{cost}\prime}\gamma^z + x_{at}'\gamma^x + \nu_{at}^P; \quad p_{at}^R = z_{at}^{\text{cost}\prime}\gamma^{Rz} + x_{at}'\gamma^{Rx} + \nu_{at}^R.$$

Then, we predict these prices, obtaining $\hat{p}_{at}, \hat{p}_{at}^R$. Finally, we construct quadratic IVs based on five variables; the two predicted prices and three distances: to downtown, to the nearest metro stop, and to the nearest SAP school. Our instruments are thus

$$z_{at}(x_t, p_t, p_t^R) = \left(\sum_{a \neq a'} (d_{at,a'}^{\hat{p}})^2, \sum_{a \neq a'} (d_{at,a'}^k)^2, \sum_{a \neq a'} d_{at,a'}^k \times d_{at,a'}^{\tilde{k}} \right); \quad \forall\, k \neq \tilde{k},$$

with the difference terms computed as

$$d_{at,a'}^k = x_{at}^k - x_{a't}^k.$$

3.F "Burn-in" procedure for initial conditions

We need to specify, at $t = 1$, the market size M_1 as well as the age and income distribution of active applicants μ_1. These values are obtained implicitly by estimating the model twice:

1. (Assume initial conditions.) Assume that 16,000 households were active at the beginning of the sample period. As for the initial age and income distribution, we assume a 20:80 ratio between poor and rich households across all ages; then, we assume that the number of households of ages $\tau = 2, 3, \ldots, 10$ decays at an exponential rate ρ. This exponential rate ρ is chosen such that

$$\sum_{y \in \{L,H\}} \sum_{\tau=1}^{\bar{\tau}} \exp(-\rho(\tau - 1))\mu_1(y, 1) = 1.$$

2. Using the estimation procedure described in Subsection 3.8.2, obtain the parameter estimates $\hat{\theta}_0$ and recovered unobserved development qualities $\hat{\xi}_0$. Let the empirical distribution of unobserved development qualities be $\hat{F}_\xi$.

3. (Select a sequence of periods for "burn in.") Given the development characteristics in periods $0, -1, -2, \ldots, -4$ before the start of the sample, find a sequence of periods $t^b, t^b + 1, \ldots, t^b + 4$ in sample with development characteristics most similar to that of periods $-4, -3, \ldots, 0$.

   (a) For each candidate 5-period sequence, compute the means and variances of observable development characteristics.

   (b) Compute the mean squared distance between each candidate 5-period sequence and that of the "pre-period."

   (c) Select the sequence of BTO cycles that has the lowest mean squared distance.

4. Append the candidate sequence to the beginning of the sample. For each development j in the candidate sequence in pre-period t, draw an associated unobserved development quality $\tilde{\xi}_{jt} \sim \hat{F}_\xi$.

5. Repeat steps 1-2 with the augmented sample, holding the pre-period unobserved development qualities $\tilde{\xi}_{jt}$ fixed.

We augment the data because we see the development characteristics before our sample period, but we do not see the number of applicants with the correct priority level, i.e. "first-timers."

3.G Counterfactuals: Computational details

3.G.1 Rule changes: Computation

Our rule changes shut down intertemporal risk for BTO applicants. Given a guess of prices on the secondary market, each applicant submits a rank-ordered list over developments offered over all periods in which she is active, as well as exiting in each period. Then, apartments in developments are assigned to applicants via Random Serial Dictatorship (RSD). Finally, the algorithm searches over paths of secondary market prices that clear the market for existing apartments. Since the problem is non-differentiable, we use the Nelder-Mead algorithm to search for secondary market prices.

The computational procedure proceeds as follows:

1. Simulate $I \times K$ sets of household idiosyncratic shocks to the attractiveness of each development (a, t), waiting and exiting. For each simulation $k = 1, 2, \ldots, K$, let the first $1, 2, \ldots, I/5$ agents be poor and the remainder be rich. Also simulate K sets of priorities for each agent.

2. Guess a path of prices on the secondary market $\vec{p}_t^S$. Compute the quantity of secondary market apartments supplied in each period, $Q_t^S(\vec{p}_t^S)$.

3. In each simulation, for each agent (i, y, τ_0), compute the utility of each development and exiting; as well as the flow utility of waiting one period in periods $\tau_0, \tau_0+1, \ldots, \tau_0+\bar{\tau}-1$.

4. Given the computed utilities, for each agent, form preference lists over each development and exiting.

5. Given the preference lists and drawn priorities, run RSD to obtain the realized allocation of applicants to developments and exiting.

6. In each simulation k, given the realized number of exiters and secondary market supply, compute the excess supply of apartments on the secondary market.

7. If, in each period, the average excess supply of apartments on the secondary market across simulations is zero, stop and return the realized secondary market prices $\vec{p}_t^S$. Else update $\vec{p}_t^S$ and repeat steps 1-7 till convergence.

3.G.2 Building more: Computation

Suppose the government built a different number of apartments. Households would then react to the policy changes, resulting in different implied success rates of applying to each apartment type. At the margin, some of these households may elect to exit the mechanism, driving up the prices of apartments currently on the resale market. To credibly evaluate these counterfactual policies, we will have to compute the equilibrium consistent with household reoptimization. Conditional on parameters, counterfactual subsidies, as well as apartment quantities, computing an equilibrium amounts to searching for a fixed point in apartment success probabilities q and mean secondary market prices $\bar{p}^S$.

From an algorithmic standpoint, finding a fixed point is difficult, as it is a high-dimensional non-convex root-finding problem. Computation proceeds as follows:

1. Guess success probabilities $q^{(0)}$ and secondary market prices $p^{S(0)}$. Denote by $\omega^{(0)} = (q^{(0)}, p^{S(0)})$ the putative fixed point.

2. At the hypothesized secondary market prices, compute secondary apartment supply $Q_t^S(p^{S(0)})$ for all t.

3. Given $\omega^{(0)}$, iterate each agent's value functions till convergence, obtaining choice-specific values $v(\cdot; \omega^{(0)}, \cdot)$ and conditional choice probabilities $s(\cdot; \omega^{(0)}, \cdot)$.

4. Given the new choice-specific values and probabilities, forward simulate the model from the initial market size and type distribution, obtaining a path of hypothesized market sizes $M(\omega^{(0)})$ and type distributions $\mu(\omega^{(0)})$.

5. Given the CVFs, CCPs, market sizes and type distributions, compute the demand for each development $D_t(a;\omega^{(0)})$ and exiting $Q_t^D(\omega^{(0)})$.

6. Obtain the implied odds of success at each development $\hat{q}(\omega^{(0)})$ and excess demand $Q_t^D(\omega^{(0)}) - Q_t^S(p^{S(0)})$.

7. If

   (a) the implied odds of success are sufficiently close to the hypothesized ones, i.e. $||\hat{q}(\omega^{(0)}) - q^{(0)}|| < \epsilon$ for a pre-specified tolerance ϵ; and

   (b) excess demand is sufficiently close to zero, i.e. $||Q_t^D(\omega^{(0)}) - Q_t^S(p^{S(0)})|| < \epsilon$,

   then return the putative fixed point $\omega^{(0)}$. Else take a Newton step; obtain a new guess $\omega^{(1)}$; and repeat steps 2 to 7 till convergence.

Here we discuss equilibrium existence and uniqueness. We observe that success probabilities q lie in $[0,1]^{|\mathcal{A}|}$, a compact and convex set. Moreover, mean secondary market prices $\bar{p}^S$ can be constrained to lie in $[0, \bar{p}^{S,\max}]^T$, also a compact, convex set. Furthermore, the mapping from ω to implied success probabilities and excess demand is a continuous function, by a standard large-market assumption and interpolation procedure. By Brouwer's Fixed Point Theorem (Brouwer, 1911), this mapping has a fixed point ω^*.

However, there may be multiple fixed points consistent with model primitives.[61] Therefore, we select the equilibrium arising from starting our optimizer at the observed success

[61] We give a trivial example of multiplicities here. Suppose there are two apartments A and B, each giving utility zero, and the continuation value is also zero. If there are 3 households, application success probabilities $(0.5, 1)$ and $(1, 0.5)$ are two fixed points consistent with optimal choices.

probabilities and secondary market prices ω^{obs}.[62] In our robustness checks, when we start the optimizer at random feasible points, we obtain similar results for our counterfactual quantities of interest below.

