Title
Advances in the Use of Microdata for Modeling Urban Systems: Application to Causal Analysis, Regional Forecasting, and Population Synthesis

Author
Gardner, Max

Publication Date
2022

Advances in the Use of Microdata for Modeling Urban Systems: Application to Causal Analysis, Regional Forecasting, and Population Synthesis

by

Max A. Gardner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Paul Waddell, Co-chair
Professor Joan Walker, Co-chair
Professor Daniel Chatman
Professor Marta Gonzalez

Summer 2022

Advances in the Use of Microdata for Modeling Urban Systems: Application to Causal Analysis, Regional Forecasting, and Population Synthesis

Abstract

Advances in the Use of Microdata for Modeling Urban Systems: Application to Causal Analysis, Regional Forecasting, and Population Synthesis

by

Max A. Gardner

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Paul Waddell, Co-chair

Professor Joan Walker, Co-chair

The research contained herein is comprised of three studies, each of which utilizes a unique set of population-wide microdata to address a few of the many challenges that arise when statistical models are used to study the complex dynamics that govern the function and evolution of urban systems. These challenges range from the technical to the political, but they are all highly relevant to the management of the modern metropolis in which "data-driven" decision-making is increasingly the norm. Taken together, these studies draw upon relevant literature from the fields of urban geography, behavioral economics, and computational statistics, among others, to synthesize a body of work that questions what can and cannot be achieved through urban systems modeling while simultaneously making several contributions to its study and its practice.

The first study demonstrates how thoughtful model design can be harnessed to make powerful inferences about the policy origins of an economic phenomenon which places many of the most vulnerable inhabitants of our cities at greater risk of displacement. Specifically, this study identifies the first causal estimate of the effect of rent control status on eviction filing rates in the scientific literature. A 10-year dataset of eviction notices (n=21,806) is combined with a complete history of property tax records (n=1,978,687) in a regression discontinuity design to estimate a local average treatment effect of 0.013 evictions per residential unit per year conditioned on positive rent control status. Compared to the baseline rate of eviction notices over this same period, this translates to a 240% increase in the likelihood of eviction for tenants living in controlled units. I argue that this finding is best understood not as an inherent inefficiency of rent control policy in general, but rather as the result of specific state-wide laws, passed in the years following the adoption of rent control in San Francisco, which granted rent controlled property owners an economic incentive to evict and the legal

means to do so. The chapter concludes by making specific policy recommendations that city officials can enact today in order to protect its most vulnerable residents.

In the chapter that follows I address the issue of simulation error due to sampling of alternatives in discrete choice models. The issue of is an important one, not only because it represents a longstanding criticism of the statistical validity of many state-of-the-art forecasting models, but also because it points to a significant disconnect in the development of methods for estimation and for simulation. This disconnect reflects a larger rift that exists between theory and practice in urban systems modeling. The purpose of this chapter is to help bridge that gap by defining a novel measure of forecast error and using it to quantify the extent of the problem as it manifests in a disaggregate model of discretionary location choice, typical of those that are commonly found in use today. The definition of this metric is itself a valuable contribution to a discipline which is often maligned for its inability to assess the accuracy of its methods. In addition, I use this metric to identify two key findings. First, I show that the proportion of aggregate demand that is misallocated due to sampling of alternatives is actually reduced as the size of the universe of alternatives increases (i.e. becomes more disaggregate). Second, I find that in most scenarios, simple random sampling actually outperforms an approach based on importance sampling. Both results contradict the traditional wisdom about best practices in microscopic models of travel and land use demand.

The final study included here presents a method designed to make it easier for researchers and practitioners alike to acquire the kinds of data required to perform microscopic urban modeling. Unlike households and persons, no public repository of establishment-level microdata currently exists for businesses in the United States. Work in this domain has therefore been primarily limited to those with the resources to purchase expensive commercial datasets. As a result, the development of disaggregate models of business and firm dynamics has lagged behind their person- and household-based counterparts, hindering the development of fully integrated transportation and land use microsimulation systems as a whole. Drawing on recent advances in the application of Bayesian networks to population synthesis and data privacy preservation, I estimate a series of probabilistic models on a dataset of proprietary business establishment listings and use them to generate synthetic populations which match the joint distributions of key characteristics from the original data but contain none of the original records themselves. In a second analysis I show how aggregate Census data can be used as control totals for sampling from the fitted models to create synthetic populations that match the aggregate Census counts but have a much richer set of features than what is made publicly available by the Census. In theory, these data, along with the fitted models themselves, can be shared freely among collaborators without fear of copyright violation or disclosure of private data. These results demonstrate the great potential of Bayesian networks, and probabilistic models in general, to democratize access to microdata and facilitate greater scientific collaboration in the field of disaggregate urban systems modeling.

I conclude with a brief summary of these findings and discuss their relation to current issues in urban systems modeling. I identify several opportunities to improve and build on the work presented here, while also addressing the inherent limitations of model-based research to meet the most pressing needs of our urban communities at this particular moment in history.

To Savana, the love of my life.

Contents

List of Figures

List of Tables

Acknowledgments

I must first thank Professor Paul Waddell, whose unwavering support at a critical moment made the difference between the completion of my doctoral studies and an unceremonious early exit. His impact on my work since that time can be found throughout this manuscript.

I am eternally grateful to Professor Joan Walker for welcoming me into her research group about midway into my studies. She made sure I had a home in McLaughlin Hall when I needed one, and gave me the space to stretch out when I didn't.

I cannot thank Professor Dan Chatman enough for introducing me to the world of transportation and land use planning. Thank you for taking seriously the ideas of an intellectual interloper, and for being so generous with your time.

To Professor Marta Gonzalez I continue to be inspired by you and your work, and feel so fortunate that you arrived at Berkeley when you did.

I thank Professor Carolina Reid and Dr. Timothy Thomas for their advice and encouragement regarding the study that became Chapter 2 of this manuscript. Thank you also to Professor Angelo Guevara who provided valuable insight that shaped the analysis in Chapter 3, and to Dr. Srinath Ravulaparthy for inspiring the research included here as in Chapter 4.

Of course, none of this would have been possible without the family I had the great fortune of being assigned at birth. Thank you, Bubbi, for never forgetting to ask me if I was done with my "book" even though I think you knew the answer had not changed since the week before. To my parents, thank you for your quiet and steady encouragement and patience, and your earnest interest in whatever I was working on at the time. Thank you to my sister, Lauren, for your help proofreading, and to my brother, Jacob, whose own doctoral pursuits provided me with the motivation to make sure I finished first.

Lastly, I thank my wife and best friend, Savana, who deserves a second acknowledgement and many, many more for contributions that are too numerous to name here. I love you.

Chapter 1

Introduction

1.1 Background

Long before the purveyors of the "smart city" arrived on the scene offering algorithmic solutions to all of our civic woes, researchers, planners, and engineers have relied on statistical models to improve our understanding of the dynamics which govern the evolution of urban systems. For example, since the 1960s many urban areas in the United States have been required to produce long-range forecasts of travel and land use demand in order to receive state and federal funds for the construction and maintenance of transportation infrastructure. These models have not all been useful, and many of them have been used to justify wrongheaded policies and decisions whose effects are still negatively impacting the quality of life for communities today, particularly those that have historically been marginalized in other ways as well. While the purpose of the research presented here is neither to defend nor denounce this legacy, it is important to recognize the historical context from which it emerges.

1.2 The Canonical Urban Modeling Workflow

Urban modelers, like many practitioners in disciplines that traffic in statistical models, often conceive of their work as occurring in two distinct phases: estimation and simulation. During the estimation phase, a sample of observations are used to fit or "train" models that attempt to quantify the relationship between a dependent variable (i.e. an "outcome") and one or more exogenous factors that are assumed to influence that outcome in some way. These models are often parametric in nature, and the specification and design of these parameters allow the modeler to make different kinds of analytic inferences about the relative influence of each exogenous variable on the outcome of interest. The usefulness of a model, however, does not end here. A model, once estimated, can also be used to predict outcomes for different sets of input data which may or may not correspond to actual observations. This second phase goes by many names including simulation, prediction, forecasting, or projection. Although there are semantic differences which distinguish these concepts from one another, in practice this phase involves some kind of scenario analysis which is performed to help decision-makers evaluate policy alternatives or allocate resources according to anticipated needs.

In recent years, disaggregate urban modeling techniques have begun to supplant more traditional approaches in which analysis is conducted at the level of aggregate zonal geographies. Disaggregate models instead take individual "agents" or "choosers" (e.g. persons, households, or firms) as their fundamental units of analysis. This means that in order to simulate system-wide conditions, population-scale datasets of disaggregate entities are required. As a result, population synthesis has emerged as a third and often requisite step in the modeling workflow that occurs after a model is estimated but before it can be used to perform simulation.

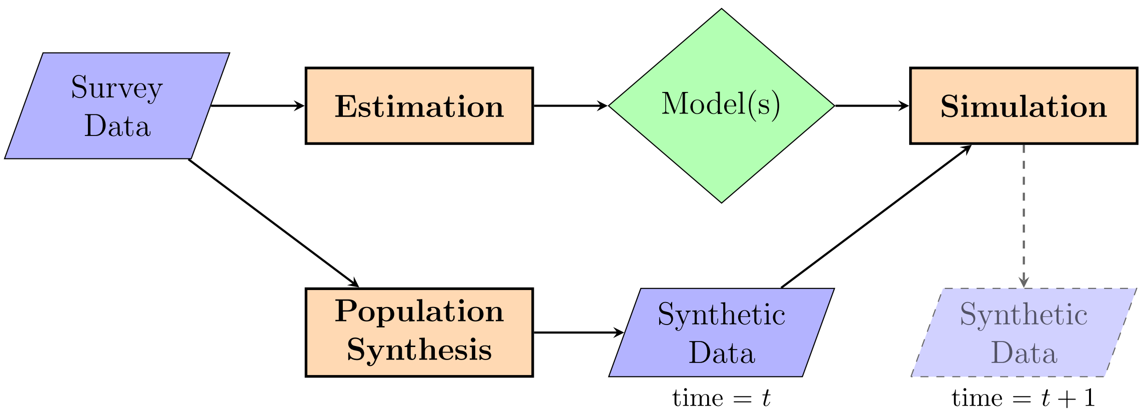


Figure 1.1: The canonical workflow for disaggregate urban modeling.

Together these three phases constitute what I call the "canonical workflow for disaggregate urban modeling" (Fig. 1.1). Incidentally, this three-phase structure also loosely defines the structure of the research presented below. In each of the following chapters I utilize a unique set of population-wide microdata to address an issue pertaining to a different phase of the canonical workflow: first estimation, then simulation, and finally population synthesis.

1.3 Motivation and Contributions

The first of the three studies collected here stands out for a number of reasons. First, it is the only one that relies on historic, observational data to make inferences about the present. It is also, therefore, the only study whose chief contribution is policy-related rather than methodological. However, the most unique aspect of this work is that it is the only study which was inspired by real events in the life of the author. In 2011, I was served with an eviction notice at the rent-controlled apartment in San Francisco that I shared with two friends. Although we ultimately succeeded in fighting the eviction, it was clear that the process could have played out much differently had we not been a trio of white young men equipped with the spare time and educational backgrounds to study the San Francisco Rent Ordinance and convincingly argue in a written petition to the Rent Board that our eviction notice had been filed in bad faith. Around the same time, I became aware of efforts by local tenant advocacy groups to document the widespread use of eviction by landlords as a means of bypassing rent control[1] [1]. This important work provided me with valuable context about what I had experienced firsthand, and led directly to the research I present in the Chapter 2. Although the phenomenon of rent control-induced eviction is well understood by those who have lived with rent control in California, evidence of its existence has been primarily anecdotal, making it very difficult to quantify the extent of the problem. However, by drawing upon my own experience, along with the work of local organizers and a relatively recent literature on the design of regression discontinuity models for causal inference, I was able to estimate a model that, provides strong evidence of a significant causal effect of rent control on eviction rates. The magnitude of this effect suggests that, all else being equal, rent controlled tenants in San Francisco are more than twice as likely to be served with an eviction notice than their non-rent controlled counterparts. By recentering tenants in a scientific debate on rent control, Chapter 2 also calls attention to what urban models can achieve when they are taken out of the hands of neutral technicians and employed in the fight for more just and equitable cities.

In the second study I turn my attention to one of the challenges that arise in the simulation phase of the workflow, when a model having already been estimated, is used to make inferences about the future rather than the past. The issue of simulation error due to sampling of alternatives in discrete choice models is an important one, not only because it represents a longstanding criticism of the statistical validity of many state-of-the-art microsimulation models, but also because it points to a significant disconnect that exists in urban systems modeling between best practices for estimation and simulation. Although I have described these two phases as complementary components of a holistic process, their representation in the scientific literature is extremely lopsided in favor of the former. No doubt this imbalance reflects the fact that simulation is sometimes dismissed as the domain of practitioners and technicians rather than scholars. But it also speaks to the difficulty of assessing the accuracy of predictions for scenarios which have not and may never come to pass [2]. Both of these factors have contributed to the increasing suspicion with which the practice of regional demand forecasting is viewed [3]. And yet, the demands for cost-benefit analyses and impact assessments in transportation and land use planning remains. This study aims to help bridge the gap between theory and practice in urban modeling by defining a novel measure of forecast error and using it to quantify the extent of the problem as it manifests in models of disaggregate location choice. Two key findings are identified which contradict the prevailing wisdom about best practices in microsimulation for travel and land use demand. First, I show that the proportion of aggregate demand that is misallocated due to sampling of alternatives is actually reduced as the size of the universe of alternatives

[1] See https://www.tenantstogether.org/ellisreport, http://www.antievictionmappingproject.net/ellis.html

increases (i.e. becomes more disaggregate). Secondly, I find that in most scenarios, simple random sampling actually outperforms an approach based on importance sampling. This latter result, in particular, has important implications for practitioners of urban microsimulation who on more than one occasion have been accused of mistaking complexity for validity (e.g. [4, 5]).

As important as it is for practitioners of urban modeling to continuously interrogate the accuracy of their own methods, it is just as important that we question who in society has access to this technology and the resources to wield it. After all, even the most accurate model is only as useful as the questions that are asked of it, and the character of these questions is in turn heavily informed by whoever is doing the asking. In the context of disaggregate urban modeling, access to reliable sources of microdata is often the limiting factor. The United States Census Bureau, for example, maintains many databases of disaggregate records collected from various censuses and surveys, but access is typically restricted in order to prevent the misuse of potentially sensitive or personally identifiable information. While such privacy concerns are important, in the absence of legislation that *prevents* the collection and dissemination of disaggregate data by other means it has merely shifted demand to the commercial sector. As a result, researchers' ability to participate in this field of study is increasingly determined by the size of their budgets. Thus, Chapter 4 addresses the competing needs of data democratization and data privacy in the realm of urban modeling. Drawing on recent advances in the use of Bayesian networks for privacy-preserving data publishing I demonstrate the potential of probabilistic models trained on proprietary business establishment microdata to generate synthetic populations that can be distributed freely among collaborators and beyond. As the state-of-the-art in urban systems modeling continues to advance in the direction of data-hungry microscopic models, it is critical to the continued integrity of the discipline that scientific discovery in this domain is not restricted to those who can afford the price of entry.

Taken together, the research contained in the following three chapters synthesizes a body of work which at once questions what can and cannot be achieved through urban systems modeling while also making important contributions to the discipline. In doing so, it can also be considered an attempt to embody an approach to practice for which it also advocates: one that is scientifically rigorous, comprehensive in scope, and grounded in humility.

Chapter 2

The Effect of Rent Control Status on Eviction Filing Rates: Causal Evidence from San Francisco

2.1 Introduction

In this chapter I present causal evidence of a significant, positive effect of rent control on evictions rates in San Francisco. In order to identify a causal mechanism, two panel datasets of 1) eviction notice filings and 2) tax assessor records are combined in a regression discontinuity (RD) design to estimate the local average treatment effect (LATE) of building-level rent control status on annual eviction filing rates per residential unit. Four models are specified with successively greater numbers of covariates, and each model identifies a statistically significant LATE between 0.009 and 0.0141 eviction notices per unit per year. These results correspond to a rate of eviction filings for rent controlled tenants that is approximately 2-3x that of their non-rent-controlled counterparts.

At first glance this finding seems to contradict the academic consensus about the benefit of rent control on tenant stability, as well as popular intuition about the outcomes that rent control policy is designed [6]. However, the causal relationship identified in this study makes perfect sense when considered in the context of certain provisions of San Francisco's Rent Ordinance that not only create an economic incentive for property owners to evict their rent controlled tenants but also provide them with specific legal processes for doing so. If in fact these provisions of San Francisco's Rent Ordinance, described in greater detail below, are to blame for the higher eviction rates observed at rent controlled properties rather than the concept of Rent Control in general, then the pressing problem identified herein can likely be remedied with policy-based solutions. I conclude this chapter by discussing a few such potential solutions and identifying opportunities for future work that can address the limitations inherent to this study.

Policy Background

In 1979 the passage of the Residential Rent Stabilization and Arbitration Ordinance, also known as the Rent Ordinance[1], did two important things: 1) established rent control for all existing multi-unit rental housing[2]; and 2) fully enumerated the “just causes” that are grounds for lawful eviction. Although the term “rent control” is often used colloquially to describe the San Francisco Rent Ordinance in its entirety, it is critical to this discussion that we distinguish between the Rent Control, whose primary aim is to keep rental housing affordable for incumbent tenants, and the tenant protections, which are designed to protect tenants from unnecessary hardship due to landlords acting in bad faith. As such, a failure of rent control policy to prevent evictions should not be mistaken for evidence of a failure of rent control itself.

The rent control provisions of the Rent Ordinance are typical of second-generation controls, which are often (and more accurately) described as rent stabilization rather than rent control[3] because they include explicit mechanisms by which rents can be increased over time.Under second-generation controls, annual allowable rent increases are often pegged to a macroeconomic indicator, which in the case of San Francisco is the Consumer Price Index (CPI). Because the Ordinance stipulates that these provisions only apply to buildings in existence at the time that the legislation was passed, it is possible to use the date of construction (built-year) as a proxy for building-level rent control status. For the purposes of this study, the most consequential change to the original San Francisco rent control law occurred in 1995 when the State of California passed the Costa Hawkins Rental Housing Act, which among other things imposed the policy of vacancy decontrol on all existing and future rent control regimes in the state.

Vacancy decontrol is a common feature of second-generation rent regulations that allows owners of controlled properties to return their rents to market rate at the start of each new tenancy. Since the inception of vacancy decontrol in California, renters there have worried about the incentive it creates for rent controlled landlords to keep tenant durations short and tenant turnover high [7]. The issue has not gone unnoticed by academics, either. present theoretical proof of a preference for shorter-duration tenants among landlords under vacancy decontrol, while Asquith (2019) provides quantitative evidence that this preference could be impacting displacement rates for rent controlled tenants. Thus far, however, no study has been unable to disentangle this effect from other regional dynamics.

In addition to rules governing rent control, the Rent Ordinance also specifies certain tenant protection measures. In particular, it fully enumerates the 16 circumstances under which tenants can be lawfully evicted, known as “just causes”. The 16 just cause provisions of the Ordinance are reproduced here in Appendix A and annotated with five columns that provide additional information deemed relevant to this study. The first three of these

[1] San Francisco Administrative Code Chapter 37

[2] Owner-occupied buildings of four units or less were excluded until 1995. As this change predates any of the eviction data used in this study, this fact is largely inconsequential to the results of the study.

[3] I will use both terms – “stabilization” and “control” – in reference to the San Francisco Rent Ordinance.

columns describe the eviction type ("no-fault" vs. "breach of lease"), the eviction level ("unit" vs. "whole building"), and whether or not a relocation payment is required. The last two columns relate each eviction type to rent control, distinguishing between evictions which result in a simple reset (vacancy decontrol) from those that result in full (permanent) decontrol, and stipulating the duration of time before those changes are allowed to go into effect.

One type of just cause eviction deserves special attention in the context of this study: the Ellis Act eviction. The Ellis Act is a California state law which guarantees landlords the right to "go out of business" by removing all units in a building from the rental market at once. Ellis Act evictions are notable for three reasons: 1) the Ellis Act was not passed until 1985 and thus they had to be amended to the original list of just cause evictions; 2) they are the only type of just cause eviction (apart from the exceedingly rare "Good Samaritan" eviction) that is specific to the policy context of San Francisco and the state of California; and 3) a body of evidence exists which indicates that Ellis Act evictions in particular have been exploited by landlords seeking to circumvent rent control restrictions. In particular, [1] report that 60% of Ellis Act evictions in San Francisco occur within the first year of ownership, and 78% occur within the first five years. This is compelling evidence that the Ellis Act is primarily being used for means other than long-term landlords exiting the rental market.

Although incomplete data made it difficult to incorporate eviction type segmentation into the analysis presented in this study, the Ellis Act, when considered in conjunction with the vacancy decontrol provisions of Costa-Hawkins Act, provides important context for understanding the ways in which specific California state laws, imposed after the adoption of rent control by the City of San Francisco, paved the way for the causal relationship identified herein.

Literature Review

Few policies in the realm of housing and urban economics occupy as prominent a position in the popular consciousness as rent control. Yet until very recently, the number of new, empirical findings on its effectiveness as a regulatory tool have been few and far between[4]. Theoretical models proving rent control's many inefficiencies formed the basis of a decades-long consensus among economists who treated the science as not merely settled but self-evident [18–20]. At the same time, and perhaps paradoxically, a lack of detailed data on tenant and landlord outcomes made it very difficult to disentangle any of the empirical effects of rent control, good or bad, from the other market forces operating in complex urban systems.

The publication of [21] seems to mark a turning point in the academic literature on rent control. In that paper Arnott argued that modern, "second-generation" rent controls were so nuanced and malleable – compared to the hard-line rent freezes imposed by their

[4]Notable exceptions include [8–17]

first-generation predecessors – that they defied a priori characterization as either good or bad policy. Arnott instead advocated for the use of empirical evidence to evaluate the effects of rent control on a case-by-case basis. Recently, as concerns over gentrification and displacement in America's "superstar cities" [22] have sparked renewed interest in the topic of rent control, a new body of empirical research has emerged which takes Arnott's call as a common point of reference. Instead of asking whether or not rent control "works", this new literature is focused on more targeted assessments like how rent control affects commute times in New Jersey [16], or how the supply of controlled rental housing changes in response to local demand shocks in San Francisco [23]. This study asks a similarly targeted question about one important aspect of rent regulation: does rent control contribute to higher rates of eviction in San Francisco? The answer to this question, and the findings presented here, have potentially far-reaching ramifications for how we think about a policy designed specifically to increase housing stability for incumbent tenants.

The present study makes two significant contributions to the literature. First, despite the fact that tenants are the presumptive beneficiaries of rent control, most quantitative studies of the policy have tended to focus on diffuse, market-wide effects like housing quality, supply, and affordability [17, 23–25]. Instead, I attempt to re-centre tenants and tenant outcomes in the debate over rent control by providing the first causal estimate in the peer-reviewed literature of its effect on eviction rates. Secondly, this study presents a novel approach for studying the direct effects of rent control by employing a popular causal inference method – the regression discontinuity (RD) – in an entirely new setting. Although the results of this particular study are specific to eviction rates in San Francisco, the same methodology can be used to investigate other effects of rent control in the many other jurisdictions where rent control eligibility is determined by similar criteria.

The work presented here builds on two other recent studies of rent control in San Francisco, both of which employ quasi-experimental designs. The first is [26], which stands out as one the first observational studies to focus explicitly on measuring outcomes for existing tenants under rent control[5]. Despite the authors' conclusion that rent control has likely worsened the effects of gentrification by reducing the supply of affordable housing available to future residents, they do find that incumbent tenants "benefit on net" as a result of reduced rates of displacement and below-market rents. This latter finding, however, does not directly test for differential rates of forced displacement among controlled and uncontrolled tenants, nor does it account for the potentially greater costs of a forced relocation relative to one that is voluntary or economically induced. In contrast, in thus study I approach rent control solely as a housing stability measure, and in doing so offers strong evidence that any attempt to quantify the benefit of rent control must also take into account the effect of rent control-induced eviction on incumbent tenants. Another significant difference is that this study covers the entire stock of rent controlled properties in the city of San Francisco, whereas the identification strategy employed by [26]restricts their analysis to small

[5][13], [13], [16], and [27] could be considered exceptions, but these rely on census data rather than disaggregate tenant observations.

multifamily properties ($\leq$ 4 units). As such, the findings presented here are more general.

The paper most closely resembling this one methodologically is [23]. Asquith leverages very similar datasets of eviction notices and tax assessor records, along with a selection-on-unobservables design to show that landlords decrease the supply of rent-controlled housing via evictions in response to local demand shocks. Rather than observing these shocks, however, the author relies on a secondary model to estimate the hedonic price effects of newly sited transit amenities targeted towards high-income knowledge workers. This two-stage design makes the results of the primary instrumental variables (IV) model difficult to interpret and limits their relevance to the context of local demand shocks. The results of the IV model also depend heavily on the validity of the estimated shocks, which the author concedes are implausibly large. The biggest factor distinguishing this study from Asquith's, however, is that the actual treatment variable of interest in Asquith's study is the demand shock, from which it impossible to disentangle the direct effect of rent control. This design works well for the purposes of that study, which is primarily concerned with evictions as a channel through which landlords can manipulate the supply of controlled rental housing. In this sense, the findings of Asquith are not that dissimilar from previous research demonstrating a depressive effect of rent control on housing supply [17, 24, 26]. In contrast, this study treats eviction as a cost that is primarily borne by incumbent tenants rather than a housing market. By centring tenant outcomes instead of market effects, my findings are more relevant to an evaluation of rent control as tool for promoting housing stability.

2.2 Data

The main data source is a database of eviction notices filed with the San Francisco Rent Board between 2007 and 2016. These data (n=21,806) represent the full universe of eviction filings by landlords against tenants in San Francisco during that ten-year period. The detailed data used for this study are made available by written request from the San Francisco Rent Board, but a geographically anonymized version of this dataset can be downloaded directly from the city's open data portal[6]. Most of the eviction records include the eviction type (e.g. failure to pay rent, owner move-in, etc.) but many are not specified (see Table 2.2 for a summary of the eviction data). Although eviction notices do not necessarily result in an eviction, the notice itself and the threat of eviction can be enough to cause tenants to pre-emptively vacate their residences [28, 29]In this way, eviction notice filing rates might actually be a better measure of forced displacement pressure than unlawful detainers or writs of restitution[7].

The dependent variable of interest in this study is eviction notices per residential unit per year across the entire population of San Francisco properties with two or more residential

[6]In this case, the eviction records were graciously provided to the author, unaltered, courtesy of the Anti-Eviction Mapping Project ([1]).

[7]In this chapter I use both "eviction" and "eviction notice" in reference to the Rent Board data.

units[8]. The full dataset was assembled by matching eviction notice records against annual parcel-level tax assessor records published by the City and County of San Francisco over the same time period as the eviction notices. After cleaning and standardizing the assessor data, 1,978,687 parcel records are aggregated by year and street address to arrive at a total population of 1,553,397 unique year/address combinations. Unit/apartment numbers are dropped (e.g. "123 Main St #4" becomes "123 Main St"), and the total units are aggregated along with the eviction counts, square footage, and assessed value. After dropping four eviction records due to incomplete or malformed data, I am able to find an exact match in the assessor records for 92.4% of the eviction notices (n=20,136).

Next, the sample is further restricted to include only those addresses that can be reasonably identified as "rent control eligible" according to their assessor designated building class codes. Table 2.1 describes these class codes and their inferred rent control eligibility, but in general I deem any property with 2+ residential units to be eligible.

Table 2.1: San Francisco Assessor-Recorder Class Codes & Inferred Rent Control Eligibility. Adapted from City and County of San Francisco Assessor-Recorder Secured Property Tax Roll Data.

Use Code	Use Definition	Class Code	Class Definition	Rent Control Eligibile
COMH	Commercial Hotel	H	Hotel	N
COMH	Commercial Hotel	H1	Hotel	N
COMH	Commercial Hotel	H2	Hotels - Other	N
COMH	Commercial Hotel	HC	Hotel Commercial (H2w/Com)	N
COMH	Commercial Hotel	M	Motels	N
COMH	Commercial Hotel	RH	Residential Hotel & SRO	Y
COMH	Commercial Hotel	RH1	Retail & Hotel	N
COMM	Commercial Misc	AC	Apartment & Commercial Store	N
COMM	Commercial Misc	E	Schools	N
COMM	Commercial Misc	G	Garages (Commercial)	N
COMM	Commercial Misc	GC	Golf Course	N
COMM	Commercial Misc	GCU	Golf Course	N
COMM	Commercial Misc	GZ	Garage Condominium	N
COMM	Commercial Misc	MIX	Mixed use	N
COMM	Commercial Misc	N1	Hospitals	N

[8]Properties with a tenancy-in-common (TIC) use code are excluded from the analysis.

