The Economics of Zoning

by

Tate Twinam

BA, Economics and Mathematics, New College of Florida, 2010

MA, Economics, University of Pittsburgh, 2011

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences

in partial fulfillment of the requirements for the degree of

PhD in Economics

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Tate Twinam

It was defended on

April 14th, 2015

and approved by

Allison Shertzer, Assistant Professor, Department of Economics

Dennis Epple, Professor, Tepper School of Business, Carnegie Mellon University

and dissertation co–directors:

Arie Beresteanu, Associate Professor, Department of Economics

Randall Walsh, Associate Professor, Department of Economics

The Economics of Zoning

Tate Twinam, PhD

University of Pittsburgh, 2015

Abstract

This dissertation consists of three chapters. Chapter 1 examines the identification power of assumptions that formalize the notion of complementarity in the context of a nonparametric bounds analysis of treatment response. I extend the literature on partial identification via shape restrictions by exploiting cross–dimensional restrictions on treatment response when treatments are multidimensional; the assumption of supermodularity can strengthen bounds on average treatment effects in studies of policy complementarity. I combine this restriction with a statistical independence assumption to derive improved bounds on treatment effect distributions, aiding in the evaluation of complex randomized controlled trials. I show how complementarities arising from treatment effect heterogeneity among subpopulations can be incorporated through supermodular instrumental variables to strengthen identification of treatment effects in studies with one or multiple treatments. I use these results to examine the long–run effects of zoning on the evolution of land use patterns.

Chapter 2 considers the determinants of land use regulation. Zoning has been cited as a discriminatory policy tool by critics, who argue that ordinances are used to deter the entry of minority residents into majority neighborhoods through density restrictions (exclusionary zoning) and locate manufacturing activity in minority neighborhoods (environmental racism). However, identifying discrimination in these regulations is complicated by the fact that land use and zoning have been co-evolving for nearly a century. We employ a novel approach to overcome this challenge, studying the introduction of comprehensive zoning in Chicago. We find evidence of a pre-cursor to exclusionary zoning as well as inequitable treatment in industrial use zoning.

Chapter 3 examines the impact of residential density and mixed land use on crime using a unique high-resolution dataset from Chicago over the period 2008-2013. I employ a novel instrumental variable strategy based on the city's 1923 zoning code. I find that commercial uses lead to more street crime in their immediate vicinity, with relatively weak spillovers. However, this effect is strongly offset by density; dense mixed use areas are actually safer than typical residential areas. Additionally, much of the commercial effect is driven by liquor stores and late-hour bars. I discuss the implications for zoning policy.

Contents

List of Tables

List of Figures

Chapter 1

Complementarity and identification

1.1 Introduction

Complementarities arise naturally in many economic problems, often manifesting as policy interactions or treatment effect heterogeneity among observed subgroups of a population. This paper examines how assumptions that formalize the notion of complementarity can aid in the identification of treatment effects. I employ a nonparametric bounds approach, where identification is driven by qualitative restrictions rooted in economic theory or empirical evidence rather than strong functional form or unconfoundedness assumptions. This approach will yield interval estimates of parameters of interest; however, informative bounds are often preferable to precise (but wrong) estimates obtained under incorrect assumptions. Partial identification tools have been fruitfully applied to a wide range of empirical problems.[1]

In particular, I explore the identification power yielded by assuming that individual treatment response functions exhibit *supermodularity* when treatments are multidimensional. This assumption allows one to construct more informative bounds in studies of policy com-

[1] Examples of applied partial identification studies include: Giustinelli (2011) and Tsunao and Usui (2014) on the returns to education, Kreider and Pepper (2007) on disability and employment, Bhattacharya, Shaikh and Vytlacil (2008, 2012) on the mortality effects of Swan–Ganz catheterization, Kreider and Hill (2009) on the effect of universal health insurance on medical expenditures, Pepper (2000) on the intergenerational transmission of welfare receipt, Manski and Nagin (1998) on sentencing and recidivism, and Gundersen, Kreider and Pepper (2012) on the health effects of the National School Lunch Program.

plementarity, which are typically stymied by the absence of pseudo–experimental variation in the assignment of multiple treatments. I also show how complementarities arising from interactions between treatment effects and observable covariates can be formalized as *supermodular instrumental variables* to improve bounds on average treatment effects. This novel instrumental variable approach is broadly applicable to studies with one or multiple treatments. Complementarity is frequently invoked in economics, but studies of its identification power have been limited to very specific contexts. This paper develops general results applicable to program evaluation in a wide range of empirical situations. I illustrate the use of my results in an empirical application on the long–run effects of zoning on land use patterns.

Typically, empirical studies seek to estimate the effect of a single treatment on one or more outcome variables. However, the effect of a treatment may vary substantially with the value of other (endogenously–determined) treatment variables. When policymakers have multiple tools at their disposal, understanding how different policies enhance or offset each other is crucial. If the positive impact of some policy intervention is substantially larger when combined with a second (costly) intervention, a measure of the magnitude of this difference is necessary for a proper cost–benefit analysis. The supermodularity and submodularity assumptions I propose can aid in quantifying how policy impacts differ with the associated policy environment.[2]

For example, unemployment relief is a multidimensional policy, involving a choice of both potential benefit duration and the wage replacement rate. Lalive, Van Ours and Zweimüller (2006) show both theoretically and empirically that these two dimensions are complementary, with simultaneous increases in both the replacement rate and potential benefit duration leading to an increase in unemployment duration substantially larger than the sum of the effects measured individually for particular subgroups. The Lalive et al. study exploits

[2]As more treatments are considered, the data are necessarily less informative about each individual treatment. Thus, the researcher faces a trade–off where richer treatment spaces allow for more interesting questions but generally lead to less precise answers.

variation in both dimensions of unemployment relief that has the characteristics of a natural experiment, but such opportunities are very rare. Pseudo–experimental variation along multiple policy dimensions is far less common than similar variation in individual policies. This has arguably led to the overwhelming focus on the effects of policies in isolation. The partial identification tools developed here, which are applicable in the absence of any unusual pseudo–experimental policy variation, should enhance the ability of researchers to measure treatment effect heterogeneity due to policy complementarities in a wide range of contexts.[3] I illustrate the use of the shape restrictions developed here in a study of how the long–run effect of commercial zoning on land use patterns varies with different restrictions on building density.

Relatedly, responses to a treatment may differ among subpopulations defined by observable covariates. Many recent experimental studies have discussed the importance of treatment effect heterogeneity between subgroups (Bitler, Gelbach and Hoynes 2006, 2008, 2014, Djebbari and Smith 2008, Feller and Holmes 2009). I show how qualitative information about such treatment effect heterogeneity leads naturally to supermodular instrumental variables, which can help narrow the bounds on average treatment effects in the same manner as a traditional instrumental variable or a monotone instrumental variable.[4] Supermodular instrumental variables can be applied in the case of a single treatment or multiple treatments, making them a potentially valuable addition to the range of identifying assumptions available to applied researchers. I demonstrate their utility in the empirical illustration in section 1.6.

While the bulk of the paper focuses on identification using non–experimental data, the assumptions developed in this paper can be applied in the evaluation of complex randomized controlled trials (RCTs) involving multiple treatments. In a discussion of program evalua-

[3] The sensitivity of effects to the surrounding policy environment may partly explain the wide variation in estimates of treatment effects for similar policies in different contexts found in many literatures; see, for example, the discussion in Lalive et al. (2006) on the effects of unemployment benefit policies on re–employment rates. See also Gelman (2013) for a related discussion.

[4] See Manski and Pepper (2000) and section 1.4.

tion, Heckman, Smith and Clements (1997) note that, even in an RCT, many parameters of interest are not point–identified, such as the proportion of the population receiving a treatment who benefit from the treatment. Heckman et al. observe that classical probability inequalities like the Fréchet-Hoeffding copula bounds are not very informative. The structural supermodularity and submodularity assumptions I introduce have implications for the entire distribution of treatment effects, so they can be used to obtain stronger bounds. Since average treatment effects are identified in this context, the supermodularity or submodularity of average effects can be established, and this can be used to provide some justification for the stronger structural assumptions. Similarly, the validity of supermodular instrumental variable assumptions can be established and used to justify stronger *quantile supermodular instrumental variable* assumptions, which can also be applied in the case of a single treatment.

The literature on partial identification is extensive.[5] Many of the contributions of Charles Manski and coauthors are relevant to the results developed below; I review them as appropriate. The literature on complementarity and identification is relatively small. Molinari and Rosen (2008) connect supermodularity to identification in the context of game estimation. They show that the approach of Aradillas-Lopez and Tamer (2008) applies to games with supermodular payoff functions. Eeckhout and Kircher (2011) find that they cannot identify (using wage data alone) whether or not the technology of a firm is supermodular, i.e., whether or not more productive workers sort towards more productive jobs. Graham, Imbens and Ridder (2014) analyze how reallocations of indivisible heterogeneous inputs across production units (leaving a potentially complementary input fixed) may affect average output. They discuss identification and estimation of the effects of a variety of correlated matching rules. Lazzati (2014) uses monotone comparative statics to partially identify treatment response in the presence of endogenous social interactions. Supermodularity arises naturally in this context when individual outcomes are increasing with the outcomes of oth-

[5] See Manski (2003) for a comprehensive overview.

ers. The shape restrictions I propose have been used in the context of estimation to improve efficiency; Beresteanu (2005, 2007) considers the efficiency gains from imposing a variety of restrictions, including supermodularity and submodularity.

The remainder of the paper is organized as follows. In section 1.2, I outline the formal setup used throughout the paper. In section 1.3, I present novel shape restrictions and the resulting bounds on average treatment effects. In section 1.4, I discuss instrumental variable assumptions and derive bounds on average potential outcomes and average treatment effects. In section 1.5, I combine shape restrictions and instrumental variables with statistical independence assumptions to derive bounds on cumulative distribution functions of treatment effects. I conclude with an empirical illustration on the long–run effects of zoning on land use patterns in section 1.6.

1.2 Notation and Setup

Individuals are drawn from a population I. The set I, the Borel σ–algebra of subsets of I denoted by $\mathcal{I}$, and the probability measure P together form a probability space $(I, \mathcal{I}, P)$. Every individual $i \in I$ is associated with a vector of covariates $x^i \in X$ and a vector of realized treatments $z^i \in T$, where T is the treatment set.[6] Since I focus on the identification of treatment effects in the presence of multiple treatments, I will discuss in detail the structure I adopt for the treatment space.

Definition. A nonempty partially ordered set V is a *lattice* if, for any $v, v' \in V$,

- V contains the join (least upper bound) of v and v', denoted by $v \vee v'$, and
- V contains the meet (greatest lower bound) of v and v', denoted by $v \wedge v'$.

Examples of lattices include $\mathbb{R}^2$, $\mathbb{Z} \times \mathbb{R}$, and $\{0, 1\}^n$ for $n \in \mathbb{N}$. The meet and join operations depend on the particular order imposed on the lattice; for example, the join

[6] I use superscripts to refer to individuals and reserve subscripts to denote vector components.

of $(2,0)$ and $(1,1)$ in $\mathbb{R}^2$ is equal to $(2,1)$ under the product order and $(2,0)$ under the lexicographic order. An element v of a lattice V is the top (bottom) of V if $v' \leq v$ ($v \leq v'$) for all $v' \in V$; if v is not the top or bottom, it is in the interior. If the top (or bottom) of a lattice exists, it is unique.

Definition. For a lattice V, a nonempty subset $U \subseteq V$ is a *sublattice* of V if, for any $u, u' \in U$, U contains the meet and join of u and u' in V.

Sublattices will be useful when I consider assumptions that do not hold globally on T. The following assumption, which I maintain throughout the paper, describes the structure imposed on the treatment space:

Assumption. The treatment space T is such that

- $T \subseteq \mathbb{R}^L$ with $L \in \mathbb{N}$,
- T is partially ordered under the product order, and
- T is a nonempty lattice.

The product order on T implies that $t \leq t'$ iff $t_l \leq_l t'_l$ for each l. If $t, t' \in T$ are incomparable, i.e., $t_l < t'_l$ and $t'_{l'} < t_{l'}$ for some l, l', I write $t \parallel t'$. The advantage of the lattice assumption is the notational clarity it provides when I employ supermodularity and submodularity to formalize how the marginal effect on the response variable of changes in some dimensions of the treatment depend on the values of other dimensions of the treatment.

This specification is flexible enough to allow for a wide variety of treatment types. In this paper, I restrict attention to discrete treatments, as these are most commonly encountered in practice. Dimensions of the treatment may be binary or multi–valued (Cattaneo 2010). Most of the results extend straightforwardly to the case of continuous treatments. In practice, however, the application to continuous treatments is hampered by the fact that, as the number of treatments increases, the data alone are increasingly uninformative about the effect of each individual treatment. The relationship between the complexity of the treatment set and the amount that can be learned from the data is an issue I will discuss further in the next section.