3.H Channels driving our results

3.H.1 Under our current mechanism, could expanding supply ever "work?"

We find that, under the current BTO rules, expanding supply fails to substantially reduce wait times. We ascribe this failure to substitution: instead of "settling" for an apartment today, marginal agents apply for a competitive development tomorrow.

To "reverse" our results, we need to eliminate this substitution effect. Focus on a stationary environment with ex ante identical developments, only one income type and no exit. Agents have perfect foresight and discount the future by factor $\delta = 0.96$. Each period, 1 development is offered, with associated utility $u(a) = 0$. Suppose houses are always short, so success probabilities are $q_t(a) = 0.8$ for all t. Then the agent's choice-specific value functions over applying and waiting are

$$v_{it}(a) = 0.8 \times u(a) + 0.2 \times \delta V + \epsilon_{it}(a); \quad v_{it}(w) = \delta V + \epsilon_{it}(w).$$

Assume that the shocks $\epsilon_{it}(\cdot)$ are distributed iid Type 1 Extreme-Valued with location $-\gamma^E \approx -0.577$. Then, the agent's (stationary) inclusive value satisfies the recursive relation

$$V = \ln\left[\exp(0.2\delta V) + \exp(\delta V)\right],$$

[62] To improve convergence, we use a derivative-based Interior Point Optimizer (`Ipopt`, Wächter and Biegler (2006)) with the `ma97` linear solver (Hsl, 2021).

or $V \approx 2.7844$. For all t, the market share of the development is then

$$s_t(a) = \frac{\exp(0.2\delta V)}{\exp(0.2\delta V) + \exp(\delta V)} \approx 0.1084.$$

Suppose that tomorrow, i.e. in time $t+1$, an unanticipated (MIT) supply expansion occurs, so $q'_{t+1}(a) = 1$. However, supply remains the same in all other periods: $q'_\tau(a) = 0.8$ for all $\tau \neq t+1$. Contrary to our findings, we claim that more people apply to **today's** development: $s'_t(a) > s_t(a)$. To see this result, we compute the new choice-specific value functions tomorrow:

$$v'_{it+1}(a) = 1 \times u(a) + 0 \times \delta V + \epsilon'_{it+1}(a) = \epsilon'_{it+1}(a); \quad v'_{it+1}(w) = \delta V + \epsilon'_{it+1}(w).$$

Abusing notation, tomorrow's inclusive value can be computed as

$$V' = \ln[1 + \exp(\delta V)] \approx 2.7398,$$

so the market share of today's development becomes

$$s'_t(a) = \frac{\exp(0.2\delta V')}{\exp(0.2\delta V') + \exp(\delta V')} \approx 0.1087 > 0.1084 = s_t(a).$$

Thus, if developments do not vary much in quality over time, we "reverse" our supply expansion result.

3.I Tables and Figures

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Period	195			2012-01-12	2015-05-27
Rooms per apartment	195	3.90769	0.774392	3	5
Quantity	195	307.616	197.637	44.1	1176.4
Applications	195	563.666	578.08	30.305	3034.97
Sticker price	195	3.09324	1.02415	1.505	6.505
Forward resale price	195	3.9447	1.02574	2.07182	7.59
Distance to downtown	195	13.6718	3.69131	4.01692	19.36
Distance to metro stop	195	0.989254	0.396679	0.0616367	1.97208
Distance to SAP school	195	4.31581	2.76348	0.293455	10.397

Table 3.I.1: Summary statistics by development in the BTO mechanism between January 2012 and May 2015.
Notes: Quantities and applications are scaled relative to the supply reserved for and the inferred number of first-timer applicants. All prices are expressed in hundreds of thousands of nominal Singapore dollars. Distances to downtown, the nearest metro stop and the nearest Special Assistance Plan (SAP) school are expressed in kilometers.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
BTO cycle	5239	27.4951	15.5898	1	54
Year	5239	2014.41	3.18921	2010	2021
Apartment type	5239	4.4539	1.10335	3	6
Price	5239	5.17865	1.66824	2.2	10.6015
Quantity transacted	5239	42.0053	46.0015	0.0	403.0
Housing stock	5239	8580.97	7552.4	0	36934

Table 3.I.2: Summary statistics of secondary market transactions by (neighborhood, apartment type, BTO cycle) from August 2010 to February 2021.
Notes: Apartment types in $\{3, 4, 5, 6\}$ correspond to 3-room, 4-room, 5-room and Executive/Multi-Generational housing respectively. All prices are expressed in hundreds of thousands of nominal Singapore dollars.

Number of rooms in apartment	2	3	4	5
Approx. floor area (m^2)	40	70	90	110

Table 3.I.3: Government apartments in Singapore by number of rooms and approximate size.

	(1) **LHS (O)**	(2) **LHS (O)**	(3) **LHS (O)**	(4) **LHS (O)**
Sale price (S$100k)	0.011 (0.035)	0.012 (0.035)	0.710** (0.309)	0.712** (0.310)
Estimator	OLS	OLS	IV	IV
Neighborhood FE Size FE	N N	Y Y	N N	Y Y
N First-stage F	4,843 -	4,843 -	4,843 11.99	4,843 11.96

Table 3.I.4: Table of results for the secondary market regressions.
Notes: Standard errors are heteroskedasticity-robust. The first two specifications are ordinary least squares (OLS) regressions of adjusted mover shares on prices, while the last two are IV regressions, instrumenting for endogenous prices with BTO eligibility changes in 2015 and 2019, as well as a shock to the maximum loan-to-value ratio in 2018. The second and fourth regressions include fixed effects for each neighborhood and apartment type.

Parameters	**Estimate**	**S.E.**
Constant	-2.698***	1.050
Distance to downtown	-0.143***	0.0470
Distance to metro stop	-0.729***	0.163
Distance to SAP school	0.0291	0.0389
4-room flat	1.987***	0.305
5-room flat	1.426***	0.499
Price	-0.855***	0.291
Resale price	0.569**	0.232
Subsidized development price	-3.433**	1.373
Wait cost (Poor)	-1.972***	0.346
Wait cost (Rich)	-2.611***	0.202
N	195	

Table 3.I.5: Table of results from estimating Equation (3.12).
Notes: "Subsidized development price" is the price of an apartment in the development, net of explicit subsidies to the poor. Standard errors are from the GMM procedure. Asterisks indicate significance at various levels (*: 10%; **: 5%; ***: 1%)

Amenity/Value (S$)	Poor	Rich
Living 100m closer to...		
Downtown	400**	1,600***
	(200)	(400)
Subway Stop	2,100**	8,500***
	(1,000)	(2,500)
Wait cost (per person per month)	9,600*	50,900***
	(5,400)	(18,400)

Table 3.I.6: Table of willingness to pay (WTP) for several amenities, in equivalent discounts on an apartment in the BTO mechanism.
Notes: For each income level, WTP is computed by dividing the estimated coefficient on the corresponding amenity by the estimated price coefficient on inside apartments. Standard errors in parentheses are computed by applying the Delta method to the variance-covariance matrix from the estimation of Equation (3.12). Asterisks indicate significance at various levels (*: 10%; **: 5%; ***: 1%)

Targeted	Target	Value (S.D.)	Untargeted	Target	Value (S.D.)
$\mathbb{E}[\xi_{at}]$	0.0	-0.03 (1.44)	$\mathbb{E}[\xi_{at} \cdot t]$	0.0	-0.81 (20.54)
$\mathbb{E}[\xi_{at} \cdot d_{\text{MBFC},at}]$	0.0	-0.82 (16.57)	$\mathbb{E}[\xi_{at} \cdot d_{\text{school},at}]$	0.0	-0.03 (0.94)
$\mathbb{E}[\xi_{at} \cdot d_{\text{MRT},at}]$	0.0	-0.09 (1.20)	$\mathbb{E}[\mathbf{1}\{\#\text{ apps} \leq 3\}]$	$> 50\%$	90%

Table 3.I.7: Table of targeted and untargeted moments in the GMM estimation of young household demand.
Notes: Distances to downtown, the nearest metro stop, and the nearest school are in kilometers. For each moment, where applicable, the standard deviation over all observations is reported in parentheses beside the value of said moment. The percentage of agents with 3 or fewer applications is computed by simulating applicant behavior in our baseline counterfactual.