COMM	Commercial Misc	N2	Convalescent/Nursing Homes	N
COMM	Commercial Misc	PD	PUD (Planned Unit Development)	N
COMM	Commercial Misc	PL	Parking Lot	N
COMM	Commercial Misc	PZ	Parking Stall Condominium	N
COMM	Commercial Misc	S	Gas Station	N
COMM	Commercial Misc	T	Theatres	N
COMM	Commercial Misc	TS	Timeshare	N
COMM	Commercial Misc	TSF	Time Share Fractional	N
COMM	Commercial Misc	TSU	Time Share Unsegregated	N
COMM	Commercial Misc	U	Clubs,Lodges,Fraternal Organizations	N
COMM	Commercial Misc	W	Churches,Convents,Rectories	N
COMO	Commercial Office	O	Office	N
COMO	Commercial Office	O35	Office Portion Leased of 35 or More	N
COMO	Commercial Office	OAH	Office - High Class A	N
COMO	Commercial Office	OAL	Office - Low Class A	N
COMO	Commercial Office	OAT	Office - "Trophy" Class A	N
COMO	Commercial Office	OBH	Office - High Class B	N
COMO	Commercial Office	OBM	Office - Middle Class B	N
COMO	Commercial Office	OCH	Office - High Class C	N
COMO	Commercial Office	OCL	Office - Low Class C	N
COMO	Commercial Office	OCM	Office - Middle Class C	N
COMO	Commercial Office	OMD	Medical- dental Office Building	N
COMO	Commercial Office	OZ	Office - Condominium	N
COMO	imputed	OZE	imputed	N
COMR	Commercial Retail	B	Bank	N
COMR	Commercial Retail	BZ	Bank Condominium	N
COMR	Commercial Retail	C	Commercial Stores	N
COMR	Commercial Retail	C1	Shopping Center	N
COMR	Commercial Retail	CD	Commercial Department Stores	N
COMR	Commercial Retail	CM	Commercial/Mixed use	N

COMR	Commercial Retail	CZ	Commercial Store Condo	N
COMR	Commercial Retail	EC	Entertainment Complex	N
COMR	Commercial Retail	OC	Office with Major Retail Units	N
GOVT	Government	CP	City Property	N
GOVT	Government	P	Public Buildings (Govt)	N
GOVT	imputed	RDA	imputed	N
GOVT	Government	RDAP	Redevelopment Agency Property	N
GOVT	Government	SP	State of California Property	N
GOVT	Government	UCP	University of California Property	N
GOVT	Government	USP	U.S. Government Property	N
GOVT	Government	Y	Port Commission Property	N
IND	Industrial	I	Industrial	N
IND	Industrial	IDC	Industrial Data Center	N
IND	Industrial	IW	Industial Warehouse	N
IND	Industrial	IX	Industrial Mixed/Other Use	N
IND	Industrial	IZ	Industrial Condominium	N
MISC	Miscellaneous/Mixed-Use	MB	Mission Bay	N
MISC	Miscellaneous/Mixed-Use	UWL	Under Water Lot	N
MISC	Miscellaneous/Mixed-Use	V	Vacant Lot	N
MISC	imputed	VA1	imputed	N
MISC	Miscellaneous/Mixed-Use	VA15	Vacant land- residential 15+ units	N
MISC	Miscellaneous/Mixed-Use	VCI	Vacant Lot Comm and Ind	N
MISC	Miscellaneous/Mixed-Use	VCIX	Vacant Lot Comm and Ind w/ Restriction	N
MISC	imputed	VPU	imputed	N
MISC	Miscellaneous/Mixed-Use	VPUB	Vacant Lot Public Use	N

MISC	Miscellaneous/Mixed-Use	VR	Vacant Lot - Restrictions	N
MISC	Miscellaneous/Mixed-Use	VRX	Vacant Lot Residential w/ Restriction	N
MISC	Miscellaneous/Mixed-Use	X	Misc	N
MRES	Multi-Family Residential	A	Apartment	Y
MRES	Multi-Family Residential	A15	Apartment 15 Units or more	Y
MRES	Multi-Family Residential	A5	Apartment 5 to 14 Units	Y
MRES	Multi-Family Residential	CO	Coop Units Unsegregated	Y
MRES	Multi-Family Residential	DA	Dwellings - Apartments	Y
MRES	imputed	DA1	imputed	Y
MRES	Multi-Family Residential	DA15	Dwellings - Apt 15 units or more	Y
MRES	Multi-Family Residential	DA5	Dwellings - Apt 5 to 14 units	Y
MRES	imputed	DCO	imputed	N
MRES	Multi-Family Residential	DCON	Legal Multi-Family Con to SFR	N
MRES	Multi-Family Residential	DD	2 Dwellings on One Parcel	Y
MRES	Multi-Family Residential	DD5	2 Dwellings on 1 Parcel 5 to 14 units	Y
MRES	Multi-Family Residential	DF	1 Flat & Dwelling-1 Parcel	Y
MRES	Multi-Family Residential	F	Flats & Duplex	Y
MRES	Multi-Family Residential	F15	Flats 15 units +	Y
MRES	Multi-Family Residential	F2	Flat & Store	Y
MRES	Multi-Family Residential	F5	Flats 5 to 14 units	Y
MRES	Multi-Family Residential	FA	1 Flat & 1 Apt Bldg-1 Parcel	Y
MRES	Multi-Family Residential	FA5	1 Flat & 1 Apt - 1 Parcel 5 to 14 units	Y
MRES	Multi-Family Residential	FS	Flat & Store 4 units or less	Y
MRES	imputed	FS1	imputed	Y
MRES	Multi-Family Residential	FS15	Flat & Store 15 units +	Y

MRES	Multi-Family Residential	FS5	Flat & Store 5 to 14 units	Y
MRES	Multi-Family Residential	OA	Office and Apartments	Y
MRES	imputed	OA1	imputed	Y
MRES	Multi-Family Residential	OA15	Office and Apartments 15 units +	Y
MRES	Multi-Family Residential	OA5	Office and Apartments 5 to 14 units	Y
MRES	imputed	TI1	imputed	N
MRES	Multi-Family Residential	TI15	TIC Bldg 15 units +	N
MRES	Multi-Family Residential	TIA	TI Apartment	N
MRES	Multi-Family Residential	TIC	TIC Bldg 4 units or less	N
MRES	Multi-Family Residential	TIC5	TIC Bldg 5 to 14 units	N
MRES	Multi-Family Residential	TIF	TI Flats & Duplex	N
MRES	Multi-Family Residential	XV	Single Struct on Multi Lot(D & F's only)	Y
SRES	Single Family Residential	COS	Coop Units Segregated	N
SRES	Single Family Residential	D	Dwelling	N
SRES	Single Family Residential	DBM	Dwelling BMR	N
SRES	Single Family Residential	LZ	Live/Work Condominium	N
SRES	imputed	LZB	imputed	N
SRES	Single Family Residential	LZBM	Live/Work Condominium BMR	N
SRES	Single Family Residential	OZEU	Office Condo Economic Unit	N
SRES	Single Family Residential	TH	Town House	N
SRES	imputed	THB	imputed	N
SRES	Single Family Residential	THBM	Town House BMR	N
SRES	Single Family Residential	Z	Condominium	N
SRES	Single Family Residential	ZBM	Condominium BMR	N
SRES	Single Family Residential	ZEU	Condominium Economic Unit	N

Of these (n=349,607) I drop an additional 5,094 parcel records (and their associated 296 evictions) due to inconsistent unit counts in the assessor records (e.g. 0 units for a parcel with a multifamily class code). In the end, the sample of observations used for analysis is comprised of 344,513 annual address records, which in total account for 13,963 eviction notices. Table 2.3 provides descriptive statistics of this dataset, and Figure 2.1 shows their

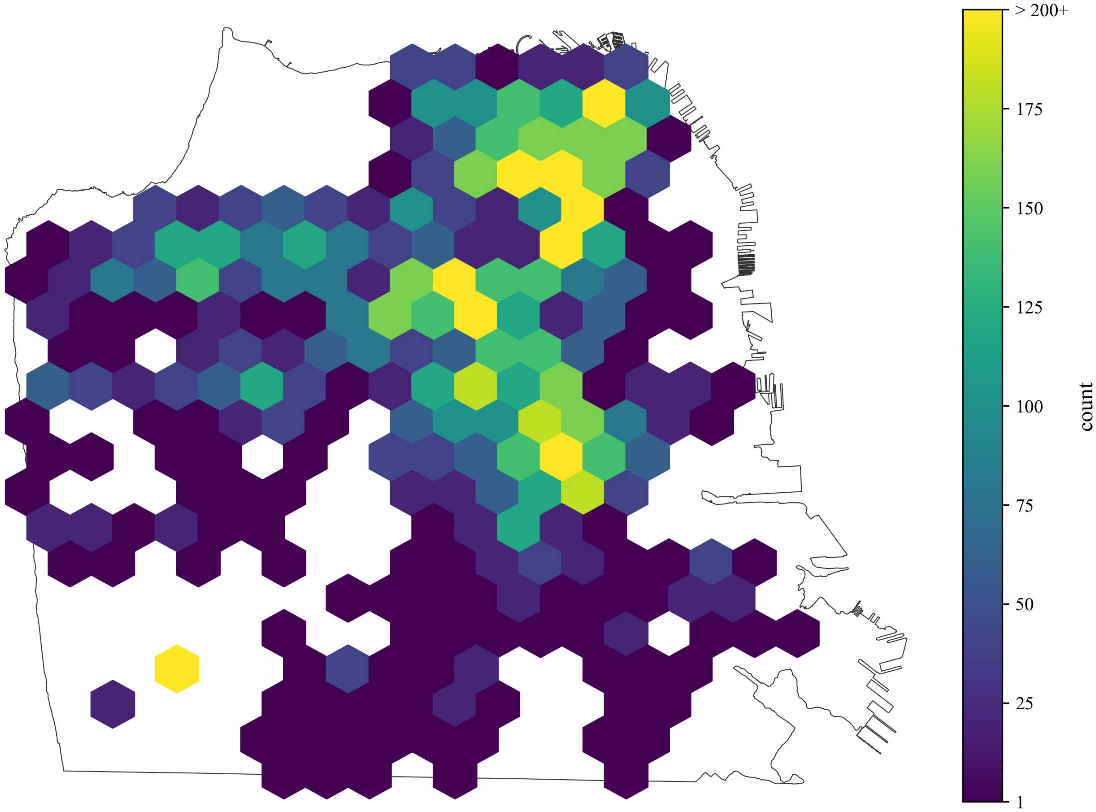


Figure 2.1: Map of San Francisco Eviction Notices (2007-2016). Only eviction notices matched to assessor records are shown (n=13,969).

geographic distribution.

After restricting the sample to properties with rent control eligible use codes, rent control status can be inferred solely by the year in which a property was built. Given that the Rent Ordinance was passed on June 16, 1979 and that it applied only to structures in existence at that time, rent control eligibility was (and continues to be) extended only to properties built on or before that date[9]. In the following analysis, this arbitrary but well-known delineation between "treated" (i.e. rent controlled) and "control" (i.e. market rate) groups of properties forms the basis for a pseudo-natural experiment in which a treatment effect (i.e. change in eviction rates) can be estimated and causality can be inferred.

[9]Since the assessor records only provide the property built year, and not month or day, I consider all properties built in 1979 to be rent controlled. By potentially including uncontrolled properties in the sample of controlled properties, it is possible that the estimated treatment effect is conservative in magnitude. However, repeated tests of the models with and without properties built in 1979 did not significantly alter the results.

2.3 Methodology

To estimate the treatment effect of rent control, I use a regression discontinuity (RD) design, exploiting the 1979 built-year cut-off eligibility requirement described above. The RD design has the dual benefit of directly estimating the treatment effect of rent control on eviction rates, and also making the identification of that treatment extremely transparent and easily understood. The use of RD dates back to 1960 [30] but its popularity as a causal inference method has gained significantly since the 1990s. Numerous studies comparing the statistical power of the RD against randomized controlled trial (RCT) experimental designs have served to bolster its reputation as an effective substitute in cases where true RCT designs are infeasible, as is often the case in policy analysis. A 2018 meta-analysis of 15 "within-study comparisons", each of which compared causal estimates obtained from both RD and RCT analysis conducted within the same study, found that the bias of the RD estimates was distributed tightly and symmetrically around zero (within 0.07 standard deviations of the RCT values in a given study on average), concluding that RD is "robustly internally valid in research practice" [31]. RD also benefits from an extremely transparent identification mechanism relative to other selection-on-unobservables designs like IV. Standard methods of graphical analysis make interpretation of both the design and its results easily understood by a wide variety of audiences [32].

The most significant shortcoming of RD relative to other causal inference methods is perhaps the limited set of circumstances in which the design is appropriate. In particular, RD requires a treatment assignment mechanism that depends either wholly or partially on a characteristic threshold value that a participant either exceeds or does not. If treatment R_i can be predicted based on whether a variable Y_i lies above or below a threshold value c, then the effect of R_i on the outcome E_i can be identified given that the relationship between Y_i and E_i is smooth and continuous for values of Y_i above and below c. The premise of RD is that if this latter assumption holds, then the causal effect of R_i on E_i can be estimated by measuring the size of the "jump" or discontinuity in E_i at $Y_i - c$.

In this study I implement the "sharp" RD design [32], where treatment assignment is completely deterministic based on the threshold. The basic functional form is given in Equation 2.1:

$$\begin{gathered} E_i = \alpha + \beta R_i + \gamma(Y_i - c) + \lambda(Y_i - c) \cdot R_i \\ \text{given} \quad R_i = \mathbf{1}\{Y_i < c\}, \quad c - h < Y_i < c + h \end{gathered} \tag{2.1}$$

where the dependent variable E is the annual evictions per unit for property i, Y is the built-year of the property, also known as the "running variable" in RD parlance, c is the threshold value (1980 in our case) along the dimension of the running variable, and R is a rent-control "treatment" indicator that evaluates to one for properties built prior to 1980 and zero otherwise. The bandwidth parameter h identifies the maximum distance between the running variable and the cut-off threshold, beyond which observations are excluded from the sample. Many methods exist to identify an optimal bandwidth, but in repeated tests I found that my estimates of the treatment effect were robust to variation of this value (see

Figure 2.4). I ultimately settled on a bandwidth of 27 because this limits the sample to buildings constructed between 1953 and 2007, which ensures that only buildings with a full 10 years of history over the period of observation (2007-2016) are included in the analysis.

2.4 Results

Empirical Analysis

Eviction Notices

Table 2.2 summarizes the eviction records (n=13,963) used in the main analysis after segmenting by eviction type category and their built-year relative to the 1980 cut-off. The left

	Built-year < 1980		**Built-year ≥ 1980**	
Eviction Category	*Count*	*%*	*Count*	*%*
Breach of Lease	6,432	46.7	124	65.6
No-fault	2,168	15.7	3	1.6
Unknown/Other	5,174	37.6	145	32.8
Total	13,774		189	

Table 2.2: Eviction frequency (2007-2016) by category and built-year cut-off for rent control eligible addresses

two columns represent eviction notices filed in rent-controlled properties, while the right two columns represent those filed in uncontrolled properties. The first thing that stands out is that there are nearly two orders of magnitude (73x) more notices filed in rent-controlled properties. More than anything this number reflects the fact that of the 344,513 property records used in the analysis only 14,132 (4.1%) were for properties built after 1979, a fact which itself is explained by the diminishing construction rate of multifamily housing in San Francisco over time (Figure 2.2).

More interesting, however, is the fact that No-fault evictions constitute a much higher percentage of eviction notices at rent-controlled properties compared to the uncontrolled sector[10]. These findings are consistent with [23], and support the idea that San Francisco's rent control laws may be incentivizing controlled landlords to evict law-abiding tenants. Also of interest is the fact that Breach of Lease or At-fault evictions constitute a smaller portion of evictions in rent controlled properties, which may suggest that rent control is

[10]Here I use "No-fault" to describe any of the following nine eviction types: owner move-in (OMI), capital improvement, Ellis Act, condo conversion, substantial rehabilitation, lead remediation, good Samaritan tenancy ends, development agreement, and demolition. All other evictions, except for those where the eviction type was not indicated, are considered "At-fault" or "Breach of lease" evictions. See Appendix A For a full enumeration of eviction types, categories, and observations.

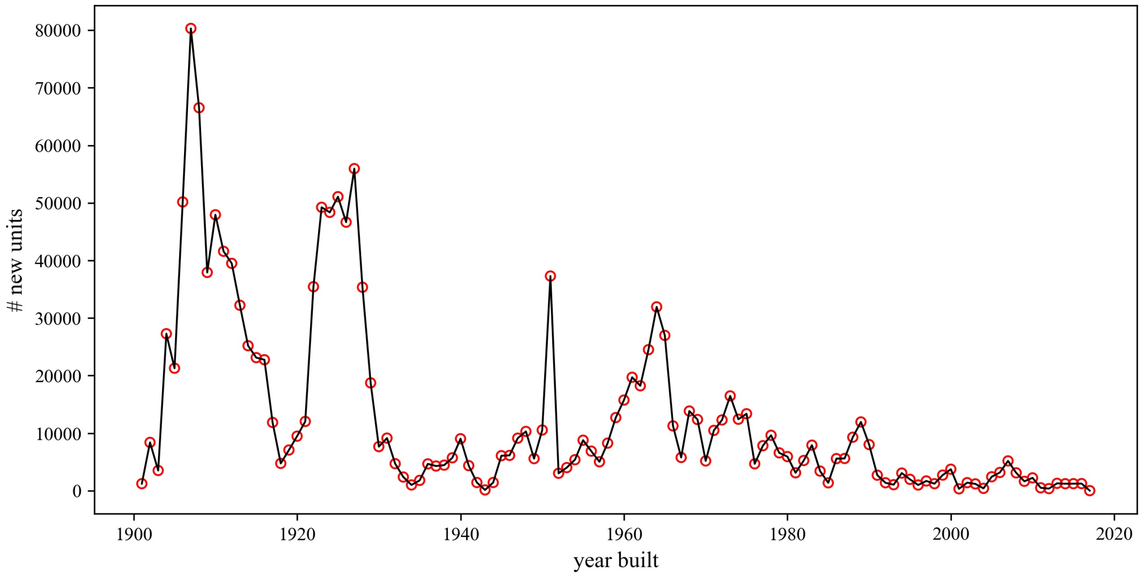


Figure 2.2: Histogram of unit counts in rent control-eligible buildings by year built. Construction of multi-family housing has slowed considerably since the 1960s, but there is no obvious discontinuity in 1980 when the rent control ordinance took effect.

actually achieving one of its primary objectives of keeping tenants from failing to pay their rent. Unfortunately, there are too many uncategorized eviction notices in the dataset to say anything more conclusive on the subject.

Mean Differences

Table 2.3 compares the counts and averages of the aggregate assessor records by rent control eligibility and built-year threshold. The first two rows of the table correspond to uncontrolled properties which can be identified as such based on their Assessor "Use Code" classification alone. Conversely, all properties in the bottom two categories have use codes which make them "eligible" for rent control. As such, their rent controlled status can be identified by built-year alone. Only these properties from these two categories are included in the regression discontinuity models. By normalizing the data according to the number of observations in each category, average outcomes can be compared between the treatment and control groups to obtain a reliable first indication of what a more detailed model might reveal. The results show a mean difference of +0.77% in the rate of eviction notices for rent controlled addresses relative to their uncontrolled counterparts. Though small in magnitude, this difference corresponds to an eviction rate that is 2.4x higher for rent controlled addresses on an annual, per-unit basis.

	Rent Control Eligible			
	N		**Y**	
	Built before 1980		**Built before 1980**	
	N	*Y*	*N*	*Y*
total addresses	70,104	1,020,259	**14,132**	**330,381**
total units	331,191	1,445,753	**114,759**	**1,591,854**
avg. units per address	4.72	1.42	**8.12**	**4.82**
total evictions	243	4,993	**189**	**13,774**
prob. 1+ eviction	0.0021	0.0031	**0.0064**	**0.0194**
avg. evictions per address	0.0035	0.0049	**0.0134**	**0.0417**
avg. evictions per unit	0.0016	0.0039	**0.0055**	**0.0131**
mean difference	+0.0023		**+0.0077**	

Table 2.3: A comparison of eviction rates by rent control eligibility and built-year cut-off. Values in the third column correspond to properties that have a rent control-eligible use code but are not rent controlled because they were built after 1979. The fourth column (bolded) represents rent controlled properties. Only properties in the bottom two rows (n=344,513) are included in the RD analysis. The mean difference in bold italics (+0.0077) should roughly approximate the value of the local average treatment effect (LATE) estimated from the regression discontinuity model.

Graphical Analysis

Figure 2.3 shows the traditional RD plot, comparing average observed outcomes on the Y-axis across binned values of the running variable (built-year) on the X axis for the range defined by the bandwidth parameter h. The data points on either side of the built-year threshold are used to fit two linear models that highlight the discontinuous nature of this relationship. Visual inspection shows the size of the discontinuity in Figure 2.3 closely approximates not only the mean difference computed from Table 2.3 (0.77%), but also the size of treatment effects estimated by the RD models below (0.09% - 1.41%). The data presented in Figure 2.3 also clearly show that a significant and unique discontinuity exists at the threshold, adding substantive, visual evidence of a real causal effect. They also suggest that apart from the discontinuity itself, there exists almost no correlation between eviction rates and built-year. This lends additional credibility to a causal interpretation because it increases the likelihood of treatment assignment (i.e. rent control status) being the only channel through which systematic variation in eviction rates explains variation in the property built-year.

Model Results

Four RD models are fit according to the approach described above. Each of the four models estimates a positive treatment effect on rent control significant at the 0.002 level or below.

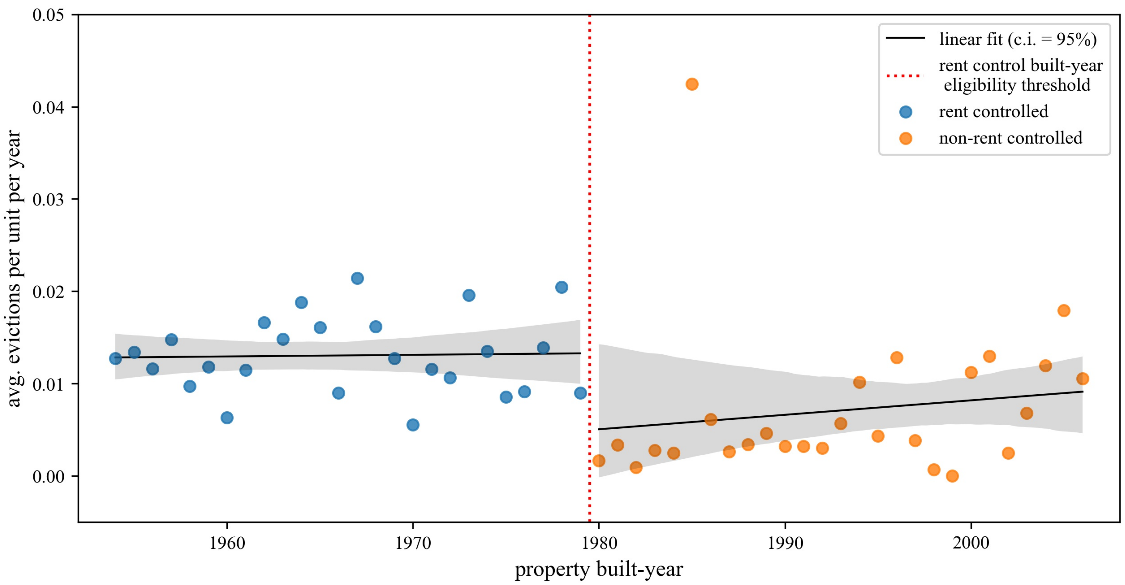


Figure 2.3: Eviction Rates by Property Built-Year (n=54,388). Only multi-family properties with rent-control eligible use-codes built after 1953 and before 2008 are included. Only eviction notices filed between 2007 and 2016 are counted.

After applying the bandwidth parameter and filtering out records with malformed or incomplete covariate data, a total of 53,493 observations are used to estimate these models. The results are summarized in Table 2.4. Using the specification described in Equation 1, Model 1 estimates an average treatment effect of 0.9% (p=0.002). This estimate is slightly larger than the mean difference-based estimate (0.77%), but still very much in line with the size of the discontinuity observed in Figure 2.3.

Model 2 adds two property-level characteristics derived from the assessor records: 1) an interaction term between the log of the assessed value and the log of the assessed square footage of the property; and 2) the log of the total number of units at the property. [33, 34] and others have demonstrated the relationship between The value-per-square-foot interaction term is found to positively correlate with eviction rates, while the coefficient on total units is negative, suggesting eviction rates are higher in more valuable properties with fewer units. This result makes intuitive sense, as more valuable properties can likely fetch higher market rate rents, and in buildings with fewer units the potential benefit of evicting one tenant represents a higher proportion of a property's total value to the landlord. Both of these variables are found to be significant at the level of < 0.001. Their inclusion in the model produces a larger treatment effect estimate of 1.36% and improves the significance level from 0.002 to < 0.001.

Model 3 adds Census tract-based demographic characteristics from the 2009-2013 5-year American Community Survey. A positive coefficient ($p \leq 0.001$) on the percent of occupied

	(1)	(2)	(3)	(4)
Rent Control	0.0097** (0.003)	0.0136*** (0.003)	0.0135*** (0.003)	0.0141*** (0.003)
Property Characteristics:	*N*	*Y*	*Y*	*Y*
Log(total value) / Log(total area)	– –	0.0010 (< 0.001)	0.0010 (< 0.001)	0.0010 (< 0.001)
Log(total units)	– –	-0.0204 (0.001)	-0.0210 (0.001)	-0.0216 (0.001)
Census Tract Demographics:	*N*	*N*	*Y*	*Y*
% Latino population	– –	– –	0.0371 (0.007)	0.0406 (0.015)
% housing units rental	– –	– –	0.0184 (0.005)	0.0199 (0.009)
Neighborhood Fixed Effects:	*N*	*N*	*N*	*Y*
R^2	< 0.001	0.008	0.008	0.01

Table 2.4: Coefficient on rent control for four RD models (n=53,493). Standard errors in parentheses. Significance codes indicate p-values as follows: '***' < 0.001 < '**' < 0.01 < '*' < 0.05.

units that are rented rather (non-owner occupied) suggests that eviction rates are higher in areas with lower rates of home ownership. Percent Latino also has a positive coefficient ($p \leq 0.001$), indicating a greater probability of eviction in areas with greater concentrations of Latinos. The addition of these sociodemographic characteristics did not significantly affect the estimated coefficient on rent control (1.35%) which remained significant at the $p < 0.001$ level. Alternative model specifications were tested using other Census-based covariates, including median household income and median move-in year. The latter term was included in order to test for the potential effect of tenancy duration, which in the context of vacancy decontrol means more heavily discounted rents and thus a greater incentive to evict. I found no evidence to suggest that such an effect exists, although it is possible that tract-level Census data is simply not granular enough to capture this relationship. It is also possible that landlords may be less willing to initiate economically motivated evictions against tenants with whom they have long-standing relationships.

Model 4 adds neighbourhood fixed effects to the equation according to the 72 assessor-designated neighbourhoods that appear in the assessor records. These terms are a blunt instrument designed to account for any other geographic variation not captured by the Census-based sociodemographic variables. The addition of neighbourhood fixed effects in a slightly higher estimated treatment effect of 1.41%, still significant at the < 0.001 level.

2.5 Discussion

Limitations of the Data

Before drawing any conclusions from this study it is worth re-examining the distinction between eviction filings and actual evictions. As mentioned earlier, filings are, in some respects, a better measure of displacement pressure than actual evictions due to the prevalence of informal evictions. Recent work on the phenomenon of "serial filing", however, has shown that eviction filings are often issued repeatedly against the same tenants by landlords seeking to achieve outcomes other than eviction, including the collection of rent and late payment fees [28, 34]. The literature suggests that serial filing is more common in larger buildings owned by larger-scale landlords, and higher rents, but also that its prevalence varies greatly from state to state, with the lowest rates observed "in areas where legal and regulatory barriers increase the cost of eviction" [35]. Although the models presented above do control for building size, rental price, and neighbourhood fixed effects, lack of tenant and landlord identifiers makes it difficult to determine the degree to which serial filing could be influencing the results; little is known about the relationship between serial filing and rent control. Future work should explore this issue by matching assessor records to property ownership data.

The other major weakness of the filings is that we do not observe disaggregate characteristics of tenants or landlords. This makes it difficult to control for landlord property owner types (e.g. "mom-and-pop" vs. property management companies). [33] shows that larger, more professional landlords are more likely to begin eviction proceedings against their tenants, but also that they are more likely to rent to tenants who go on to miss a rent payment.

Internal Validity

Common tests of Validity for Regression Discontinuity

One of the biggest limitations of the RD design is that it is applicable under only a very limited set of experimental conditions. Fortunately, because the identification strategy is so transparent, there are many well-documented methods available for investigating the internal validity of a given RD design. I will now briefly mention these, which in addition to the small standard errors and p-values reported in the previous section suggest that this study has avoided some of the biggest pitfalls of the RD method.

First, RD assumes that the treatment effect is the only discontinuity in an otherwise smooth functional form describing the relationship between the running variable and the outcome. Figure 2.3 demonstrates this fact visually. The strength of this assumption is further supported by the fact that Figure 2.3 shows a nearly flat response in the y-axis, and also that each of the four RD models fails to reject the null hypothesis that the coefficient on year-built is zero. Together, these results suggest that apart from treatment effect itself, the relationship between built-year and eviction rates is effectively random. Although RD does not require that the running variable be uncorrelated with the outcome, the fact that in this study the treatment assignment itself seems to be the only channel through which variation in eviction rates is related to the property built-year significantly strengthens the case for a causal interpretation.

Another common source of bias in RD designs occurs when treatment assignment can be manipulated by study participants. This occurs when subjects are aware of the cut-off threshold that determines treatment assignment (e.g. policy eligibility), and are able to nudge themselves past that threshold in the dimension of the running variable in order to qualify for treatment. Manipulation of this kind would introduce a structural imbalance in the sample population immediately above and below the threshold value, invalidating the experimental design. Although neither landlord nor tenant has the means to change the construction date as recorded by the county assessor's office, it is reasonable to assume that under normal circumstances both parties are aware of the rent control status of a property. It is therefore worth considering whether landlord and/or tenant selection bias might be able to explain our findings. We discuss this issue in greater detail below in the section entitled "Selection Bias".