Every individual i is associated with a (measurable) response function $y^i(\cdot) : T \to Y \in \mathbb{R}$ mapping treatments into outcomes $y^i(t)$.[7] $z^i \in T$ is the treatment that i actually receives, so $y^i(z^i)$ is individual i's realized outcome, $\{y^i(t)\}_{t \neq z^i}$ are individual i's counterfactual outcomes, and $\{y^i(t)\}_{t \in T}$ are individual i's potential outcomes. Throughout the paper, I maintain the stable unit treatment value assumption,[8] which says that individuals' potential outcomes $\{y(t)\}_{t \in T}$ do not depend on other individuals' realized treatments (Rubin 1978).

1.3 Shape Restrictions

In this section, I explore the identifying power of shape restrictions that formalize complementarity and substitutability, with an emphasis on the identification of average treatment effects. I review shape restrictions proposed in the previous literature before moving on to the novel restrictions I propose. Using these assumptions, I derive bounds on average treatment effects for both simple and complex treatment spaces.

Throughout, I assume that there exist $\underline{K}, \overline{K} \in \mathbb{R}$ such that $\underline{K} \leq y(t) \leq \overline{K}$ for all t; these are global bounds on response functions. As Manski (1990) observes, this is not as restrictive as it seems; for example, if y is a probability, it is naturally bounded between zero and one. In the absence of these global bounds, the results below will generally be uninformative. All well–defined expectations are assumed to exist; if an expectation $\mathrm{E}[\, y(t) \mid z = t' \,]$ is ill–defined because the event $z = t'$ is off the support of z, I establish the convention that $\mathrm{E}[\, y(t) \mid z = t' \,] \, P(z = t') \equiv 0$.

Manski (1989) introduced the no–assumption bounds on $E[\, y(t) \,]$. The no–assumption upper bound is the average of $E[\, y(t) \mid z = t \,]$ and the global upper bound $\overline{K}$, weighted respectively by $P(z = t)$ and $P(z \neq t)$; likewise for the lower bound. Since they are typically

[7] I suppress i when referring to arbitrary response functions, covariates, or realized treatments.

[8] This assumption is alternatively referred to as noninterference by Cox (1958) and individualistic treatment response by Manski (2013).

wide, research has focused on other credible assumptions that yield additional identifying power.

Manski (1997) studied the identification power of assumptions on the shape of individual response functions; in particular, he considered restricting response functions to be monotone, semi–monotone, or concave–monotone. I reproduce the semi–monotone treatment response assumption here, in my notation:

Assumption SMTR (Semi–monotone treatment response)**.** Response functions exhibit *semi–monotone treatment response* on $S \subseteq T$ if, for all $t, t' \in S$,

$$t \leq t' \Longrightarrow y\left(t\right) \leq y\left(t'\right)$$

If S is a chain,[9] then this assumption is referred to as *monotone treatment response* (MTR).

Manski (1997) uses this assumption to derive bounds on numerous quantities, including average and quantile treatment effects. He motivated MTR by considering traditional demand analysis, where researchers often make strong parametric assumptions but do not exploit the less–controversial assumption that demand curves are downward sloping. SMTR removes the need for a totally ordered T. It has the same identification power regardless of whether $T \subseteq \mathbb{R}$ or $T \subseteq \mathbb{R}^L$ for $L > 1$, except that in the latter case, it is possible that $t \parallel t'$. Bhattacharya et al. (2008) derives bounds using SMTR without assuming a particular direction of monotonicity. Tsunao and Usui (2014) study the identification power of concave–monotone treatment response combined with monotone treatment selection (discussed in section 1.4).

MTR and SMTR are within–dimension restrictions on the response functions. Additional identification power can be obtained from cross–dimension restrictions, where the marginal effect of a change in some dimensions of the treatment variable depends on the values of the other dimensions:

[9] A subset S of a partially ordered set T is a *chain* if it is totally ordered under the inherited order.

Assumption SPM (Supermodularity)**.** Response functions are *supermodular* on a sublattice $S \subseteq T$ if, for all $t, t' \in S$,

$$y(t') + y(t) \leq y(t \vee t') + y(t \wedge t') \tag{1.3.1}$$

They are *strictly supermodular* when the inequality is strict.

Assumption SBM (Submodularity)**.** Response functions are *submodular* on a sublattice $S \subseteq T$ if, for all $t, t' \in S$,

$$y(t') + y(t) \geq y(t \vee t') + y(t \wedge t') \tag{1.3.2}$$

They are *strictly submodular* when the inequality is strict.

SPM is a formalization of the notion of complementarity. If two dimensions of a treatment, say t_1 and t_2, are complementary, then the magnitude of the change in the response variable due to an increase in t_1 is increasing with t_2. Thus, the two dimensions of the treatment act to amplify each others marginal effects. In the case of a linear model

$$y = \alpha + \beta t_1 + \delta t_2 + \gamma t_1 t_2 \tag{1.3.3}$$

supermodularity is equivalent to the sign restriction $\gamma > 0$. SBM is a formalization of substitutability, the case where elements of the treatment may mitigate each others effects. Returning to (1.3.3), submodularity is equivalent to the sign restriction $\gamma < 0$. If both supermodularity and submodularity hold, response functions are said to be *modular*. Since these assumptions can be applied on sublattices of T, it is possible to allow some dimensions of a treatment to be complements while those same dimensions are substitutes with other dimensions.

As I discussed in the introduction, Lalive et al. (2006) study the Austrian labor market and find that the two dimensions of unemployment relief, potential benefit duration and the wage replacement rate, are complementary (strongly for some groups, weakly for others). This finding could motivate the use of SPM in assessments of these and similar policies in other contexts where the pseudo–random variation they exploit is absent.

Neumark and Wascher (2011) provide another example of policy complementarity in a study on the interaction between the Earned Income Tax Credit (EITC) and the minimum wage. They find that a higher minimum wage enhances the positive effect of the EITC on the labor supply of single mothers; they find the opposite effect for childless individuals, suggesting a crowding–out effect. These findings suggest that assumptions SPM and SBM, respectively for each subgroup, could be applied in other studies on how the effect of minimum wage changes are influenced by the EITC or similar programs.

Another naturally multidimensional policy is zoning. Zoning laws typically regulate many aspects of the built environment; most broadly, they regulate both what types of uses are allowed (commercial, industrial, etc.) and how densely land can be developed (lot coverage of buildings, maximum height, etc.). The effects of specific zoning policies vary widely with the overall policy bundle. Shertzer, Twinam and Walsh (2014b) study the long–run impact of the initial zoning of Chicago on a variety of modern land use outcomes. The long–run impact of historical commercial zoning on present–day commercial land use turns out to hinge critically on the associated density restrictions; commercial zoning has a substantially larger effect when paired with low–density zoning. This motivates the assumption of SBM in the empirical application in section 1.6.

Since assumptions SPM and SBM can be applied on sublattices of T, it is possible to allow some dimensions of a treatment to be complements while those same dimensions are substitutes with other dimensions. For example, consider $T = \{0,1\}^3$ under the product order. Assumption SPM on the two sublattices

$$S_1 = \{(1,1,0),(1,0,0),(0,1,0),(0,0,0)\}$$

$$S_2 = \{(1,1,1),(1,0,1),(0,1,1),(0,0,1)\}$$

combined with assumption SBM on the five sublattices

$$S_3 = \{(1,0,0),(1,0,1),(0,0,0),(0,0,1)\}$$
$$S_4 = \{(1,1,0),(1,1,1),(0,1,0),(0,1,1)\}$$
$$S_5 = \{(0,1,0),(0,1,1),(0,0,0),(0,0,1)\}$$
$$S_6 = \{(1,1,0),(1,1,1),(1,0,0),(1,0,1)\}$$
$$S_7 = \{(1,1,0),(1,1,1),(0,0,0),(0,0,1)\}$$

yields complementarity between the first two dimensions of the treatment but substitutability between the first two (individually and jointly) and the third.

An inspection of the no–assumption bounds reveals that the amount one can learn about $E[\,y(t)\,]$ or $E[\,y(t) - y(t')\,]$ from the data alone depends on $P(z = t)$ and, in the latter case, $P(z = t')$. If $P(z = t)$ is small, the data are practically uninformative about $E[\,y(t)\,]$.[10] Thus, the researcher faces a trade–off where richer treatment spaces (which entail a larger number of treatments) allow for more interesting questions but generally lead to less precise answers. Adding "nuisance" dimensions to the treatment space that allow for the application of additional SPM or SBM assumptions will generally not aid in the identification of treatment effects of interest.

In propositions 1 and 2, I show how SPM and SBM can be used to compute bounds on the expectations of average treatment effects. In general, these bounds will improve upon the no–assumption bounds in the case of multidimensional treatments; with only a single treatment, SPM and SBM have no identifying power. The simplest nontrivial lattice treatment space is $T = \{(0,0),(1,0),(0,1),(1,1)\}$, which corresponds to a two–dimensional binary treatment. The following result shows the implications of supermodularity for identification on this simple treatment space:

[10] In the case where one or more of the dimensions of the treatment are continuous, the data are necessarily uninformative about almost all of the treatments. This motivates my restriction to discretely–valued treatments.

Proposition 1. *Assume that* $T = \{(0,0), (1,0), (0,1), (1,1)\}$. *Assume that SPM holds on* T. *Then, the bounds*

$$\begin{aligned}
&\mathrm{E}\left[\, y(1,0) \mid z = (1,0) \,\right] P(z = (1,0)) + \underline{K} P(z \neq (1,0)) \\
&- \mathrm{E}\left[\, y(0,0) \mid z = (0,0) \,\right] P(z = (0,0)) - \overline{K} P(z \neq (0,0)) \\
&\leq \mathrm{E}\left[\, y(1,0) - y(0,0) \,\right] \leq \qquad (1.3.4) \\
&\mathrm{E}\left[\, y(1,1) \mid z = (1,1) \,\right] P(z = (1,1)) + \mathrm{E}\left[\, y(1,0) \mid z = (1,0) \,\right] P(z = (1,0)) \\
&+\overline{K} P(z \in \{(0,0), (0,1)\}) - \mathrm{E}\left[\, y(0,1) \mid z = (0,1) \,\right] P(z = (0,1)) \\
&- \mathrm{E}\left[\, y(0,0) \mid z = (0,0) \,\right] P(z = (0,0)) - \underline{K} P(z \in \{(1,0), (1,1)\})
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{E}\left[\, y(1,1) \mid z = (1,1) \,\right] P(z = (1,1)) + \mathrm{E}\left[\, y(1,0) \mid z = (1,0) \,\right] P(z = (1,0)) \\
&+\underline{K} P(z \in \{(0,0), (0,1)\}) - \mathrm{E}\left[\, y(0,1) \mid z = (0,1) \,\right] P(z = (0,1)) \\
&- \mathrm{E}\left[\, y(0,0) \mid z = (0,0) \,\right] P(z = (0,0)) - \overline{K} P(z \in \{(1,0), (1,1)\}) \\
&\leq \mathrm{E}\left[\, y(1,1) - y(0,1) \,\right] \leq \qquad (1.3.5) \\
&\mathrm{E}\left[\, y(1,1) \mid z = (1,1) \,\right] P(z = (1,1)) + \overline{K} P(z \neq (1,1)) \\
&- \mathrm{E}\left[\, y(0,1) \mid z = (0,1) \,\right] P(z = (0,1)) - \underline{K} P(z \neq (0,1))
\end{aligned}$$

are sharp.[11] *The no–assumption bounds remain sharp for* $\mathrm{E}\left[\, y(1,1) - y(0,0) \,\right]$ *and each average potential outcome* $\mathrm{E}\left[\, y(\cdot) \,\right]$ *defined on* T.

Proof of proposition 1. First, I show that SPM does not improve upon the no–assumption bounds on potential outcomes. SPM implies that

$$y^i(1,0) + y^i(0,1) \leq y^i(1,1) + y^i(0,0)$$

[11] There is no guarantee that these bounds will be nonempty; if an assumption implies that the bounds on the parameter of interest are empty, the assumption is falsified by the data. This caveat applies to all the results that follow.