Policy	Intervention
Change rules	Allocate by random serial dictatorship
Add supply	Build 10% more in oversubscribed developments
Combined	Change rules + Add supply

Table 3.I.8: Summary of counterfactual policies.

Outcome	Base	Build 10% more	Change rules	Combined
Vacancies	12%	16%	7%	10%
Wait times	Poor: 1.5 years Rich: 10 months	↑0.4% ↓3%	↓8% ↓23%	↓8% ↓23%
2° market prices	S$460k	↓0.02%	↑12%	↑9%

Table 3.I.9: Table of key outcome comparisons between the actual mechanism and three alternate policies in the sample period of January 2012 and May 2015.
Notes: The three alternative policies are: when 10% more apartments are built in each oversubscribed development during the sample period; when the mechanism is changed to a dynamic variant of Random Serial Dictatorship; and when the previous two policies are simultaneously implemented. Secondary market prices are means over the sample period. For the actual mechanism, vacancies and secondary market prices are computed from the data. Baseline wait times and all other quantities are computed from counterfactual simulations.

Neighborhood	Flat type	Units	Apps	Application rate
Non-mature				
Punggol	3-room	112	367	3.3
Punggol	4-room	519	1001	1.9
Punggol	5-room	225	658	2.9
Mature				
Clementi	4-room	229	1995	8.7
Clementi	5-room	156	2173	13.9

Table 3.I.10: Application rates (selected) for the May 2015 BTO cycle.
Notes: These rates are computed by dividing the number of applications received by the number of units on offer.

	(1) **rate**	(2) **rate**
mbfc	0.156***	0.063
	(0.029)	(0.097)
mrt	-1.057***	-0.651***
	(0.073)	(0.220)
hawker	-0.803***	-0.722***
	(0.182)	(0.196)
super	-1.406	-1.645**
	(1.894)	(0.851)
sch	-0.520	-0.820**
	(0.311)	(0.435)
sap	-0.292***	
	(0.032)	
mature	1.278***	0.773**
	(0.189)	(0.395)
sap ¡= 2		1.932***
		(0.141)
size FE	Y	Y
time FE	Y	Y
N	233	233
R2	0.399	0.353

Table 3.I.11: Regression results for hedonic (1) and boundary discontinuity (2) regressions for amenity access.

Notes: We regress the number of BTO applications per apartment in each development, "rate", on amenity covariates. Distances to the Marina Bay Financial Centre, the nearest metro stop, the nearest hawker center, the nearest supermarket, the nearest non-SAP school and the nearest SAP school are expressed in kilometers. Standard errors are estimated using the method of Spatial Correlation Principal Components (Müller and Watson, 2021). Stars indicate statistical significance. **: 5%; ***: 1%.

Neighborhood	Flat type	Units	Apps	Application rate
Non-mature				
Punggol	3-room (Income ceiling $6,000)	25	195	7.8
	3-room (Income ceiling $12,000)	32	70	2.2
	4-room	238	618	2.6
	5-room / 3Gen / Executive	106	354	3.3
Sembawang	3-room (Income ceiling $6,000)	39	210	5.4
	4-room	49	383	7.8
	5-room / Executive	23	169	7.3

Table 3.I.12: (Selected) clearinghouse application rates for the May 2018 Sale of Balance Flats cycle.
Notes: These rates are computed by dividing the number of applications received by the number of units on offer.

As of...	Ceiling (3-room and up)
08/2011	$120,000
08/2015	$144,000
09/2019	$168,000

Table 3.I.13: Household income ceilings over time in nominal Singapore dollars per year.
Notes: Households are typically married couples applying to live in an apartment together. Their combined income must be lower than the specified amount to apply for a BTO apartment.

# of Rooms	2	3	4	5
H/h income ceiling (p.a.)	$84,000	$168,000	$168,000	$168,000

Table 3.I.14: Household income ceilings by apartment size in nominal Singapore dollars per year, as of September 2019.
Notes: A few three-room apartments in non-mature estates may have an income ceiling of S$84,000.

Town Type/ Number of Rooms	2	3	4	5
11/2011-05/2015				
Non-mature	$60,000	$50,000	$25,000	$10,000
Mature	NA	$30,000	$15,000	$10,000
08/2015-05/2019				
Non-mature	$80,000	$75,000	$60,000	$5,000
Mature	$40,000	$20,000	$5,000	$0
09/2019-present				
Non-mature	$80,000	$75,000	$60,000	$45,000
Mature	$80,000	$60,000	$45,000	$30,000

Table 3.I.15: Maximal household explicit subsidies by apartment size and town type over time.
Notes: Subsidies are denominated in nominal Singapore dollars per year. We did not specify a maximal subsidy for 2-room flats between November 2011 to August 2015 because the precursor to 2-room "flexi" flats starting November 2015 is not comparable with flats of the same size during the earlier time period.

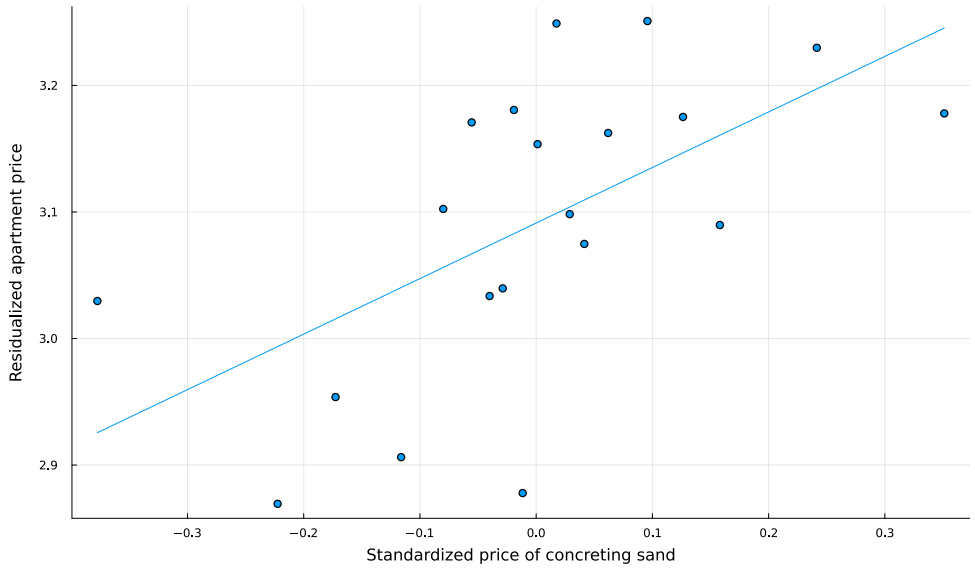


Figure 3.I.1: Binned scatter plot of BTO prices against prices of concreting sand, residualized on controls including distances to downtown and to the nearest metro stop.

	08/11-05/15		08/15-05/19		09/19-
Income (Mth)	Full	Restr.	Full	Restr.	
$\leq$ \$1.5k	\$60k	\$40k	\$80k	\$40k	\$80k
\$1.5k–\$2k	\$50k+	\$35k+	\$75k	\$35k	\$75k
\$2k–\$2.5k	\$30k+	\$30k+	\$70k	\$30k	\$70k
\$2.5k–\$3k	\$25k	\$25k	\$65k	\$25k	\$65k
\$3k–\$3.5k	\$20k	\$20k	\$60k	\$20k	\$60k
\$3.5k–\$4k	\$15k	\$15k	\$55k	\$15k	\$55k
\$4k–\$4.5k	\$10k	\$10k	\$50k	\$10k	\$50k
\$4.5k–\$5k	\$5k	\$5k	\$45k	\$5k	\$45k
\$5k–\$5.5k	\$0	\$0	\$35k	\$0	\$40k
\$5.5k–\$6k	\$0	\$0	\$30k	\$0	\$35k
\$6k–\$6.5k	\$0	\$0	\$25k	\$0	\$30k
\$6.5k–\$7k	\$0	\$0	\$20k	\$0	\$25k
\$7k–\$7.5k	\$0	\$0	\$15k	\$0	\$20k
\$7.5k–\$8k	\$0	\$0	\$10k	\$0	\$15k
\$8k–\$8.5k	\$0	\$0	\$5k	\$0	\$10k
\$8.5k–\$9k	\$0	\$0	\$0	\$0	\$5k
\$9k–\$10k	\$0	\$0	\$0	\$0	\$0
\$10k–\$12k	ineligible	ineligible	\$0	\$0	\$0
\$12k–\$14k	ineligible	ineligible	ineligible	ineligible	\$0

Table 3.I.16: Explicit subsidies by monthly income, apartment size and town type over time. Notes: Subsidies are denominated in nominal Singapore dollars. Full subsidies are applied to 2- and 3-room flats in non-mature estates in all time periods, as well as 4-room flats in non-mature estates starting November 2015. Restricted subsidies are applicable to all other apartments. "Ineligible" means that households of that income level are not permitted to apply for any BTO apartment at all.