Sensitivity to bandwidth selection is another common specification test used in RD analysis. Whereas overly narrow bandwidths might overestimate the significance of the variation observed at the discontinuity, too-wide of a bandwidth might bias the results by including observations that are irrelevant to behaviour at the discontinuity. Accordingly, RD estimates that are robust to this somewhat-arbitrarily chosen parameter are much more credible than those that are heavily dependent upon it [32]. The results of a bandwidth sensitivity analysis are presented in Figure 2.4, which shows that the treatment effect estimates of all four models are largely insensitive to variation in bandwidth.

In the analysis, all four models were re-estimated at 13 different evenly spaced bandwidths between 5 and 65. All four models consistently produced positive estimates of the LATE at all bandwidths, the majority of which fall within one standard error of the LATEs reported in Table 2.1. With the exception of those estimated with the smallest bandwidth (h=5), all LATEs are statistically significant at the $p <= 0.1$ level. Smaller bandwidths mean fewer observations are included in the model, likely contributing to the larger standard errors observed for the smaller bandwidths. In fact, below h=20 the algorithm is effectively discarding more than 50% of the observations. On the other hand, above h=20, all of the LATE estimates in all four models are significant at $p < 0.05$. There does appear to be a positive correlation between the size of the bandwidth and the size of the estimated LATE. The fact that the models with more observations (larger bandwidths) produce larger and more significant LATE estimates offers further evidence that the observed effect is real, and also that the estimates reported in Table 2.1 are perhaps conservative. Above h=27, however, data imbalance issues arise from 1) the inclusion of buildings with less than 10 years of assessor records; and 2) an asymmetrical RD design that occurs above h=37 because based on observations from the most recent tax assessor records included in the study (2017), there is a maximum possible distance of 37 years above the 1980 threshold, while no such lower limit exists.

In the context of an extremely contentious and highly visible policy like rent control, the

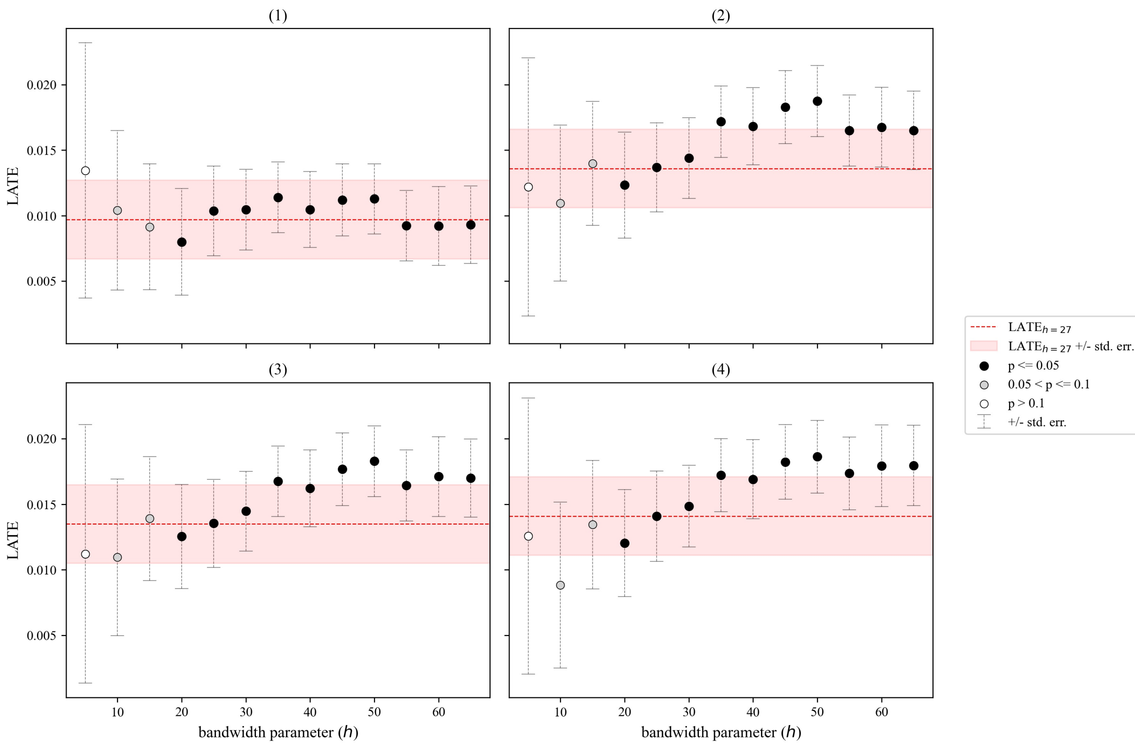


Figure 2.4: The red dashed line and the area highlighted in pink represent the estimated treatment effect and standard errors, respectively, of the results reported in Table 3.

importance of having a transparent identification strategy and easily interpreted experimental results cannot be overstated.

Selection Bias

Let us first consider selection bias due to landlord preference. While it is theoretically possible that higher rates of eviction for rent controlled tenants could reflect a preference of rent controlled landlords for more vulnerable, easily evicted tenants, this is not what the literature shows. Navarro [19], for example, cites evidence of rent controlled landlords in Cambridge, Massachusetts discriminating against minority, low-income households to compensate because rent control prevents them from implicitly discriminating based on price alone. Other studies agree, including [36], who interprets landlords preference for smaller households with older heads of household as evidence of landlords rationing controlled units to keep out occupants they believe could lead to quicker depreciation of their units, and [37], who along similar lines argues that rent control exacerbates economic segregation due to landlord preference for higher income tenants. Similarly, there exists evidence to suggest that due to vacancy decontrol landlords are actually able to target higher income tenants by leveraging the rent controlled status of a property to extract higher initial rents [12]. Indeed, one of the most often cited weaknesses of rent control policy is its inefficient targeting of benefits, and the resulting misallocation of rent controlled housing stock [15, 17, 19, 20, 38, 39]. If in fact rent controlled landlords were more likely to target more vulnerable tenants, it is unlikely that evidence of such misallocation would be so widespread. It is worth mentioning here that recent work by [33] does show that certain types of landlords, specifically larger-scale owners and property management companies, are more likely to rent to tenants who go on to miss a rent payment. While there is no evidence to suggest such a pattern exists among landlords of rent controlled properties, the results presented here would be strengthened by additional analyses to control for the different tenant selection tendencies between "mom-and-pop" landlords and property management firms.

But what about tenant self-selection? It is well documented that certain demographics are associated with higher eviction rates, including women, particularly Black women, and households with children [40–42]. Could the causal mechanism identified here perhaps explained by an overrepresentation of more "eviction-prone" tenants in rent controlled housing? The evidence here is somewhat mixed. Studies show that low income, minority households were overrepresented in rent controlled housing in New York City between 1965 and 1968 [38, 43]. In Cambridge, Massachusetts, however, [19], citing [44] and [45], found that the distribution of household income, elderly population, students, and household size were not meaningfully different between the controlled and uncontrolled housing sectors. More relevant perhaps is [26] which shows no statistical difference between the age and tenure duration of residents in controlled and uncontrolled properties in San Francisco (1993 – 2016), though their study is limited to buildings with fewer than four units. Unfortunately, the eviction filing data used to conduct the analysis presented here do not contain disaggregate demographic characteristics. Instead, Table 5 presents a supplemental comparative analysis of controlled and uncontrolled renters using the Public Use Microdata Sample (PUMS) data from the 2012 American Community Survey. In general, the results suggest that the demographic groups which are typically associated with higher eviction rates are not overrepresented in rent controlled housing in San Francisco. In fact, the data show that black female tenants are significantly less likely to be found living in controlled housing than uncontrolled, while White and non-Hispanic renters are significantly overrepresented. Similarly, rent controlled tenants have fewer children, especially young children. Most importantly, however, Table 5 shows that while rent controlled households do report lower incomes, they also benefit from lower rents, and so the mean gross rent per household income is nearly the same for both groups. This finding suggests that rent control policy in San Francisco has been at least modestly successful in targeting its beneficiaries.

		All Renters				Renters in Multi-unit Buildings Built 1950-2007			
		Rent Control				Rent Control			
		N		Y		N		Y	
		Value	*S.E.*	*Value*	*S.E.*	*Value*	*S.E.*	*Value*	*S.E.*
Person	% Asian	28.4	2.3	24.0	1.3	30.9	3.0	26.1	2.3
	% Black	9.2	1.4	4.4	0.5	9.9	1.8	5.7	1.0
	% Other	13.3	2.1	8.0	1.2	12.8	3.0	7.2	2.0
	% Two + Races	5.0	0.9	5.2	0.5	5.6	1.2	6.1	1.0
	% White	44.1	3.0	58.4	1.5	40.8	3.5	54.9	2.9
	% Hispanic	24.8	2.2	15.9	1.0	22.3	2.9	16.3	2.2
	% Non-hispanic	75.2	2.2	84.1	1.0	77.7	2.9	83.7	2.2
	% Female	48.0	1.7	47.8	0.9	48.7	2.4	49.9	1.8
	% Male	52.0	1.7	52.2	0.9	51.3	2.4	50.1	1.8
	% Asian, Female	14.8	1.5	13.5	0.9	17.2	1.9	14.7	1.7
	% Asian, Male	13.6	1.2	10.5	0.7	13.6	1.6	11.3	1.3
	% Black, Female	5.2	1.0	2.0	0.3	5.6	1.2	3.3	0.8
	% Black, Male	4.0	0.8	2.4	0.3	4.3	1.1	2.3	0.5
	% Other, Female	6.1	1.1	3.7	0.6	6.2	1.6	3.2	1.0
	% Other, Male	7.2	1.4	4.3	0.8	6.6	1.9	4.0	1.2
	% Two+ Races, Female	2.7	0.7	2.9	0.4	2.8	0.7	3.5	0.7
	% Two+ Races, Male	2.4	0.5	2.4	0.3	2.8	0.7	2.6	0.6
	% White, Female	19.2	1.7	25.7	0.9	16.9	2.1	25.2	2.0
	% White, Male	24.9	1.8	32.7	1.1	23.9	2.3	29.8	2.2
	% Non-hispanic, Female	36.9	1.5	40.9	1.0	38.8	2.2	43.1	2.1
	% Non-hispanic, Male	38.3	1.8	43.3	0.9	38.9	2.5	40.7	2.0
	% Hispanic, Female	11.2	1.2	6.9	0.6	9.9	1.6	6.9	1.2
	% Hispanic, Male	13.6	1.6	9.0	0.7	12.4	2.1	9.4	1.4
	% Disabled	11.0	1.8	11.1	0.8	11.8	1.5	13.2	1.6
	% Not Disabled	89.0	1.8	88.9	0.8	88.2	1.5	86.8	1.6
	% No Public Assistance Income	98.4	0.4	98.0	0.4	98.7	0.5	99.0	0.4

	% Public Assistance Income > $0	1.6	0.4	2.0	0.4	1.3	0.5	1.0	0.4
Household	% Household SNAP Benefits	8.1	1.4	5.6	0.7	7.7	1.7	4.6	0.9
	% No Household SNAP Benefits	91.9	1.4	94.4	0.7	92.3	1.7	95.4	0.9
	Mean Monthly Rent ($)	1,578	50	1,452	23	1,519	72	1,418	42
	Mean Gross Monthly Rent ($)	1,684	52	1,507	23	1,603	73	1,469	43
	Mean Household Income ($)	88,114	5,174	79,872	3,209	83,953	6,111	71,167	5,466
	Mean Gross Monthly Rent per Household Income ($)	38	2	38	1	36	2	38	2
	Mean Age of Head	45.8	0.8	44.9	0.5	45.6	1.1	46.9	1.0
	% No children	76.9	1.9	90.5	0.9	81.1	2.5	90.2	1.6
	% Only children under 6	7.6	1.4	2.3	0.4	5.6	1.5	2.4	0.8
	% Only children 6 to 17	11.0	1.7	5.1	0.7	9.2	2.3	5.0	1.3
	% Children under 6 and 6+	4.5	1.1	2.2	0.5	4.1	1.4	2.4	0.8
	Mean Number of Own Children	0.34	0.04	0.13	0.01	0.26	0.04	0.14	0.03

Table 2.5: Comparison of San Francisco Renter Demographics by Treatment Status. Source: 2012 American Community Survey Public Use Microdata Sample (PUMS)

Policy Implications (External Validity)

To the best of this author's knowledge, the findings presented here constitute the first rigorous estimate of the causal effect of rent control on eviction rates in the peer-reviewed literature. That being said, it is important to understand the context in which these findings are most relevant.

The treatment effect estimated via RD is a local average treatment effect (LATE), rather than the more general average treatment effect (ATE) associated with a true RCT design. In other words, the RD estimate is not guaranteed to be unbiased for observations outside of the subpopulation where the treatment effect is measured ($Y_i = c$). In the present study that means one must be careful in extending the validity of the model results to properties built in the 1920s, for example, which depends on the degree to which a homogeneous treatment effect can be assumed. However, given the seemingly stochastic nature of the relationship between built-year and eviction rates in this study, this does not seem like an unreasonable assumption to make.

Additionally, recent research indicates that certain precautions can reduce the likelihood that RD estimates are biased, many of which – including large sample sizes, the use of non-parametric tests and estimators, and careful bandwidth selection – were implemented in this study [31, 46]. It is also worth noting that regardless of whether or not the results are relevant to very old or very new buildings in San Francisco, the estimated treatment effect is clearly only valid for the city of San Francisco. That being said, the implications of these findings are highly relevant to current and future policy in San Francisco and beyond.

In 2020 California voters had the chance to repeal the 1995 law restricting municipalities from enacting rent control on residential units constructed after February 1995[11]. Had it passed, Proposition 21 would have been replaced the 1995 restriction with a rolling 15-year window, allowing San Francisco legislators to expand rent control eligibility to buildings built before 2006 by 2021, 2007 by 2022, and so on. Even though Prop 21 was ultimately rejected by voters, the pressure to expand rent control in California does not appear to be dissipating, as indicated most recently by the passage of Assembly Bill No. 1482[12], the Tenant Protection Act of 2019. If and when the time comes for San Francisco to reassess its rent control eligibility requirements, the results presented here will be directly applicable for both policymakers, residents, and property owners wishing to evaluate the impact of such a policy change.

Increased housing stability is one area in which experts typically agree that rent control offers real benefits [15, 26]. In fact, a recent review of the rent control literature published by the USC Equity Research Institute (ERI) found that "nearly every academic study finds that rent stabilization [...] increases housing stability for rent-stabilized residents" [6]. However, the findings presented here show that the effects of rent control on housing stability are more heterogeneous than the current academic literature suggests. While the ERI review

[11]See California Civil Code 1954.50-1954.535

[12]Full text available at `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1482`

offers strong evidence in favour of re-evaluating rent control as an anti-displacement measure rather than anti-gentrification, the reality is that even under this more appropriately defined rubric its effectiveness in San Francisco is unclear.

As we have seen, [26] were able to estimate a net benefit of $393 million per year for incumbent rent controlled tenants in San Francisco, but failed to account for the cost of rent control-induced evictions. And while it may seem unlikely that a 1-2% increase in the rate of evictions would do much to offset such a large benefit, the severity of the impact of an eviction must not be underestimated. Recent data from the City of San Francisco itself suggests that eviction is the third-leading cause of homelessness there, representing 13% of survey respondents (n=1,039), up from just 4% in 2011 [47]. Additionally, recent work by Desmond and others has demonstrated the deleterious, sometimes trans-generational effects of eviction on health outcomes, homelessness, and job retention [40–42]. Long term effects like these make it a very difficult and fraught task to estimate the true cost of rent control-induced eviction to rent controlled tenants.

Whatever those costs may be, it would be a mistake to interpret them as a failure that is inherent to Rent Control as a policy. Instead, the causal effect of rent control on eviction rates as measured in this study represents a failure of San Francisco's Rent Ordinance to adequately protect tenants from the incentive to evict that exists under vacancy decontrol. Fortunately, a solution to this problem exists that does not require abolishing vacancy decontrol altogether, which would almost certainly be politically impossible in San Francisco today. Instead, the city can enact targeted eviction protections aimed at decoupling the eviction process from rent control avoidance. For example, breach-of-lease type evictions should be treated as a disqualifying event, preventing a property owner from taking advantage of vacancy decontrol at the start of their next lease. If the goal of an at-fault eviction is not, in fact, to take advantage of vacancy decontrol, but to remove a problem tenant, then the evicting landlord should be willing to continue to rent the unit in question at the same rate. San Francisco policymakers have the ability to close these eviction-based rent control loopholes, and have previously made strides in this direction with respect to no-fault evictions in particular [48]. With local and state-wide COVID-19 eviction moratoriums expiring soon, it is now more important than ever that City officials re-examine the role that evictions play in undermining the success of rent stabilization.

Chapter 3

Numerical Analysis of Error due to Sampling of Alternatives in Logit-Based Demand Forecasting Models with Massive Choice Sets

3.1 Introduction

Use of the multinomial logit (MNL) discrete choice model is widespread in transportation and land use demand modeling, both in the research literature and in applied settings. As the state-of-the-art continues to evolve away from aggregate forecasting and towards activity-based models and microsimulation, MNL models are being asked to accommodate increasingly large choice sets [49, 50]. Different strategies exist to address the challenge of estimating models with very large choice sets, but perhaps none is as commonly employed as sampling of alternatives. The effect of sampling of alternatives on model estimation has received considerable attention in the scientific literature [51–53]. Yet comparatively little quantitative research exists which examines the issues that arise when the same sampling strategies are applied in the forecasting or simulation phase of the modeling process.

Despite a lack of empirical evidence, it is often assumed that sampling strategies which improve parameter estimates must also improve the accuracy of demand forecasts as well. This paper tests the validity of this assumption by defining a new empirical measure of prediction error and using it to evaluate outcomes under different modeling scenarios and sampling strategies. The mathematical definition of this prediction error metric is itself a valuable contribution to the scientific literature on a topic which is rarely studied in part because of how challenging it is to measuring model error without the benefit of observed choices, as is almost always the case for forecasting. Just as important, however, are the experimental results which can guide practitioners as they must navigate the tradeoff between sample rate and aggregation of alternatives due to limited computational resources.

Of particular interest is quantifying the impact of sampling of alternatives for logit-based forecasts at the scale of microsimulation, in which choice sets can number into the millions of alternatives.

Literature Review

The use of sampling of alternatives in logit models with large choices dates back to McFadden [54] who showed that one can obtain consistent MNL estimates of parameter vector β^* using only a sample of non-chosen alternatives and adding an alternative-specific correction term to the standard logit choice model:

$$p_n(i|D_n) = \frac{e^{\mu\beta^\top X_{in} + \ln\pi_n(D_n|i)}}{\sum_{j\in D_n} e^{\mu\beta^\top X_{jn} + \ln\pi_n(D_n|j)}}, \quad \forall \quad D_n \subset C \tag{3.1}$$

The term $\log\pi_n(D_n|i)$ is called the correction factor and represents the likelihood of constructing a choice set conditional on chooser selecting alternative. Moreover, under simple random sampling (SRS) it follows that

$$\pi_n(D_n|i) = \pi_n(D_n|j) \quad \forall \quad i,j \in D_n \tag{3.2}$$

which McFadden calls the "uniform conditioning property" of the logit model. This property allows the correction factors in the numerator and denominator of Equation 3.1 to cancel out, which means the original, simplified logit form can be used instead:

$$p_n(i|D_n) = \frac{e^{\mu\beta^\top X_{in}}}{\sum_{j\in D_n} e^{\mu\beta^\top X_{jn}}} \tag{3.3}$$

The simplicity of this model, and the fact that $\hat{\beta}$ has a closed form solution under maximum likelihood estimation, has led to its proliferation among practitioners of travel and land use demand modeling.

Consistency of parameter estimates does not imply that they are also efficient, which, as it relates to sampling of alternatives, depends on the size and composition of the sample. As Train puts it, "Comparing a person's chosen alternative with a group of undesirable alternatives provides little information about the reason for a person's choice" [55, p70]. As such, considerable time and effort have been spent attempting to quantify the relationship between sampling of alternatives and model error [51–53, 56]. Researchers have also found success developing and applying alternative sampling strategies which are able to improve parameter efficiency by oversampling alternatives deemed most relevant to a particular choice context [51, 57–60].

Although the effect of sampling of alternatives in discrete choice models seems to be well understood, the problem space is much murkier for practitioners who, having estimated a model, must actually use it to perform scenario analyses or to generate regional forecasts of demand. The key distinguishing factor is that in the context of prediction there are no "observed" choices. This fact has two important consequences which serve as the motivation for this study. The first is that sampling of alternatives may produce choice sets composed entirely of unattractive or "uninformative" alternatives, particularly in the case of SRS. Moreover, the likelihood of generating such a sample increases with the size of the universe of alternatives [51, 61]. The issue is therefore likely to be of increasing concern as the trend towards microsimulation continues. Despite this, it is found that many such models in practice rely on random sampling of alternatives [49, 62–64].

The second consequence is that there is no straightforward way to assess performance in the forecasting phase of the modeling lifecycle. Intuitively, there can be no "ground truth" for a model designed to predict outcomes under scenarios which have not yet come to pass. This is especially true when the choices being modeled are those of a synthetic population, as is often the case in microsimulation and activity-based models. As a result, the study of forecast error in discrete choice models has rarely been treated separately from estimation. Two exceptions are worth mentioning, even though neither is explicitly concerned with sampling of alternatives. The first is [65] which studies the effect of aggregation of alternatives on forecasting error for location choice models. In that study, the authors explicitly describe the why variation in parameter estimates cannot reliably measure predictive performance in the forecasting phase. The principal reason cited is that error in the individual parameter estimates can balance itself out across parameters, and across alternatives, too. Furthermore, if in practice a modeler is primarily concerned with the spatial distribution of demand, then it is possible for this error to balance itself out across choosers as well. In this sense, the net effect could therefore be significantly less than the estimation error would indicate.

The second study is by Guevara & Ben-Akiva [66] who investigate methods for reducing forecasting error due to endogeneity bias. Of particular relevance is a passage in that study in which the authors observe that "predictions at a disaggregated level are almost always meaningless," and advocate instead for microsimulation performance to be assessed at an aggregate level. Although aggregating the results of a disaggregate model may, at first glance, seem self-defeating, it is not. Since the inception of microsimulation for policy analysis [67], practitioners have stressed that its value over aggregate modeling lies in its ability to more accurately predict the distribution of demand across a population of individuals, not chooser-level outcomes [68]. This is in large part due to the highly stochastic nature of microsimulation, which, far from being a shortcoming of the approach, is actually one of the features that allows it to incorporate the effects of population heterogeneity, emergent behaviors, and other low probability outcomes. In practice, this stochastic variation (also called Monte Carlo error [69]) is typically handled by running the model(s) multiple times with different seeds supplied to the random number generators, and taking the average of the results [50]. It follows, therefore, that any metric used to quantify forecasting error must focus on aggregate demand for alternatives across a population, and must utilize the entire distribution of choice probabilities for each chooser rather than just that of the highest utility alternative or the alternative selected via Monte Carlo simulation.

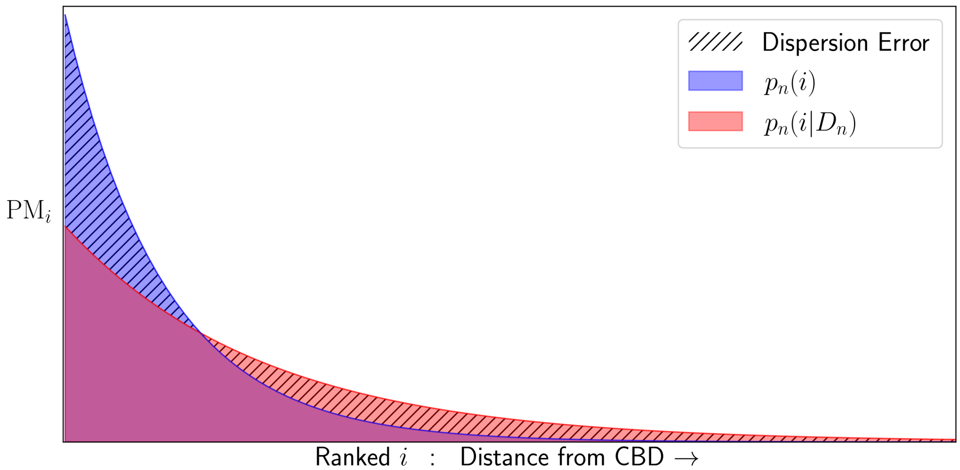


Figure 3.1: Graphical representation of dispersion error in a location choice model for an idealized monocentric city with random sampling of alternatives for a single chooser. The area of both the blue and red distributions is equal to one since each is a probability density function. The theoretical upper limit of dispersion error is therefore two, which would occur if the distributions had zero overlap.

Key Questions and Contributions of this Study

In addition to the definition of an error metric itself, the primary contributions of this study are made by using the metric as means for investigating the following three research questions: 1) How does error due to sampling of alternatives change as the size of the universe of alternatives increases by orders of magnitude? 2) Does error due to random sampling of alternatives balance itself out in terms of aggregate demand for alternatives across choosers? 3) Can forecast error be reduced by replacing simple random sampling is replaced with a strategically weighted sampling approach?

In regards to the first question, I hypothesize that error will be largest at the most disaggregate scales, where choice sets can number in the millions of alternatives. This hypothesis is informed by previous work that has shown that it is easier to sample "non-competitive alternatives" when choice sets are largest [51]. In the estimation phase, this leads to greater parameter error. In forecasting, however, I expect inefficient choice sets to result in over-assignment of choosers to low-utility alternatives, and therefore an over-dispersion or diffusion of aggregate probability across alternatives relative to the true distribution one would obtain from a 100% sample. In a model of residential location choice, for example, this might manifest as under-predicted population densities for the most attractive locations. Figure 3.1 illustrates an idealized version of this phenomenon which we term "dispersion error".

It is more difficult to speculate about the second question since it has no analogue in the literature on model estimation. Research has shown that even with random sampling of alternatives, discrete choice microsimulations can achieve greater predictive accuracy than traditional aggregate models [70]. But no study until now has managed to isolate the effect of sampling of alternatives on forecasting error in a way that can guide practitioners who must navigate the trade-off between computation time and sample size. If in fact the sample error that is associated with microsimulation is balanced out (i.e. mean zero) when aggregating demand across choosers, then the marginal benefit of using larger and larger sample sizes should decrease past a certain point. By using an aggregate metric like dispersion error to compare the performance of a single disaggregate model at multiple sample rates, this study is able to test the hypothesis and determine if such a threshold exists.

The expected outcome of the third research question follows from the first, in that importance sampling is expected to increase the likelihood of constructing choice sets with high-utility alternatives, thereby reducing dispersion error. I also hypothesize that the benefit of importance sampling will be greatest for the largest choice set scenarios, since I expect that strategic sampling can ameliorate the effect of the size of the universe of alternatives.

The rest of the paper is structured as follows. I first provide a mathematical definition of dispersion error in multinomial logit predictions with sampling of alternatives. I next describe the experimental design, including the generation of synthetic data from MNL models of my own design, and the recent advances in GPU-based computing which have made this study possible. The use of synthetic data means no model estimation is required, and the effect of sampling of alternatives on parameter efficiency can be ignored. Instead, dispersion error is measured by repeatedly simulating choice probabilities using different sample sizes between 2 and $J-1$, where J is the number of alternatives in the universal choice set, and comparing the distribution of aggregate choice probabilities to that obtained from a 100% sample. I also define the boundary conditions of the experiment through the use of three experimental controls: 1) the size of J; 2) the variance of the model error as defined in the data generation process; and 3) sample rate. This procedure is then repeated using an important sampling strategy to explore how alternative techniques to random sampling of alternatives can be used to reduce dispersion error. Lastly, I contextualize these findings by estimating a location choice model using real data taken from the California Household Travel Survey (CHTS). This serves the dual purpose of both validating the experimental results, and demonstrating how they can be used to guide modeling decisions in practice.

3.2 Methodology

Definition of Forecast Error Metric

Let us first define the probability of chooser n selecting alternative i from the universal choice set C with J elements using the standard multinomial logit model formulation

$$p_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C} e^{\mu V_{jn}}} = \frac{e^{\mu \beta^\top X_{in}}}{\sum_{j \in C} e^{\mu \beta^\top X_{jn}}}, \quad i = 1, \ \ldots, \ J \tag{3.4}$$

where V_{in} represents the structural component the utility function

$$U_{in} = V_{in} + \epsilon_{in}, \quad \epsilon \sim \text{Gumbel}(0, \mu) \tag{3.5}$$

and is itself defined as a linear combination of parameter vector X_{in} with a vector of fixed coefficients β

$$V_{in} = \beta^\top X_{in} \tag{3.6}$$

I define the aggregate demand for alternative i across the population of choosers N as the *probability massing*:

$$\text{PM}_i = \sum_n p_n(i), \quad 0 \leq \text{PM}_i \leq N \tag{3.7}$$

An aggregate metric like this incorporates the total choice probability for each alternative regardless of whether a chooser would be predicted to select it under Monte Carlo simulation.

From this perspective, it follows that the best outcome one can hope to achieve from simulation with sampling of alternatives is to replicate the empirical distribution of aggregate probability obtained when choosers have access to the full universe of alternatives. Therefore, I define the *probability massing error* (PME) due to sampling of alternatives for alternative i as the difference between the best-case-scenario probability massing (no sampling) and that obtained from the conditional logit probabilities:

$$\begin{gathered} \text{PM Error}_i = \sum_n p_n(i|D_n) - \sum_n p_n(i) \\ \text{where} \quad p_n(i|D_n) = \begin{cases} \frac{e^{\mu V_{in}}}{\sum_{j \in D_n} e^{\mu V_{jn}}} & \text{if} \quad i \in D_n \\ 0 & \text{otherwise} \end{cases}, \quad D_n \subset C \end{gathered} \tag{3.8}$$

An additional benefit of aggregation is that by comparing demand across an entire population, this metric avoids the issue of scale that arises at the chooser level, where

$$p_n(i|D_n) > p_n(i) \quad \forall \quad i \in D_n \tag{3.9}$$

for any given chooser due only to the fact that that denominator in the conditional logit is necessarily smaller. This fact should be even more apparent when I sum across alternatives below, since the total probability mass of a population $\sum_i^J \sum_n p_n(i)$ must always equal N, regardless of the sample size.