For each i, exactly one of these outcomes is observed. The unobserved terms may take any value in $\left[\underline{K}, \overline{K}\right]$. When $z^i \neq (1,0)$, there are three cases to consider. If $z^i = (1,1)$, then SPM implies

$$y^i\,(1,0) \leq \overline{K} \leq y^i\,(1,1) + \overline{K} - \underline{K}$$

If $z^i = (0,1)$, then

$$y^i\,(1,0) \leq \overline{K} \leq \overline{K} + \overline{K} - y^i\,(0,1)$$

If $z^i = (0,0)$, then

$$y^i\,(1,0) \leq \overline{K} \leq \overline{K} + y^i\,(0,0) - \underline{K}$$

Thus, it follows that

$$y^i\,(1,0) \in \begin{cases} \left\{y^i\,(1,0)\right\} & \text{if } z^i = (1,0) \\ \left[\underline{K}, \overline{K}\right] & \text{if } z^i \in \left\{(0,0)\,, (0,1)\,, (1,1)\right\} \end{cases}$$

Taking expectations yields the no–assumption bounds. A similar argument applies to the other elements of T.

The SPM inequality does permit strengthened identification results for treatment effects. In the no–assumption case, if $z^i = (1,1)$ or $z^i = (0,1)$, then $y^i\,(1,0) - y^i\,(0,0) \in \left[\underline{K} - \overline{K}, \overline{K} - \underline{K}\right]$. Under SPM, the fact that we observe one of $\left\{y^i\,(1,1)\,, y^i\,(0,1)\right\}$ allows us to further reduce this upper bound. Sharp bounds for the treatment effects $y^i\,(1,0) - y^i\,(0,0)$, $y^i\,(1,1) - y^i\,(0,1)$, and $y^i\,(1,1) - y^i\,(0,0)$ are given below.

$$y^i\,(1,0) - y^i\,(0,0) \in \begin{cases} \left[\underline{K} - y^i\,(z^i)\,, \overline{K} - y^i\,(z^i)\right] & \text{if } z^i = (0,0) \\ \left[y^i\,(z^i) - \overline{K}, y^i\,(z^i) - \underline{K}\right] & \text{if } z^i = (1,0) \\ \left[\underline{K} - \overline{K}, \overline{K} - y^i\,(z^i)\right] & \text{if } z^i = (0,1) \\ \left[\underline{K} - \overline{K}, y^i\,(z^i) - \underline{K}\right] & \text{if } z^i = (1,1) \end{cases} \tag{1.3.6}$$

$$y^i(1,1) - y^i(0,1) \in \begin{cases} \left[\underline{K} - y^i(z^i), \overline{K} - \underline{K}\right] & \text{if } z^i = (0,0) \\ \left[y^i(z^i) - \overline{K}, \overline{K} - \underline{K}\right] & \text{if } z^i = (1,0) \\ \left[\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)\right] & \text{if } z^i = (0,1) \\ \left[y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}\right] & \text{if } z^i = (1,1) \end{cases} \tag{1.3.7}$$

$$y^i(1,1) - y^i(0,0) \in \begin{cases} \left[\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)\right] & \text{if } z^i = (0,0) \\ \left[\underline{K} - \overline{K}, \overline{K} - \underline{K}\right] & \text{if } z^i = (1,0) \\ \left[\underline{K} - \overline{K}, \overline{K} - \underline{K}\right] & \text{if } z^i = (0,1) \\ \left[y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}\right] & \text{if } z^i = (1,1) \end{cases} \tag{1.3.8}$$

Taking expectations in equations (1.3.6) and (1.3.7) yields the bounds in (1.3.4) and (1.3.5), respectively. Equation (1.3.8) shows that the no–assumption bounds remain sharp for $\mathrm{E}\left[\, y(1,1) - y(0,0)\,\right]$.

□

In proposition 1, assumption SPM improves the upper bound on $\mathrm{E}\left[\, y(1,0) - y(0,0)\,\right]$ and the lower bound on $\mathrm{E}\left[\, y(1,1) - y(0,1)\,\right]$ by establishing a monotonicity relationship between the two treatment effects. This monotonicity relationship implies that this assumption will allow improvements in the bounds on one treatment effect (due to the imposition of other assumptions) to further improve the bounds on other treatment effects. Because of this, bounds computed jointly under SPM and other assumptions like monotone or supermodular instrumental variables[12] will generally be strictly contained within the intersection of the bounds computed under these assumptions separately. The empirical application in section 1.6 illustrates this phenomenon. Thus, while SPM may have substantial identifying power on its own, it may yield even more identifying power when combined with other assumptions.

In the special case where y is bounded between zero and one, SPM can establish that

[12]See section 1.4.

$\mathrm{E}\left[\, y\,(1,0) - y\,(0,0) \,\right] \in [-1,0]$ or $\mathrm{E}\left[\, y\,(1,1) - y\,(0,1) \,\right] \in [0,1]$ if the observed expectations in (1.3.4) and (1.3.5) take certain boundary values. In general, however, SPM is not sufficient to identify the sign of a treatment effect in the absence of other assumptions.

Sharp bounds can be derived on general treatment spaces using the same approach, as I show in proposition 2:

Proposition 2. *Let $\{S_\gamma\}_{\gamma\in\Gamma}$ be the collection of all sublattices of T such that, for every $\gamma \in \Gamma$, S_γ is not a chain and $|S_\gamma| = 4$. Define $\Gamma^{SPM} \subseteq \Gamma$ to be the set of γ such that SPM holds on S_γ and SBM does not hold on S_γ iff $\gamma \in \Gamma^{SPM}$; likewise, define $\Gamma^{SBM} \subseteq \Gamma$ to be the set of γ such that SBM holds on S_γ and SPM does not hold on S_γ iff $\gamma \in \Gamma^{SBM}$. Define $\Gamma^{MOD} \subseteq \Gamma$ to be the set of γ such that both SPM and SBM hold on S_γ iff $\gamma \in \Gamma^{MOD}$. Let $\Gamma^{SPM}_{t,t'} \subseteq \Gamma^{SPM}$ be the set of γ such that $t, t' \in S_\gamma$ and $\gamma \in \Gamma^{SPM}$; likewise for $\Gamma^{SBM}_{t,t'}$ and $\Gamma^{MOD}_{t,t'}$. Then, for $t' < t$,*

$$
\begin{aligned}
&\left[\mathrm{E}\left[\, y\,(t) \mid z = t \,\right] - \overline{K}\right] P\,(z=t) + \left[\underline{K} - \mathrm{E}\left[\, y\,(t') \mid z = t' \,\right]\right] P\,(z=t') \\
&\quad + \sum_{t''\in\Lambda_1\cup\Lambda_3} \left[\mathrm{E}\left[\, y\,(t'') \mid z = t'' \,\right] - \overline{K}\right] P\,(z=t'') \\
&\quad + \sum_{t''\in\Lambda_4\cup\Lambda_5} \left[\underline{K} - \mathrm{E}\left[\, y\,(t'') \mid z = t'' \,\right]\right] P\,(z=t'') \\
&\quad + \sum_{t''\in\Lambda_2\cup\Lambda_6\cup\Lambda_7} \left[\underline{K} - \overline{K}\right] P\,(z=t'') \\
&\qquad \le \mathrm{E}\left[\, y\,(t) - y\,(t') \,\right] \le \qquad\qquad (1.3.9)\\
&\left[\overline{K} - \mathrm{E}\left[\, y\,(t') \mid z = t' \,\right]\right] P\,(z=t') + \left[\mathrm{E}\left[\, y\,(t) \mid z = t \,\right] - \underline{K}\right] P\,(z=t) \\
&\quad + \sum_{t''\in\Lambda_1\cup\Lambda_2} \left[\mathrm{E}\left[\, y\,(t'') \mid z = t'' \,\right] - \underline{K}\right] P\,(z=t'') \\
&\quad + \sum_{t''\in\Lambda_4\cup\Lambda_6} \left[\overline{K} - \mathrm{E}\left[\, y\,(t'') \mid z = t'' \,\right]\right] P\,(z=t'') \\
&\quad + \sum_{t''\in\Lambda_3\cup\Lambda_5\cup\Lambda_7} \left[\overline{K} - \underline{K}\right] P\,(z=t'')
\end{aligned}
$$

where $\Lambda_1, \ldots, \Lambda_7$ are defined in (1.3.11). These bounds are sharp.

Proof of proposition 2. By the Law of Iterated Expectations,

$$\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \right] = \sum_{t'' \in T} \mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid z = t'' \right] P\left(z = t''\right) \tag{1.3.10}$$

Sharp bounds for the unidentified expectations on the right hand side of (1.3.10) will yield sharp bounds on $\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \right]$. I proceed by finding the sharp identification region for an arbitrary $y^i\left(t\right) - y^i\left(t'\right)$ and every possible z^i. These can be averaged to find sharp bounds on $\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid z \right]$ for all z. Define the following sets:

$$\begin{aligned}
\Lambda_1 = &\left\{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma^{MOD}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t' < t < t'' \text{ and} \\
&\text{either } \exists \gamma \in \Gamma^{MOD}_{t,t'} \text{ s.t. } t'' \in S_\gamma \text{ or } \exists \gamma \in \Gamma^{SPM}_{t,t'}, \gamma' \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma \cap S_{\gamma'}\Big\} \\
\Lambda_2 = &\left\{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t' < t < t'', \\
&\nexists \gamma \in \Gamma^{MOD}_{t,t'} \cup \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\Big\} \\
\Lambda_3 = &\left\{t'' \mid t'' < t, t'' \parallel t', \text{ and } \exists \gamma \in \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t' < t < t'', \\
&\nexists \gamma \in \Gamma^{MOD}_{t,t'} \cup \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\Big\} \\
\Lambda_4 = &\left\{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma^{MOD}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t'' < t' < t \text{ and} \\
&\text{either } \exists \gamma \in \Gamma^{MOD}_{t,t'} \text{ s.t. } t'' \in S_\gamma \text{ or } \exists \gamma \in \Gamma^{SPM}_{t,t'}, \gamma' \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma \cap S_{\gamma'}\Big\} \\
\Lambda_5 = &\left\{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t'' < t' < t, \\
&\nexists \gamma \in \Gamma^{MOD}_{t,t'} \cup \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\Big\} \\
\Lambda_6 = &\left\{t'' \mid t' < t'', t'' \parallel t, \text{ and } \exists \gamma \in \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\right\} \cup \{t'' \mid t'' < t' < t, \\
&\nexists \gamma \in \Gamma^{MOD}_{t,t'} \cup \Gamma^{SPM}_{t,t'} \text{ s.t. } t'' \in S_\gamma, \text{ and } \exists \gamma \in \Gamma^{SBM}_{t,t'} \text{ s.t. } t'' \in S_\gamma\Big\} \\
\Lambda_7 = &\left(\bigcup_{j=1}^{6} \Lambda_j\right)^c
\end{aligned} \tag{1.3.11}$$

The four distinct orderings presented in $\Lambda_1, \ldots, \Lambda_6$ in (1.3.11) include every possible ordering that is compatible with the restriction that $t' < t$ and that also allows at least one of SPM or SBM to have some implications for identification. Each of $\Lambda_1, \ldots, \Lambda_6$ is a union of two sets. When t'' is not comparable with exactly one of t, t', there can be at most one four–point sublattice containing t, t', and t'', since the incomparable treatments define a unique meet and join. This simplifies the construction of the first set in each of these six two–set unions. The first set in Λ_1 isolates the t'' which belong to a sublattice containing t and t' where both SPM and SBM hold and where t'' is not strictly larger or smaller than t', so it must be the case that $t' \vee t'' = t$. Since both SPM and SBM hold on this sublattice, it follows that

$$y^i(t) - y^i(t') = y^i(t'') - y^i(t' \wedge t'')$$

so that, when $z^i = t''$, the bounds

$$y^i(t) - y^i(t') \in \left[y^i\left(z^i\right) - \overline{K}, y^i\left(z^i\right) - \underline{K}\right]$$

are sharp. A similar argument can be made for the first set in each of $\Lambda_2, \ldots, \Lambda_6$, and these sets are mutually exclusive due to the particular combinations of order restrictions and γ memberships along with the fact that

$$\Gamma^{MOD}_{t,t'} \cap \Gamma^{SPM}_{t,t'} = \Gamma^{MOD}_{t,t'} \cap \Gamma^{SBM}_{t,t'} = \Gamma^{SPM}_{t,t'} \cap \Gamma^{SBM}_{t,t'} = \emptyset$$

by definition.

The construction of the second set in each of the six two–set unions $\Lambda_1, \ldots, \Lambda_6$ is complicated by the fact that the orderings $t' < t < t''$ and $t'' < t' < t$ are compatible with multiple four–point sublattices containing t, t', and t'', since there may be multiple t''' such that $t \vee t''' = t''$ and $t \wedge t''' = t'$ (in the former case) and $t' \vee t''' = t$ and $t' \wedge t''' = t''$ (in the latter case). Each set is constructed to capture the t'' whose sublattice membership(s) yield the same implications for identification as the set it is paired with. The particular combinations of order restrictions and γ memberships imply that they are mutually exclusive.