	(1) price	(2) price	(3) price
Standardized price of concreting sand	0.163** (0.069)	0.140** (0.070)	0.293** (0.132)
Interacted with apartment size? Other controls?	N N	Y N	Y Y
N R2 F	195 0.025 5.533	195 0.525 55.761	195 0.880 146.084

Table 3.I.17: Table of first stage regressions of BTO price on the standardized price of concreting sand.
Notes: The second specification includes interactions of the price of concreting sand with the number of rooms of each apartment in the BTO development. The third specification further adds, as controls, distances to downtown and the nearest metro stop, interacted with apartment size and the instrument. We interact said distances with the price of concreting sand because developments closer to downtown and to metro stops are likely located on smaller land parcels, necessitating less concreting sand to build each development. Stars represent statistical significance at various levels. *: 10%; **: 5%; ***: 1%.

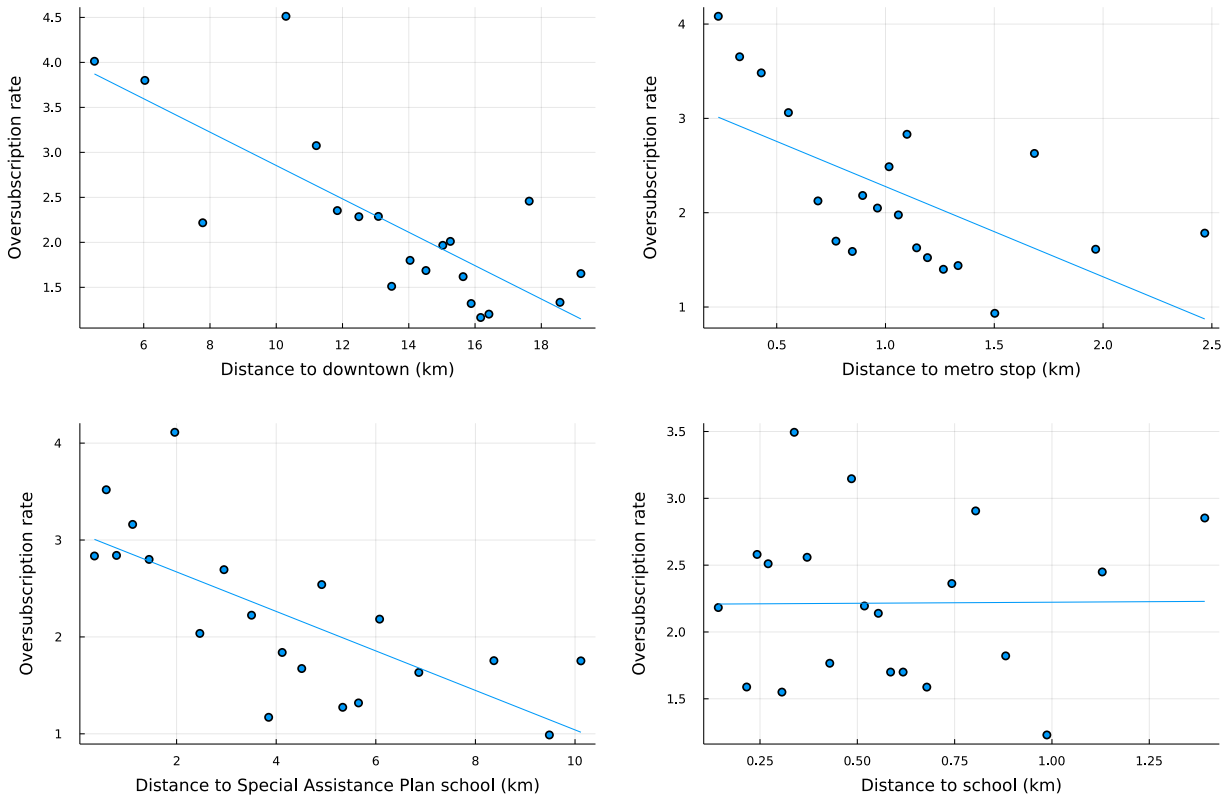


Figure 3.I.2: Binned scatter plots of oversubscription rates by development against the distance of said development to various amenities.
Notes: The oversubscription rate of a development is the ratio of its number of applicants to the quantity supplied. The plotted developments span the January 2012 to May 2019 BTO cycles.

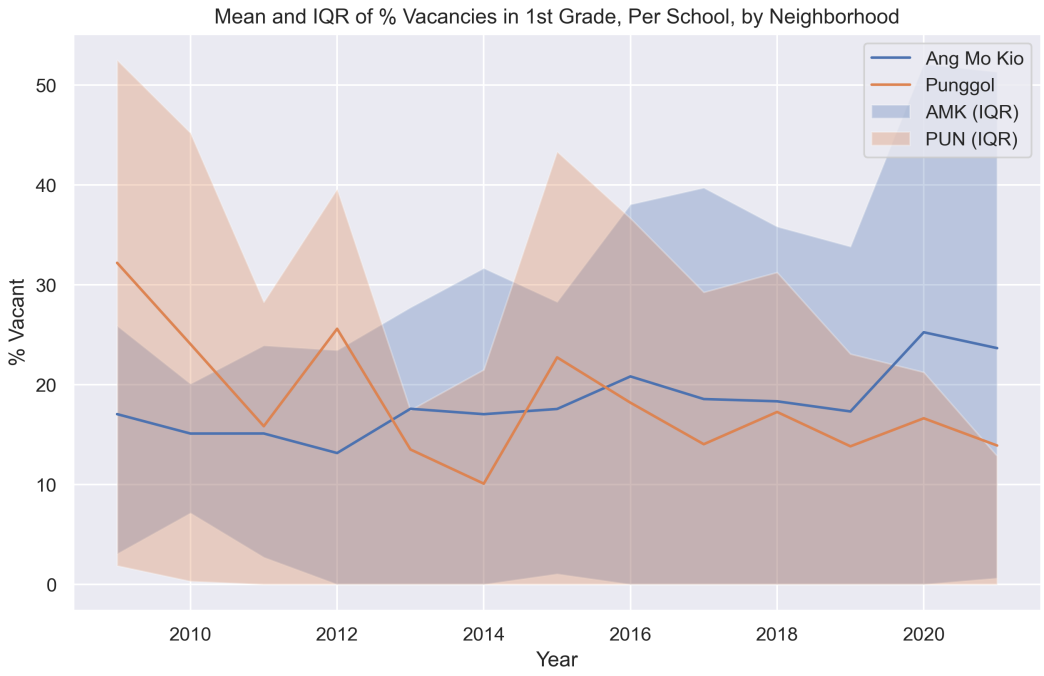


Figure 3.I.3: Means and interquartile ranges of percentage vacancies per school by neighborhood by year.
Notes: Percentage vacancies in a year correspond to the number of vacant places after Phase 2C of the Primary One Registration Exercise in that year. Ang Mo Kio is a representative mature estate (without much new BTO apartment construction) and Punggol is a representative non-mature estate (with planned future BTO construction).