Equation 3.10 defines *dispersion error* (DE) as the sum of PME$_i$'s over the universe of alternatives, normalized by twice the total probability mass (N). The absolute value of the inner term is taken to prevent positive and negative PME values of different alternatives from canceling each other out:

$$\mathrm{DE} = \frac{1}{2N} \sum_{i \in C} \left| \sum_n p_n(i|D_n) - p_n(i) \right|, \quad \text{s.t.} \quad 0 \leq \mathrm{DE} \leq 1 \tag{3.10}$$

I call this metric "dispersion error" because it measures the degree to which the distribution of aggregate demand for a set of choosers is under- or over-dispersed relative to the "true" distribution that is observed when no sampling is performed (Figure 3.1). Equation 3.11 follows from the assumption that when sampling without replacement, $D_n = C$ for any D_n that is constructed using a 100% sample.

$$\lim_{\widetilde{J}_n \to J} p_n(i|D_n) = p_n(i) \tag{3.11}$$

By substituting Equation 3.11 into Equation 3.10, it can be shown that that dispersion error should trend towards zero as the size of the sample ($\widetilde{J}_n$) approaches J:

$$\lim_{\widetilde{J}_n \to J} \mathrm{DE} = 0 \tag{3.12}$$

Thus by varying the size of D_n and measuring the rate at which dispersion error approaches zero, I am able to quantify the marginal benefit of using larger and larger sample rates. I hypothesize that the magnitude of this benefit will depend on both the size of the universe of alternatives (J) and the precision (error) of the model itself.

Data Generation

The data generation procedure used in this study builds on that of [52]. I modify both the model specification and the scale of the simulated data in order to more accurately approximate the form of a location choice model that one might encounter in a modern microsimulation platform like UrbanSim [71]. I describe the changes to the model specification here, and address the scale of the data in the context of the experiment design in the section that follows.

The model is structured as a linear-in-parameters MNL specification with five independent variables and the coefficients of each fixed to one. Three of the five variables are generated according to the procedure described in [52]: drawing values from the standard univariate normal distribution using a mean value of 1.0 for the first half of the alternatives and 0.5 for the second. For the fourth variable I introduce an interaction term between an alternative-specific attribute and a chooser-level attribute drawn Lognormal$(1, 0.5)$ and

Lognormal$(0, 0.5)$, respectively. Interaction terms are a common feature of location choice models because they capture chooser preference heterogeneity that arises from real-world relationships like rent-to-income ratio or home-to-work distance. The fifth independent variable is sampled from the distribution Lognormal(0, 1) and is meant to approximate a spatial measure like distance to a central business district.

Experiment Design

A series of simulations are performed in which I compute choice probabilities are generated for synthetic datasets and dispersion error while approximating different modeling scenarios along three experimental controls: 1) sample rate ($\frac{\widetilde{J}_n}{J}$); 2) the scale parameter μ of the logit model; and 3) the scale of the data (N and J). A total of 1,750 scenarios are modeled using 10 sample rates, seven scale parameters, and twenty-five dataset sizes. I aggregate the results of each scenario over 10 runs for a total of 17,500 simulations.

For each chooser n in each dataset I construct ten choice sets. The first of these is simply C, the universe of alternatives. The other nine choice sets $D_{n1} \cdots D_{n9}$ are assembled through simple random sampling (SRS) without replacement from C using nine equally spaced sample rates on the interval from 10 to 90%. To construct each D_{ns} of size $\widetilde{J}_{ns}$, I first partition the alternatives in C into ten random subsets with $\frac{J}{10}$ elements each and then, setting D_{n1} equal to S_{n1}, reincorporate the remaining subsets one at a time:

$$D_{ns} = D_{ns-1} \cup S_{ns} \quad s = \{1, \ldots, 9\} \tag{3.13}$$

such that $\widetilde{J}_{ns} = \frac{J}{10} \times s$. Although this approach means that $D_{n1} \cdots D_{n9}$ are not drawn independently for a given chooser, it reduces the number of times I need to draw random samples without replacement by a factor of 10. Given that the compute time of the sampling algorithm I use is $\mathcal{O}(J + J \log J)$, this was a necessary compromise to ensure I could simulate over the largest choice sets ($J = 2 \times 10^6$) in a reasonable amount of time [72, 73].

The scale parameter (μ) of the logit equation is the second critical axis of inquiry in this study. Since μ is confounded with the vector of coefficients β and all ϵ_{in}'s are assumed to be distributed i.i.d., common practice is to simply set $\mu = 1$ [74]. In this case, however, we know a priori that each of the coefficients is 1. This creates a unique opportunity to use as an experimental control for model precision. From Equation 3.1, it follows that as the model becomes completely random with all alternatives having equal probability:

$$\lim_{\mu \to 0} p_n(i) = \frac{1}{J} \quad \forall \quad i = \{1, \ldots, J\} \tag{3.14}$$

Intuitively, sampling of alternatives should have no material impact on a truly random model (i.e. a model with no explanatory power) and thus I expect DE to trend toward zero in this case. Likewise, as $\mu \to \infty$, the model becomes completely deterministic, which should produce the highest values of DE. Thus, to investigate the effect of model precision on DE, I repeat each simulation for seven values of $\mu = \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75\}$.

This is analogous to the approaches of [51, 75] to assess the effect of noisy data on parameter estimation.

Lastly, I generate 25 synthetic populations of N choosers and J alternatives representing all possible combinations of $N = \{7.5 \times 10^2, \dots, 7.5 \times 10^6\}$ and $J = \{2 \times 10^2, \dots, 2 \times 10^6\}$. This makes it possible to assess the change in DE as the scale of the model approaches that of microsimulation in terms of number of choosers and alternatives. All 25 datasets are regenerated for every value of μ for each of the 10 runs for a grand total of 1750 distinct datasets.

Importance Sampling

The final component of the analysis is an investigation of the benefits of alternatives sampling strategies over SRS in terms of dispersion error. In theory, a strategic sampling strategy like importance sampling should reduce the sample rate required to obtain choice sets with relevant (i.e. high utility) alternatives. Since the experiment described above necessarily involves computing the true, unconditional choice probabilities $p_n(i)$ for all choosers and alternatives, I conduct a second set of simulation runs in which each choice set D_{ns} is constructed using the unconditional choice probabilities as sample weights. The use of importance sampling means choice sets are no longer constructed from truly random draws. As such, the uniform conditioning property (Equation 3.2) does not hold, and the more generalized form of the conditional probability $p_n(i|D_n)$ defined in Equation 3.1 must be used instead [74, p268].

Computation

The entirety of each simulation, from sampling of alternatives to the calculation of choice probabilities, is performed on a single NVIDIA GeForce RTX 3090 graphics processing unit (GPU). I leverage the open source JAX library to perform extremely fast linear algebra operations on the GPU [72]. Although the duration of each simulation is still quite long for the largest datasets (Figure 3.2), initial tests indicated that run times with a CPU-based approach would have been prohibitively slow. It is possible that massive parallelization across a sufficient number of CPU's could achieve run times comparable to the GPU (choice probabilities are computed independently across choosers) but such an approach was determined to be cost prohibitive given the number of processors required and the current cost of on-demand cloud computing resources.

The use of a single GPU does impose strict limits on the amount of data that can be processed at once. I experimentally determined that the full table of $N \times J$ probabilities would exceed the 24GB memory capacity of the GPU for scenarios where $N \times J \geq 1.5 \times 10^9$. In scenarios above this threshold, I compute choice probabilities in batches of choosers, with larger datasets requiring more batches, and fewer choosers per batch as J increases. The inefficiency of batched processing is reflected in Figure 3.2, which shows that the average and per-chooser runtimes are essentially invariant to the number of choosers until the dataset

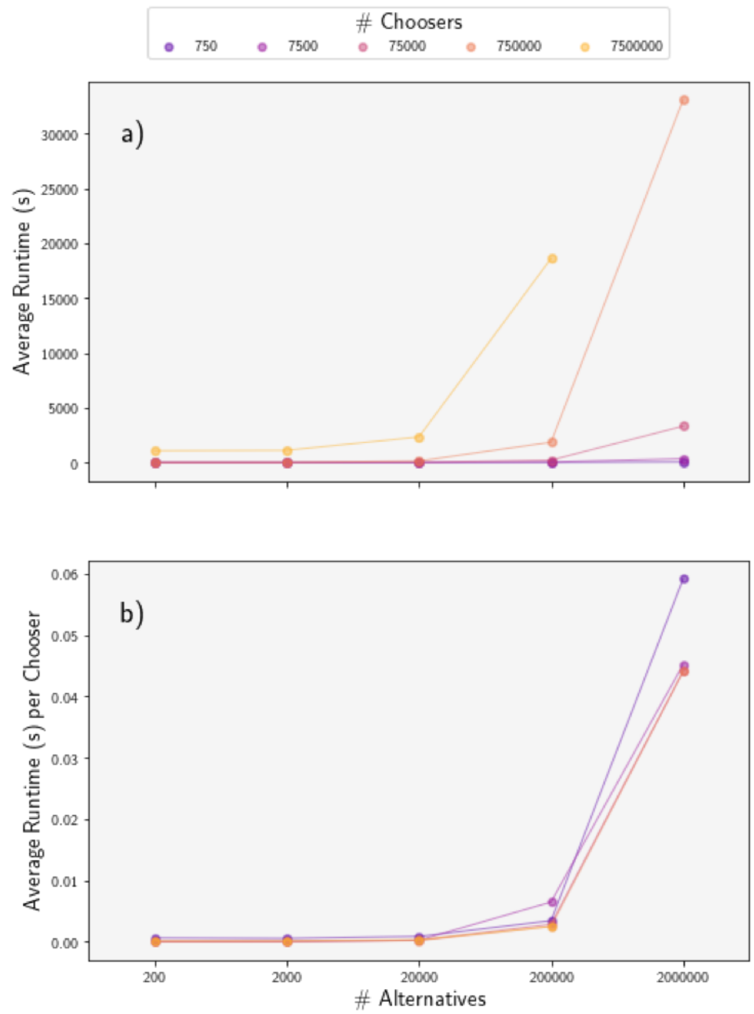


Figure 3.2: Average runtime for each of the 1750 datasets by population size (N and J). A single runtime represents the total duration to simulate probabilities and compute dispersion error at all nine sample rates.

sizes exceed the batching threshold. This means that although the analysis is quite slow to run for the largest datasets, runtimes could be drastically reduced by processing batches simultaneously across a cluster of GPUs rather than in a queue on a single processor. However, such optimization is not the intent of this research.

3.3 Results

Experimental Results

In general I find that dispersion error is highly correlated with lower sample rates, model precision (larger μ), and smaller choice sets (J). Figures 3.3, 3.4, and 3.5 highlight these relationships across each of the 25 dataset sizes. For almost every scenario, the highest overall error is observed for the highest value of μ. The effect of sample rate aligns with expectations, with dispersion error trending towards zero as sample rate approaches 100%. In nearly every case we observe a trend of diminishing benefits for marginal increases in sample rate, with the greatest reduction in dispersion error observed on the interval from 10 to 20%.

Lower model precision is consistently associated with lower levels of dispersion error, lending support to the initial hypothesis about the attenuating effect of model noise on error due to sampling of alternatives. Notably, in many cases the effect of model noise is found to have a greater impact on dispersion error than sample rate. For example, Figure 3b shows that that reducing the precision by a factor of two ($\mu = 1$ to $\mu = .5$) has a greater impact than tripling the sample rate (10 to 30%). Obviously in practice a modeler does not have the ability to exogenously define or even observe the scale parameter separately from the parameter coefficients, but this finding is nevertheless significant. In particular, it suggests that in low-precision models, there is less to be gained from using larger and larger sample rates.

The size of the universe of alternatives (J) is also observed to have a significant effect on dispersion error, albeit the opposite effect of what was anticipated, with the highest levels of error observed for models with the fewest alternatives, and comparatively little dispersion error observed for scenarios with the largest choice sets. For example, at $J = 2e6$, the range of dispersion error observed for a 10% sample rate across all values of μ is reduced to that of a 70-80% sample for the equivalent $J = 200$ scenario. Although the cause of this trend is not immediately clear, it does align with results reported by both [51] and [52], though neither paper makes explicit mention of it. This finding is discussed in greater detail in the discussion section below.

Importance Sampling

The results of the importance sample simulation runs, summarized in Figure 3.6, show that the benefit of importance sampling over SRS is limited only to those models with the highest levels of precision (i.e. very low variance of error). More surprisingly, however, is

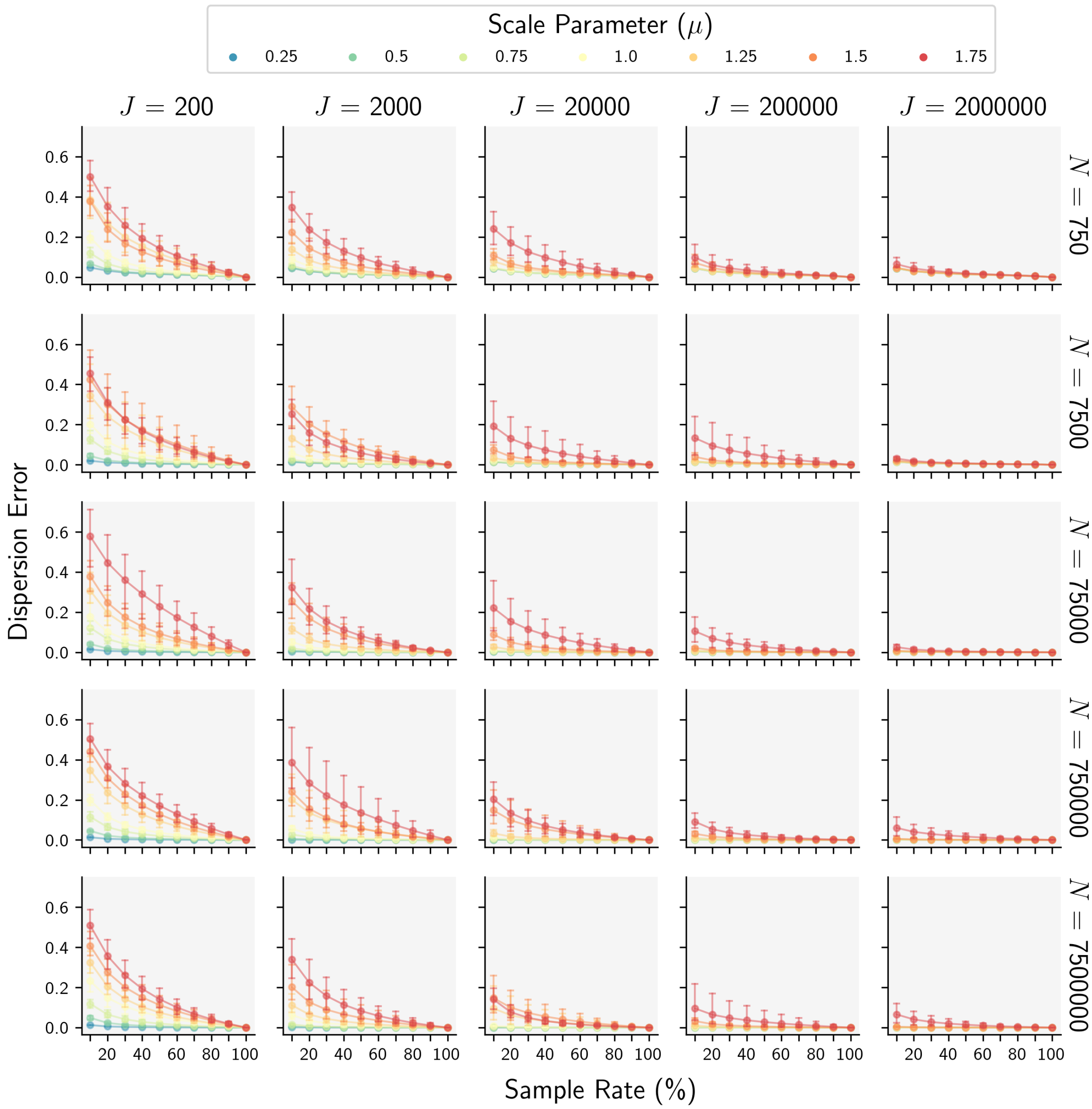


Figure 3.3: Mean dispersion error across 10 runs for each of the 1,750 model scenarios. Curves highlight the rate at which dispersion error decreases with larger sample rates.

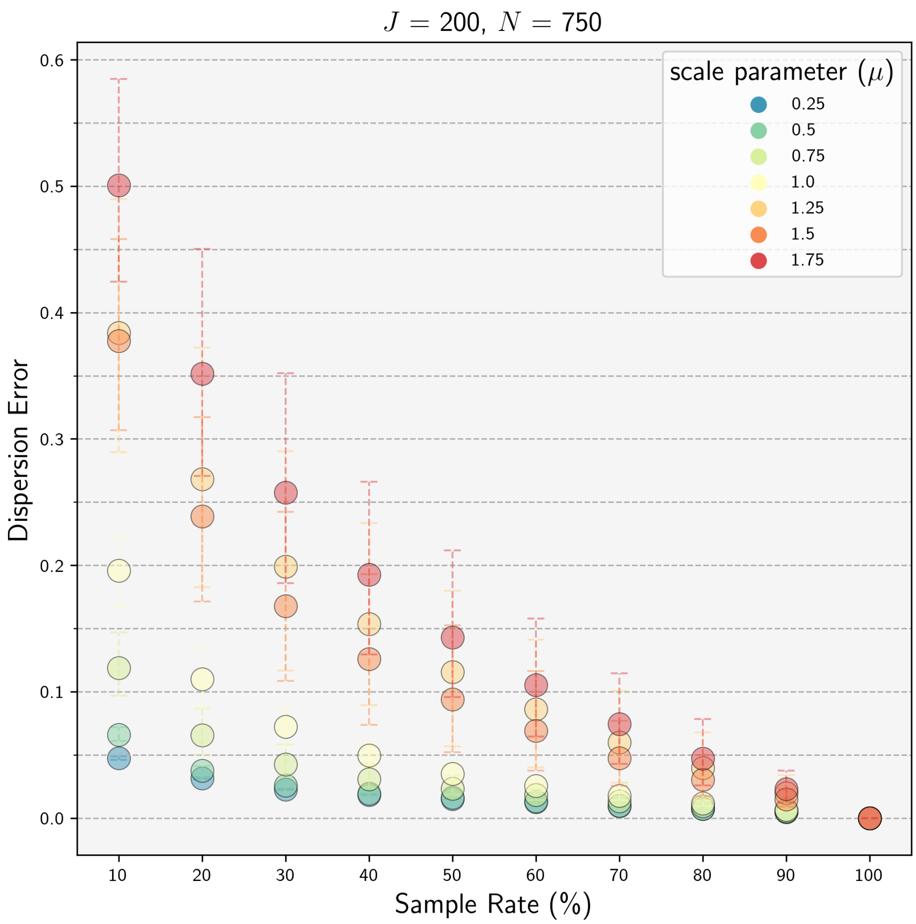


Figure 3.4: Detail view of the top left grid cell from Figure 3.3

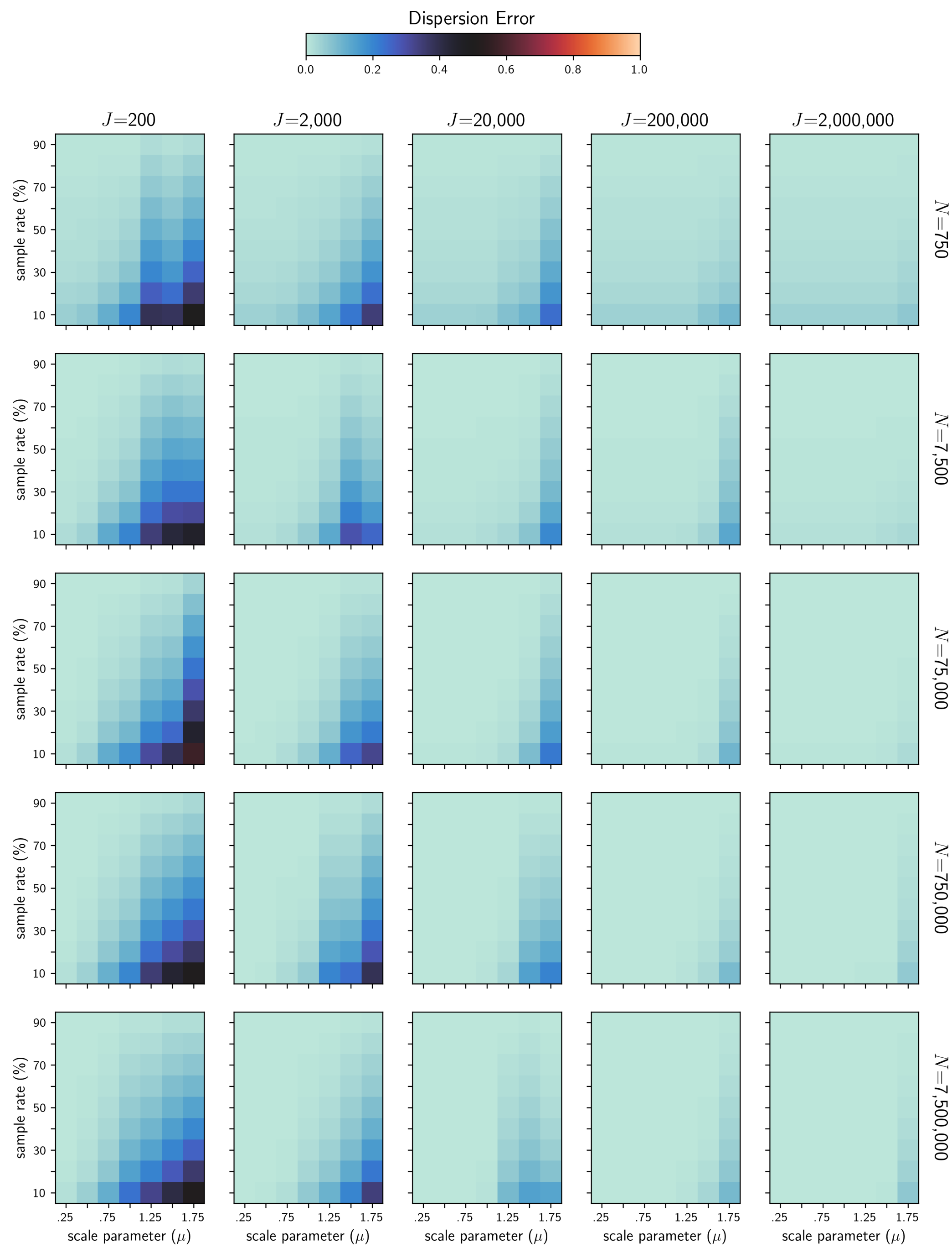


Figure 3.5: Heatmaps show differences in mean dispersion error across scenarios.

that importance sampling actually performs worse than SRS for the majority of scenarios tested here, particularly for large choice set ($J \leq 2000$) scenarios and low precision models ($\mu < 1$). This represents an important and novel finding of the study, which challenges traditional wisdom about best practices for sampling from large choice sets.

With respect to the size of the universe of alternatives, I find that dispersion error is largely invariant to the scale of J under importance sampling. Meanwhile dispersion error under SRS is shown to decrease dramatically as J increases. Comparatively, this means that any advantage offered by importance sampling for small J scenarios is quickly lost as the size of the universe of alternatives approaches the most disaggregate scales and the performance of SRS improves. The performance of importance sampling is also found to vary with model precision, with the most precise models demonstrating the greatest benefit. This makes intuitive sense, since Equation 3.5 shows that when ϵ_{in} is smaller, the information contained in the model V_{in} necessarily represents a greater portion of the total utility U_{in} for each choice. Although Lemp & Kockelman [51] found a similar correlation between model precision and benefit of strategic sampling, they are unable to identify this threshold beyond which SRS is able to out-perform importance sampling, owing to the limited number of scenarios they test and their explicit focus on parameter estimation instead of aggregate error.

Less intuitive is why importance sampling would actually exacerbate dispersion error for low precision models. By examining the individual probability massings (Equation 3.7) for a number of scenarios, I was able to determine that the vast majority of this error is actually the result of *underdispersion*, a kind of statistical overfitting where so much of the aggregate probability mass is allocated to the highest utility alternatives that the lowest probability alternatives are essentially ignored. Figure 3.7 shows the results of this analysis for a scenario with 750 choosers and 200 alternatives. One possible explanation as to why this phenomenon has gone unnoticed until now is that underdispersion does not lead to choosers selecting "bad" alternatives. This means that any metric which relies on the simulated choices of individual choosers, such as those used in [52] and [51], rather than the full distribution of choice probabilities, cannot capture this type of error.

Application to Real Data

In order to validate my experimental results, I estimate a location choice model with real data from the 2012 California Household Travel Survey (CHTS). I limit this model to include only in-region, out-of-home discretionary activities completed by residents of the nine County San Francisco Bay Area during the survey period (n=28,880). I aggregate activity locations to the Census Block Group (n=106,910), and estimate a model by randomly sampling 1,000 alternatives per chooser (Table 3.1). Given the already extremely small standard errors obtained from this model, it was deemed unnecessary to use a larger sample size.

For the prediction phase I generate a synthetic population of choosers using the open source SynthPop software [76]. From this I randomly sample 4e5 choosers such that the ratio of choosers to alternatives is comparable to that of the original Nerella and Bhat [52] experiment. I perform the same series of simulations that were carried out on the synthetic

Description	Coeff.	Std. Err.	z	$p > \|z\|$
Log(% retail)	2.2708	0.032	71.36	0.000
Log(total jobs)	0.6120	0.003	233.45	0.000
Log(total pop.)	0.0377	0.002	15.92	0.000
jobs-to-housing ratio	-0.0001	0.000	-11.06	0.000
Log(dist. to CBD) : household vehicles/person	0.1165	0.001	110.50	0.000
Log-likelihood: -167,200.577, ***Pseudo R^2:***, 0.160				

Table 3.1: Multinomial logit model of discretionary activity location choice estimated from California Household Travel Survey data (n=28,880)

data, varying the sample rate and scale parameter (J is fixed), and compute dispersion error for each scenario. A comparison of these results against the experimental data shows that the dispersion error in the CHTS model closely approximates the expected error of the low precision ($\mu < 1$) experimental models for a similarly sized dataset (Figure 3.8).

This pattern holds true for all nine sample rates, and suggests that the experimental findings described above are more than just an artifact of the data generation process. Importantly, it also provides contextual information about the relative precision of the CHTS model, at least insofar as it relates to prediction error. I discuss the significance of this below.

3.4 Discussion

The results of this analysis raises several critical issues that to date have not been adequately addressed in the scientific literature on sampling of alternatives in discrete choice models. I will briefly describe each of these here, and then discuss their implications both for practitioners and for future research.

The Inadequacy of Sample Rate Recommendations

The first two issues arise from the observation that in many scenarios, both model precision and the size of the universe of alternatives are found to be more significant determinants of the bounds of dispersion error than sample rate alone. This is significant because it suggests that the validity of many commonly cited sample rate guidelines may be more limited than previously thought. For example, the authors of [52] recommend a minimum sample rate of one-eighth (12.5%) for MNL using SRS. They do not, however, test the robustness of this result against varying levels of model precision, making it impossible to determine how well their findings might generalize to a model estimated on real data. Furthermore, their findings are based on experiments in which the total number of alternatives is fixed to 200, which given the evidence presented here, would likely not hold for choice scenarios in which the universe of alternatives is many orders of magnitude larger, such as those found in even modestly disaggregate travel or land use demand models. The good news for practitioners is that larger choice scenarios are associated with smaller aggregate error due to sampling of alternatives and thus the recommended 12.5% sample rate recommended is probably much larger than necessary, particularly for the most disaggregate scenarios. This is made clear from Figure 3.8, which shows that average dispersion error is reduced by approximately two orders of magnitude as the size of the universe of alternatives increases from 200 to 2e6, regardless of the scale parameter and sample rate.

The sensitivity of dispersion error to model precision is particularly problematic for establishing sampling guidelines because in practice one cannot explicitly observe the scale parameter separately from estimated parameters. However, the simulation results from the CHTS model described above make it possible to infer something about the expected range of precision one might encounter in practice. Specifically, Figure 3.8 shows that the dispersion error observed in the CHTS model aligns quite nicely with the experimental results for models in the low-precision regime ($\mu < 1$). This is a strong indication that the CHTS model, and other highly disaggregate location choice models like it, may have relatively low levels of precision. If this is the case, then based on evidence presented here, practitioners should expect only modest gains from the use of sample rates greater than 10% under simple random sampling (Figures 3.3, 3.4, 3.5). This is consistent with what we observe in the CHTS model, in which dispersion error decreases from 0.004 to .0025 when doubling the sample rate from 10 to 20%. A 40% reduction in error might seem significant, but in absolute terms it corresponds to a mere 0.15% of the total probability mass of the population. More work should be done to confirm these findings using other sources of estimation data and simulated choices.