The sets $\Lambda_1, \ldots, \Lambda_6$ define every sublattice membership pattern for t and t' for which SPM and SBM may have any implications; this follows from proposition 1 and its straightforward extension to the case of SBM. The set Λ_7 contains those t'' such that either 1. any sublattice S_γ containing t, t', and t'' must have t' as the bottom and t as the top, 2. t'' does not belong to any four–point sublattice containing t and t', or 3. t'' obeys one of the orderings from $\Lambda_1, \ldots, \Lambda_6$ but does not belong to any sublattice containing t and t' on which at least one of SPM and SBM hold .

The focus on four–point sublattices is without loss of generality, since the implications of assumptions SPM and SBM only appear on four–point sublattices. SPM and SBM have no implications on chains, so sublattices that are chains can be ignored. Restricting attention to elements of $\{S_\gamma\}_{\gamma \in \Gamma_{t,t'}} \subseteq \{S_\gamma\}_{\gamma \in \Gamma}$ is without loss of generality as well. This follows from the fact that SPM and SBM have no implications for potential outcomes under the maintained assumptions, and any implications for the treatment effect $y^i(t) - y^i(t')$ from another treatment effect which are mediated by a third treatment effect are realized directly on a sublattice containing the treatments from the first two treatment effects. To see this concretely, suppose that $S_\gamma = \{t', t, t'', t'''\}$ and $S_{\gamma'} = \{t'', t''', t'''', t'''''\}$ where $t \parallel t''$, $t''' \parallel t''''$, $t' = t \wedge t''$, $t''' = t \vee t''$, $t'' = t''' \wedge t''''$, and $t''''' = t''' \vee t''''$. Suppose that SPM holds on both S_γ and $S_{\gamma'}$. This implies

$$y^i(t) - y^i(t') \leq y^i(t''') - y^i(t'') \leq y^i(t''''') - y^i(t'''')$$

$$\Longrightarrow y^i(t) - y^i(t') \leq y^i(t''''') - y^i(t'''')$$

I show that $\{t', t, t'''', t'''''\} \in \{S_\gamma\}_{\gamma \in \Gamma_{t,t'}^{SPM}}$; this follows directly from lemma 1 and the definition of $\Gamma_{t',t}^{SPM}$:

Lemma 1. *Assume that $t \parallel t''$, $t''' \parallel t''''$, $t' = t \wedge t''$, $t''' = t \vee t''$, $t'' = t''' \wedge t''''$, and $t''''' = t''' \vee t''''$. Then, $t'''' \wedge t = t'$ and $t'''' \vee t = t'''''$.*

Proof. See appendix. □

A similar argument applies for SBM.

The sets defined in (1.3.11) along with the arguments of proposition 1 yield the following sharp identification regions for $y^i(t) - y^i(t')$ and each possible z^i:

$$y^i(t) - y^i(t') \in \begin{cases} \left[y^i(z^i) - \overline{K}, y^i(z^i) - \underline{K}\right] & \text{if } z^i \in \{t\} \cup \Lambda_1 \\ \left[\underline{K} - \overline{K}, y^i(z^i) - \underline{K}\right] & \text{if } z^i \in \Lambda_2 \\ \left[y^i(z^i) - \overline{K}, \overline{K} - \underline{K}\right] & \text{if } z^i \in \Lambda_3 \\ \left[\underline{K} - y^i(z^i), \overline{K} - y^i(z^i)\right] & \text{if } z^i \in \{t'\} \cup \Lambda_4 \\ \left[\underline{K} - y^i(z^i), \overline{K} - \underline{K}\right] & \text{if } z^i \in \Lambda_5 \\ \left[\underline{K} - \overline{K}, \overline{K} - y^i(z^i)\right] & \text{if } z^i \in \Lambda_6 \\ \left[\underline{K} - \overline{K}, \overline{K} - \underline{K}\right] & \text{if } z^i \in \Lambda_7 \end{cases} \tag{1.3.12}$$

Since the sets $\{t\}, \{t'\}, \Lambda_1, \ldots, \Lambda_7$ are mutually exclusive and exhaustive, averaging the bounds in (1.3.12) across i yields sharp bounds on $\mathrm{E}[y(t) - y(t')]$ via (1.3.10). These sharp bounds are given in (1.3.9).

□

Proposition 2 generalizes proposition 1 by allowing for a much richer set of treatments. The treatment may have any finite number of dimensions, and each may be binary or multi-valued. Some dimensions of the treatment may be complements while others are substitutes; the result allows for arbitrary combinations of SPM and SBM as appropriate. The complexity of the result is due to two factors. First, the treatment pair t, t' may belong to multiple sublattices. Second, the position of the treatment pair within a lattice, i.e., whether it includes the top and/or bottom of the sublattice, differs across sublattices. The position of the treatment pair within a sublattice combined with the assumptions that hold on the sublattice determine whether the upper and/or lower bound (or neither) are improved.

A number of the assumptions made in proposition 2 are primarily for ease of exposition and interpretation and do not limit the generality of the result. For example, the assumption that each S_γ has a cardinality of four is without loss of generality, since the implications of assumptions SPM and SBM only appear on four–point sublattices. Similarly, no generality is sacrificed by excluding sublattices that are chains, as SPM and SBM have no implications on chains.

I have focused on bounding expectations of treatment effects using only supermodularity and submodularity assumptions, but in practical applications these will often be paired with other monotonicity and instrumental variable assumptions. Deriving sharp bounds under combinations of assumptions is nontrivial. Applying results from section 1.4 to bound $\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid z = t'' \,\right]$ for each $t, t', t'' \in T$ before applying proposition 2 will yield bounds that contain the true value but are not necessarily sharp. However, these bounds may be much simpler to compute than the sharp bounds (which remain an open problem).

1.4 Instrumental Variables

Traditional instrumental variable (IV) analysis of treatment response relies on the existence of a variable that is correlated with the treatment variable of interest but is mean–independent or independent of the distribution of response functions. Whether or not such independence assumptions are justified in a particular context is often the subject of vigorous debate. This has motivated researchers to find weaker and more credible forms of these assumptions that still retain some identification power. A leading example is the notion of a monotone instrumental variable (Manski and Pepper 2000, 2009):

Assumption MIV. x_k is a *monotone instrumental variable* if, for all $t, t' \in T$ and all x_{-k},

$$x_k \leq x_k' \Longrightarrow \mathrm{E}\left[\, y\left(t\right) \mid z = t', x = \left(x_k, x_{-k}\right) \right] \leq \mathrm{E}\left[\, y\left(t\right) \mid z = t', x = \left(x_k', x_{-k}\right) \right]$$

Manski and Pepper motivated MIV by considering the problem of determining the returns to schooling. Average wages should be weakly increasing with observable measures of ability (such as test scores or realized years of schooling), so such measures can be used as MIVs but not IVs. Giustinelli (2011) analyzes the returns to education in Italy using a similar monotonicity restriction on the quantile function.

Assumption MIV can be generalized to allow x_k to be partially ordered. Manski and Pepper then refer to x_k as a semi–monotone instrumental variable (SMIV). Another special case of MIV occurs when the realized treatment z is itself an MIV; Manski and Pepper refer to this as the monotone treatment selection (MTS) assumption. MIV and its generalizations impose restrictions on functionals of potential outcome distributions. Restrictions can also be imposed directly on functionals of treatment effect distributions:

Assumption SPMIV (Supermodular instrumental variable). x_k is a *supermodular instrumental variable* for $\mathrm{E}\left[\, y(t) - y(t') \mid x_k, x_{-k} \,\right]$ with $t' \leq t$ if

$$x_k \leq x_k' \Longrightarrow \mathrm{E}\left[\, y(t) - y(t') \mid x_k, x_{-k} \,\right] \leq \mathrm{E}\left[\, y(t) - y(t') \mid x_k', x_{-k} \,\right] \tag{1.4.1}$$

for all x_{-k}.[13]

SPMIV is an alternative formulation of complementarity where treatment effects vary monotonically (on average) with an observed covariate x_k.[14] An advantage of these assumptions is that evidence for their validity may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. This evidence can motivate the application of these assumptions in other contexts where similar identification strategies are not available. This contrasts with traditional IV assumptions, which tend to be highly context–specific.

[13]The weak inequality in (1.4.1) can be reversed, in which case x_k would be a *submodular instrumental variable*. If the inequality is replaced with equality, x_k becomes a *modular instrumental variable*.

[14]The SPM/SPMIV distinction is analogous to the MTR/MIV distinction discussed in Manski and Pepper (2009).

The Djebbari and Smith (2008) study of the heterogeneous impacts of the PROGRESA conditional cash transfer program provides some examples of potential SPMIVs. PROGRESA provided payments to households conditional on regular school attendance by the household's children as well as visits to health centers. Djebbari and Smith find that the impact of this program on per capita consumption is substantially larger for poorer households and households in more "marginal" villages, i.e., villages with greater rates of illiteracy, more limited infrastructure, and a greater dependence on agricultural activities. Evaluations of cash transfer programs in other contexts could make use of this information by using household poverty or village marginality as SPMIVs.

Further examples are provided by the Bitler et al. (2014) study of the impact of the Connecticut Jobs First experiment. This program substantially lowered the marginal tax rate on earnings below the poverty line for families on relief, relative to the existing Aid to Families with Dependent Children (AFDC) program. In the Jobs First program, the entire benefit package is terminated once earnings rise above the poverty line; this is in contrast to the AFDC, where benefits decline linearly with earnings. Labor supply theory clearly suggests that the impact of this alternative budget scheme should boost earnings and employment much more for those who were previously out of work or whose earnings left them far below the poverty line. These hypotheses are strongly borne out by the data, suggesting that measures of pre–program earnings and employment could serve as SPMIVs in studies of similar programs which are not implemented experimentally.

For the remainder of this section, let $\underline{B}\,(t,x)$ and $\overline{B}\,(t,x)$ be defined as

$$\underline{B}\,(t,x) = \mathrm{E}\left[\, y\,(t) \mid z=t,x \,\right] P\,(z=t \mid x) + \underline{K} P\,(z \neq t \mid x) \quad \forall t \in T, x \in X$$

and

$$\overline{B}\,(t,x) = \mathrm{E}\left[\, y\,(t) \mid z=t,x \,\right] P\,(z=t \mid x) + \overline{K} P\,(z \neq t \mid x) \quad \forall t \in T, x \in X$$

The following bounds can be derived using SPMIV:

Proposition 3. *Assume that x_k is an SPMIV for* $\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid x_k, x_{-k} \,\right]$ *with* $t, t' \in T$. *Then, the bounds*

$$
\begin{aligned}
&\sup_{x'_k \leq x_k} \left\{ \underline{B}\left(t, x'_k, x_{-k}\right) - \overline{B}\left(t', x'_k, x_{-k}\right) \right\} \\
&\leq \mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid x_k, x_{-k} \,\right] \leq \\
&\inf_{x_k \leq x'_k} \left\{ \overline{B}\left(t, x'_k, x_{-k}\right) - \underline{B}\left(t', x'_k, x_{-k}\right) \right\}
\end{aligned}
\tag{1.4.2}
$$

are sharp.

As is the case for bounds derived under IV or MIV assumptions, inference is complicated by the sup and inf operators in equation (1.4.2) (Manski and Pepper 2009). Analog estimators of the bounds in (1.4.2) are consistent but biased in finite samples; the estimated bounds will generally be too narrow. Fortunately, the methods developed by Chernozhukov, Lee and Rosen (2013) can be applied to find bias–corrected estimates and associated confidence intervals. Chernozhukov et al. discuss in detail the special cases of estimating nonparametric bounds using instrumental variables and MIVs; the bounds in (1.4.2) are essentially identical for the purposes of estimation, so their results can be applied directly to my estimation problem. The theoretical extension allowing for multiple SPMIVs is straightforward, and presents no novel estimation challenges besides those associated with high–dimensional nonparametric conditioning.

Returning to assumption SPMIV: If the second inequality in (1.4.1) is reversed, x_k becomes a *submodular instrumental variable.* If x_k is a supermodular and submodular instrumental variable, i.e., average treatment effects are constant across different values of x_k, then x_k is a *modular instrumental variable.* While this may seem like a strong assumption, it is routinely employed in applied work that assumes both exogeneity of the treatment and no interactions.

SPMIVs may also improve the bounds on functionals of potential outcome distributions, as the following proposition illustrates:

Proposition 4. *Assume that x_k is an SPMIV for* $\mathrm{E}\left[\, y\left(t\right) - y\left(t'\right) \mid x_k, x_{-k} \,\right]$ *with* $t, t' \in T$.