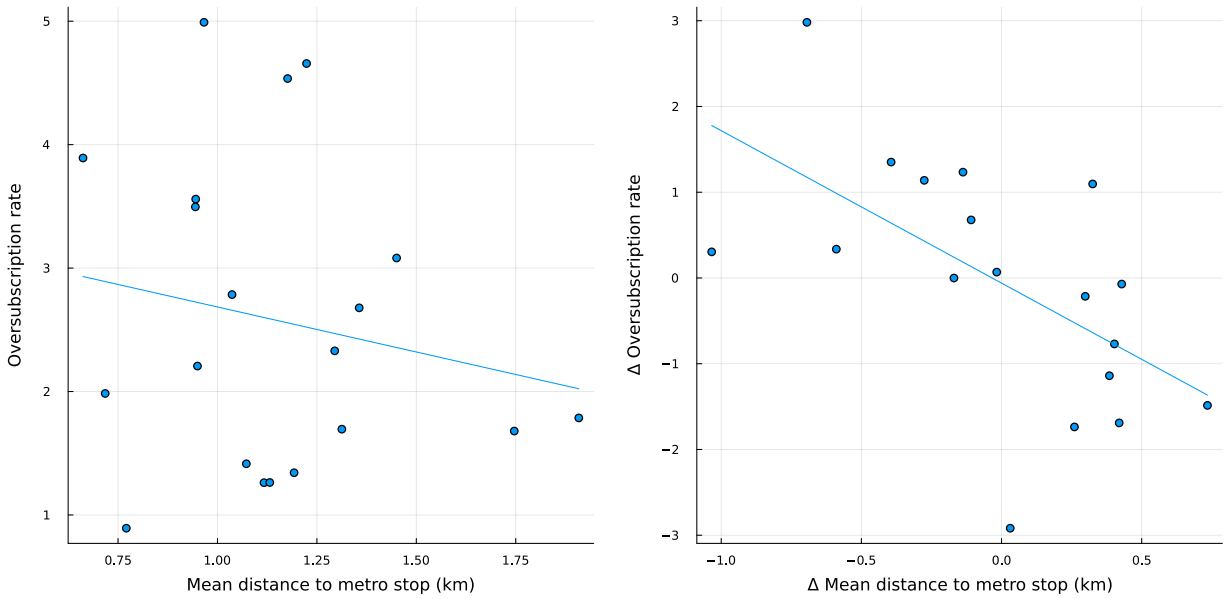


Figure 3.I.4: Binned scatter plots of oversubscription rates in a BTO cycle against the mean distance to the nearest metro stop of all developments offered in that cycle.
Notes: The left panel displays the plot in levels; the right displays first differences in oversubscription rates and mean distances over BTO cycles $(t, t+1)$. The plotted BTO cycles span January 2012 to May 2019.

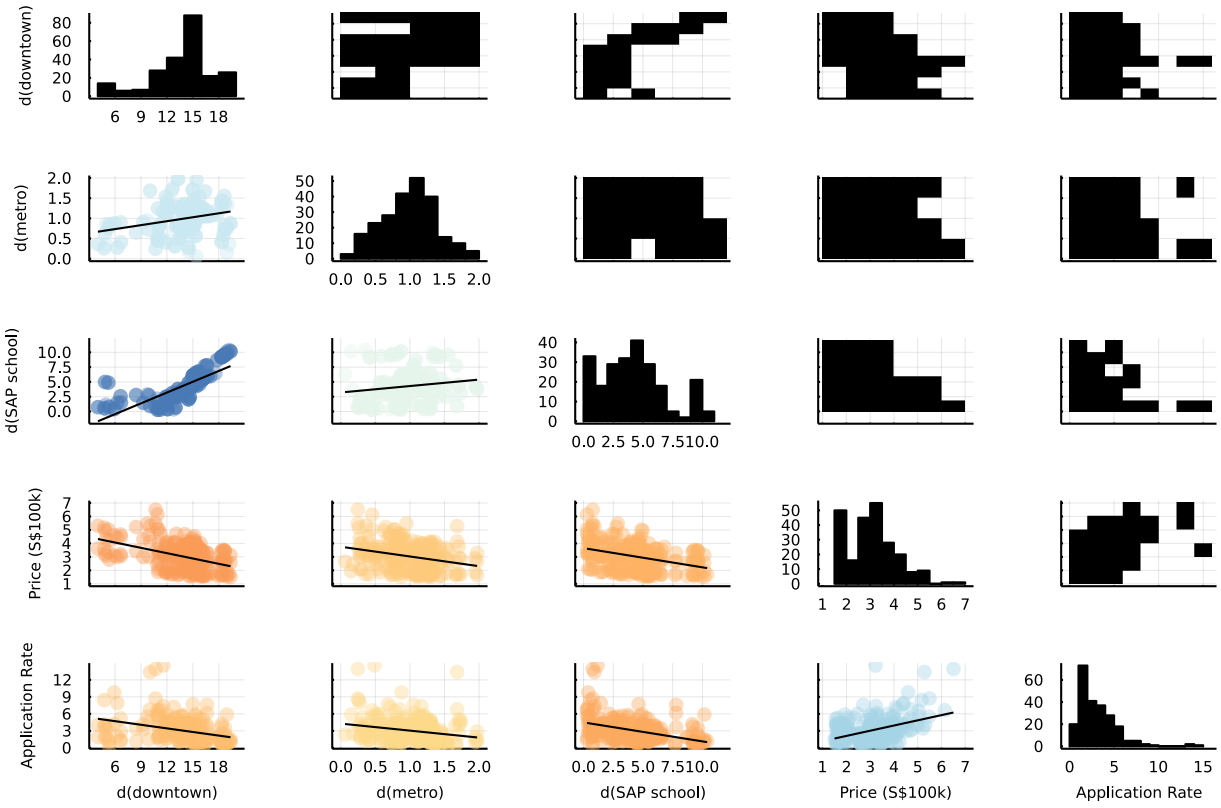


Figure 3.I.5: Correlation plot of selected variables against application rates (# applications / # quantity).
Notes: Distances to downtown, the nearest metro stop and the nearest Special Assistance Plan (SAP) school are in kilometers. Redder (bluer) hues in each scatter plot indicate a stronger negative (resp. positive) correlation between the two plotted variables.

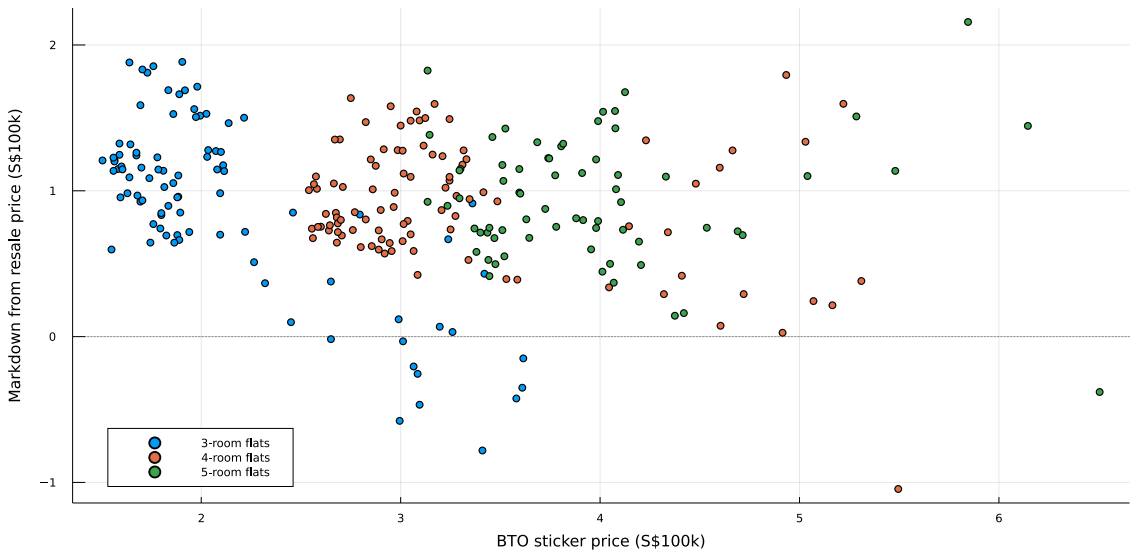


Figure 3.I.6: Implicit discount, by apartment type, of the BTO sticker price relative to the five-year-ahead resale price of similar apartments in the same neighborhood.
Notes: Several of the smaller, cheaper apartments in mature estates actually see a negative markdown. Therefore, households opting for these units would pay for their apartment than they would have gotten had they liquidated their property immediately after living in it for 5 years.