Compared to model precision, the fact that the size of the universe of alternatives is a known quantity should make it easier for practitioners to incorporate its observed effect into their modeling decisions. However, given the surprising finding that large J scenarios actually exhibit *less* overdispersion as a proportion of total probability mass, some additional discussion is merited. Figure 3.9 shows the results of a secondary analysis in which we compare the empirical distribution of true probability massings (100% sample) at two different values of J. These results show that as J increases, the aggregate choice probabilities do become more concentrated in the peak and tails of the distribution, as measured by higher values of skewness and kurtosis, respectively. However, both the standard deviation and the maximum value of probability mass for a given alternative are on average smaller. The net effect of these dynamics is that the peak of the distribution of $\frac{1}{N}\sum_n p_n(i)$ does not grow as fast as the tail, meaning that the number of bad alternatives increases faster than the aggregate probability that is concentrated in the good ones. In one sense, then, the initial hypothesis about the effect of J was correct: as the size of the universe of alternatives increases, simple random sampling is less likely to produce choice sets containing the highest probability alternatives. So why then does error due to random sampling of alternatives actually improve for larger values of J?

To answer this question one must recognize that the goals of model estimation and forecasting are not the same. Specifically, in a choice scenario where alternatives are not subject to capacity constraints, forecast performance is not optimized by increasing the within-sample variance among alternatives like it is for estimation. Rather, if the accuracy of a forecast under sampling of alternatives is defined by the ability to replicate the aggregate demand for alternatives obtained without sampling alternatives, it follows that optimal results are obtained when sampled choice sets are representative of the universe of alternatives. Thus, if the variance of $\frac{1}{N}\sum_n p_n(i)$ shrinks for larger values of J, dispersion error due to random sampling of alternatives should decrease as well. It is also worth noting that even though we have artificially controlled for model precision in this secondary analysis, in practice precision will be strongly correlated with the variance of the utility of alternatives. This idea is supported by the results of the CHTS model, which showed that a typical microsimulation-scale location choice model belongs to the regime of low-precision models in which dispersion error is relatively small.

Simple Random Sampling: Not So Bad After All?

The distinction between estimation and forecasting also explains why the widely held belief about the effect of importance sampling proved incorrect in the context of this study. As shown in Figure 3.7, importance sampling does in fact produce distributions of aggregate demand where a greater proportion of probability mass is allocated to the higher utility alternatives. Since the utility functions themselves are unchanged between the sampling strategies, this simply means that importance sampling is functioning as intended, with high utility alternatives showing up in sampled choice sets more frequently. This is what makes importance sampling beneficial for parameter estimation. However, the results shown here suggest in the context of forecasting aggregate demand for alternatives which are not capacity constrained (e.g. discretionary activity location or route choice), this very same mechanism can actually lead to dispersion error of another kind: under-dispersion, or an overallocation of probability mass to high-utility alternatives. For choice scenarios which are capacity-constrained (e.g. building-level residential location choice), or models which treat demand as a market-clearing process, over-selection of high-probability alternatives is not possible and the effect of this under-dispersion is likely immaterial. However, if the choice model design is not self-regulating in this way, then importance sampling may in fact produce demand forecasts which are less accurate than what would be obtained with SRS, particularly if model precision is low or the size of the universe of alternatives is large. It is possible that a less precise importance sampling approach, one that is not based on the true probabilities $p_n(i)$ for each alternative, would mitigate the under-dispersion effect observed here. Future work should investigate whether heuristic-based importance sampling strategies, for example one based on the distribution of observed trip distances, might perform better.

3.5 Conclusions

The results of this study strongly suggest that forecast error due to sampling of alternatives in discrete choice models is less significant than previously hypothesized. In many cases, simple random sampling is shown to produce choice probabilities which match the true distribution of aggregate demand quite well, particularly when model precision is low. Since model precision is expected to decrease as the size of the universe of alternatives increases, it follows that the impact of random sampling of alternatives should lessen as choice scenarios approach the microsimulation-scale. Empirical evidence from a model estimated on real survey data lent further support to the notion that microsimulation-scale models will tend to have low levels of precision relative to the possible scenarios tested here. Additionally, simple random sampling is found to outperform an importance sampling strategy for all but the smallest, most precise choice scenarios. Although the validity of this finding is limited to unconstrained choice problems, it is nonetheless significant in that it challenges the prevailing belief that optimal sampling strategies for model estimation are necessarily optimal for demand forecasting as well.

A secondary contribution of this study is that it has demonstrated of the potential of GPU-based computing for simulation of discrete choice models with massive choice sets. It is quite possible that as GPU technology becomes more pervasive and the costs of purchasing or renting time on a cluster of GPUs becomes more affordable, sampling of alternatives may no longer be necessary at all. In the meantime, land use and travel demand modelers would do well to mind the differences between estimation and forecasting as it relates to sampling of alternatives, and think twice before turning to importance sampling as a one-size-fits-all solution.

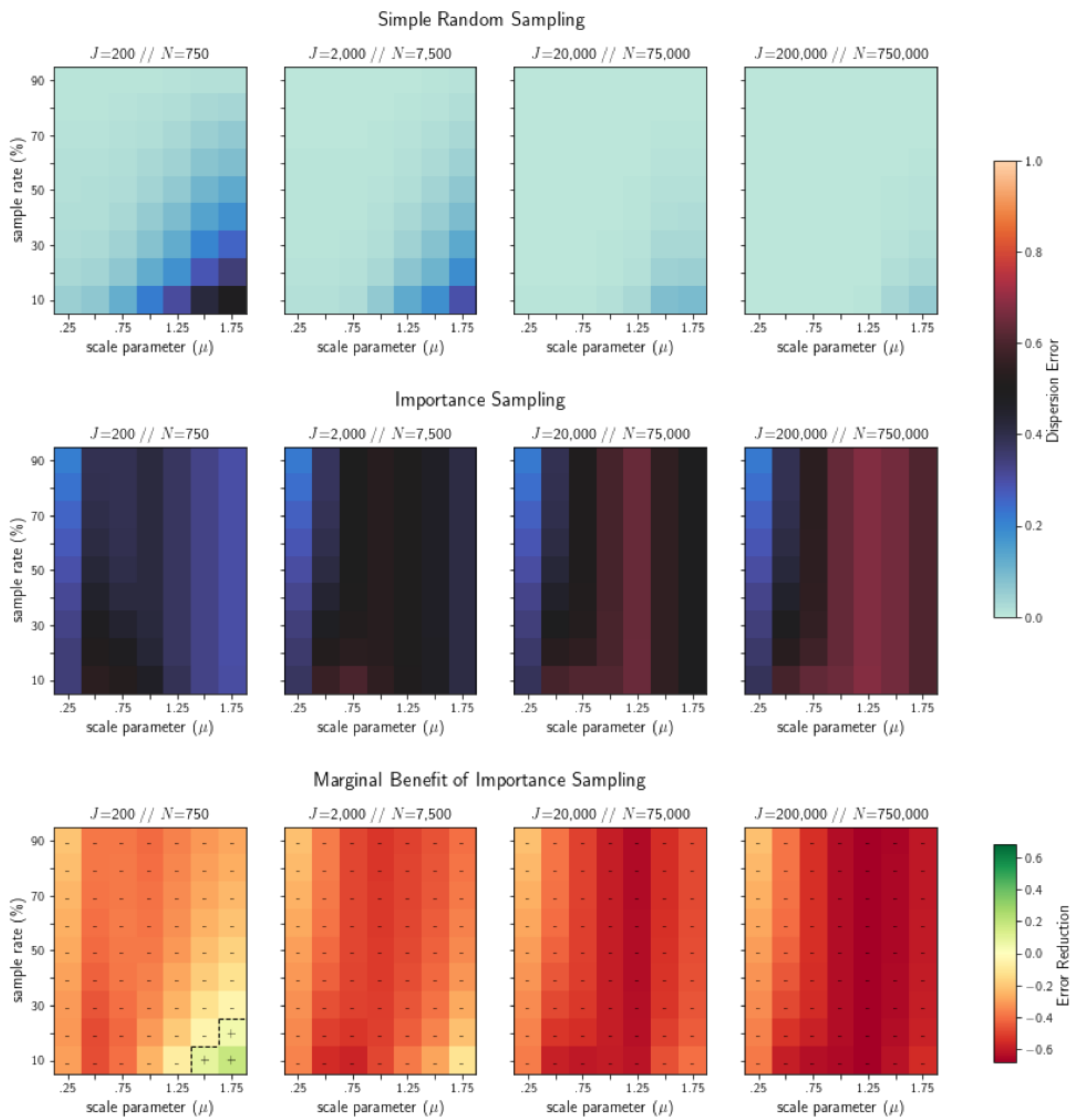


Figure 3.6: A comparison of mean dispersion error for 10 runs under simple random sampling and importance sampling. Only 3 of the 252 (1.2%) of scenarios tested exhibit reduced dispersion error under importance sampling, identified in the third row of heatmaps with a "+". Otherwise, SRS provides a much better fit to the true distribution of aggregate demand.

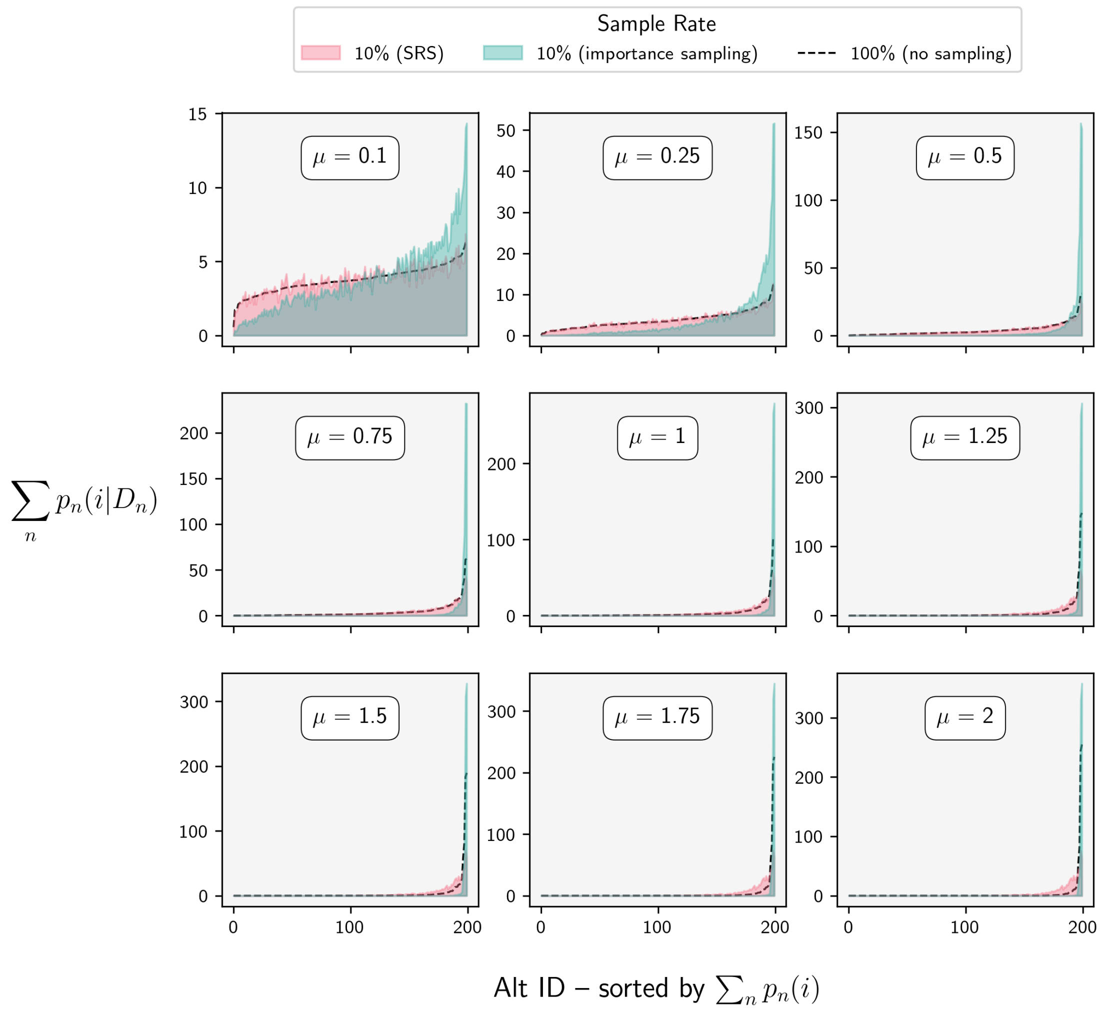


Figure 3.7: A comparison of the distribution of probability massings (PM_i) for a population of 750 choosers and 200 alternatives generated using a 10% sample rate with simple random sampling (pink) and importance sampling (turquoise). Importance sampling produces a more accurate distribution of aggregate choice probability when model precision is high ($\mu > 1.25$) and the true distribution of probability mass is concentrated among fewer alternatives, where SRS shows evidence of overdispersion. Conversely, when model precision is low, importance sampling leads to *underdispersion* of aggregate choice probability relative to the true distribution. In such cases SRS seems to produce a better fit.

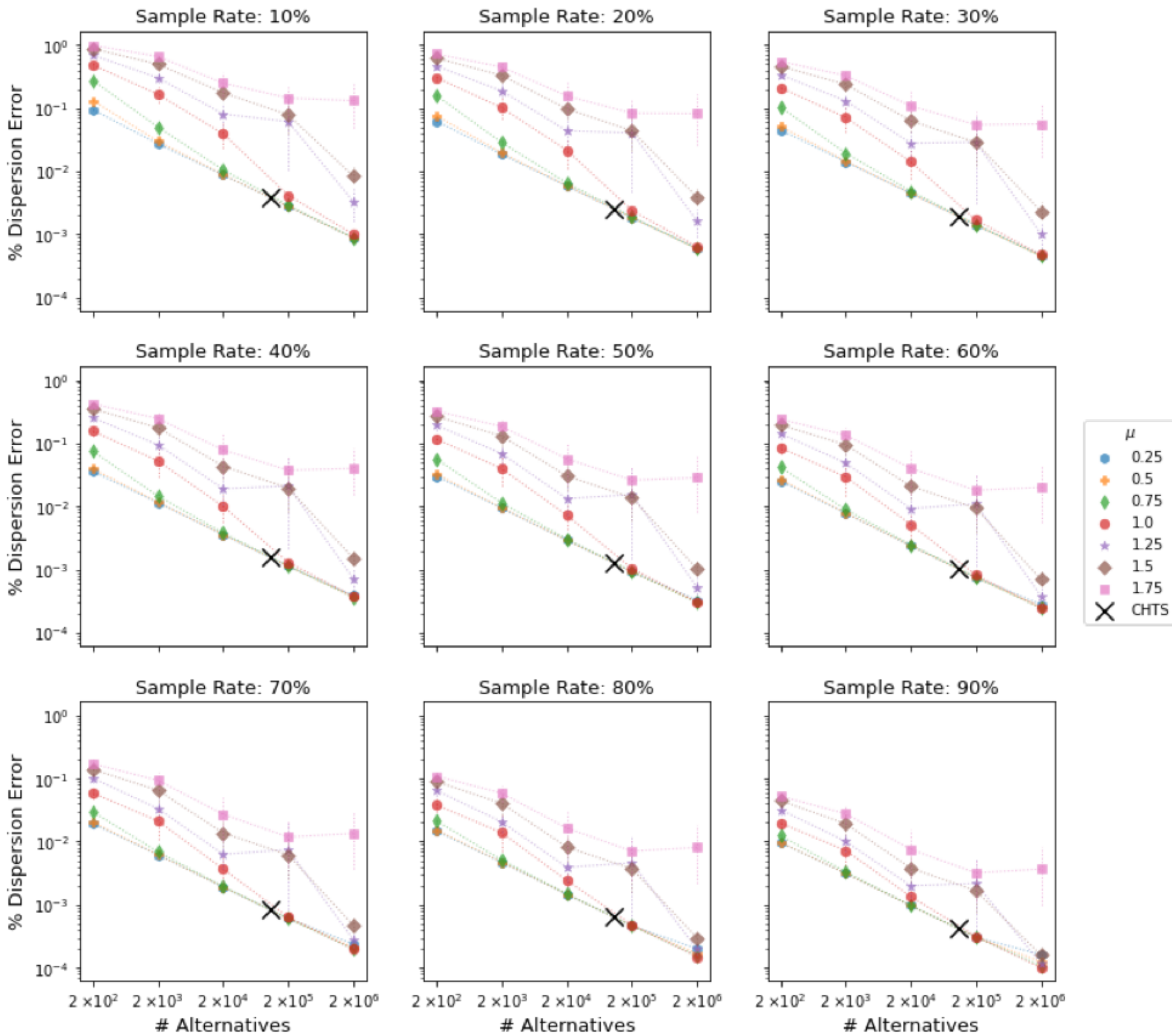


Figure 3.8: A comparison of experimental results to those obtained from a model estimated on data from the California Household Travel Survey (CHTS). Each grid cell plots the relationship between average dispersion error and the size of the universe of alternatives for a given scale parameter (μ) and sample rate. The black '$\times$' represents the CHTS results in which the true scale parameter is unknown.

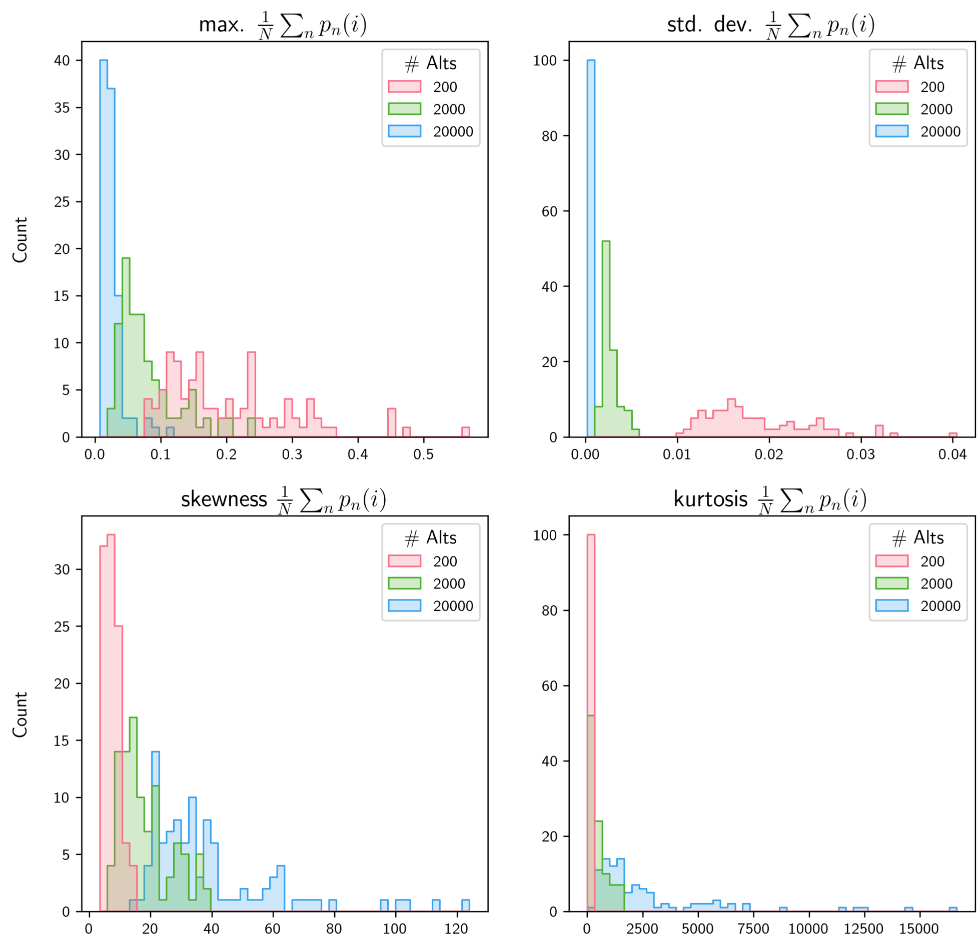


Figure 3.9: Descriptive statistics of the true distribution of probability mass by number of alternatives J for 100 iterations with $\mu = 1$.

Chapter 4

Population Synthesis of Business Establishments with Bayesian Networks

4.1 Introduction

One of the chief benefits of using microscopic models is the increased behavioral realism that results when short- and long-term mobility decisions are determined at the level of the individual chooser. However, the ability to construct models of this nature depends on the availability of disaggregate data to feed into them. Thus, as the use of Activity-Based Models (ABMs) and microsimulation has grown in recent years, so too has the need for highly detailed datasets. For certain applications, like the estimation of behavioral models, sampled population records of the kind found in the typical household travel survey are sufficient, provided that the sample is representative (random) or that sample weights are well-known. The simulation phase of the urban modeling workflow, however, is much more challenging in this regard. In order to generate a forecast of regional demand, choices must be generated for every chooser in the region, and thus a fully enumerated dataset of disaggregate population records is required. Population synthesis is a well-established practice by which this can be achieved through the use of algorithms or models to scale up a sample of disaggregate observations.

In the case of business establishment data, even sample records can be difficult to acquire. In the United States, this is due to the fact that most of these data are derived from tax records, and federal law prohibits publishing even the fact that a business has filed [77]. A lack of publicly available "firmographic" microdata has meant that the application of methods for population synthesis to the study of regional business dynamics has lagged behind. As a result, some urban microsimulation systems do not characterize business establishment populations with the same level of fidelity as persons or households counterparts, while others simply treat businesses as static, supply-side amenities (e.g. jobs) that may attract trips but do not operate as agents within the system [78–80].

The purpose of this chapter is to describe two ways in which recent advances in probabilistic modeling can make business establishments more accessible through population synthesis. Both approaches outlined below use a set of proprietary business establishment microdata to estimate Bayesian network-based models which encode the joint distributions of all possible combinations of establishment attributes. In the first approach, the estimated models are used to generate a replica of the original dataset comprised entirely of synthetic records, thereby preventing the disclosure of confidential data. The second method shows how the same models can be used to enrich publicly available aggregate data sources to produce similarly detailed synthetic populations without the use of control totals based on proprietary data. By helping democratize access to business establishment microdata these efforts stand to improve the internal validity of many microsimulation systems in use today while also advancing the field of study as a whole by encouraging scientific collaboration and the adoption of these methods.

Population Synthesis for Households and Persons

Typically synthetic population data is created by oversampling a population of real microdata (e.g. from a survey or Public Use Microdata Sample (PUMS)) and then allocating those records geographically to match known distributions of aggregate sociodemographic characteristics [76]. A rich body of scientific literature exists which is dedicated to this type of population synthesis, and methods are constantly being improved upon to generate more accurate, representative synthetic populations. Because the requisite microdata inputs for household and person populations are often publicly available at no cost, and because once generated, synthetic records can be shared freely among collaborators without fear of disclosing personally identifiable information (PII) or illegally distributing proprietary data, recent advances in population synthesis have dramatically lowered the barrier to entry for demographic microsimulation. Although these data can typically be accessed remotely, restrictive data licensing agreements often make collaboration just as difficult, and may even limit the types of findings that a researcher can publish.

According to a 2014 survey, most methods of population synthesis fall into one of two categories: analytic or combinatorial [81]. Analytic methods are typically based on some kind of iterative balancing of marginal distributions of population attributes until the population converges towards a known control. Because this control is often known only for a very high level geography, spatial allocation or upsampling is often performed as a post-processing step. Iterative proportional fitting (IPF), first proposed in 1940 [82], is the most commonly employed analytic method, and many other popular techniques are based upon it, including iterative proportional fitting for synthetic reconstruction (IPFSR) [83] and iterative proportional updating (IPU) [76]. Combinatorial optimization (CO) methods, on the other hand, refer to an approach in which an initial population is generated by oversampling a set of microdata and then successively replacing individual records with new draws in order to improve the fit [84]. These methods are less common, but recent studies have shown CO to perform better than analytic methods in a number of cases [85, 86]. More recently, Sun and Erath [87] have proposed a third "alternative modeling paradigm" based on modeling causality and dependence among population attributes, which they demonstrate through the use of Bayesian networks. The research presented here aims to contribute to the growing body of literature on this new paradigm in population synthesis by evaluating its performance in the context of business establishments.

The Need for Better Businesses Data

The vast majority of the scholarship on population synthesis for urban modeling focuses on populations of persons and households rather than businesses or jobs. One major reason for this discrepancy is the availability of data. In the United States, for example, there is no publicly available set of national business listing microdata comparable to what the American Community Survey PUMS data provides for person and household-based populations[1]. In the case of the United States, disaggregate business establishment data actually *is* collected [2,3], but is typically only made available to the public in aggregate form[4,5]. Researchers can apply for access to "restricted use" data[6], but the process is cumbersome and even successful applicants may be deterred by on-site access restrictions which preclude scientific collaboration and can impose significant travel expenses[77]. As a result, commercial products purporting to offer population-level coverage have become a popular alternative for researchers with the funds to purchase these expensive datasets (e.g. [88]). Although these data can typically be accessed remotely, restrictive data licensing agreements often make collaboration just as difficult, and may even limit the types of findings that a researcher can publish. Even more prohibitive, however, are the costs associated with actually purchasing these data. Thus the refusal of the government to publish a PUMS product for business establishments has not prevented the acquisition of these data by anyone with the means to do so. Instead it has allowed the free market to determine who is able to contribute to scientific discovery in this domain.

One recent development that is worth mentioning here is SynLBD[7]: a comprehensive effort by researchers from the United States Census Bureau and affiliated institutions to create a synthetic version of the Longitudinal Business Database microdata deemed appropriate for public consumption [77, 89]. Although this work should be applauded, in practice it seems that access is still heavily restricted. Namely, an application process is still required, albeit streamlined, and while data can be *accessed* remotely they cannot be downloaded . Even the

[1]See https://www.census.gov/programs-surveys/acs/microdata.html

[2]https://www.census.gov/programs-surveys/ces/data/restricted-use-data/longitudinal-business-database.html

[3]https://www.census.gov/programs-surveys/ces/data/restricted-use-data/lehd-data.html

[4]https://www.census.gov/programs-surveys/cbp.html

[5]https://lehd.ces.census.gov/data/

[6]For example, see https://www.census.gov/about/adrm/ced/apply-for-access.html

[7]See https://www2.vrdc.cornell.edu/news/data/lbd-synthetic-data/

results of analysis performed on the remote access server can only be retrieved by placing additional requests for "specific files that are a result of their analyses"[8].

Bayesian Networks

Bayesian networks, also known as Bayesian belief networks, refer to a class of models which describe the joint distribution of random variables as a Directed Acyclic Graph (DAG) of conditional independencies. In the graphical structure of a Bayesian network, each node corresponds to a random variable and stores the probability distribution of the possible values that it can take conditional on the values of other nodes upon which it depends. These conditional independencies are encoded as the edges of the graph, connecting nodes where such a relationship exists. Because the graph is directional, the edges imply a causal flow of conditional independence, where the marginal probability of a given node depends only on its parents. This means that each node is considered to be conditionally independent of all other non-descendent nodes given its parents [90]. Thus in the example Bayesian network in Figure 4.1, nodes b and c are conditionally independent given a, as are a and d given c. Likewise, a node without parents is interpreted to be independent from all other variables in the network.

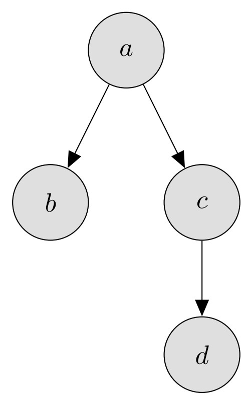


Figure 4.1: An example of a directed acyclic graph.

By storing information about conditional independencies in this way, the structure of a Bayesian network allows for the extremely efficient computation of joint probabilities. Another advantage of the approach is that apriori knowledge can be encoded into the structure of the graph, which significantly reduces the search space during the structural learning (i.e. model estimation) phase. Bayesian networks also make it easy to leverage apriori knowlege to perform evidence-based sampling on a model that has already been estimated.

Although the development of Bayesian networks dates back to the late 1970s [91], these and other attractive features have contributed to the growth of their popularity across a

[8]https://www2.census.gov/ces/synlbd/SynLBDapplication.pdf

number of fields of study. Recently, a number of studies have demonstrated the use of Bayesian networks as a tool for privacy-preserving data publishing [92, 93]. In the present study I draw upon this work as well as that of [87] to show how Bayesian networks afford modelers the ability to strike a balance between data obfuscation for privacy preservation and predictive accuracy for effective population synthesis.

4.2 Data

The primary data source used in this study is the National Establishment Time Series (NETS) [94]. NETS is a longitudinal database of establishment-level observations designed to track business starts, relocations, and dissolutions for the entire universe of businesses in the United States. Business characteristics recorded in the database include the age, sales, industry sector, number of employees, and geographic location over time. The NETS database is one of the more popular sources of establishment-level data used by researchers across a variety of disciplines [95–99].

NETS data is best understood as a combination of employer and nonemployer business establishments. The union of Census County Business Patterns (CBP) and Nonemployer Statistics (NES) thus provides the closest approximation of the population of businesses covered by NETS in terms of publicly available Census data products[9]. Although the NETS records fall short of the full universe of establishments one would expect from the union of these two sub-universes (CBP + NES), numerous studies have shown that the NETS microdata is reliable for detailed analysis of business activity [100, 101]. Below, Table 4.1 characterizes the different Census data products which offer (aggregate) establishment counts and attributes.

Product	**Universe**	**Geography**	**Sector**	**Size**	**Age**	**Sales**	**Payroll**
CBP	Payroll	County	√	√	–	–	√
NES	Nonemployer	County	√	√	–	√	–
ZBP	Payroll	ZIP Code	√	√	–	–	–
BDS	Payroll	MSA	√	–	√	–	–
BDS	Payroll	County	–	–	√	–	–
BDS	Payroll	County	√	–	–	–	–

Table 4.1: Summary of public establishment count data products from the US Census.