Then, the bounds

$$\begin{gathered}\max\left\{\underline{B}\left(t,x\right),\sup_{x_k'\leq x_k}\left\{\underline{B}\left(t,x_k',x_{-k}\right)-\overline{B}\left(t',x_k',x_{-k}\right)\right\}+\underline{B}\left(t',x\right)\right\}\\ \leq \mathrm{E}\left[\,y\left(t\right)\mid x_k,x_{-k}\,\right]\leq \\ \min\left\{\overline{B}\left(t,x\right),\inf_{x_k\leq x_k'}\left\{\overline{B}\left(t,x_k',x_{-k}\right)-\underline{B}\left(t',x_k',x_{-k}\right)\right\}+\overline{B}\left(t',x\right)\right\}\end{gathered} \tag{1.4.3}$$

and

$$\begin{gathered}\max\left\{\underline{B}\left(t',x\right),\underline{B}\left(t,x\right)-\inf_{x_k\leq x_k'}\left\{\overline{B}\left(t,x_k',x_{-k}\right)-\underline{B}\left(t',x_k',x_{-k}\right)\right\}\right\}\\ \leq \mathrm{E}\left[\,y\left(t'\right)\mid x_k,x_{-k}\,\right]\leq \\ \min\left\{\overline{B}\left(t',x\right),\overline{B}\left(t,x\right)-\sup_{x_k'\leq x_k}\left\{\underline{B}\left(t,x_k',x_{-k}\right)-\overline{B}\left(t',x_k',x_{-k}\right)\right\}\right\}\end{gathered} \tag{1.4.4}$$

are sharp.

As in the case of proposition 3, analog estimators of the bounds in (1.4.3) and (1.4.4) are consistent but biased in finite samples; the Chernozhukov et al. approach can be applied here as well.

1.5 Independence

Independence assumptions have been used to operationalize the belief that individuals' realized treatments are unrelated to any individual characteristics which may influence responses. This should be the case, for example, in a randomized controlled trial. I show how statistical independence can be combined with shape restrictions and instrumental variables assumptions to narrow the bounds on entire treatment effect distributions.

The familiar assumption of statistical independence of treatments and response functions is defined in my notation as follows:

Assumption SI (Statistical independence)**.** Potential outcomes are *statistically independent* of realized treatments if

$$P\left(y\left(t\right)\mid z\right)=P\left(y\left(t\right)\right)\quad\forall t\in T$$

Assumption SI implies that the marginal distribution of $y\left(t\right)$, denoted F_t, is point identified for all $t\in T$ such that $P\left(z=t\right)>0$. However, the distribution of $y\left(t\right)-y\left(t'\right)$, whose cumulative distribution function is denoted by $F_{t,t'}$, is only partially identified. Makarov (1982) was the first to derive pointwise sharp bounds on the distribution of the sum of two random variables with fixed marginal distributions. Frank, Nelsen and Schweizer (1987) derived these bounds in a simpler manner and extended them to allow for other operations such as differences and products as well as more than two variables. However, as Kreinovich and Ferson (2006) show, these bounds are not sharp in the case of more than two variables. The following result, taken from Theorem 2 of Williamson and Downs (1990), gives the pointwise sharp bounds on the distribution $F_{t,t'}$ for any $t,t'\in T^R$:

$$\begin{aligned}\underline{F}_{t,t'}\left(w\right)=\sup_{u+v=w}\left\{\max\left\{F_t\left(u\right)-F_{t'}\left(-v\right),0\right\}\right\}\\ \leq F_{t,t'}\left(w\right)\leq\\ 1+\inf_{u+v=w}\left\{\min\left\{F_t\left(u\right)-F_{t'}\left(-v\right),0\right\}\right\}=\overline{F}_{t,t'}\left(w\right)\end{aligned}\tag{1.5.1}$$

Fan and Park (2010) discuss consistent nonparametric estimation of these bounds.

SI can be combined with SPM to refine (1.5.1), as the following result shows:

Proposition 5. *Assume that SI holds and that* $T=\left\{t\wedge t',t,t',t\vee t'\right\}$ *with* $t\wedge t'<t,t'<t\vee t'$. *Assume that SPM holds on* T. *Then, the bounds*

$$\underline{F}_{t\vee t',t}\left(w\right)\leq F_{t\vee t',t}\left(w\right)\leq\min\left\{\overline{F}_{t\vee t',t}\left(w\right),\overline{F}_{t',t\wedge t'}\left(w\right)\right\}\tag{1.5.2}$$

and

$$\max\left\{\underline{F}_{t,t\wedge t'}\left(w\right),\underline{F}_{t\vee t',t'}\left(w\right)\right\}\leq F_{t,t\wedge t'}\left(w\right)\leq\overline{F}_{t,t\wedge t'}\left(w\right)$$

are sharp, where $\underline{F}, \overline{F}$ *are defined as in* (1.5.1)*.*

Similar results can be derived for SBM, and these results can be used to obtain narrower bounds on functionals of treatment effect distributions.[15] These shape restrictions could be justified by theoretical arguments; alternatively, since average treatment effects are point–identified in this context, the supermodularity or submodularity of average effects could be used to provide some justification for stronger structural assumptions. Extending these results to general lattices is problematic due to the fact that sharp bounds on the distribution function of a sum of more than two variables are an open question. Nonetheless, it is straightforward to collect all possible stochastic dominance relations implied by the maintained assumptions, and bounds which contain the true value (but are not necessarily sharp) can be obtained in a manner similar to that in proposition 5. Such bounds may be useful in policy evaluation.

A reformulation of the SPMIV assumption can also be applied in this setting. SPMIV itself is unhelpful, since conditional average treatment effects are point–identified. However, if the distribution of treatment effects conditional on x is thought to obey a stochastic dominance relationship in one or more covariates, this can be used to derive improved bounds. The formal statement of the assumption is as follows:

Assumption Q–SPMIV (Quantile supermodular instrumental variable)**.** x_k is a *quantile supermodular instrumental variable* for $y(t) - y(t')$ if

$$x_k \leq x'_k \Longrightarrow F_{t,t'}(w \mid x_k, x_{-k}) \geq F_{t,t'}(w \mid x'_k, x_{-k})$$

for all x_{-k}.

The following proposition computes the bounds derived under this assumption:

[15] However, see Firpo and Ridder (2010) for a discussion of pointwise vs. uniform sharpness and the implications for deriving sharp bounds on functionals of the distribution of treatment effects.

Proposition 6. *Assume that SI holds. Assume that x_k is a Q–SPMIV for $y(t) - y(t')$ with $t, t' \in T$. Then, the bounds*

$$\sup_{x_k \leq x_k'} \left\{ \underline{F}_{t,t'}\left(w \mid x_k', x_{-k}\right) \right\} \leq F_{t,t'}\left(w \mid x_k, x_{-k}\right) \leq \inf_{x_k' \leq x_k} \left\{ \overline{F}_{t,t'}\left(w \mid x_k', x_{-k}\right) \right\}$$

are sharp.

Again, since conditional average treatment effects are point–identified, they can provide some evidence to support the validity of the stronger Q–SPMIV assumption. The improved bounds on $F_{t,t'}$ derived using this result can be combined with SPM or SBM to yield even stronger bounds.

1.6 Empirical Illustration

To illustrate the use of the identification results developed in this paper, I reanalyze data from Shertzer et al. (2014b). That study examines the extent to which Chicago's first zoning ordinance, passed in 1923, influenced the evolution of the spatial distribution of commercial, industrial, and residential activity in the city. That study found evidence of substantial treatment effect heterogeneity, which motivates the use of SBM and SPMIV assumptions in the analysis below.

Chicago's 1923 zoning ordinance regulated land by restricting uses and density; for details on the ordinance, consult Shertzer et al. (2014b). Here, I bound the effects of 1923 commercial zoning on the probability that a city block will contain any commercial activity in 2005, focusing on the outlying (largely residential) portions of the city that were zoned into the two lowest density categories. As discussed in section 1.3, zoning is a multidimensional policy and the long–run effect of commercial zoning likely varies substantially with the associated density restrictions. Since both use and density zoning are endogenous policy variables, quantifying the heterogeneous effects of commercial zoning with respect to density requires a multidimensional treatment variable; simply conditioning on assigned density zoning would not yield correct estimates of how the commercial zoning effect varies with density zoning.

Formally, the outcome variable $y_i(\cdot)$ is an indicator equal to 1 iff city block i contains any commercial activity in 2005. y_i is a function of a treatment $t \in T = \{0,1\} \times \{1,2\}$. The first dimension of t is equal to 1 if the block received any commercial zoning in 1923 and 0 otherwise. The second dimension of t is equal to 1 if the block was zoned for the lowest density development (3 or fewer stories) and 2 if it was zoned for higher density development (8–10 stories).

Areas zoned for lower densities will be more residential in character and contain a larger proportion of single–family homes (Shertzer et al. 2014b). It is well documented that residential property owners (especially single–family homeowners) generally oppose the encroachment of commercial uses and have substantial power to block such development (Fischel 2001). It is likely that the early establishment of commercial activity through zoning will be a more important determinant of future commercial land use in areas also zoned for lower densities. This assumption is also consistent with previous literature showing that mixed use areas are more likely to see conversion to completely non–residential use than strictly residential use (McMillen and McDonald 1991). This motivates the assumption that y exhibits SBM on T.

One may also expect commercial zoning to have more persistent effects when it does not conflict with the existing land use pattern. The data identifies blocks which had commercial activity prior to the introduction of zoning; an indicator for the presence of pre–zoning commercial activity is a natural SPMIV.

Table 1.1 shows a series of bounds computed under different assumptions. It is noteworthy that the bounds under the combination of SBM and SPMIV are not simply the intersection of the bounds computed under these assumptions separately. This illustrates the fact that the shape restrictions I introduce can magnify the identifying power of other assumptions. While the sign of the treatment effect is not identified using only these assumptions, there

Table 1.1: Bounds on long–run zoning impact

Assumptions:	Bounds on: $\mathrm{E}[\,y\,(1,1) - y\,(0,1)\,]$	$\mathrm{E}[\,y\,(1,2) - y\,(0,2)\,]$
None	$[-0.686, 0.934]$	$[-0.573, 0.807]$
SBM	$[-0.259, 0.934]$	$[-0.573, 0.741]$
SPMIV	$[-0.532, 0.934]$	$[-0.573, 0.747]$
SBM & SPMIV	$[-0.231, 0.934]$	$[-0.573, 0.512]$
Observations	8,572	8,572

The outcome variable y is an indicator for the presence of commercial activity on the block in 2005. The first dimension of the treatment is an indicator that equals one iff the block received any commercial zoning in 1923. The second dimension of the treatment indicates the density level the block was zoned for in 1923.

is up to a 28% reduction in the width of the bounds.

1.7 Conclusion

In this paper, I contribute to the literature on the partial identification of treatment effects by developing and applying assumptions that formalize the notion of complementarity. I examine the identification power of these assumptions and discuss how they can be justified. The supermodularity and submodularity assumptions I propose can be used to narrow bounds on treatment effects in studies of policy complementarity, which have traditionally been stymied by a lack of pseudo–experimental variation in multiple policies simultaneously. In proposition 1, I show how these shape restrictions can improve bounds on average treatment effects in the simple case of a two–dimensional binary treatment. Proposition 2 extends this result to a more general treatment set with an arbitrary finite number of (possibly multivalued) treatments and the possibility of complex combinations of supermodularity and submodularity.

Complementarity may also stem from differential treatment response among subpopulations defined by observed covariates. Subgroup heterogeneity in treatment effects is an increasingly widely recognized phenomenon, and can often be motivated directly from economic theory (see, e.g., Bitler et al. (2014)). Propositions 3 and 4 show how qualitative information about treatment effect heterogeneity embodied in supermodular instrumental variables can be used to improve bounds on average treatment effects and average potential outcomes. Supermodular instrumental variables can be used in studies with one or many treatments, making them a versatile and potentially powerful addition to the arsenal of applied econometricians.

The assumptions I propose can be useful in the experimental context as well. Proposition 5 shows how supermodularity can be combined with an assumption of statistical independence between assigned treatments and responses to yield improved bounds on the cumulative distribution function of a treatment effect. These results can be applied to the evaluation of outcomes in complex (multi–treatment) randomized controlled trials, which are increasingly prevalent in many fields, including development economics. Since average treatment effects are point–identified in this context, one can determine if average responses exhibit supermodularity or submodularity. This can provide evidence that individual response functions are supermodular or submodular. Similarly, the behavior of (point–identified) conditional average treatment effects can motivate the use of a quantile supermodular instrumental variable; in proposition 6, I show how this assumption can strengthen the bounds on the CDF of a treatment effect distribution.