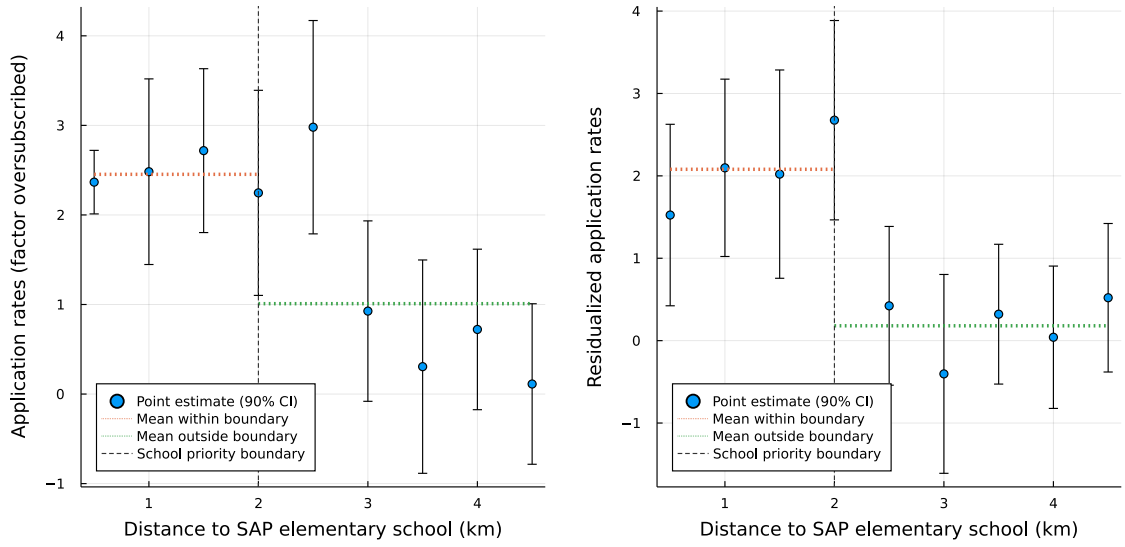


Figure 3.I.7: BTO application rates by distance to SAP elementary school.
Notes: In our exercise, we examine the school priority boundary at the 2-kilometer mark.





BUILD-TO-ORDER

Choose your home from 3 new Build-To-Order (BTO) projects that offer a total of 3,095 units of 2-room Flexi, 3-, 4-, 5-room, and 3Gen flats:

- **Canberra Vista** in Sembawang
- **Kim Keat Ripples** and **Toa Payoh Ridge** in Toa Payoh

Apply for your choice town and flat type by 17 February 2020, Monday, 11:59pm.

Apply Online Now

Information on the BTO projects is available on the HDB InfoWEB. You can apply online using your mobile phone, tablet, or computer without visiting the HDB Hub. If you are feeling unwell or on Leave of Absence Order, please refrain from visiting HDB offices and Branches. You can visit Ministry of Health's website for the latest updates. Successful applicants will be determined by a computer ballot, and not on a first-come, first-served basis.

February 2020 BTO Prices

Town	Project	Flat Type	Selling Price (Excluding Grants)	Selling Price (Including Grants ^)
Non-Mature Town				
Sembawang	Canberra Vista	2-room Flexi*@	From $89,000	From $9,000
		3-room	From $177,000	From $102,000
		4-room	From $272,000	From $212,000
		5-room	From $350,000	From $305,000
		3Gen	From $355,000	From $310,000
Mature Town				
Toa Payoh	Kim Keat Ripples	2-room Flexi*^^ (40-year Lease)	From $90,000	From $10,000
	Toa Payoh Ridge	3-room	From $351,000	From $291,000
		4-room	From $395,000	From $350,000

Notes:

* 2-room Flexi flats come in two sizes - 36 sqm (Type 1) and 45/46 sqm (Type 2).

^ The starting prices of flats in the table above are based on 99-year leases. For illustration purposes, the assumed Enhanced CPF Housing Grant (EHG) amounts are:

i) 2-room Flexi flat: $80,000

ii) 3-room flat: $75,000 in non-mature towns and $60,000 in mature towns

iii) 4-room flat: $60,000 in non-mature towns and $45,000 in mature towns



Block Number	Number of Storeys	2 Room Flexi Type 1	2 Room Flexi Type 2	4 Room	Total	Lift opens at
268A	16/25/31	–	–	236	236	Every Storey
269A	16/25/31	–	–	233	233	Every Storey
270A	31	30	120	89	239	Every Storey
Total		30	120	558	708	

Figure 3.I.8: Information available to BTO applicants.

Notes: Clockwise from top left: Intro page for February 2020 BTO sales launch; Prices for that sales launch; Site plan for Kim Keat Ripples, a development in the February 2020 sales launch; Locality map for Kim Keat Ripples.

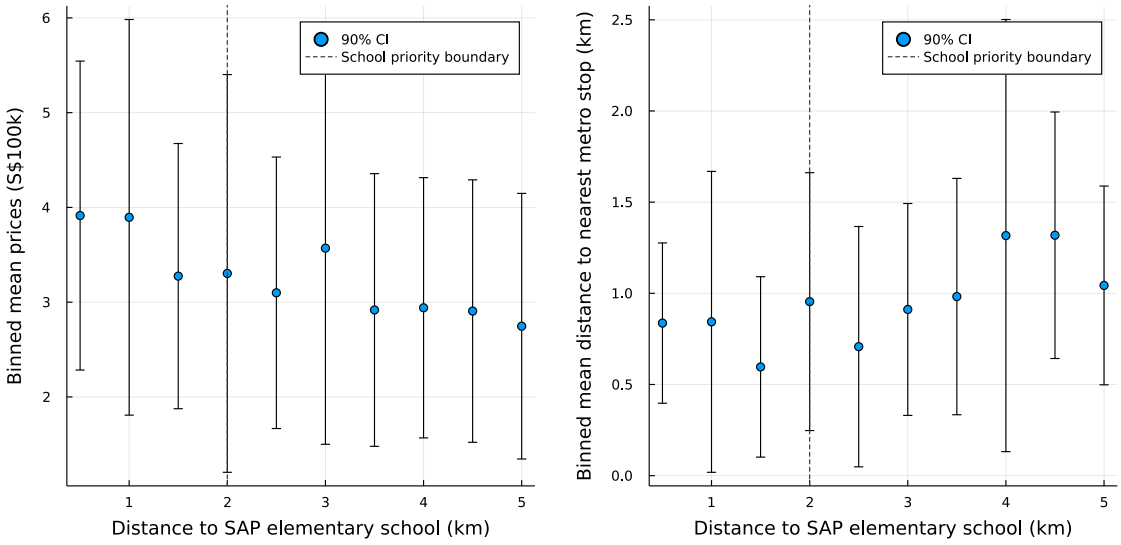


Figure 3.I.9: Prices and distance to metro stop seem evenly distributed across both sides of the 2-kilometer school priority boundary.

Bibliography

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School choice: A mechanism design approach," *American Economic Review*, 2003, *93* (3), 729–747.

_ **and** _ , "Matching markets: Theory and practice," *Advances in Economics and Econometrics*, 2013, *1*, 3–47.

_ **, Nikhil Agarwal, and Parag A Pathak**, "The welfare effects of coordinated assignment: Evidence from the New York City high school match," *American Economic Review*, 2017, *107* (12), 3635–89.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis using Strategic Reports: An application to a school choice mechanism," *Econometrica*, 2018, *86* (2), 391–444.

_ **, Itai Ashlagi, Michael A Rees, Paulo J Somaini, and Daniel C Waldinger**, *An empirical framework for sequential assignment: The allocation of deceased donor kidneys*, National Bureau of Economic Research, 2019.

_ , _ , _ , **Paulo Somaini, and Daniel Waldinger**, "Equilibrium allocations under alternative waitlist designs: Evidence from deceased donor kidneys," *Econometrica*, 2021, *89* (1), 37–76.

Agarwal, Sumit, Yi Fan, Wenlan Qian, and Tien Foo Sing, "Tying the Knot in a New Home: Consumption Responses to a Pro-Marriage Housing Policy," *Available at SSRN 3975080*, 2021.

Aguirregabiria, Victor and Aviv Nevo, "Recent developments in empirical IO: Dynamic demand and dynamic games," *Advances in economics and econometrics*, 2013, *3*, 53–122.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan, "Thickness and information in dynamic matching markets," *Journal of Political Economy*, 2020, *128* (3), 783–815.