Figures 4.2-4.4 compare the frequency distributions of business attributes between the NETS records and the CBP + NES universe by county and two-digit NAICS sector code.

[9]Because NETS does not distinguish between employer and nonemployer establishments, and because every establishment in the database shows at least one employee, it is assumed that employment counts include the business owner. Thus 1 is subtracted from each employment count in order to compare NETS and Census totals, in line with the approaches of [100, 101]

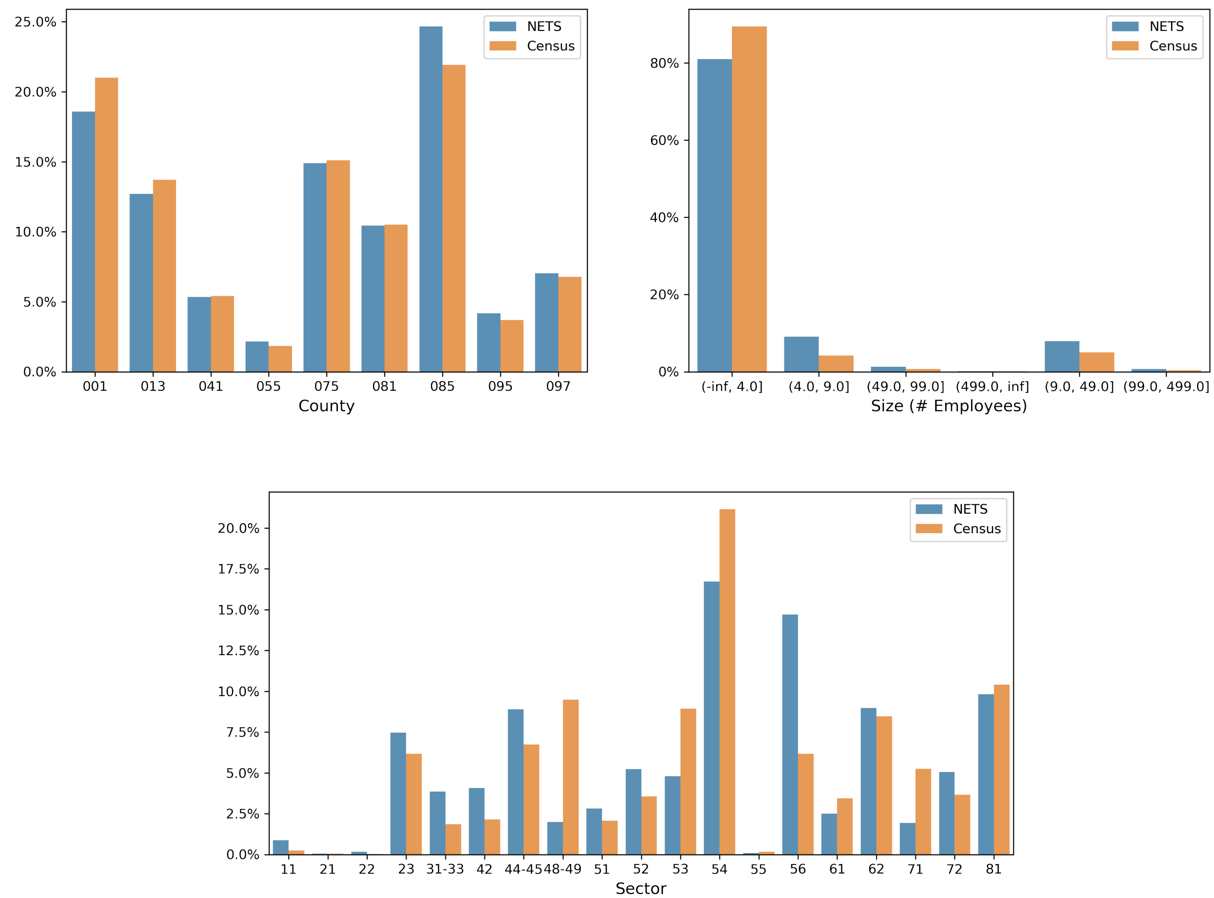
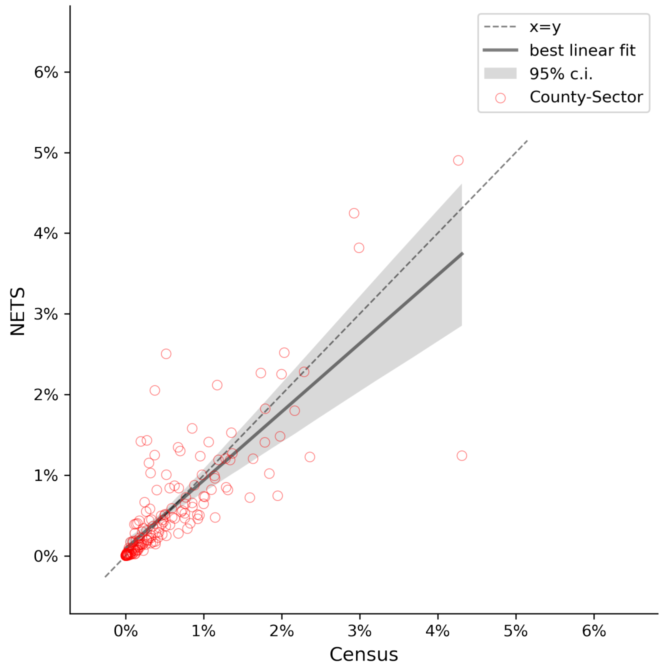


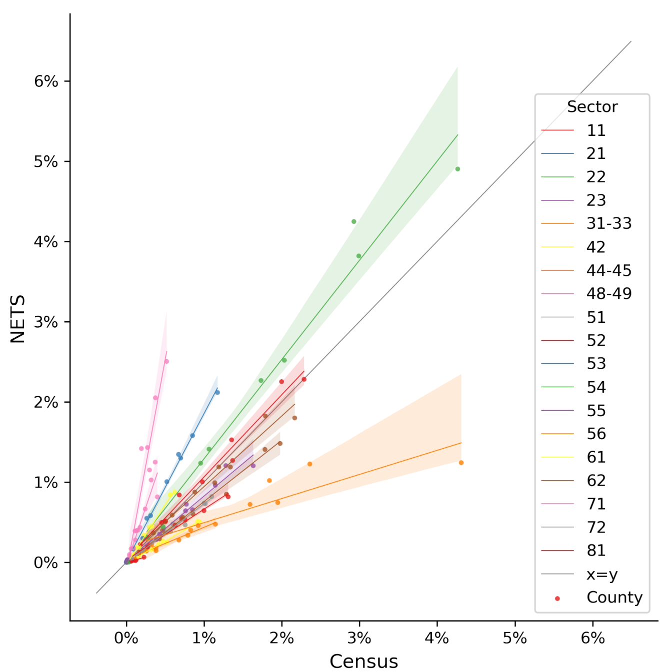
Figure 4.2: NETS vs. Census (CBP + NES): marginal distributions of establishments by county, size, and sector

Because each population contains a different number of total business establishments, rates are used to compare them instead of absolute counts. Figure 4.2 suggests that the marginal distributions of establishments by county, sector, and size are quite similar between the two data sets, while 4.3a shows that the joint distribution by county and sector is somewhat noisier. However, most of this noise appears to be due to sector-level variation as shown in Figure 4.3b.

Figure 4.4 presents a different view of the joint distribution of each attribute pair (county, sector, and size) which suggests that the vast majority of the mismatch between the NETS and Census data is being driven by sector-level discrepancies in smallest class business establishment sizes. This finding is consistent with the assessment of NETS data by Barnatchez et al. [101] which identified the smallest class of establishments as the primary source of error between NETS and official Census-based sources.



(a)



(b)

Figure 4.3: Nets vs. Census (CBP + NES): Sector frequency rates by County

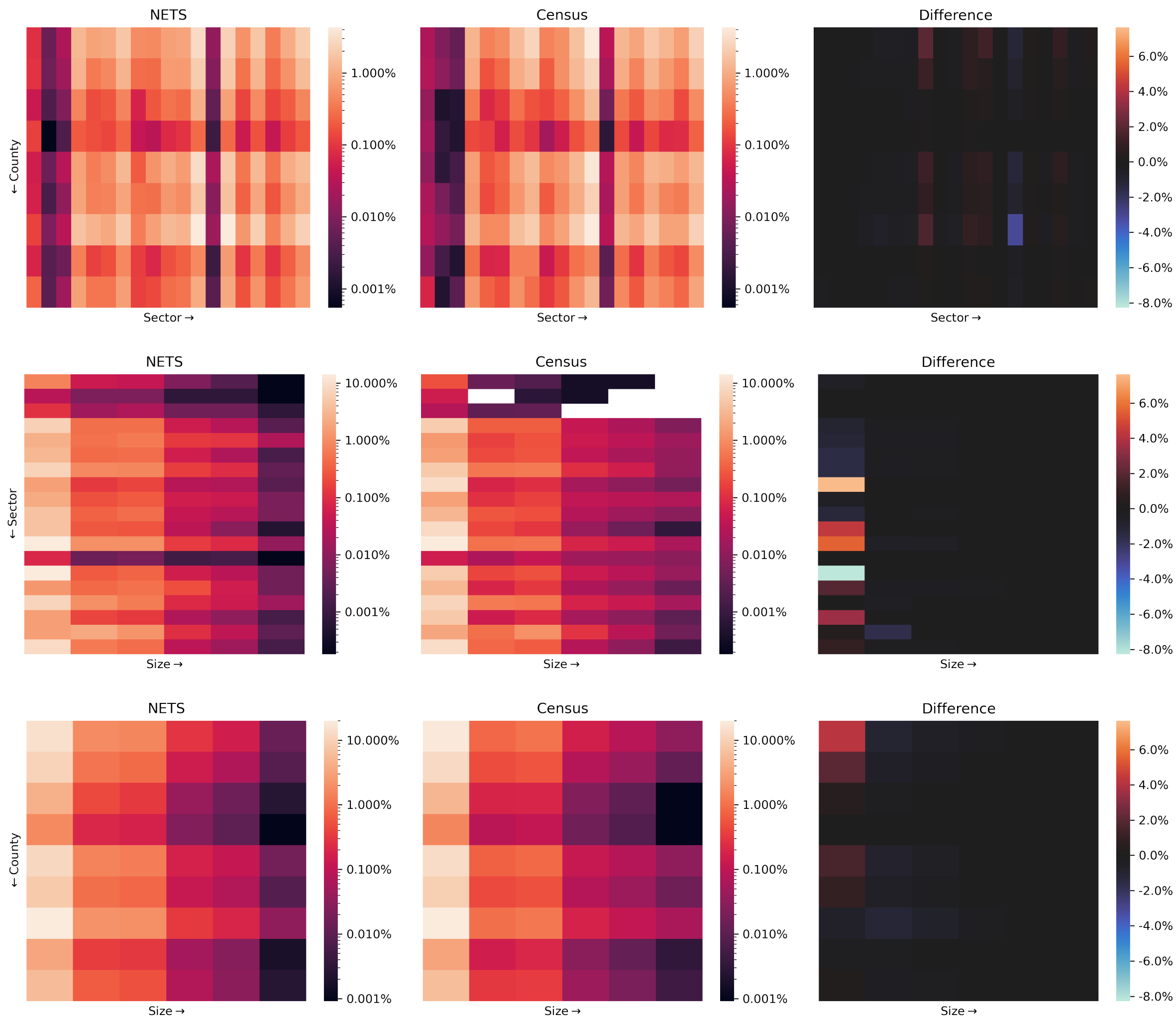


Figure 4.4: NETS vs. Census (CBP + NES): Attribute pair frequency rates and error

4.3 Methodology

The general approach applied here follows that of a typical machine learning workflow. A set of training data is first identified, prepared, and used to estimate a model, and that model is then used to generate predictions. Below I describe the process of applying this approach to generate synthetic populations of business establishments for the nine county San Francisco Bay Area region.

	Stratification	**Constraints**	**Variables (Nodes)**	**# Categories**
(1)	Pooled	–	COUNTY × SECTOR × SIZE × AGE × SALES	30,780
(2)	Pooled	root=COUNTY	COUNTY × SECTOR × SIZE × AGE × SALES	30,780
(3)	County	–	SECTOR × SIZE × AGE × SALES	3,420
(4)	County	root=SECTOR	SECTOR × SIZE × AGE × SALES	3,420
(5)	County	–	SECTOR × SIZE × AGE × SALES × TRACT	136,800 - 1,395,360
(6)	County	root=SECTOR	SECTOR × SIZE × AGE × SALES × TRACT	136,800 - 1,395,360
(7)	Sector	–	COUNTY × SIZE × AGE × SALES	1,620
(8)	Sector	root=COUNTY	COUNTY × SIZE × AGE × SALES	1,620
(9)	County-Sector	–	SIZE × AGE × SALES	180
(10)	County-Sector	root=AGE	SIZE × AGE × SALES	180
(11)	County-Sector	–	SIZE × AGE × SALES × TRACT	7,200 - 73,440
(12)	County-Sector	root=AGE	SIZE × AGE × SALES × TRACT	7,200 - 73,440

Table 4.2: Bayesian network models for generating County-level synthetic businesses. The number of total categories in Models 5, 6, 11, and 12 are ranges because the number of tracts varies across counties.

Data Extraction and Cleaning

The vast majority of NETS database is comprised of longitudinal data. And although it may be desirable for certain analytic applications to synthesize longitudinal dynamics [77], only single-year "snapshot" is necessary for the purposes of creating a set of base-year population inputs for microsimulation. Thus the first step in the approach outlined here is to extract from the NETS database only those records which correspond to establishments not categorized as "OutOfBusiness" for the year in question. In this case 2017 is defined as the base-year. Of these records, only a small subset of business attributes are used for population synthesis. These attributes are as follows: COUNTY, TRACT, SECTOR, AGE, SIZE, SALES. Table 4.3 describe each of these attributes as well as the schemes used for binning continuous variables into discrete categories.

field	**description**	**categories / breaks**	**# classes**
COUNTY	3-digit FIPS code	001, 013, 041, 055, 075, 081, 085, 095, 097	9
SECTOR	2-digit NAICS code	11, 21, 22, 23, 31-33, 42, 44-45, 48-49, 51, 52, 53, 54, 55, 56, 61, 62, 71, 72, 81	19
AGE	Years since start	[< 1, 5, 10, 15, 25+]	6
SIZE	# employees	[< 4, 9, 49, 99, 499+]	6
SALES	Annual sales ($)	[< 77300, 122200, 200000, 400000]	5

Table 4.3: Business attributes used for population synthesis.

While Bayesian networks are not theoretically limited to categorical variables, there are few advantages offered by using continuous data in the context of population syntheses. For example, if the behavioral models that comprise an urban microsimulation are estimated from discrete data, any continuous variables in the synthetic population may wind up being discretized anyways. Secondly, as previously mentioned, most potential sources of validation data are only available in an aggregate, categorical format, thus making it difficult to assess the validity of any additional precision offered by continuous variables. Lastly, restricting the dataset to categorical variables makes it possible to leverage existing open-source software solutions for estimating Bayesian networks.

The NETS database records a total of 619,047 unique business establishments that were active as of the year 2017 across the nine counties included in this study. Of these, 15,253 observations (2.5%) have no SALES values recorded. However, one attractive feature of Bayesian networks is that it is possible to estimate models from inconsistent or incomplete data. The only records which are discarded for incompleteness are the 172 entries with geographic coordinates that position them outside of the study area. At this stage, 618,875 unique establishments remain. For the sake of comparison, Census County Business Patterns and Nonemployer Statistics data identify 204,230 and 686,361 establishments for 2017, respectively, for a total of 890,591 establishments across all nine counties. This finding is consistent with other studies that have suggested NETS data reflects a combination of employer and nonemployer businesses that is less comprehensive than the true union of these two universes [101, 102]. Finally, observations with NAICS codes 92 (Public Administration and Government) and 99 (Unclassifiable) are dropped, the former because none of the Census data products used for validation contain establishments in this sector, and the latter because these records are not useful for the purposes of constructing a synthetic population. Thus, a total of 552,868 establishment records are used to train the models described below.

Model Estimation

Multiple estimation strategies are tested based on four levels of stratification and two approaches to structural learning. The four stratification levels tested are "pooled" (unstratified), County, Sector, and County-Sector. The two methods used for learning the optimal structure of Bayesian networks are both based on the dynamic programming and A* (DP/A*) algorithm as implemented in the `pomegranate` package for probabilistic programming in Python [103]. The first method learns the structure of each network without any user input, while the second method makes use of user-defined constraints.

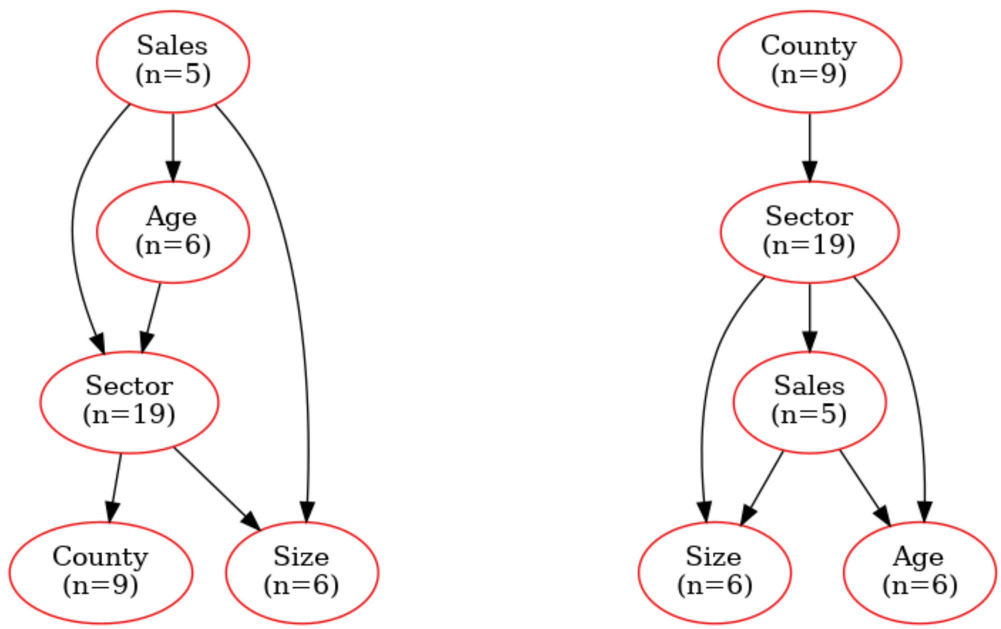
The use of constraints is a feature of Bayesian network learning that encodes apriori knowledge into the model specification. This has the dual benefit of drastically reducing the search space over which a model is optimized, and also allowing the user to take advantage of domain knowledge they may have about the relationships between variables in the network. In this study, the only constraints that are imposed are the specification of the root node in the model. Figure 4.5 shows the different network structures that result when constraints such as these are imposed. See Table 4.2 for more detail.

For each of the models which include county-level stratification (County and County-Sector levels) I estimate a secondary specification that includes an additional TRACT node as a part of the network structure, whose possible values correspond to the Census Tracts in each county. Including TRACT as a variable allows the Bayesian network models to perform a kind of spatial allocation in the same step as population synthesis.

It is also worth noting that Bayesian networks make it easy to work with sparse datasets. Because they are at their core simply a series of conditional probability tables, the "zero-bin" issue that plagues other approaches to population synthesis can be avoided by adding pseudocounts to each set of estimation data if there exist combinations of attributes for which no observations are recorded. Pseudocounts can even be fractional in order to avoid biasing data when the total number of observations is small. In the models below, I use a pseudocount value of 1 unless stated otherwise.

Simulation

Synthetic populations are generated by drawing samples from each of the fitted Bayesian networks according to two approaches. The two approaches differ only in the source of the control totals which are used to constrain the sampling process, with the former relying on the



(a) Constraints: `NONE` (b) Constraints: `root=COUNTY`

Figure 4.5: Learned network structures with and without constraints for pooled models (1) and (2)

original NETS data, and latter using only publicly available Census data. The first approach can be viewed as a technique for data obfuscation, where Bayesian networks are employed to create extremely accurate synthetic representations of the original input data that can be shared without the risk of disclosing proprietary data or PII. The second method, on the other hand, is perhaps better described as a kind of data enrichment process or multiple imputation [104], where the models are used to enhance the fidelity of an existing data set by appending new attributes to it. Notably, this method also represents an even more conservative approach in regards to data privacy since not even the control totals are based on proprietary data.

In the synthetic populations that are described below, only COUNTY and SECTOR variables have been employed as controls, while the rest of the variables are drawn probabilistically from the fitted models. The way control totals are imposed depends on the level of stratification that is used. With pooled models, for example, in which both COUNTY and SECTOR are both represented as nodes in the Bayesian network, COUNTY and SEC-

TOR values are assigned as "evidences", and samples are drawn according to the number of times each unique (COUNTY, SECTOR) tuple appears in the observed population of record. When evidence for a variable is supplied to a model during sampling, the the corresponding node in network is temporarily fixed to that value. For county-level models, however, there is no COUNTY node, so only SECTOR can used as evidence, and COUNTY is instead controlled for by drawing samples from a given county-level model and appending a value corresponding to that county for each sample that is drawn. Thus in the case of county-sector level models, no evidence is passed to the model at all, and control totals are instead enforced by the number of samples that are drawn from each county-sector model.

Lastly, I perform an auxiliary analysis to demonstrate how different combinations of model design and Census-based control totals can be used to perform sub-County spatial allocation of synthetic business establishment.

4.4 Validation

Assessing Goodness-of-fit

Each synthetic population is validated against the dataset from which the marginal distributions of establishment counts (control totals) were drawn to generate the records. In addition to visual assessment, two common measures of goodness-of-fit for synthetic data are used. The first is the Standard Root Mean Squared Error (SRMSE)

$$SRMSE = \sqrt{\sum_{m_1=1}^{M_1} \cdots \sum_{m_N=1}^{M_N} \left(f_{m_1,\ldots,m_N} - \hat{f}_{m_1,\ldots,m_N} \right)^2 (M_1, \ldots, M_N)} \tag{4.1}$$

where M_N is the number of categories in variable N, $f_{m_1,\ldots,m_N}$ and $\hat{f}_{m_1,\ldots,m_N}$ are the frequency counts for all possible attribute combinations in the observed and synthetic populations, respectively [105]. The second metric is the Freeman-Tukey statistic (FT^2)

$$FT^2 = 4 \times \sum_i \sum_j \left(\sqrt{T_{ij}} - \sqrt{E_{ij}} \right)^2 \tag{4.2}$$

where T_{ij} and E_{ij} represent the observed and synthetic frequency counts for the cell in row i and column j in two-way contingency table of population attributes [106]. The FT^2 statistic is distributed χ^2, with degrees of freedom equal to one less than the total number of categories being matched between two populations. This makes it possible to compute a critical value for rejecting the null hypothesis that two populations are drawn from the same underlying distribution. All but two of the scores reported in Tables 4.4 and 4.5 allow the null hypothesis to be rejected a 95% level of confidence or higher. Nevertheless, both the SRMSE and FT^2 metrics are useful for comparing relative levels of accuracy between models.

Population synthesis is repeated ten times for each of the twelve models and each source of control totals. Tables 4.4 and 4.5 present the average metrics across each of these ten runs for the NETS- and Census-based approaches, respectively. Because the magnitude of the scores depends on the number of categories that are available in the reference data, I include two sets of scores in Table 4.4: one for the full set of five variables (COUNTY, SECTOR, SIZE, AGE, and SALES) and one for the subset (COUNTY, SECTOR, and SIZE) that are available from the Census. The former ("All Vars.") is used to assess the accuracy of the NETS-based synthetic populations to the original NETS data, while the latter ("Census Vars.") makes it possible to compare the accuracy of the synthetic populations to those generated from Census-based controls.

NETS-based control totals

Table 4.4 describes the results of population syntheses based on control totals taken from the NETS data. Looking first at the "All Vars." scores, the County-Sector level models appear to perform the best, particularly as judged by mean SRMSE. The trend is less obvious for the FT^2 scores where there is little variation across models with the exception of all four sector-level models which perform nearly 50% worse than the next worst performer. In almost every case the unconstrained models appear to fit the data better, although the difference is not large. Interestingly, model performance does not seem to suffer from the inclusion of a sixth TRACT variable compared to the respective reference model (e.g. 5 vs 3, 6 vs 4). This could indicate that the TRACT variables are being incorporated as descendent nodes rather than parents, and therefore do not exert much influence over the conditional probabilities of the other five variables. But even in the worst case scenario of a completely disconnected tract node, the marginal distribution of establishments by tract will encode useful information for the initial allocation of of synthetic businesses to tract geometries.

The last two columns of the table show goodness-of-fit metrics when only COUNTY, SECTOR, and SIZE are considered. Since all but two of the models control for COUNTY and SECTOR, either through stratification or evidence passing, this means that the scores really only reflect variation in SIZE. In general the trends are not very different from the scores computed using all five variables, but there is much greater variation between the scores for the best- and worst-performing models. The County-level models are shown to perform the worst, and by a significant margin. This holds true regardless of the metric that is used, suggesting that important information is lost by removing COUNTY from the structural learning process. Meanwhile the County-Sector scores are roughly $5x$ better. The unconstrained models in particular (9 and 11) are the only two for which the FT^2-based null hypothesis cannot be rejected, suggesting that the difference between the synthetic and reference distributions are not significantly different for these models, at least in regard to COUNTY, SECTOR, and SIZE.

One additional observation worth noting is that for the pooled models, the absence of COUNTY $\times$ SECTOR evidences does not seem to impact the fit of the synthetic populations that are generated. This is important because it means that the models can generate

					All Vars.		Census Vars.	
Model	**Stratification**	**Constrained**	**Tract**	**Evidences**	**SRMSE**	**FT^2**	**SRMSE**	**FT^2**
(1)	Pooled	–	–	–	0.887	58,940	0.113	5,671
(1)	Pooled	–	–	COUNTY, SECTOR	0.893	61,168	0.105	5,514
(2)	Pooled	√	–	–	0.893	59,561	0.111	5,620
(2)	Pooled	√	–	COUNTY, SECTOR	0.897	59,444	0.105	5,464
(3)	County	–	–	SECTOR	0.815	89,796	0.133	10,652
(4)	County	√	–	SECTOR	0.819	90,082	0.133	10,584
(5)	County	–	√	SECTOR	0.816	90,140	0.133	10,672
(6)	County	√	√	SECTOR	0.815	90,201	0.133	10,597
(7)	Sector	–	–	COUNTY	0.643	55,331	0.064	3,553
(8)	Sector	√	–	COUNTY	0.647	55,384	0.064	3,845
(9)	County-Sector	–	–	–	0.386	55,337	0.024	846**
(10)	County-Sector	√	–	–	0.409	59,941	0.025	1,224
(11)	County-Sector	–	√	–	0.384	55,323	0.024	860**
(12)	County-Sector	√	√	–	0.406	59,877	0.026	1,227

Table 4.4: Mean (n=10) goodness-of-fit metrics for synthetic populations generated using each of the twelve estimated models and NETS-based control totals (evidences). FT^2 scores marked with '**' are sufficient for *accepting* the null hypothesis that the observed and estimated populations are drawn from the same underlying distribution with 95% certainty or greater.

reasonably accurate synthetic populations without the use of *any* proprietary data apart from the total count of businesses (and the models themselves). Figure 4.6 shows that the distribution of fully joint frequency counts for one population generated in this way are tightly correlated with the original NETS data, with an R^2 value that is not significantly worse than that of a population generated from the most highly stratified and best fitting model (11). The synthetic population generated with Model (11) is itself shown to fit the NETS data remarkably well ($R^2 = 0.994$), and in Figure 4.7 a series of heatmaps shows very low error in even the pairwise distributions of the three variables which are *not* explicitly controlled for. Together these results demonstrate how both the design of Bayesian networks and the protocol that is used to draw samples from them can be tailored by the modeler depending on whether accuracy or data privacy is the priority.

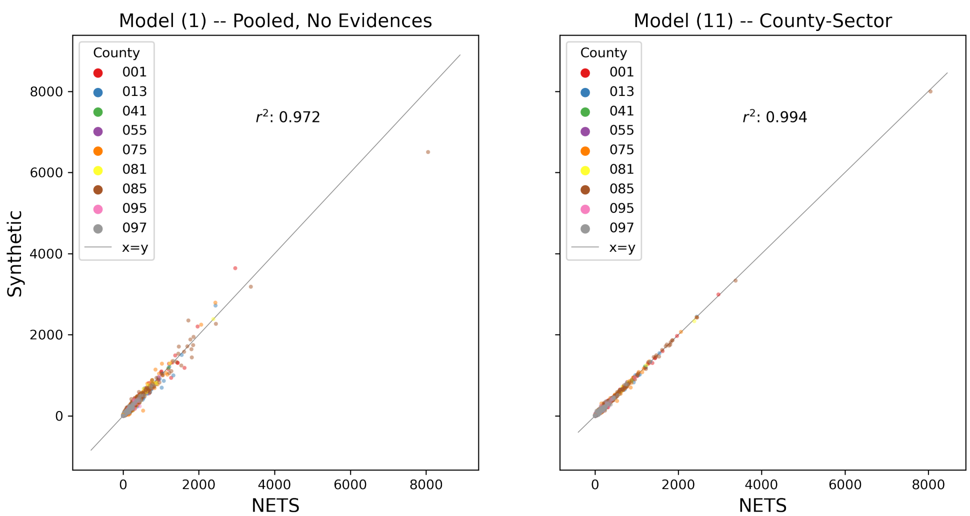


Figure 4.6: Fully joint (COUNTY × SECTOR × SIZE × AGE × SALES) frequency count comparisons for two synthetic populations generated from the best (right) and worst (left) performing models with NETS-based controls.