Bounds derived under the assumptions I propose here are of interest only to the extent that such assumptions are considered credible. Where might evidence for their validity come from? Arguments for policy complementarity may be provided by economic theory, as in Lalive et al. (2006), or they may come from multi–treatment randomized controlled trials. Evidence on subgroup heterogeneity in treatment effects may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. In such studies, conditional average treatment effects are point–identified, so the validity of the assumptions I propose can be established. This can motivate their use in other contexts where similar identification strategies are not available. This distinguishes supermodular IV assumptions from traditional IV assumptions, since the latter tend to be context–specific.

The empirical illustration in section 1.6 employs assumptions SBM and SPMIV to study the impact of historical zoning on the evolution of land use in Chicago. Of particular interest is the fact that the bounds computed under both SBM and SPMIV are substantially narrower than the intersection of the bounds computed under each assumption separately. This demonstrates the general fact that assumptions SPM and SBM can magnify the identification power of other assumptions.

1.8 Appendix

Proof of lemma 1. I first show that $t'''' \wedge t = t'$:

$$\begin{aligned} & t'''' \wedge t''' = t'' \\ \implies & (t'''' \wedge t''') \wedge t = t'' \wedge t \\ \implies & t'''' \wedge (t''' \wedge t) = t' \\ \implies & t'''' \wedge t = t' \end{aligned}$$

where the last implication follows from the fact that $t \leq t'''$. Now, I show that $t'''' \vee t = t'''''$:

$$\begin{aligned} & t''' \vee t'''' = t''''' \\ \implies & (t \vee t'') \vee t'''' = t''''' \\ \implies & t \vee (t'' \vee t'''') = t''''' \\ \implies & t \vee t'''' = t''''' \end{aligned}$$

where the last implication follows from the fact that $t'' \leq t''''$. $\square$

Proof of proposition 3. In the absence of other assumptions, the bounds

$$\underline{B}\,(t,x) \leq \mathrm{E}\,[\,y\,(t) \mid x\,] \leq \overline{B}\,(t,x)$$

and

$$\underline{B}\,(t',x) \leq \mathrm{E}\,[\,y\,(t') \mid x\,] \leq \overline{B}\,(t',x)$$

and thus

$$\underline{B}\,(t,x) - \overline{B}\,(t',x) \leq \mathrm{E}\,[\,y\,(t) \mid x\,] - \mathrm{E}\,[\,y\,(t') \mid x\,] \leq \overline{B}\,(t,x) - \underline{B}\,(t,x)$$

are sharp for all $x \in X$. The assumption that x_k is an SPMIV for $\mathrm{E}\,[\,y\,(t) - y\,(t') \mid x_k, x_{-k}\,]$ implies that

$$\underline{B}\,(t,x'_k,x_{-k}) - \overline{B}\,(t',x'_k,x_{-k}) \leq \mathrm{E}\,[\,y\,(t) - y\,(t') \mid x_k, x_{-k}\,]$$

for all $x'_k \leq x_k$ and

$$\mathrm{E}\,[\,y\,(t) - y\,(t') \mid x_k, x_{-k}\,] \leq \underline{B}\,(t,x'_k,x_{-k}) - \overline{B}\,(t',x'_k,x_{-k})$$

for all $x_k \leq x'_k$. The result follows. □

Proof of proposition 4. Proposition 3 implies that the bounds

$$\sup_{x'_k \leq x_k} \left\{\underline{B}\,(t,x'_k,x_{-k}) - \overline{B}\,(t',x'_k,x_{-k})\right\}$$
$$\leq \mathrm{E}\,[\,y\,(t) - y\,(t') \mid x_k, x_{-k}\,] \leq$$
$$\inf_{x_k \leq x'_k} \left\{\overline{B}\,(t,x'_k,x_{-k}) - \underline{B}\,(t',x'_k,x_{-k})\right\}$$

are sharp. Thus, $\mathrm{E}\,[\,y\,(t) \mid x\,]$ and $\mathrm{E}\,[\,y\,(t') \mid x\,]$ must simultaneously satisfy the no–assumption bounds

$$\underline{B}\,(t,x) \leq \mathrm{E}\,[\,y\,(t) \mid x\,] \leq \overline{B}\,(t,x) \tag{1.8.1}$$

and

$$\underline{B}(t', x) \le \mathrm{E}[\, y(t') \mid x \,] \le \overline{B}(t', x) \tag{1.8.2}$$

as well as

$$\begin{gathered}\sup_{x_k' \le x_k} \left\{ \underline{B}(t, x_k', x_{-k}) - \overline{B}(t', x_k', x_{-k}) \right\} + \mathrm{E}[\, y(t') \mid x \,] \\ \le \mathrm{E}[\, y(t) \mid x \,] \le \\ \inf_{x_k \le x_k'} \left\{ \overline{B}(t, x_k', x_{-k}) - \underline{B}(t', x_k', x_{-k}) \right\} + \mathrm{E}[\, y(t') \mid x \,]\end{gathered} \tag{1.8.3}$$

and

$$\begin{gathered}\mathrm{E}[\, y(t) \mid x \,] - \inf_{x_k \le x_k'} \left\{ \overline{B}(t, x_k', x_{-k}) - \underline{B}(t', x_k', x_{-k}) \right\} \\ \le \mathrm{E}[\, y(t') \mid x \,] \le \\ \mathrm{E}[\, y(t) \mid x \,] - \sup_{x_k' \le x_k} \left\{ \underline{B}(t, x_k', x_{-k}) - \overline{B}(t', x_k', x_{-k}) \right\}\end{gathered} \tag{1.8.4}$$

From (1.8.1)–(1.8.4), it is clear that

$$\begin{gathered}\max \left\{ \underline{B}(t, x), \sup_{x_k' \le x_k} \left\{ \underline{B}(t, x_k', x_{-k}) - \overline{B}(t', x_k', x_{-k}) \right\} + \underline{B}(t', x) \right\} \\ \le \mathrm{E}[\, y(t) \mid x_k, x_{-k} \,] \le \\ \min \left\{ \overline{B}(t, x), \inf_{x_k \le x_k'} \left\{ \overline{B}(t, x_k', x_{-k}) - \underline{B}(t', x_k', x_{-k}) \right\} + \overline{B}(t', x) \right\}\end{gathered}$$

and

$$\begin{gathered}\max \left\{ \underline{B}(t', x), \underline{B}(t, x) - \inf_{x_k \le x_k'} \left\{ \overline{B}(t, x_k', x_{-k}) - \underline{B}(t', x_k', x_{-k}) \right\} \right\} \\ \le \mathrm{E}[\, y(t') \mid x_k, x_{-k} \,] \le \\ \min \left\{ \overline{B}(t', x), \overline{B}(t, x) - \sup_{x_k' \le x_k} \left\{ \underline{B}(t, x_k', x_{-k}) - \overline{B}(t', x_k', x_{-k}) \right\} \right\}\end{gathered}$$

must hold. I show that these bounds are feasible, i.e., consistent with (1.4.2), whence it follows that they are sharp. Consider the following events:

$$\begin{aligned}&\max\left\{\underline{B}\left(t,x\right),\sup_{x'_k\leq x_k}\left\{\underline{B}\left(t,x'_k,x_{-k}\right)-\overline{B}\left(t',x'_k,x_{-k}\right)\right\}+\underline{B}\left(t',x\right)\right\}\\&=\sup_{x'_k\leq x_k}\left\{\underline{B}\left(t,x'_k,x_{-k}\right)-\overline{B}\left(t',x'_k,x_{-k}\right)\right\}+\underline{B}\left(t',x\right)>\underline{B}\left(t,x\right)\end{aligned}\tag{1.8.5}$$

$$\begin{aligned}&\min\left\{\overline{B}\left(t,x\right),\inf_{x_k\leq x'_k}\left\{\overline{B}\left(t,x'_k,x_{-k}\right)-\underline{B}\left(t',x'_k,x_{-k}\right)\right\}+\overline{B}\left(t',x\right)\right\}\\&=\inf_{x_k\leq x'_k}\left\{\overline{B}\left(t,x'_k,x_{-k}\right)-\underline{B}\left(t',x'_k,x_{-k}\right)\right\}+\overline{B}\left(t',x\right)<\overline{B}\left(t,x\right)\end{aligned}\tag{1.8.6}$$

$$\begin{aligned}&\max\left\{\underline{B}\left(t',x\right),\underline{B}\left(t,x\right)-\inf_{x_k\leq x'_k}\left\{\overline{B}\left(t,x'_k,x_{-k}\right)-\underline{B}\left(t',x'_k,x_{-k}\right)\right\}\right\}\\&=\underline{B}\left(t,x\right)-\inf_{x_k\leq x'_k}\left\{\overline{B}\left(t,x'_k,x_{-k}\right)-\underline{B}\left(t',x'_k,x_{-k}\right)\right\}>\underline{B}\left(t',x\right)\end{aligned}\tag{1.8.7}$$

$$\begin{aligned}&\min\left\{\overline{B}\left(t',x\right),\overline{B}\left(t,x\right)-\sup_{x'_k\leq x_k}\left\{\underline{B}\left(t,x'_k,x_{-k}\right)-\overline{B}\left(t',x'_k,x_{-k}\right)\right\}\right\}\\&=\overline{B}\left(t,x\right)-\sup_{x'_k\leq x_k}\left\{\underline{B}\left(t,x'_k,x_{-k}\right)-\overline{B}\left(t',x'_k,x_{-k}\right)\right\}<\overline{B}\left(t',x\right)\end{aligned}\tag{1.8.8}$$

It is easy to show that (1.8.5) $\implies \neg$(1.8.7); thus, the lower bounds in (1.4.3) and (1.4.4) are consistent with (1.4.2). Similarly, (1.8.6) $\implies \neg$(1.8.8), and so the upper bounds in (1.4.3) and (1.4.4) are consistent with (1.4.2). $\square$

Proof of proposition 5. For a lattice $T=\{t,t',t\vee t',t\wedge t'\}$ which is not a chain, SPM implies the following inequalities:

$$y^i\left(t'\right)-y^i\left(t\wedge t'\right)\leq y^i\left(t\vee t'\right)-y^i\left(t\right)$$

$$y^i\left(t\right)-y^i\left(t\wedge t'\right)\leq y^i\left(t\vee t'\right)-y^i\left(t'\right)$$

$$y^i\left(t\right)+y^i\left(t'\right)-2y^i\left(t\wedge t'\right)\leq y^i\left(t\vee t'\right)-y^i\left(t\wedge t'\right)\leq 2y^i\left(t\vee t'\right)-y^i\left(t\right)-y^i\left(t'\right)$$

Since these inequalities hold for all $i\in I$, they imply the following first–order stochastic dominance relationships:

$$F_{t\vee t',t}\left(w\right) \leq F_{t',t\wedge t'}\left(w\right) \tag{1.8.9}$$

$$F_{t\vee t',t'}\left(w\right) \leq F_{t,t\wedge t'}\left(w\right) \tag{1.8.10}$$

$$F_{t\vee t',t,t\vee t',t'}\left(w\right) \leq F_{t\vee t',t\wedge t'}\left(w\right)$$

$$F_{t\vee t',t\wedge t'}\left(w\right) \leq F_{t,t\wedge t',t',t\wedge t'}\left(w\right)$$

for all $w \in \mathbb{R}$. Here, $F_{t\vee t',t,t\vee t',t'}$ is the cdf of $2y\left(t\vee t'\right) - y\left(t\right) - y\left(t'\right)$ and $F_{t,t\wedge t',t',t\wedge t'}$ is the cdf of $y\left(t\right) + y\left(t'\right) - 2y\left(t\wedge t'\right)$. Pointwise sharp bounds on $F_{t\vee t',t}$, $F_{t\vee t',t'}$, $F_{t',t\wedge t'}$, $F_{t,t\wedge t'}$, and $F_{t\vee t',t\wedge t'}$ in the absence of SPM are given by (1.5.1). Combining SPM with the inequalities (1.8.9) and (1.8.10) yields the results.

□

Proof of proposition 6. Trivial. □

Chapter 2

Race, ethnicity, and discriminatory zoning

(with Allison Shertzer and Randall Walsh)

2.1 Introduction

Few local policies are as controversial or as frequently linked to discrimination as zoning. Critics argue that zoning is used as a tool to deter entry of poorer households into wealthier neighborhoods, often through the imposition of minimum lot sizes.[1] According to this view, low–income minority households become trapped in poor neighborhoods as a result of "exclusionary" zoning, contributing to racial segregation and disparities (Schlay and Rossi, 1981; Rothwell and Massey, 2009). Scholars and policy makers also argue that zoning is used to steer industrial activity towards minority neighborhoods, leading to disproportionate toxic exposure and depressed land values (Maantay, 2001; Wilson, Hutson, and Mujahid, 2008). "Environmental racism" associated with zoning could thus serve as a channel through which minorities remain disadvantaged and isolated.[2]

[1]For reviews of the exclusionary zoning literature, see Ihlanfeldt (2004) and Pogodzinski (1991).