Almagro, Milena and Tomás Domínguez-Iino, "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Mimeo*, 2019.

Altmann, Sam M, "Choice, Welfare, and Market Design," *mimeo*, 2023.

Anderson, Axel and Lones Smith, "Dynamic matching and evolving reputations," *The Review of Economic Studies*, 2010, *77* (1), 3–29.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro, "Transparency in structural research," *Journal of Business & Economic Statistics*, 2020, *38* (4), 711–722.

Arcidiacono, Peter and Robert A Miller, "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 2011, *79* (6), 1823–1867.

Arnosti, Nick and Peng Shi, "How (Not) to Allocate Affordable Housing," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 204–08.

_ **and** _ , "Design of Lotteries and Wait-Lists for Affordable Housing Allocation," *Management Science*, 2020.

Ashlagi, Itai, Maximilien Burq, Patrick Jaillet, and Amin Saberi, "Maximizing efficiency in dynamic matching markets," *arXiv preprint arXiv:1803.01285*, 2018.

Azevedo, Eduardo M and Jacob D Leshno, "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 2016, *124* (5), 1235–1268.

Baccara, Mariagiovanna and Leeat Yariv, "Dynamic matching," *Online and Matching-Based Market Design*, 2021.

_ **, SangMok Lee, and Leeat Yariv**, "Optimal dynamic matching," *Theoretical Economics*, 2020, *15* (3), 1221–1278.

Baum-Snow, Nathaniel and Lu Han, "The microgeography of housing supply," Working Paper 2021.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

_ **, Robert McMillan, Alvin Murphy, and Christopher Timmins**, "A dynamic model of demand for houses and neighborhoods," *Econometrica*, 2016, *84* (3), 893–942.

Becker, Dan, "Non-wage job characteristics and the case of the missing margin," *Available at SSRN 1805761*, 2011.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak, "Measuring utility by a single-response sequential method," *Behavioral science*, 1964, *9* (3), 226–232.

Bergemann, Dirk and Juuso Välimäki, "Learning and strategic pricing," *Econometrica: Journal of the Econometric Society*, 1996, pp. 1125–1149.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

_ **and Philip A Haile**, "Identification in differentiated products markets using market level data," *Econometrica*, 2014, *82* (5), 1749–1797.

Brouwer, Luitzen Egbertus Jan, "Über abbildung von mannigfaltigkeiten," *Mathematische annalen*, 1911, *71* (1), 97–115.

Budish, Eric and Estelle Cantillon, "The multi-unit assignment problem: Theory and evidence from course allocation at Harvard," *American Economic Review*, 2012, *102* (5), 2237–71.

Bulow, Jeremy and Paul Klemperer, "Auctions versus negotiations," *The American Economic Review*, 1996, *86* (1), 180–94.

Cason, Timothy N and Charles R Plott, "Misconceptions and game form recognition: Challenges to theories of revealed preference and framing," *Journal of Political Economy*, 2014, *122* (6), 1235–1270.

Chade, Hector, Jan Eeckhout, and Lones Smith, "Sorting through search and matching models in economics," *Journal of Economic Literature*, 2017, *55* (2), 493–544.

Chen, Yi-Chun and Gaoji Hu, "Learning by matching," *Theoretical Economics*, 2020, *15* (1), 29–56.

Chin, Daryl and Rachel Chang, "HDB will be the price-setter: Khaw Boon Wan," *The Straits Times*, 4 2013.

Choo, Daryl, "The big read: Affluent and craving space to grow, more single millennials leave the nest for greater freedom," *Today*, 9 2021.

Christensen, Bent Jesper, Rasmus Lentz, Dale T Mortensen, George R Neumann, and Axel Werwatz, "On-the-job search and the wage distribution," *Journal of Labor Economics*, 2005, *23* (1), 31–58.

Chyn, Eric, "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, 2018, *108* (10), 3028–56.

Combe, Julien, Olivier Tercieux, and Camille Terrier, "The design of teacher assignment: Theory and evidence," *Unpublished paper, University College London.[1310]*, 2018.

Currie, Janet and Firouz Gahvari, "Transfers in cash and in-kind: Theory meets the data," *Journal of Economic Literature*, 2008, *46* (2), 333–83.

Dagsvik, John, Boyan Jovanovic, and Andrea Shepard, "A foundation for three popular assumptions in job-matching models," *Journal of Labor Economics*, 1985, *3* (4), 403–420.

Diao, Mi, Delon Leonard, and Tien Foo Sing, "Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values," *Regional Science and Urban Economics*, 2017, *67*, 64–77.

Dijk, Winnie Van, "The socio-economic consequences of housing assistance," *Mimeo*, 2019.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer price search and platform design in internet commerce," *American Economic Review*, 2018, *108* (7), 1820–59.

Doraszelski, Ulrich and Ariel Pakes, "A framework for applied dynamic analysis in IO," *Handbook of industrial organization*, 2007, *3*, 1887–1966.

Doval, Laura, "Dynamically stable matching," *Theoretical Economics*, 2022, *17* (2), 687–724.

Elangovan, Navene, "Hot HDB resale market, higher building costs pushing top range BTO prices closer to million-dollar mark: Analysts," *Today Newspaper*, 9 2022.

Erdem, Tülin, Susumu Imai, and Michael P Keane, "Brand and quantity choice dynamics under price uncertainty," *Quantitative Marketing and economics*, 2003, *1* (1), 5–64.

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond truth-telling: Preference estimation with centralized school choice and college admissions," *American Economic Review*, 2019, *109* (4), 1486–1529.

Ferdowsian, Andrew, Kwok Hao Lee, and Luther Yap, "Build to order: Endogenous supply in centralized mechanisms," Working Paper 2023.

_ , Muriel Niederle, and Leeat Yariv, "Decentralized Matching with Aligned Preferences," Technical Report, mimeo 2022.

Fernandez, Marcelo Ariel, Kirill Rudov, and Leeat Yariv, "Centralized matching with incomplete information," *American Economic Review: Insights*, 2022, *4* (1), 18–33.

Galiani, Sebastian, Alvin Murphy, and Juan Pantano, "Estimating neighborhood choice models: Lessons from a housing assistance experiment," *American Economic Review*, 2015, *105* (11), 3385–3415.

Galichon, Alfred and Yu-Wei Hsieh, "Aggregate stable matching with money burning," *Available at SSRN 2887732*, 2018.

Gandhi, Amit and Jean-François Houde, "Measuring substitution patterns in differentiated products industries," Technical Report, National Bureau of Economic Research 2021.

Geyer, Judy and Holger Sieg, "Estimating a model of excess demand for public housing," *Quantitative Economics*, 2013, *4* (3), 483–513.

Gillingham, Kenneth, Fedor Iskhakov, Anders Munk-Nielsen, John Rust, and Bertel Schjerning, "Equilibrium trade in automobiles," *Journal of Political Economy*, 2022.

Gittins, John C, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1979, *41* (2), 148–164.

Gittins, John, Kevin Glazebrook, and Richard Weber, *Multi-armed bandit allocation indices*, John Wiley & Sons, 2011.

Gorry, Aspen, "Experience and worker flows," *Quantitative Economics*, 2016, *7* (1), 225–255.

Gowrisankaran, Gautam and Marc Rysman, "Dynamics of consumer demand for new durable goods," *Journal of Political Economy*, 2012, *120* (6), 1173–1219.

_ **and** _ , "A Framework for Empirical Models of Dynamic Demand," *mimeo*, 2020.

Greene, William H and David A Hensher, "A latent class model for discrete choice analysis: contrasts with mixed logit," *Transportation Research Part B: Methodological*, 2003, *37* (8), 681–698.

Gul, Faruk and Ennio Stacchetti, "The English auction with differentiated commodities," *Journal of Economic theory*, 2000, *92* (1), 66–95.

Guo, Yingni and Johannes Hörner, "Dynamic allocation without money," *TSE Working Paper*, 2020.

Hall, Robert E and Alan B Krueger, "Evidence on the incidence of wage posting, wage bargaining, and on-the-job search," *American Economic Journal: Macroeconomics*, 2012, *4* (4), 56–67.

Hansen, Lars Peter and Kenneth J Singleton, "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1269–1286.