Census-based control totals

The goodness-of-fit scores for the synthetic populations generated using Census-based control totals are summarized in Table 4.5. The reference data in this case is created from the union of the US Census CBP and NES data products. Because AGE and SALES controls are not published by COUNTY, SECTOR, or SIZE across these data, only COUNTY, SECTOR, and SIZE are used to compute the scores shown here. The first thing to notice is that the SRMSEs are on average $5x$ greater than those in the "Census Vars." column in Table 4.4, while the FT^2 are between one and two orders of magnitude larger in scale. Interestingly the pooled models here perform the best in terms of SRMSE, while the County-level models perform best in terms of FT^2. This likely reflects the fact that sector-level variation was found to be the major source of disagreement between the NETS and Census reference datasets (Fig. 4.3b). Thus, the models which impose external sector-level controls via stratification (7-12) do not perform as well here compared to those which incorporate that variation into the model structure itself (1-6) and thus avoid overfitting to the NETS data. Overall, however, there is much less variation in goodness-of-fit between the various models compared to what is observed when using the NETS-based control totals.

Two trends are notable in Figure 4.8, which shows the frequency count distribution plot for a synthentic population generated using Model (3) and Census-based control totals. On the one hand, it seems that the model is over-predicting many rare observations. At the same time, the more common observations are underpredicted relative to the reference data, resulting in the line of best fit which gradually trends below the line $x = y$. Whether the counts are over- or under-predicted appears to depend quite heavily on the value of SIZE, which in the figure is represented by the color of each dot. It is possible, of course, to improve this fit by incorporating SIZE-based control totals as well, since both CBP and NES datasets make this value available. But without AGE or SALES variables available in the reference data to incorporate into the goodness-of-fit metrics, both SRMSE and FT^2 scores would go to zero and model fit could not be assessed. The reader will notice, however, that the slope of the blue line of best fit in Figure 4.8 is very similar to the one seen in Figure 4.3a comparing the NETS and Census frequency rates, which indicates that the bias observed in the synthetic population here is likely attributable to the inherent differences between the two datasets.

Figure 4.9 shows the joint frequency count distributions and error for every (SIZE $\times$ COUNTY) and (SIZE $\times$ SECTOR) tuple. Visually, the left and middle columns appear to match quite well. The far right column shows the residual error, which again makes it clear that size class is the variable which determines whether a particular establishment type is over- or under-predicted. The vast majority of the error is generated by establishments in the smallest three size classes, which mostly reflects the fact that there are relatively few establishments in the largest three classes to begin with.

Sub-County spatial allocation with Census-based controls

Lastly, I examine how different model specifications can be combined with different sets of control totals to allocate establishments to sub-County geographies during population synthesis. The result is a kind of spatial "upsampling" which I achieve in three ways: a) ZIP code-level control totals from Census ZIP Code Business Patterns (ZBP) data; b) estimation of County-level Bayesian networks which include a Census Tract node; and c) the combination of a) and b). Goodness-of-fit metrics are not particularly useful here since there are no tract-level Census data to compare against. Instead, Figure 4.10 presents a series of maps which allow for a more qualitative visual inspection of a subset of the results. In the analyses that follow only "payroll" establishments (i.e. those with employees other than the

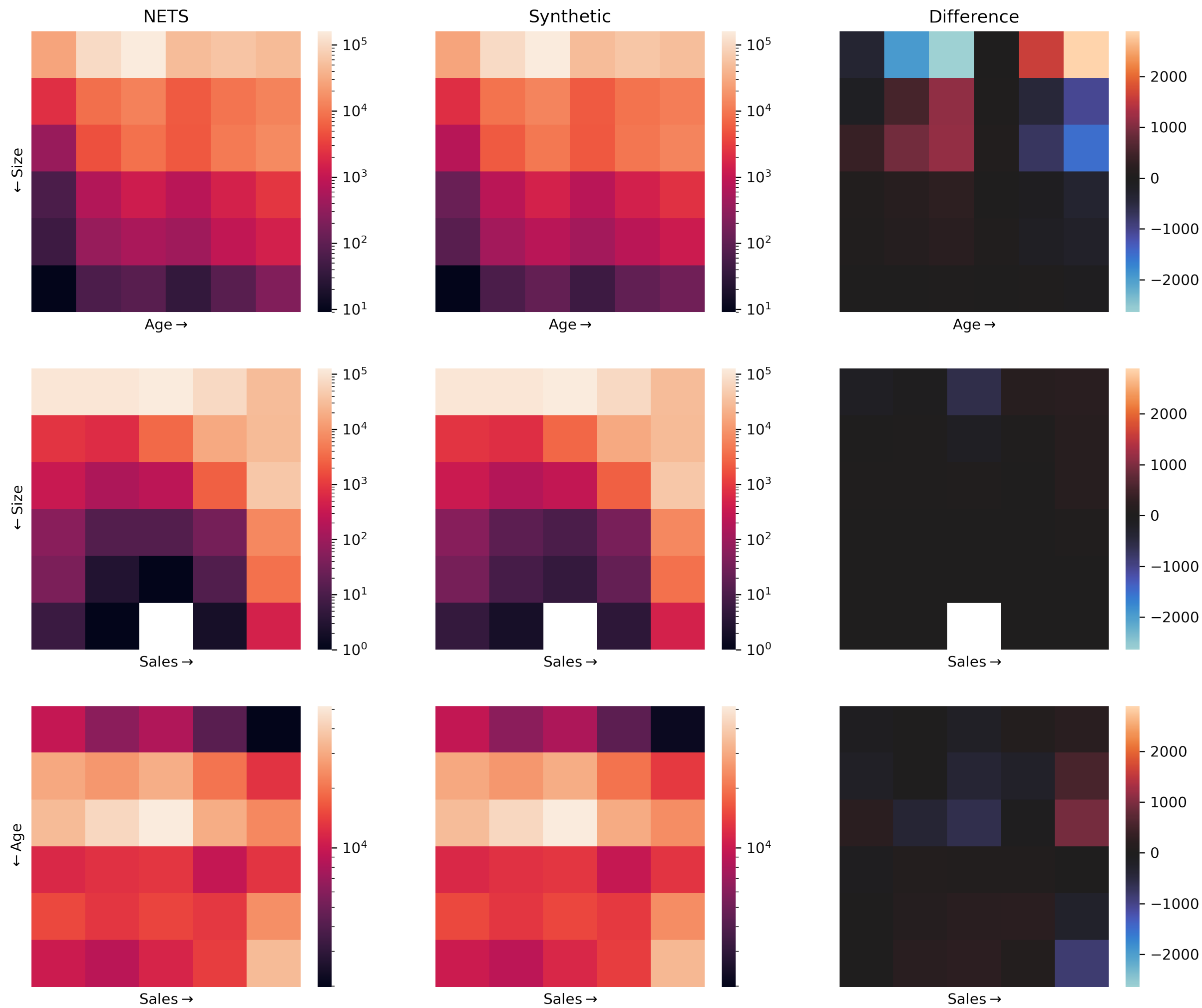


Figure 4.7: Observed (left) and synthetic (middle) frequency counts and frequency count error (right) for all pairs of non-control variables (SIZE, AGE, and SALES) in a synthetic population generated from Model (11) with NETS-based controls.

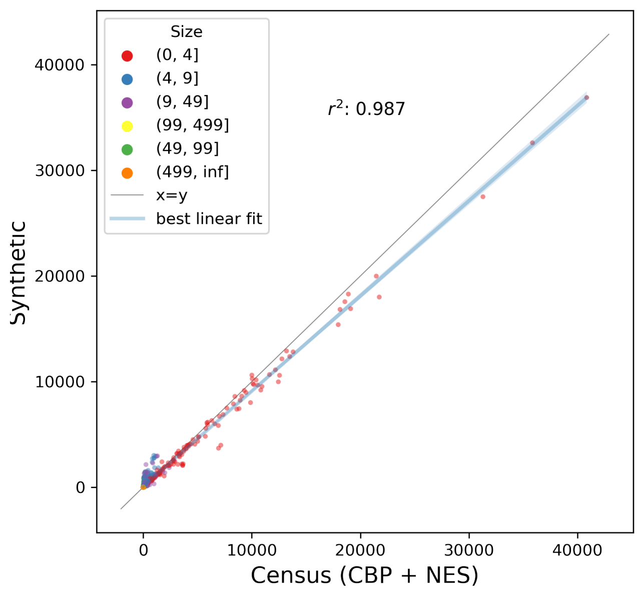


Figure 4.8: Joint (COUNTY × SECTOR × SIZE) frequency count comparison for a synthetic population generated from Model (3) with Census-based controls.

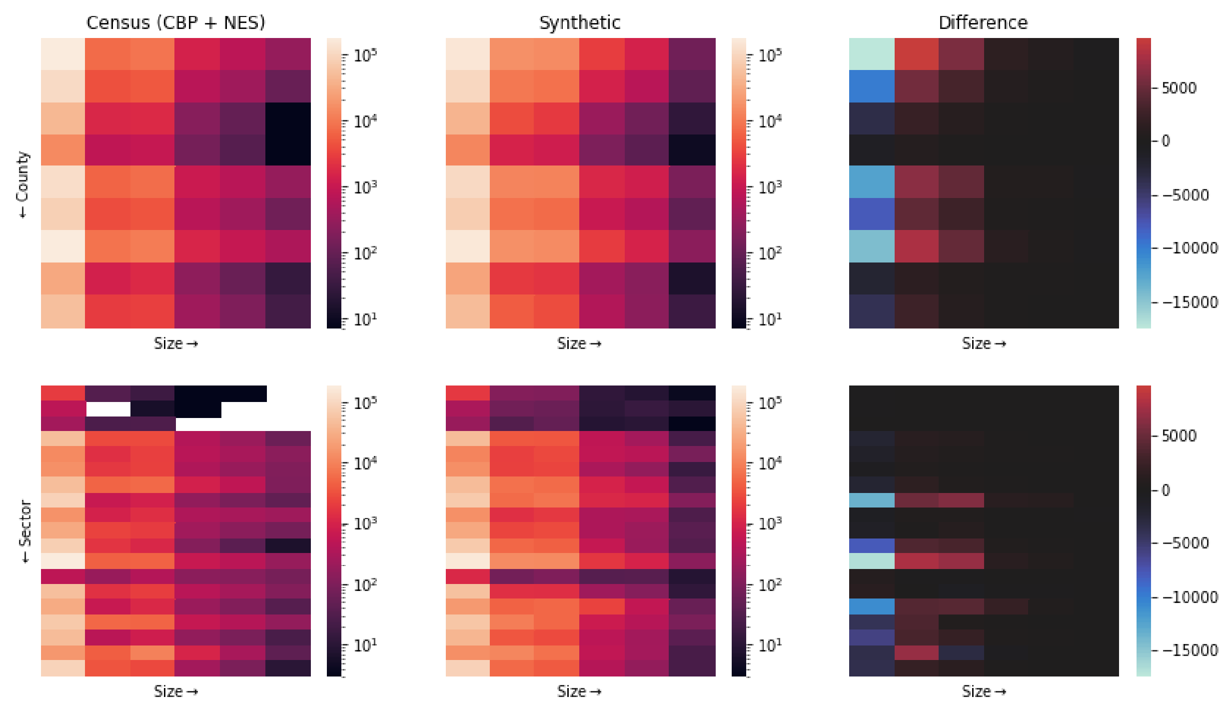


Figure 4.9: Observed (left) and synthetic (middle) frequency counts and frequency count error (right) for all pairs joint distribution of one uncontrolled variable (SIZE) with the two controlled variables (COUNTY, SECTOR) in a synthetic population generated from Model (3) with Census-based controls.

Model	**Stratification**	**Constrained**	**tract**	**Evidences**	**SRMSE**	**FT^2**
(1)	Pooled	–	–	COUNTY, SECTOR	0.496	130,761
(2)	Pooled	$\surd$	–	COUNTY, SECTOR	0.497	130,969
(3)	County	–	–	SECTOR	0.510	124,145
(4)	County	$\surd$	–	SECTOR	0.515	125,536
(5)	County	–	$\surd$	SECTOR	0.511	124,168
(6)	County	$\surd$	$\surd$	SECTOR	0.513	124,778
(7)	Sector	–	–	COUNTY	0.510	130,502
(8)	Sector	$\surd$	–	COUNTY	0.512	130,259
(9)	County-Sector	–	–	–	0.511	129,042
(10)	County-Sector	$\surd$	–	–	0.520	132,215
(11)	County-Sector	–	$\surd$	–	0.514	129,810
(12)	County-Sector	$\surd$	$\surd$	–	0.518	131,539

Table 4.5: Mean (n=10) goodness-of-fit metrics for synthetic populations generated using each of the twelve estimated models and Census-based (CBP + NES) control totals (evidences). Scoring is based on COUNTY, SECTOR, and SIZE variables only.

owner) are included. This is another limitation of the reference data due to the fact that the Census does not publish any nonemployer statistics at the ZIP code level. In order to account for this discrepancy I estimate a second set of Bayesian network models using only NETS records with non-zero employment.

In scenario a) a synthetic population is generated by drawing samples from a County-level model using ZIP code control totals from ZBP data. This is accomplished by enumerating a list of establishments by ZIP code and NAICS sector according to the ZBP totals and then assigning each establishment to its county using a ZIP-to-County crosswalk table published by the Census Longitudinal Employer-Household Dynamics (LEHD) program[10]. When a given ZIP code is associated with multiple counties, I probabilistically assign a county according to the percentage of the land area of the ZIP code contained in the county. I then generate a new set of control totals and draw samples from the appropriate County-level model according to the total count for each unique tuple of ZIP code, county, and sector values. Scenario b) simply utilizes one of the County-Sector level Bayesian networks from Table 4.5 which include Census Tract as an additional node. Thus, when samples are drawn according to CBP-based control totals, a TRACT value is automatically assigned. The third scenario incorporates both of these techniques, using ZIP-level control totals to draw samples from a model which has a TRACT node.

A selection of the results from each approach are shown in Figure 4.10, which uses a series of choropleth maps to compare the resulting distributions of construction sector businesses with between one and four employees. Since both County and Sector are used as controls, any discrepancy between the left and right columns is due to variation in Size alone. Tract-level frequency rates are shown instead of raw counts in order to account for the difference in total establishments between the NETS and Census datasets. The first row shows a very similar spatial distribution of establishments between the Scenario a) synthetic population and the ZIP-level data that were used as sample controls, although perhaps the large ZIP code in the southeast quadrant of the city has usurped establishments from some of the ZIP codes to the west. For Scenarios b) and c), however, there are no Census data to compare against, and thus tract-level data from NETS is used as reference instead. Both Scenarios b) and c) are both shown to match the NETS-based reference data quite well, but Scenario c) appears to fit the ZIP-level reference data better as evidenced by the the re-emergence of a high concentration of businesses on the western boundary of the city, and a less pronounced cluster to the east. These observations are primarily qualitative in nature, and not meant to serve as definitive characterizations of the methodologies described in this section. Rather they are meant to demonstrate the immense flexibility that Bayesian networks offer by allowing the modeler to generate reasonable population distributions at varying geographic resolutions.

[10] https://lehd.ces.census.gov/data/lodes/LODES7/ca/

4.5 Conclusions

Data access and privacy preservation are often competing goals when working with microscopic urban models. In this chapter I have demonstrated how a probabilistic approach to population synthesis can be used to resolve some of this tension. In particular, I showed how Bayesian networks trained on proprietary data can be used to create realistic populations of business establishment microdata which preserve much of the rich statistical information contained the original records but none of the sensitive data. These synthetic populations, as well as the underlying models themselves, can *in theory* be distributed among collaborators without violating data use agreements or disclosing information that can be used to re-identify individual establishments. The phrase "in theory" is emphasized here for two reasons. First, certain data providers may explicitly prohibit a user from distributing any derivatives of their proprietary data. Non-commercial endeavors like academic research are often excluded from such provisions – or at least their enforcement – however such legal questions are beyond the scope of this chapter.

The second reason is that in the approaches I have described above there is more work that would need to be done in order to guarantee that that re-indentification is not possible. The primary concern here is for business types which for certain geographies contain very few observations. As the distributions in Figure 4.2 show, establishments with more than 500 employees are particularly susceptible to this type of disclosure. One potential solution is further top-coding of the attribute classes that are used, for example combining the $(99, 499]$ and $(500+, \infty)$ size classes. The use of larger numbers of pseudocounts in the structural learning phase of network estimation is another way that variable amounts of noise can be injected into a model to meet the requirements of a particular use case. The authors of [92, 93] describe and implement a more sophisticated variation on this idea in which Laplace noise is added to the conditional probability tables in a Bayesian network that has already been estimated. More practical, perhaps, is an approach in which samples of business types with small numbers of observations are generated separately from the primary data synthesis step using less granular geographic controls. Bayesian networks give the modeler an enormous amount of control in this regard as discussed above in the section on sub-County spatial allocation.

In addition to refinements around privacy protection, there are opportunities to build on the work I have described here in order to improve its usefulness. For example, the network design can be expanded to jointly simulate populations of establishments and workers (jobs). Sun and Erath [87] describe one way this can be implemented through the use of hierarchical Bayesian networks, which they use to simultaneously synthesize a population of households and persons. The Census LEHD Origin-Destination Employment Statistics (LODES) data actually contains very detailed estimates of the distribution of jobs at the level of the Census block which could be used to calibrate such a model and to validate its results.

It is worth questioning why in the United States public business establishment microdata is so much more difficult to acquire than comparable datasets of persons and households. Federal law related to the publication of tax records may be the practical explanation, but an ethical question remains. In the meantime I have attempted to outline a path by which we can ensure that scientific discovery in this domain is not restricted to only those who can afford the price of entry.

(a)

(b)

(c)

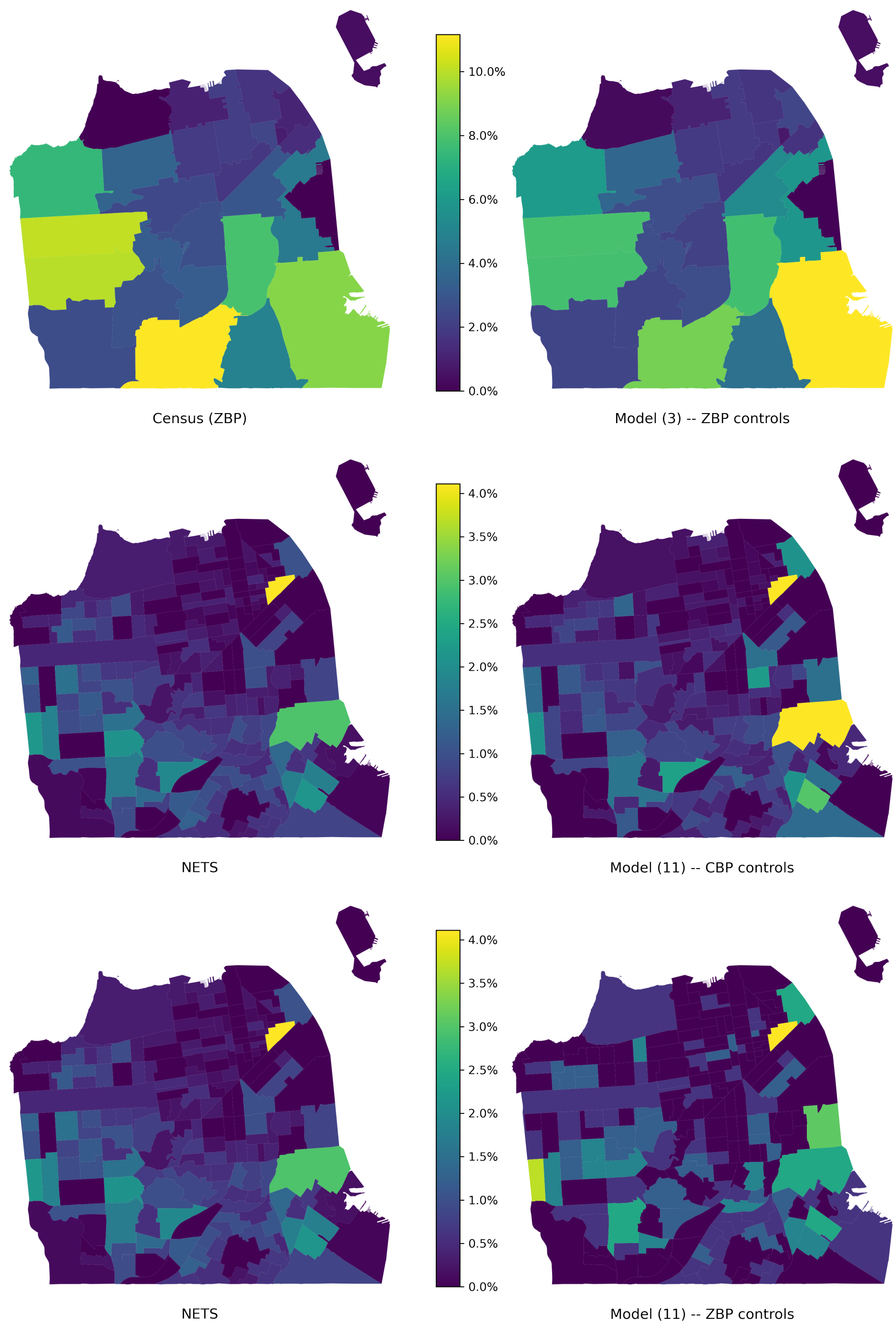


Figure 4.10: Frequency rate distribution of construction sector (NAICS=23) business establishments with 1-4 employees in San Francisco County by ZIP code (top) and census tract (middle, bottom).

Chapter 5

Conclusion

"The future is not a single grand vision or an inevitable consequence of trends, but rather an object of manipulation, discussion, debate, and eventually, perhaps, even consensus."

— Martin Wachs

"Prediction and freedom are opposites: to the extent we can predict your behavior, you are not free."

— Jarrett Walker

5.1 Summary

In the preceding collection of works, I have examined all three phases of the canonical workflow for disaggregate urban modeling, taking up issues that include rent control, simulation error, and the democratization of disaggregate data itself. In Chapter 2, I showed how the estimation of causal models can add much-needed evidence to the debate about a damaging effects of vacancy-control induced eviction, an economic phenomenon which has long been known to exist but has remained largely unexamined in the quantitative research. Chapter 3 calls attention to the significant gap between theory and practice in urban systems modeling, a gap that it attempts to bridge by proposing a novel method for quantifying forecast error due to sampling of alternatives and demonstrating its use. Lastly, Chapter 4 explores the potential of probabilistic models to resolve the conflict that exists between access to microdata and individuals' right to privacy. Although broad in scope, each of these studies shares a common purpose in contributing to the array of tools with which researchers, public servants, and planners may study the complex dynamics which govern the function and evolution of urban systems.

5.2 Modeling (in) the future

In recent years much of the initial excitement over the promise of a data-driven urbanism seems to have given way to skepticism about the ability of a system as complex as a city to be quantified at all. In the United States, at least, this skepticism has been earned the hard way. For example, in 2013 Edward Snowden's revelations about the widespread use of dragnet-style domestic surveillance programs by federal agencies caused many Americans, some perhaps for the first time, to consider the civil liberties that might be sacrificed in order to make a city quantifiable. Seven years later, the murder of George Floyd by Minneapolis police served as a stark reminder that for many others, these civil liberties have already been lost. What purpose can it possibly serve to build models of urban systems when the systems themselves appear to be so fundamentally broken? We, as planners, engineers, academics, and also community members, can no longer afford to avoid this question.

This means we must also grapple with the mounting evidence that suggests that in some cases it is the models themselves, particularly those used to peer into the future, which are to blame for our present dysfunction. As documented in [3, 5, 107–109] and elsewhere, our increasing reliance on model-based forecasting for everything from local project approvals to the procurement of federal dollars to build those projects has made them subject to manipulation. When it occurs, this manipulation takes one of two forms: i) deliberate deception by bureaucrats and corporate interests in search of "objective" data to muster support for initiatives in which they are already heavily invested; and ii) complacent modelers and technicians who encode their own cognitive biases into the models they design. Although the identification of useful correctives will depend on the type of manipulation that is being committed, both forms must be reckoned with if there is to be any place for this type of expertise in the future.

As troublesome as these revelations may be, they do not threaten the integrity of the planning-based professions so much as they provide an opportunity to re-establish that integrity moving forward. In fact, many positive developments have already begun to emerge as a result, including renewed interest in the use of *ex post* analysis to assess forecast accuracy [2, 110], reference class forecasting for reducing the influence of cognitive biases[111, 112], and the development of standards for quantifying and reporting uncertainty in forecasts [3, 113]. Emphasizing the role of uncertainty in forecasting is particularly important because it forces stakeholders and even practitioners themselves to consider not only what these models *can* do but also what they decidedly *cannot*. In the field of urban systems modeling certain technical challenges do remain that can be solved with more sophisticated methods, bigger datasets, and faster machines. Yet, at some point, we must also be willing to consider that the problem we face may not be the accuracy of our models as much as our refusal to accept that some things simply cannot be predicted. This is what Hartgen [3] describes as the dichotomy between hubris and humility which in many ways represents the choice faced by practitioners who must now decide for themselves which vision for the future of demand forecasting to embrace.

5.3 In uncertainty, possibility

Interesting things start to happen when planners allow themselves to consider an uncertain future. Suddenly, a straight-line projection of the status quo onto a point-forecast future becomes a branching network of emergent alternatives that could not have otherwise been imagined (Fig. 5.1). Yes, the incorporation of uncertainty into urban modeling is useful for

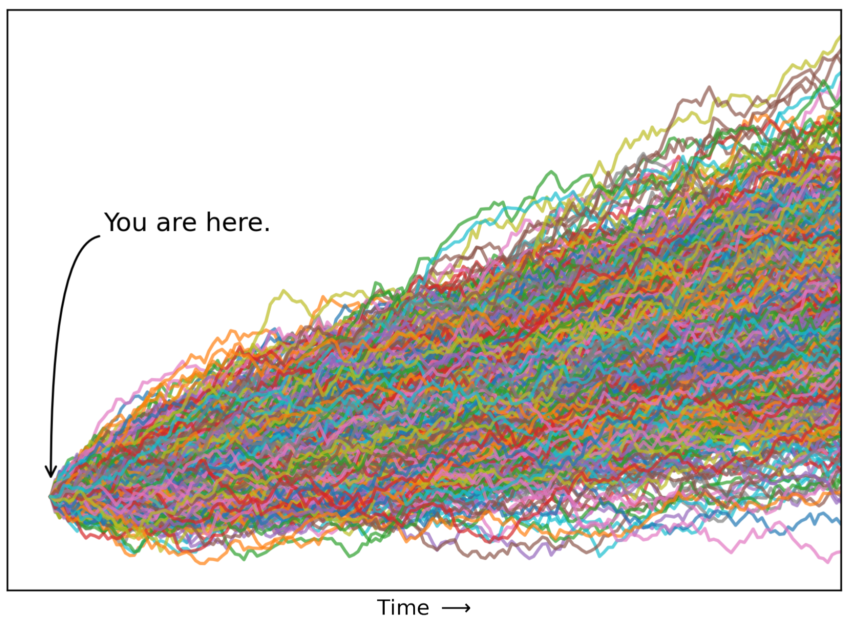


Figure 5.1: A branching network of emergent futures.[1]

encouraging stakeholders to think about forecasts in distributional terms, and to temper their expectations with caveats and confidence bands. But at a moment in history when the status quo feels so untenable, one is tempted to look beyond the 95% confidence interval and instead to the margins, and to find there possibility there rather than low-probability outcomes to ignore. In other words, whereas yesterday's planner may have viewed uncertainty with a certain contempt, the planner of today may see a way out.

Perhaps, then, in modeling uncertainty, there is also an opportunity to envision alternative futures; and maybe, working backwards from the most desirable outcomes, we can make inferences about the interventions we must make in order to achieve them. Instead of dispassionately building models to tell us what our cities *will* look in the year 2050, let us first ask ourselves what we *want* them to look like, and then use the quantitative tools at our disposal to guide us there.

[1]Figure based on [114]

Appendices

Appendix A

Table A1: "Just Causes" for Eviction Under the San Francisco Rent Ordinance. Adapted from Asquith (2019), Chapter 37 of the San Francisco Administrative Code and https://sfrb.org/topic-no-201-overview-just-cause-evictions.