[2]The term "environmental racism" was coined by Reverend Benjamin Chavis during a press release regarding the influential report "Toxic Waste and Race in the United States: A National Report on the

Research continues to demonstrate that minorities remain disproportionately isolated in poor neighborhoods and exposed to pollution (Sharkey, 2013; EPA Plan EJ 2014).[3] However, identifying the link between local land use regulations and these disparities is difficult because land use and zoning have been co–evolving for almost a century in most American cities. Existing scholarship has struggled to disentangle inequitable treatment in zoning ordinances and nuisance siting from residential mobility that is correlated with land use. For instance, the availability of affordable housing may cause low–income residents to cluster in areas with locally undesirable land uses (Been and Gupta, 1997). Nonetheless, understanding the link between zoning and disparities in access to public goods and exposure to pollution is critical for effective policymaking.

In this paper we employ a novel approach to studying how land use regulations affect minorities, focusing on the introduction of comprehensive zoning in the United States. The key innovation of our approach is that we observe detailed measures of existing land use at the city block level prior to the introduction of comprehensive zoning in Chicago. Our empirical strategy asks what impact pre–existing minority populations had on zoning outcomes, conditional on the extant land use and settlement patterns at the time of initial zoning adoption. The ability to control for ex ante density allows us to distinguish between minority neighborhoods receiving higher density zoning and the tendency of minorities to settle in neighborhoods with denser development. Similarly, the ability to observe and control for ex ante minority proximity to undesirable land uses enables us to disentangle discrimination in land use regulation from the observationally equivalent mechanism of poor minorities sorting into less expensive neighborhoods near polluting sites.[4]

We focus on the initial comprehensive zoning ordinance adopted by Chicago in 1923,

Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites" (United Church of Christ, 1987).

[3]For the EPA see "Plan EJ 2014" see http://www.epa.gov/compliance/ej/plan-ej/.

[4]Recent work by Depro, Timmins and O'Neil (2014) takes a different approach to this question, estimating a structural model of mobility by race in the presence of polluting sites. They show that race–pollution correlations can be in part explained by whites having a higher marginal willingness than Hispanics to pay to avoid pollution exposure.

one of the first and most influential policies of its kind, and ask how the racial and ethnic composition of neighborhoods influenced local zoning outcomes. A second contribution of our study is the rich detail of the microdata assembled for the analysis. We observe place of birth and parents' place of birth for the universe of individuals living in Chicago in 1920, allowing us to precisely measure the size of both first– and second–generation immigrant populations. We are also able to distinguish northern–born black populations from enclaves of southern–born blacks who had migrated to Chicago, which enables us to ask whether these groups were treated differently in the zoning process.

We first study the density component of the zoning ordinance, finding evidence of an early form of "exclusionary" zoning that was applied to black neighborhoods.[5] On the margin between the two lowest levels of density zoning, where the greatest scope for unequal treatment in density restrictions would have existed, a one standard deviation increase in the black share of a neighborhood was associated with a 16 percentage point increase in the likelihood of the neighborhood being zoned primarily for higher density buildings. For European immigrants, the relationship is reversed. Thus, at the margin, the zoning board appears to have endeavored to increase the building density in neighborhoods with high numbers of black residents and decrease the density in neighborhoods with large numbers of European immigrants.

Turning to the use component of the zoning ordinance, we find that neighborhoods with a larger share of southern–born blacks or first–generation immigrants were more likely to be zoned for industrial uses than comparable neighborhoods with white natives. Specifically, a standard deviation increase in southern black share is associated with a 8 percentage point increase in the likelihood of an enumeration district being zoned to include manufacturing uses, and a one standard deviation increase in the first–generation immigrant share is associated with a 5 percentage point increase in the likelihood of an enumeration district being

[5] The extant literature on exclusionary zoning emphasizes differences in zoning ordinances across various incorporated municipalities, not within a single city (for instance, The Homevoter Hypothesis, Fischel, 2001). However, to the extent that cities faced pressure to concentrate minorities in particular neighborhoods, we may expect to see higher density zoning in black and immigrant neighborhoods in our context.

zoned for manufacturing uses. These are quantitatively important effects given that only 26 percent of enumeration districts received any zoning for manufacturing uses.

Inequitable zoning had consequences in both the short and long run for blacks and immigrants. Minority communities receiving industrial use and higher density zoning were excluded from the economic benefit of low density, purely residential zoning in the 1923 ordinance.[6] Zoning thus served as a channel through which government action reduced the value of minority–owned homes relative to the properties owned by white native–born individuals. Discrimination in zoning ordinances translates directly into economic disparities since "for the great majority of homeowners, the equity in their home is the most important savings they have." (Fischel, 2001, p. 4). We also show that conditional on pre–zoning land use, neighborhoods that received higher density zoning in 1923 had higher housing unit and population density by 1940. This finding buttresses the claim that zoning ordinances can be used to concentrate minorities in denser neighborhoods, contributing to segregation and environmental disparities (Rothwell, 2011). Furthermore, we demonstrate that this type of discriminatory policy had emerged as early as the 1920s.

Our results cast doubt on the de jure racial blindness of comprehensive zoning ordinances, of which all but one (New York) were passed after the Supreme Court ruled explicitly racial zoning unconstitutional in the 1917 Buchanan v. Warley case. Although our evidence is historical, the results demonstrate that racial discrimination can arise even with the most general and widely used forms of land use control. Furthermore, Shertzer, Twinam, and Walsh (2014) find that these ordinances have persistent effects on a city's economic geography today. Because minority enclaves also exhibit substantial persistence over time, the results of these papers taken together indicate that observed inequities today could partially result from zoning decisions made many decades in the past.

[6] The price premium for strictly residential use zoning in the context of the Chicago ordinance is documented in McMillan and McDonald (2002). In order for blacks to be disadvantaged by the impact of the zoning ordinance on housing prices, it must be the case that some were homeowners and landlords. We cannot observe landlord status in the census, but nonetheless we see that 7 percent of blacks in our sample region were homeowners in 1920 and 10 percent in 1930.

2.2 Background on Zoning in Chicago

2.2.1 Brief History of Zoning in Chicago

The origins of comprehensive land use regulation in Chicago were rooted in public demand for "orderly" urban development, in particular the prevention of industrial and commercial encroachment on residential neighborhoods. Early twentieth century observers, including the influential Chicago Real Estate Board, expressed concern about the effect of unchecked expansion of commercial and industrial activity on property values (Schwieterman and Caspall, 2006). Others objected to the "canyon effect" created by unbroken rows of skyscrapers and the potential negative effects of the associated reduction in sunlight exposure and air flow on public health (Hall, 2002).

Chicago's city government had made previous attempts to control undesirable land uses, including an 1837 municipal code that prohibited any landowner or tenant from maintaining certain nuisances such as dead animals, dung, putrid meat, or fish entrails on their property. However, such piecemeal approaches proved insufficient for meeting public demand for controlled development, and in 1920 the newly created Chicago Zoning Commission began preparing a comprehensive zoning ordinance. The Commission, composed of eight aldermen and fourteen representatives from the Chicago community, spent eighteen months surveying existing land use in Chicago before issuing the initial statute.

Chicago's comprehensive zoning ordinance regulated land through both use districts and volume districts. Four distinct use districts were included: residential (single family housing), apartment, commercial, and manufacturing. These use districts were hierarchical, with apartment districts allowing residential uses, commercial districts allowing both apartments and single–family homes, and manufacturing districts allowing any use. Volume districts imposed restrictions on maximum lot coverage, aggregate volume, and height. The five volume districts in Chicago's ordinance were also hierarchical with district 5 allowing the tallest buildings.

Zoning statutes spread across the country in rapid order after Chicago's ordinance was passed, and by 1925 nearly 500 cities had adopted similar forms of comprehensive land use regulation (Mills, 1979). By this time, the question of whether zoning could explicitly address race and block black residents from certain neighborhoods had been settled: the U.S. Supreme Court had ruled a Louisville, Kentucky racial zoning ordinance unconstitutional in Buchanan v. Warley in 1917. This case squashed an effort by the Chicago Real Estate Board to convince the city to adopt a similar racial zoning ordinance. The realtors, led by agents from the Hyde Park, Kenwood, and Oakland neighborhoods, had argued that the dispersion of African–Americans throughout the city could lead to more than $250 million (in 1922 dollars) in property value depreciations (Chicago Commission on Race Relations, 1922).

When the move for a racial zoning ordinance failed, demand for segregation and protection from black "encroachment" led to the proliferation of private alternatives such as restrictive covenants (Brooks, 2011; Brooks and Rose, 2013). White residents were concerned by the arrival of blacks from the South, seeing them as "ignorant and rough–mannered, entirely unfamiliar with the standards of conduct in northern cities" (Chicago Commission on Race Relations, 1922). White immigrants were also concerned about competition for jobs from newly arrived African Americans and viewed the prospect of Negro neighbors as a "catastrophe equal to the loss of their homes" (Grossman, 1989, p. 175). Even longtime black residents of Chicago were hostile to the new arrivals, worrying that they would lose what social privileges they had as a result of the influx of poor and uneducated southern blacks into the city (Kennedy, 1968, p. 222).

For their part, African Americans were suspicious of the movement for comprehensive zoning, particularly so soon after the racial zoning debate. Nonetheless, the 1923 zoning ordinance passed without notable opposition from the black community in Chicago. Enthusiasm from black elites, many of whom optimistically welcomed the move for comprehensive zoning, may partly explain this outcome. For instance, a prominent African American developer on the zoning board, Charles S. Duke, championed land use regulation to the black community and is credited by historians for having shielded the wealthiest black neighborhoods from mixed–use zoning (NAACP, 1923). Secondary historical sources indicate that City Council in Chicago may have deliberating lowered zoning standards (e.g. permitted higher building density and mixed uses) in poorer black neighborhoods while maintaining strict zoning in white neighborhoods to prevent "encroachment" of blacks (Flint, 1977). However, to our knowledge there is no empirical evidence regarding the presence of racial animus in either the 1923 ordinance or subsequent amendments over the 1930s and 1940s.

2.2.2 Related Empirical Work on Zoning in Chicago

Although to our knowledge we are the first scholars to empirically ask how the spatial distribution of minority populations shaped initial zoning ordinances, comprehensive land use regulation is the subject of a large literature, and the case of Chicago has attracted particular interest. Previous work on Chicago's 1923 zoning ordinance used a sample of city blocks to determine the extent to which the ordinance followed existing uses, finding that zoning patterns were highly predictable given existing land uses, proximity to transportation networks, and distance to waterways (McMillen and McDonald, 1999). The same authors also asked how the 1923 zoning ordinance impacted land values (McMillen and McDonald, 2002). Using propensity score matching on the same sample of city blocks, they find that strictly residential zoning increased land values relative to mixed–use zoning.

2.3 Data

The dataset used in this paper has three components: 1920 census data at the enumeration district level, the comprehensive 1922 Chicago land use survey, and a map of the city's 1923 zoning ordinance. Summary statistics for key predictors and outcomes are provided in Table 2.1.

Table 2.1: Summary statistics.

Percent manufacturing	0.097
	(0.196)
Percent manufacturing if greater than 5 percent	0.371
	(0.214)
Indicator for manufacturing zoning	0.262
	(0.440)
Percent commercial zoning	0.218
	(0.181)
Indicator for volume district 2 if within 500 feet of district 1 and 2	0.587
	(0.493)
Total blacks	0.057
	(0.181)
Southern blacks	0.039
	(0.126)
Northern blacks	0.018
	(0.057)
First-gen. immigrants	0.462
	(0.221)
Second-gen. immigrants	0.208
	(0.080)
1913 land values	103.368
	(386.982)

Notes: Descriptive statistics for primary outcome and explanatory variables at the enumeration district (ED) level. Means are given with standard deviations in parentheses. Statistics are computed on the full sample unless otherwise indicated. Percentages of zoning variables are the fraction of the area of each ED covered by the specified type of zoning. Indicators equal 1 if and only if the ED includes any of the specified zoning. Demographic variables are the fraction of the total ED population attributed to each group. See Figure 1 for demographic group definitions.

2.3.1 Census Enumeration District Data

We obtained counts of the number of blacks and white ethnic group members at the census enumeration district level for a 100 percent sample of the population using a digitized version of the original 1920 Census taken from the genealogy website Ancestry.com. Enumeration districts were small administrative units used internally by the Census to divide cities into small areas that could be surveyed by one person.[7] The spatial microdata compiled for this paper represents a significant improvement over existing sources, most of which are tabulations of the population at the ward level produced by the Census Bureau.[8] The average enumeration district in Chicago had 1,182 individuals in 1920, less than two percent

[7]The Census Bureau did not switch to a mail–based survey system until 1960.