Hastings, Justine, Thomas J Kane, and Douglas O Staiger, "Heterogeneous preferences and the efficacy of public school choice," *NBER Working Paper*, 2009, *2145*, 1–46.

Hendel, Igal and Aviv Nevo, "Measuring the implications of sales and consumer inventory behavior," *Econometrica*, 2006, *74* (6), 1637–1673.

Hong Kong Housing Authority, "Hong Kong 2017 Census," 2 2017.

Housing and Development Board, *HDB Annual Report* 2014.

_ , *HDB Annual Report* 2019.

_ , "HDB History and Towns," 8 2021.

_ , *HDB Sample Household Survey 2018*, Housing and Development Board, 2021.

_ , "Rental statistics," 2022.

Hsl, "Hsl. A collection of fortran codes for large scale scientific computation.," 2021.

Huat, Chua Beng, *Liberalism disavowed: Communitarianism and state capitalism in Singapore*, NUS Press, 2017.

Immorlica, Nicole, Jacob Leshno, Irene Lo, and Brendan Lucier, "Information acquisition in matching markets: The role of price discovery," *Available at SSRN 3705049*, 2020.

Jovanovic, Boyan, "Job matching and the theory of turnover," *Journal of political economy*, 1979, *87* (5, Part 1), 972–990.

Kadam, Sangram V and Maciej H Kotowski, "Multiperiod matching," *International Economic Review*, 2018, *59* (4), 1927–1947.

Kelso, Alexander S and Vincent P Crawford, "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1483–1504.

Lee, Desmond, "Written Answer by Ministry of National Development on average waiting time for mature estate BTO flat applications for first- and second-timers, as well as other applicants who have failed balloting more than twice," 7 2021.

_ , "Written Answer by Ministry of National Development on current Build-To-Order (BTO) flat demand-supply situation, plans to meet the strong demand for public housing, and relevance of BTO system in addressing the housing needs," *Ministry of National Development*, 1 2022.

Lee, Kwok Hao and Brandon Joel Tan, "Urban transit infrastructure and inequality," *mimeo*, 2023.

_ **and Leon Musolff**, "Entry into markets shaped by recommender systems," Working Paper 2021.

_ **, Andrew Ferdowsian, and Luther Yap**, "The dynamic allocation of public housing: Policy and spillovers," *Mimeo*, 2023.

Lee, Robin S and Michael Schwarz, "Interviewing in two-sided matching markets," *The RAND Journal of Economics*, 2017, *48* (3), 835–855.

Lehmann, Sebastian, "Toward an Understanding of the BDM: Predictive Validity, Gambling Effects, and Risk Attitude," *Working Paper Series*, 2015.

Leshno, Jacob D, "Dynamic matching in overloaded waiting lists," *American Economic Review*, 2022, *112* (12), 3876–3910.

Li, Bingbing, Stephanie Wang, and Xiaohan Zhong, "Timing is everything? An experimental investigation of incomplete information in centralized matching mechanisms," *Manuscript, Univ. Pittsburgh. http://www. pitt. edu/swwang/papers/Matching. pdf*, 2016.

Liu, Lydia T, Feng Ruan, Horia Mania, and Michael I Jordan, "Bandit learning in decentralized matching markets," *Journal of Machine Learning Research*, 2021, *22* (211), 1–34.

_ , Horia Mania, and Michael Jordan, "Competing bandits in matching markets," in "International Conference on Artificial Intelligence and Statistics" PMLR 2020, pp. 1618–1628.

Liu, Qingmin, "Stability and bayesian consistency in two-sided markets," *American Economic Review*, 2020, *110* (8), 2625–66.

_ , George J Mailath, Andrew Postlewaite, and Larry Samuelson, "Stable matching with incomplete information," *Econometrica*, 2014, *82* (2), 541–587.

Mah, Bow Tan, *Reflections on Housing a Nation* 2010.

_ , *Reflections on housing a nation*, Ministry of National Development, 2011.

Menzio, Guido and Shouyong Shi, "Efficient search on the job and the business cycle," *Journal of Political Economy*, 2011, *119* (3), 468–510.

_ , Irina A Telyukova, and Ludo Visschers, "Directed search over the life cycle," *Review of Economic Dynamics*, 2016, *19*, 38–62.

Miller, Robert A, "Job matching and occupational choice," *Journal of Political economy*, 1984, *92* (6), 1086–1120.

Monderer, Dov and Lloyd S Shapley, "Potential games," *Games and economic behavior*, 1996, *14* (1), 124–143.

Müller, Holger and Steffen Voigt, "Are there gambling effects in incentive-compatible elicitations of reservation prices? An empirical analysis of the BDM-mechanism," *Working Paper Series*, 2010.

Müller, Ulrich K and Mark W Watson, "Spatial Correlation Robust Inference," Working Paper 2021.

Naik, Canishk and Neil Thakral, "A welfare analysis of public housing allocation mechanisms," Working Paper 2021.

Narita, Yusuke, "Match or mismatch? Learning and inertia in school choice," *Learning and Inertia in School Choice (June 18, 2018)*, 2018.

Network, Strada Education, "On second thought: US adults reflect on their education decisions," 2017.

Ong, Justin, "Analysts say too soon to tell if BTO applications dip means stabilising demand, as projects may have been less attractive," *Today Newspaper*, Mar 2023.

Papageorgiou, Theodore, "Large firms and within firm occupational reallocation," *Journal of Economic Theory*, 2018, *174*, 184–223.

Postel-Vinay, Fabien and Helene Turon, "On-the-job search, productivity shocks, and the individual earnings process," *International Economic Review*, 2010, *51* (3), 599–629.

Prendergast, Canice, "The Allocation of Food to Food Banks.," *EAI Endorsed Trans. Serious Games*, 2016, *3* (10), e4.

Radio Farda, "Minister says no buyers for 100,000 government housing units," 9 2018.

Reeling, Carson, Valentin Verdier, and Frank Lupi, "Valuing goods allocated via dynamic lottery," *Journal of the Association of Environmental and Resource Economists*, 2020, *7* (4), 721–749.

Rothschild, Michael, "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 1974, *9* (2), 185–202.

_ **and Joseph Stiglitz**, "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," in "Uncertainty in economics," Elsevier, 1978, pp. 257–280.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Shi, Peng, "Optimal priority-based allocation mechanisms," *Management Science*, 2022, *68* (1), 171–188.

_ **and Junxiong Yin**, "Eliminating Waste in Cadaveric Organ Allocation," Working Paper 2022.

Sieg, Holger and Chamna Yoon, "Waiting for affordable housing in New York City," *Quantitative Economics*, 2020, *11* (1), 277–313.

Thakral, Neil, "Matching with stochastic arrival," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 209–12.

Urgun, Can, "Restless Contracting," *mimeo*, 2021.

Verdier, Valentin and Carson Reeling, "Welfare effects of dynamic matching: An empirical analysis," *The Review of Economic Studies*, 2021.

_ **and** _ , "Welfare effects of dynamic matching: An empirical analysis," *The Review of Economic Studies*, 2022, *89* (2), 1008–1037.

Voorneveld, Mark and Henk Norde, "A characterization of ordinal potential games," *Games and Economic Behavior*, 1997, *19* (2), 235–242.

Wächter, Andreas and Lorenz T Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, 2006, *106* (1), 25–57.

Waldinger, Daniel, "Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation," *American Economic Review*, 2021, *111* (8), 2660–96.

Weitzman, Martin L, "Optimal search for the best alternative," *Econometrica: Journal of the Econometric Society*, 1979, pp. 641–654.

Whittle, Peter, "Multi-armed bandits and the Gittins index," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1980, *42* (2), 143–149.

Wong, Lawrence, "Oral Answer by Ministry of National Development on BTO flat prices," 2 2017.

_ , "Written Answer by Ministry of National Development on first-time BTO flat applicants being unsuccessful in balloting for five times or more," 2 2017.

Wong, Maisy, "Estimating ethnic preferences using ethnic housing quotas in Singapore," *Review of Economic Studies*, 2013, *80* (3), 1178–1214.

_ , "Estimating the distortionary effects of ethnic quotas in Singapore using housing transactions," *Journal of Public Economics*, 2014, *115*, 131–145.

Yuen, Sin, "Singapore resident population in HDB flats falls to 3.04m, with smaller households spread over more flats," *The Straits Times*, 2 2021.