Just Cause	Eviction Category	Whole-Building Eviction	Relocation Payment Required	Change in Rent Control	Change Takes Effect
"Non-payment of rent, habitual late payment of rent, or frequent bounced checks"	Breach of lease	No	No	Reset	Next lease
"Failure to cure a substantial breach of a rental agreement or lease"	Breach of lease	No	No	Reset	Next lease
"Nuisance or substantial interference with the comfort, safety, or enjoyment of the landlord or other tenants in the building, the nature of which must be severe, continuing or reoccurring in nature"	Breach of lease	No	No	Reset	Next lease
"Illegal use of a rental unit, not including (a) the mere occupancy of an unwarranted rental unit or (b) a single violation of San Francisco's short-term rental law (Chapter 41A) that is cured by the tenant within 30 days of written notice by the landlord"	Breach of lease	No	No	Reset	Next lease

"The tenant has refused, after written request by the landlord, to execute a written extension or renewal of an expired rental agreement for a further term of like duration and under terms that are materially the same as the previous agreement"	Breach of lease	No	No	Reset	Next lease
"The tenant has refused, after written notice to cease, to allow the landlord access to the rental unit as required by state or local law"	Breach of lease	No	No	Reset	Next lease
"The only tenant residing in the unit at the end of the term of the rental agreement is a subtenant not approved by the landlord. (Note that approval need not be in writing and may be implied from the landlord's conduct)"	Breach of lease	No	No	Reset	Next lease
"Owner-occupancy (OMI) or, in limited circumstances, occupancy by a member of the landlord's immediate family"	No-fault	No	Yes	Reset	3 years
"The landlord seeks to recover possession in good faith in order to sell the unit in accordance with a condominium conversion approved under the San Francisco subdivision ordinance"	No-fault	Yes	Yes	Full decontrol	Immediate
Demolition	No-fault	Likely	Yes	Full decontrol	Immediate

"To perform capital improvements or rehabilitation work that will make the unit temporarily uninhabitable while the work is performed – the tenant must be allowed to reoccupy the unit immediately after the work is completed"	No-fault	Likely	Yes	None	–
"To perform substantial rehabilitation of a building that is at least 50 years old and essentially uninhabitable, provided that the cost of the proposed work is at least 75% of the cost of new construction"	No-fault	Yes	Yes	Full decontrol	Immediate
"To withdraw all rental units in a building from the rental market under the state Ellis Act"	No-fault	Yes	Yes	Reset	5 years
"To perform lead remediation/abatement work required by San Francisco Health Code Articles 11 or 26"	No-fault	Likely	Yes	None	–
"Demolition related to development agreement entered into by the City under Chapter 56 of the San Francisco Administrative Code"	No-fault	Likely	Yes	Full decontrol	Immediate
"Expiration of 'Good Samaritan' occupancy agreement. A 'Good Samaritan' tenancy occurs when a tenant is displaced from a rental unit due to an emergency or disaster and the landlord agrees to provide the tenant a temporary replacement unit at a reduced rent"	No-fault	No	No	Reset	Next lease

Table A2: Eviction frequency by type and built-tear cut-off

Eviction Type	**Eviction Category**	**Built-year < 1980**			
		Y		*N*	
		count	%	*count*	%
unknown	unknown/other	5066	36.8	44	23.3
Breach of Lease Agreement	Breach of lease	3031	22.0	66	34.9
Nuisance	Breach of lease	1621	11.8	28	14.8
Owner Move-in (OMI)	No-fault	1177	8.5	3	1.6
Capital Improvement	No-fault	518	3.8	–	–
Non-payment of Rent	Breach of lease	487	3.5	11	5.8
Habitual Late Payment of Rent	Breach of lease	399	2.9	4	2.1
ELLIS	No fault	346	2.5	–	–
Illegal Use of Unit	Breach of lease	199	1.4	1	0.5
Breach of Lease Agreement Nuisance	Breach of lease	182	1.3	3	1.6
Roommate Living in Same Unit	Breach of lease	119	0.9	3	1.6
Unapproved Subtenant	Breach of lease	110	0.8	1	0.5
Other	unknown/other	108	0.8	18	9.5
Demolition	No fault	94	0.7	–	–
Denial of Access to Unit	Breach of lease	54	0.4	2	1.1
Nuisance Illegal Use of Unit	Breach of lease	46	0.3	–	–
Breach of Lease Agreement Illegal Use of Unit	Breach of lease	34	0.2	–	–
Breach of Lease Agreement Nuisance Illegal Use of Unit	Breach of lease	24	0.2	–	–
Non-payment of Rent Breach of Lease Agreement	Breach of lease	14	0.1	–	–
Failure to Sign Lease Renewal	Breach of lease	19	0.1	–	–
Denial of Access to Unit Breach of Lease Agreement	Breach of lease	10	0.1	–	–
Condo Conversion	No fault	17	0.1	–	–
Habitual Late Payment of Rent Breach of Lease Agreement	Breach of lease	11	0.1	–	–
Non-payment of Rent Habitual Late Payment of Rent	Breach of lease	10	0.1	1	0.5
Unapproved Subtenant Breach of Lease Agreement	Breach of lease	8	0.1	–	–
Substantial Rehabilitation	No fault	7	0.1	–	–
Breach of Lease Agreement Roommate Living in Same Unit	Breach of lease	1	0.0	–	–
Breach of Lease Agreement Other	Breach of lease	5	0.0	1	0.5

Roommate Living in Same Unit Nuisance	Breach of lease	6	0.0	–	–
Development Agreement	No fault	1	0.0	–	–
Denial of Access to Unit Breach of Lease Agreement Nuisance	Breach of lease	4	0.0	–	–
Unapproved Subtenant Breach of Lease Agreement Nuisance	Breach of lease	2	0.0	–	–
Unapproved Subtenant Breach of Lease Agreement Nuisance Illegal Use of Unit	Breach of lease	1	0.0	–	–
Unapproved Subtenant Illegal Use of Unit	Breach of lease	1	0.0	–	–
Unapproved Subtenant Nuisance	Breach of lease	1	0.0	–	–
Unapproved Subtenant Nuisance Illegal Use of Unit	Breach of lease	2	0.0	–	–
Nuisance Capital Improvement	Breach of lease	1	0.0	–	–
Non-payment of Rent Nuisance	Breach of lease	2	0.0	–	–
Non-payment of Rent Unapproved Subtenant	Breach of lease	1	0.0	–	–
Non-payment of Rent Other	Breach of lease	1	0.0	–	–
Denial of Access to Unit Other	Breach of lease	2	0.0	–	–
Non-payment of Rent Habitual Late Payment of Rent Nuisance	Breach of lease	1	0.0	–	–
Non-payment of Rent Habitual Late Payment of Rent Breach of Lease Agreement Nuisance	Breach of lease	1	0.0	–	–
Non-payment of Rent Denial of Access to Unit	Breach of lease	1	0.0	–	–
Non-payment of Rent Breach of Lease Agreement Nuisance	Breach of lease	1	0.0	–	–
Denial of Access to Unit Breach of Lease Agreement Nuisance Illegal Use of Unit	Breach of lease	3	0.0	–	–
Breach of Lease Agreement Failure to Sign Lease Renewal	Breach of lease	3	0.0	–	–
Denial of Access to Unit Breach of Lease Agreement Other	Breach of lease	1	0.0	–	–
Habitual Late Payment of Rent Nuisance	Breach of lease	4	0.0	–	–
Habitual Late Payment of Rent Breach of Lease Agreement Nuisance	Breach of lease	6	0.0	–	–
Habitual Late Payment of Rent Breach of Lease Agreement Failure to Sign Lease Renewal	Breach of lease	1	0.0	–	–
Denial of Access to Unit Nuisance	Breach of lease	2	0.0	–	–
Good Samaritan Tenancy Ends	No fault	2	0.0	–	–
Lead Remediation	No fault	6	0.0	–	–

Habitual Late Payment of Rent Other	Breach of lease	–	–	1	0.5
Habitual Late Payment of Rent Roommate Living in Same Unit Nuisance	Breach of lease	–	–	1	0.5
Nuisance Other	Breach of lease	–	–	1	0.5

Bibliography

[1] Manissa M. Maharawal and Erin McElroy. "The anti-eviction mapping project: Counter mapping and oral history toward bay area housing justice". In: *Annals of the American Association of Geographers* 108.2 (2018). Publisher: Taylor & Francis, pp. 380–389.

[2] Morten Skou Nicolaisen and Patrick Arthur Driscoll. "Ex-post evaluations of demand forecast accuracy: A literature review". In: *Transport Reviews* 34.4 (2014). Publisher: Taylor & Francis, pp. 540–557.

[3] David T. Hartgen. "Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling". In: *Transportation* 40.6 (2013). Publisher: Springer, pp. 1133–1157.

[4] Martin Wachs. "Forecasting versus envisioning: A new window on the future". In: *Journal of the American Planning Association* 67.4 (2001). Publisher: Taylor & Francis, pp. 367–372.

[5] Charles L. Marohn Jr. *Confessions of a Recovering Engineer: Transportation for a Strong Town*. en. Google-Books-ID: eKo_EAAAQBAJ. John Wiley & Sons, Aug. 2021. ISBN: 978-1-119-69925-5.

[6] Manuel Pastor, Vanessa Carter, and Maya Abood. "Rent Matters: What are the Impacts of Rent Stabilization Measures?" In: *Los Angeles: USC Dornsife Program for Environmental and Regional Equity* (2018).

[7] Elaine Herscher. *Berkeley Renters Vow to Resist End of Controls*. en-US. Section: News. July 1995. URL: https://www.sfgate.com/news/article/Berkeley-Renters-Vow-to-Resist-End-of-Controls-3027981.php (visited on 07/01/2021).

[8] Joseph Gyourko and Peter Linneman. "Rent controls and rental housing quality: A note on the effects of New York City's old controls". In: *Journal of Urban Economics* 27.3 (1990). Publisher: Academic Press, pp. 398–409.

[9] Michael P. Murray et al. "Analyzing rent control: the case of Los Angeles". In: *Economic Inquiry* 29.4 (1991). Publisher: Wiley Online Library, pp. 601–625.

[10] Choon-Geol Moon and Janet G. Stotsky. "The effect of rent control on housing quality change: a longitudinal analysis". In: *Journal of Political Economy* 101.6 (1993). Publisher: The University of Chicago Press, pp. 1114–1148.

[11] John Nagy. "Increased Duration and Sample Attrition in New York City's Rent Controlled Sector". In: *Journal of Urban Economics* 38.2 (1995). Publisher: Elsevier, pp. 127–137.

[12] John Nagy. "Do vacancy decontrol provisions undo rent control?" In: *Journal of Urban Economics* 42.1 (1997). Publisher: Elsevier, pp. 64–78.

[13] Dirk W. Early and Edgar O. Olsen. "Rent control and homelessness". In: *Regional Science and Urban Economics* 28.6 (1998). Publisher: Elsevier, pp. 797–816.

[14] Allan D. Heskin, Ned Levine, and Mark Garrett. "The effects of vacancy control: A spatial analysis of four California cities". In: *Journal of the American Planning Association* 66.2 (2000). Publisher: Taylor & Francis, pp. 162–176.

[15] Edward L. Glaeser and Erzo FP Luttmer. "The misallocation of housing under rent control". In: *American Economic Review* 93.4 (2003), pp. 1027–1046.

[16] Robert Krol and Shirley Svorny. "The effect of rent control on commute times". In: *Journal of Urban Economics* 58.3 (2005). Publisher: Elsevier, pp. 421–436.

[17] David P. Sims. "Out of control: What can we learn from the end of Massachusetts rent control?" In: *Journal of Urban Economics* 61.1 (2007). Publisher: Elsevier, pp. 129–151.

[18] Thomas Hazlett. "Rent Controls and the housing Crisis". In: *Resolving the Housing Crisis: Government Policy, Decontrol and the Public Interest* 277 (1982). Publisher: Ballinger Cambridge, MA.

[19] Peter Navarro. "Rent control in cambridge, mass". In: *The Public Interest* 78 (1985). Publisher: National Affairs, inc., p. 83.

[20] Blair Jenkins. "Rent control: Do economists agree?" In: *Econ journal watch* 6.1 (2009).

[21] Richard Arnott. "Time for revisionism on rent control?" In: *Journal of economic perspectives* 9.1 (1995), pp. 99–120.

[22] Joseph Gyourko, Christopher Mayer, and Todd Sinai. "Superstar cities". In: *American Economic Journal: Economic Policy* 5.4 (2013), pp. 167–99.

[23] Brian Asquith. "Do Rent Increases Reduce the Housing Supply under Rent Control? Evidence from Evictions in San Francisco". In: (2019).

[24] David H. Autor, Christopher J. Palmer, and Parag A. Pathak. "Housing market spillovers: Evidence from the end of rent control in Cambridge, Massachusetts". In: *Journal of Political Economy* 122.3 (2014). Publisher: University of Chicago Press Chicago, IL, pp. 661–717.

[25] Andreas Mense, Claus Michelsen, and Konstantin Kholodilin. "Empirics on the causal effects of rent control in Germany". In: (2018). Publisher: Kiel, Hamburg: ZBW-Leibniz-Informationszentrum Wirtschaft.

[26] Rebecca Diamond, Timothy McQuade, and Franklin Qian. *The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco*. Tech. rep. National Bureau of Economic Research, 2018.

[27] Joshua D. Ambrosius et al. "Forty years of rent control: Reexamining New Jersey's moderate local policies after the great recession". In: *Cities* 49 (2015). Publisher: Elsevier, pp. 121–133.

[28] Dan Immergluck et al. "Evictions, large owners, and serial filings: Findings from Atlanta". In: *Housing Studies* 35.5 (2020). Publisher: Taylor & Francis, pp. 903–924.

[29] Sam Levin. *Sick, elderly, pregnant: the California renters being evicted even during the pandemic*. en. Section: World news. Apr. 2020. URL: http://www.theguardian.com/world/2020/apr/17/sick-elderly-pregnant-the-california-renters-being-evicted-even-during-the-pandemic (visited on 06/29/2021).

[30] Donald L. Thistlethwaite and Donald T. Campbell. "Regression-discontinuity analysis: An alternative to the ex post facto experiment." In: *Journal of Educational psychology* 51.6 (1960). Publisher: American Psychological Association, p. 309.

[31] Duncan D. Chaplin et al. "The internal and external validity of the regression discontinuity design: A meta-analysis of 15 within-study comparisons". In: *Journal of Policy Analysis and Management* 37.2 (2018). Publisher: Wiley Online Library, pp. 403–429.

[32] Guido W. Imbens and Thomas Lemieux. "Regression discontinuity designs: A guide to practice". In: *Journal of econometrics* 142.2 (2008). Publisher: Elsevier, pp. 615–635.

[33] Nathaniel Decker. "How Landlords of Small Rental Properties Decide Who Gets Housed and Who Gets Evicted". In: *Urban Affairs Review* (2021). Publisher: SAGE Publications Sage CA: Los Angeles, CA, p. 10780874211041513.

[34] Philip ME Garboden and Eva Rosen. "Serial filing: How landlords use the threat of eviction". In: *City & Community* 18.2 (2019). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 638–661.

[35] Lillian Leung, Peter Hepburn, and Matthew Desmond. "Serial eviction filing: civil courts, property management, and the threat of displacement". In: *Social Forces* (2020).

[36] Dirk W. Early. "Rent control, rental housing supply, and the distribution of tenant benefits". In: *Journal of Urban Economics* 48.2 (2000). Publisher: Elsevier, pp. 185–204.

[37] Edward L. Glaeser. "Does rent control reduce segregation?" In: *Available at SSRN 348084* (2002).

[38] Richard Ault and Richard Saba. "The economic effects of long-term rent control: The case of New York City". In: *The Journal of Real Estate Finance and Economics* 3.1 (1990). Publisher: Springer, pp. 25–41.

[39] Peter Linneman. “The effect of rent control on the distribution of income among New York City renters”. In: *Journal of Urban Economics* 22.1 (1987). Publisher: Elsevier, pp. 14–34.

[40] Matthew Desmond. “Eviction and the reproduction of urban poverty”. In: *American journal of sociology* 118.1 (2012), pp. 88–133.

[41] Matthew Desmond and Carl Gershenson. “Housing and employment insecurity among the working poor”. In: *Social Problems* 63.1 (2016). Publisher: Oxford University Press, pp. 46–67.

[42] Matthew Desmond and Rachel Tolbert Kimbro. “Eviction’s fallout: housing, hardship, and health”. In: *Social forces* 94.1 (2015). Publisher: Oxford University Press, pp. 295–324.

[43] Joseph Gyourko and Peter Linneman. “Equity and efficiency aspects of rent control: An empirical study of New York City”. In: *Journal of urban Economics* 26.1 (1989). Publisher: Academic Press, pp. 54–74.

[44] Herman B. Leonard. *Regulation of the Cambridge Housing Market: Its Goals and Effects*. 1981.

[45] Kirk McClure. “An evaluation of the rent control policy of Cambridge, Massachusetts.” PhD Thesis. Massachusetts Institute of Technology, 1978.

[46] Andrew Gelman and Guido Imbens. “Why high-order polynomials should not be used in regression discontinuity designs”. In: *Journal of Business & Economic Statistics* 37.3 (2019). Publisher: Taylor & Francis, pp. 447–456.

[47] Applied Survey Research. *San Francisco Homeless Count & Survey*. Tech. rep. San Francisco Department of Homelessness and Supportive Housing, 2019. URL: `https://hsh.sfgov.org/wp-content/uploads/2020/01/2019HIRDReport_SanFrancisco_FinalDraft-1.pdf` (visited on 07/01/2021).

[48] Neal J. Riley. *Condo conversion law OKd by S.F. board*. en-US. Section: Bay Area & State. June 2013. URL: `https://www.sfgate.com/bayarea/article/Condo-conversion-law-OKd-by-S-F-board-4594985.php` (visited on 07/01/2021).

[49] Jason D. Lemp, Laura B. McWethy, and Kara M. Kockelman. “From aggregate methods to microsimulation: Assessing benefits of microscopic activity-based models of travel demand”. In: *Transportation Research Record* 1994.1 (2007). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 80–88.

[50] Michael Wegener. “From macro to micro—how much micro is too much?” In: *Transport Reviews* 31.2 (2011), pp. 161–177.

[51] Jason D. Lemp and Kara M. Kockelman. “STRATEGIC SAMPLING FOR LARGE CHOICE SETS IN ESTIMATION AND APPLICATION”. In: *Transportation Research Part A* 46.2012 (2012), pp. 602–613.

[52] Sriharsha Nerella and Chandra R. Bhat. “Numerical analysis of effect of sampling of alternatives in discrete choice models”. In: *Transportation Research Record* 1894.1 (2004). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 11–19.

[53] Minhui Zeng, Ming Zhong, and John Douglas Hunt. “Analysis of the Impact of Sample Size, Attribute Variance and Within-Sample Choice Distribution on the Estimation Accuracy of Multinomial Logit Models Using Simulated Data”. In: *Journal of Systems Science and Systems Engineering* 27.6 (2018). Publisher: Springer, pp. 771–789.

[54] Daniel McFadden. “Modeling the choice of residential location”. In: *Transportation Research Record* 673 (1978).

[55] Kenneth E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[56] Frank S. Koppelman and Chausie Chu. “Effect of sample size on disaggregate choice model estimation and prediction”. In: *Transportation Research Record* 944 (1983).

[57] Moshe Ben-Akiva and Thawat Watanatada. “Application of a continuous spatial choice logit model”. In: *Structural analysis of discrete data with econometric applications* (1981). Publisher: Mit Press Cambridge, MA, pp. 320–343.

[58] Emma Frejinger, Michel Bierlaire, and Moshe Ben-Akiva. “Sampling of alternatives for route choice modeling”. In: *Transportation Research Part B: Methodological* 43.10 (2009). Publisher: Elsevier, pp. 984–994.

[59] Min-Tang Li et al. “Geographically stratified importance sampling for the calibration of aggregated destination choice models for trip distribution”. In: *Transportation research record* 1935.1 (2005). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 85–92.

[60] Kenneth E. Train, Daniel L. McFadden, and Moshe Ben-Akiva. “The demand for local telephone service: A fully discrete model of residential calling patterns and service choices”. In: *The RAND Journal of Economics* (1987). Publisher: JSTOR, pp. 109–123.

[61] Sophie Curtis-Ham et al. “The Importance of Importance Sampling: Exploring Methods of Sampling from Alternatives in Discrete Choice Models of Crime Location Choice”. In: *Journal of Quantitative Criminology* (2021). Publisher: Springer, pp. 1–29.

[62] Joe Castiglione, Joel Freedman, and Mark Bradley. “Systematic investigation of variability due to random simulation error in an activity-based microsimulation forecasting model”. In: *Transportation Research Record* 1831.1 (2003). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 76–88.

[63] Jonas Eliasson. “The influence of accessibility on residential location”. In: *Residential location choice*. Springer, 2010, pp. 137–164.

[64] Paul Waddell et al. "Microsimulation of urban development and location choices: Design and implementation of UrbanSim". In: *Networks and spatial economics* 3.1 (2003). Publisher: Springer, pp. 43–67.

[65] Jonathan Jones, Isabelle Thomas, and Dominique Peeters. "Forecasting jobs location choices by Discrete Choice Models: A sensitivity analysis to scale and implications for LUTI models". In: *Region* 2.1 (2015), pp. 67–93.

[66] Cristian Angelo Guevara and Moshe E. Ben-Akiva. "Change of scale and forecasting with the control-function method in logit models". In: *Transportation Science* 46.3 (2012). Publisher: INFORMS, pp. 425–437.

[67] Guy H. Orcutt et al. *Policy exploration through microanalytic simulation*. The Urban Insitute, 1976.

[68] Joan L. Walker. "Making household microsimulation of travel and activities accessible to planners". In: *Transportation research record* 1931.1 (2005). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 38–48.

[69] Elizabeth Koehler, Elizabeth Brown, and Sebastien J.-PA Haneuse. "On the assessment of Monte Carlo error in simulation-based statistical analyses". In: *The American Statistician* 63.2 (2009). Publisher: Taylor & Francis, pp. 155–162.

[70] Liming Wang, Paul Waddell, and Maren Outwater. "Incremental Integration of Land Use and Activity-Based Travel Modeling: Workplace Choices and Travel Demand". In: *Transportation Research Record: Journal of the Transportation Research Board* 2255 (2011). 1, pp. 1–10. URL: `http://trrjournalonline.trb.org/doi/abs/10.3141/2255-01` (visited on 02/02/2017).

[71] Paul Waddell. "UrbanSim: Modeling urban development for land use, transportation, and environmental planning". In: *Journal of the American planning association* 68.3 (2002). Publisher: Taylor & Francis, pp. 297–314.

[72] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. 2018. URL: `http://github.com/google/jax`.

[73] Veira, Tim. *Algorithms for sampling without replacement — Graduate Descent*. 2019. URL: `https://timvieira.github.io/blog/post/2019/09/16/algorithms-for-sampling-without-replacement/` (visited on 03/12/2022).

[74] Moshe E. Ben-Akiva and Steven R. Lerman. *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press, 1985.

[75] C. Angelo Guevara and Moshe E. Ben-Akiva. "Sampling of alternatives in logit mixture models". In: *Transportation Research Part B: Methodological* 58 (2013). Publisher: Elsevier, pp. 185–198.

[76] Xin Ye et al. "A methodology to match distributions of both household and person attributes in the generation of synthetic populations". In: *88th Annual Meeting of the transportation research Board, Washington, DC*. 2009.

[77] Satkartar K. Kinney et al. “Towards unrestricted public use business microdata: The synthetic longitudinal business database”. In: *International Statistical Review* 79.3 (2011). Publisher: Wiley Online Library, pp. 362–384.

[78] Paul Salvini and Eric J. Miller. “ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems”. In: *Networks and spatial economics* 5.2 (2005). Publisher: Springer, pp. 217–234.

[79] Urbansim, Inc. *Data Overview — UrbanSim Cloud Platform 3.12.1 documentation.* URL: `https://cloud.urbansim.com/docs/general/documentation/urbansim%20block%20model%20data.html#block-agents-section` (visited on 08/12/2022).

[80] Laron Smith, Richard Beckman, and Keith Baggerly. *TRANSIMS: Transportation analysis and simulation system*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1995.

[81] Peter Vovsha et al. *New features of population synthesis*. Tech. rep. 2015.

[82] W. Edwards Deming and Frederick F. Stephan. “On a least squares adjustment of a sampled frequency table when the expected marginal totals are known”. In: *The Annals of Mathematical Statistics* 11.4 (1940). Publisher: JSTOR, pp. 427–444.

[83] Alan Geoffrey Wilson and Carol E. Pownall. “A new representation of the urban system for modelling and for the study of micro-level interdependence”. In: *Area* (1976). Publisher: JSTOR, pp. 246–254.

[84] Paul Williamson, Mark Birkin, and Phil H. Rees. “The estimation of population microdata by using data from small area statistics and samples of anonymised records”. In: *Environment and Planning A* 30.5 (1998). Publisher: SAGE Publications Sage UK: London, England, pp. 785–816.

[85] Justin Ryan, Hanna Maoh, and Pavlos Kanaroglou. “Population synthesis: Comparing the major techniques using a small, complete population of firms”. In: *Geographical Analysis* 41.2 (2009). Publisher: Wiley Online Library, pp. 181–203.

[86] Zengyi Huang and Paul Williamson. “A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata”. In: *Department of Geography, University of Liverpool* (2001).

[87] Lijun Sun and Alexander Erath. “A Bayesian network approach for population synthesis”. In: *Transportation Research Part C: Emerging Technologies* 61 (2015). Publisher: Elsevier, pp. 49–62.

[88] Rebecca Diamond, Tim McQuade, and Franklin Qian. “The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco”. In: *American Economic Review* 109.9 (2019), pp. 3365–94.

[89] Satkartar K. Kinney, Jerome P. Reiter, and Javier Miranda. “Synlbd 2.0: improving the synthetic longitudinal business database”. In: *Statistical Journal of the IAOS* 30.2 (2014). Publisher: IOS Press, pp. 129–135.

[90] Zoubin Ghahramani. “An introduction to hidden Markov models and Bayesian networks”. In: *Hidden Markov models: applications in computer vision*. World Scientific, 2001, pp. 9–41.

[91] Judea Pearl. “Bayesian networks”. In: *The handbook of brain theory and neural networks*. 1998, pp. 149–153.

[92] Jun Zhang et al. “Privbayes: Private data release via bayesian networks”. In: *ACM Transactions on Database Systems (TODS)* 42.4 (2017). Publisher: ACM New York, NY, USA, pp. 1–41.

[93] Haoyue Ping, Julia Stoyanovich, and Bill Howe. “Datasynthesizer: Privacy-preserving synthetic datasets”. In: *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. 2017, pp. 1–5.

[94] Donald W. Walls. *National Establishment Time-Series Database©: Data Overview*. en. SSRN Scholarly Paper. Rochester, NY, Nov. 2007. DOI: `10.2139/ssrn.1022962`. URL: `https://papers.ssrn.com/abstract=1022962` (visited on 07/11/2022).

[95] Srinath Ravulaparthy et al. “Exploratory Analysis of Spatial Hierarchical Clustering in Los Angeles County, California: Relationship of Opportunity-Based Accessibility, Reported Land Values, and Resident Characteristics”. In: *Transportation research record* 2307.1 (2012). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 132–140.

[96] Cynthia L. Curl et al. “Associations of organic produce consumption with socioeconomic status and the local food environment: Multi-Ethnic Study of Atherosclerosis (MESA)”. In: *PloS one* 8.7 (2013). Publisher: Public Library of Science San Francisco, USA, e69778.

[97] Amy H. Auchincloss et al. “Improving retrospective characterization of the food environment for a large region in the United States during a historic time period”. In: *Health & place* 18.6 (2012). Publisher: Elsevier, pp. 1341–1347.

[98] David I. Levine and Michael W. Toffel. “Quality management and job quality: How the ISO 9001 standard for quality management systems affects employees and employers”. In: *Management Science* 56.6 (2010). Publisher: INFORMS, pp. 978–996.

[99] David Neumark, Brandon Wall, and Junfu Zhang. “Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series”. In: *The Review of Economics and Statistics* 93.1 (2011). Publisher: The MIT Press, pp. 16–29.

[100] David Neumark, Junfu Zhang, and Brandon Wall. “Employment dynamics and business relocation: New evidence from the National Establishment Time Series”. In: *Aspects of worker well-being*. Vol. 26. Emerald Group Publishing Limited, 2007, pp. 39–83.

[101] Keith Barnatchez, Leland Dod Crane, and Ryan Decker. "An assessment of the national establishment time series (nets) database". In: (2017). Publisher: FEDS working paper.

[102] John Haltiwanger, Ron S. Jarmin, and Javier Miranda. "Who creates jobs? Small vs. large vs. young". In: *NBER working paper* 16300 (2011).

[103] Jacob Schreiber. "Pomegranate: fast and flexible probabilistic modeling in python". In: *The Journal of Machine Learning Research* 18.1 (2017). Publisher: JMLR. org, pp. 5992–5997.

[104] Trivellore E. Raghunathan, Jerome P. Reiter, and Donald B. Rubin. "Multiple imputation for statistical disclosure limitation". In: *Journal of official statistics* 19.1 (2003). Publisher: Statistics Sweden (SCB), p. 1.

[105] Daniele Casati et al. "Synthetic population generation by combining a hierarchical, simulation-based approach with reweighting by generalized raking". In: *Transportation Research Record* 2493.1 (2015). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 107–116.

[106] David Voas and Paul Williamson. "Evaluating goodness-of-fit measures for synthetic microdata". In: *Geographical and Environmental Modelling* 5.2 (2001). Publisher: Taylor & Francis, pp. 177–200.

[107] Martin Wachs. "When planners lie with numbers". In: *American Planning Association. Journal of the American Planning Association* 55.4 (1989). Publisher: Taylor & Francis Inc., p. 476.

[108] Bent Flyvbjerg. "Design by deception: The politics of megaproject approval". In: *Harvard Design Magazine, Spring/Summer* 22 (2005), pp. 50–59.

[109] Robert Bain. "Toll roads: big trouble down under". In: *Infrastructure Journal* 17 (2013).

[110] National Academies of Sciences, Engineering, and Medicine 2020. *Traffic Forecasting Accuracy Assessment Research*. English. Washington, DC: The National Academies Press., 2016. URL: `https://doi.org/10.17226/25637`.

[111] Daniel Kahneman and Amos Tversky. *Intuitive prediction: Biases and corrective procedures*. Tech. rep. Decisions and Designs Inc Mclean Va, 1977.

[112] Bent Flyvbjerg. "Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice". In: *European planning studies* 16.1 (2008). Publisher: Taylor & Francis, pp. 3–21.

[113] Yong Zhao and Kara Maria Kockelman. "The propagation of uncertainty through travel demand models: an exploratory analysis". In: *The Annals of regional science* 36.1 (2002). Publisher: Springer, pp. 145–163.

[114] Kingpie. *Creating a fanchart from a series of Monte Carlo projections in Python.* Forum post. Feb. 2021. URL: `https://stackoverflow.com/q/66146705` (visited on 08/12/2022).