[8]The IPUMS sample for 1920 (Ruggles et al, 2004) covers 1 percent of the population of Chicago and contain enumeration district identifiers; however, this small sample is insufficient for studying neighborhoods.

of the population of the average ward.

In order to investigate the relationship between the composition of the population and zoning outcomes, we digitized the 1920 enumeration district map of Chicago. We first used written descriptions of the enumeration districts available on microfilm from the National Archives. The information from these microfilms has been digitized and made available on the web due to the work of Stephen P. Morse.[9] Second, we took digital photographs of the physical map of the 1920 census enumeration districts of Chicago from the National Archives. Working primarily with a geocoded (GIS) historic base street map developed by the Early Indicators Project, we generated a GIS representation of the Chicago enumeration district map that is consistent with the historic street grid.[10]

In our empirical work we focus on four categories of racial and ethnic minorities. Given the emphasis in the historical record on the lack of cohesiveness between northern and southern blacks, we separate these two groups in much of our empirical work. We define as southern blacks those individuals who report their race as black or mulatto and their place of birth as in the South.[11] We also include in the southern black category "second–generation" blacks, that is, individuals born in the North but with southern–born fathers in order to group all blacks of southern origin together. Northern blacks are defined as black or mulatto individuals who were both born outside the South with fathers born outside the South.

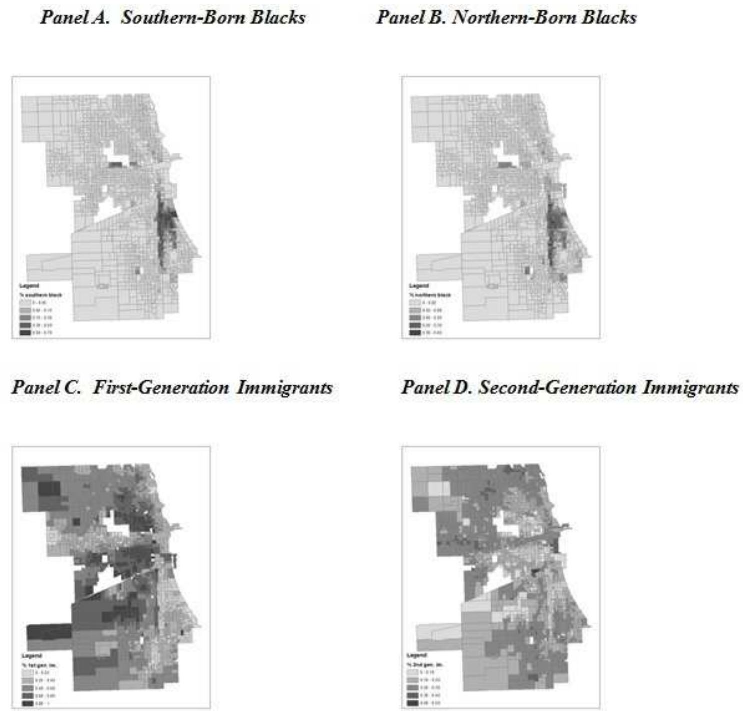
First–generation immigrants include all foreign–born individuals plus second–generation individuals under the age of 18, the latter of whom are presumably children residing in the same household as their foreign–born parents. Second–generation immigrants are defined as individuals who were born in the U.S. and who are at least 18 years old with foreign–born fathers. Using these definitions, we avoid a standard problem in the segregation literature of immigrant populations being diluted by the presence of their native–born children (see Cut-

[9]Website: http://stevemorse.org/ed/ed.php.

[10]See "Historical health conditions in major US cities: The HUE dataset" (Villareal, Bettenhausen, Hanss, Hirsch) for details on the street file construction.

[11]We use an eleven state definition of the South, defining the region to include Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia.

Figure 2.1: Spatial distribution of the minority groups.

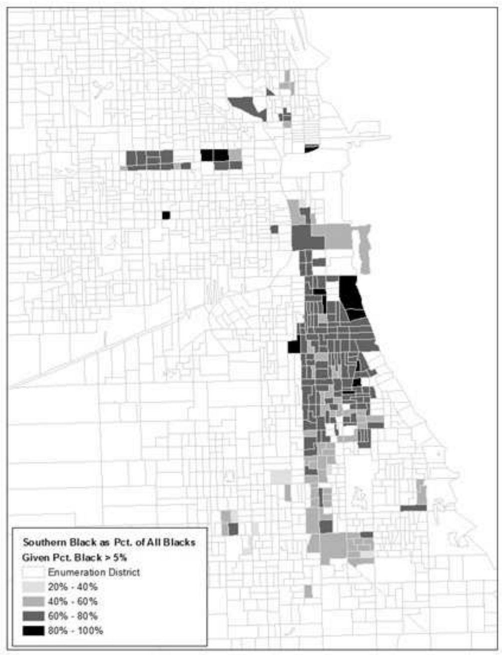


Notes: The sample covers the 1800 enumeration districts for which we have digitized land use data and census data. Southern blacks are black individuals born in the South or black individuals born in the North whose fathers were born in the South. Northern blacks are black individuals born in the North whose fathers were also born in the North. First-generation immigrants are individuals born abroad or minor second-generation immigrants (aged 18 or younger). Second-generation immigrants are individuals aged 18 and above who were born in the United States but whose fathers were born abroad.

ler, Glaeser, and Vigdor, 2008). Third–generation whites are defined as white individuals who were born in the U.S. and whose fathers were born in the U.S. As is shown in Table 2.1, the population of our study area is composed of 1.5 percent northern blacks, 2.9 percent southern blacks, 52.0 percent first–generation immigrants, and 17.9 percent second–generation immigrants in 1920. The remainder are white third–generation and beyond natives.

There are important compositional and economic differences between the first– and second–generation immigrant groups. Adult second–generation immigrants primarily traced their ancestry to Ireland and Germany and tended to be wealthier than recent arrivals. First–generation immigrants were more likely to have arrived from Poland, Italy, Russia, Bohemia (now the Czech Republic), and the other "new" sending countries of the late nineteenth and early twentieth century European immigration. The German and Irish communities also held political clout and most aldermanic seats; the larger new immigrant groups had mobilized politically but counted few aldermen among their number (Centennial List of Mayors, City Clerks, City Attorneys, City Treasurers, and Aldermen, 1937). We may thus expect first and second–generation immigrants to have been treated differently by the zoning process.

Figure 2.2: Variation in northern and southern blacks.



Notes: The figure shows the share of the percentage of each enumeration district's black population that we classify as being southern black among the sample of enumeration districts that are at least five percent black. Southern blacks are black individuals born in the South or black individuals born in the North whose fathers were born in the South.

The spatial distribution of the minority groups we study is displayed in Figure 2.1.[12] Panel A shows the concentration of southern–born blacks in the "Black Belt" south of downtown with a secondary population to the west. Northern–born blacks appear to be concentrated in the Black Belt as well, but with larger numbers living to the north and south of the most densely African American areas. Figure 2.2 graphically illustrates the variation in where northern and southern blacks lived in finer detail, with a close up view of the black

[12]The two blank areas are the result of missing data. We had to omit 84 enumeration districts (out of 1884) from our sample: 36 were missing from Ancestry.com's database and 48 had illegible or missing land use maps, leaving us with 1800 observations.

neighborhoods to the south and west of downtown. Focusing exclusively on enumeration districts that were at least 5 percent black, the figure shows the spatial distribution in the percentage of each neighborhoods' black population that we classify as being southern black. As is clear from the figure, the southern black composition of these neighborhoods ranges from a low near 20 percent to a high in excess of 80 percent. We thus find there is sufficient variation in where southern and northern blacks lived to examine their impact on zoning separately.

Turning to European immigrants, Panels C and D of Figure 2.1 show the distribution of first– and second–generation immigrants, respectively. Numerically much larger than the black population, first–generation immigrants were most concentrated in inland neighborhoods in the periphery of the central business district. Second–generation immigrants occupy the next ring of enumeration districts further out from the downtown, particularly in the northwest.

2.3.2 The 1922 Chicago Land Use Survey

The comprehensive land use survey we draw upon was conducted by the Chicago Zoning Commission in 1922 for the purposes of informing the drafting process for the zoning ordinance. Four teams, each equipped with an automobile, recorded the use of every building and lot in the city (Zoning Chicago 1922 Pamphlet). From these survey maps we obtain the location of every commercial and manufacturing use in the city; we also obtain the location and number of stories for every building with four or more stories. We geocoded the largest sample to date of this pre–zoning survey for our study. While previous work by McMillen and McDonald used a sample of 1000 blocks, we digitized nearly two–thirds of the city by land mass.[13] Our sample covers 79.4 percent of the 1920 population along with 97.8 percent of blacks and 80.8 percent of first–generation immigrants. Figure 2.3 provides a graphical illustration of the land mass covered by our sample.

[13]Our sample covers 64 percent of the 1920 area of Chicago and 56 percent of the current (2013) city area.

Figure 2.3: Sample area.

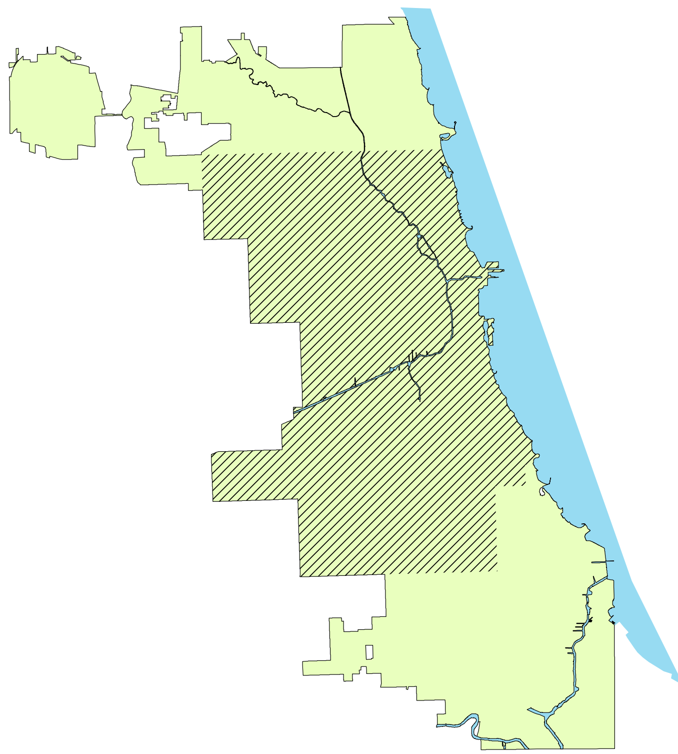
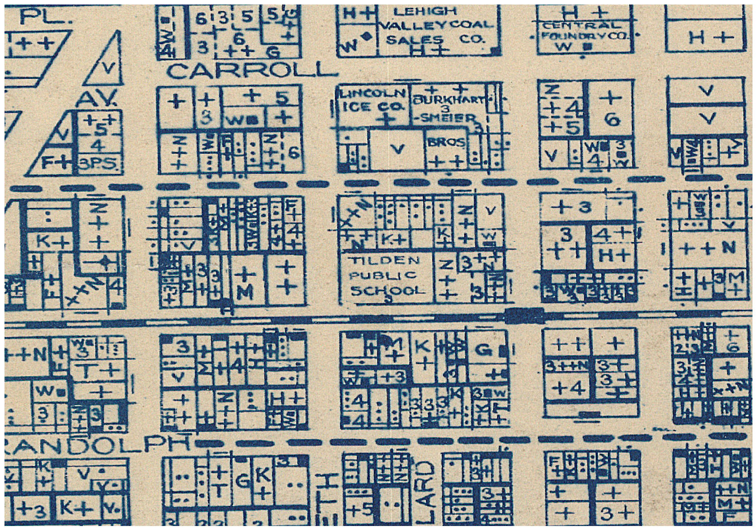


Figure 2.4 provides a map image of several blocks from the survey. The Tilden Public School in the center of the image is surrounded by noxious facilities, indicated by "++N" on the map. The building heights of all structures over four stories can also be seen (surveyors occasionally indicated three–story buildings although not consistently). The letters on buildings correspond to specific uses, which we classified as residential, commercial, or manufacturing (further distinguished by subclass) using the same system as the Chicago Zoning Commission in 1922. Of particular interest to our study are the various manufacturing classes: A and B include general manufacturing that does not cause a nuisance but may require yard storage, class S includes large–scale industrial facilities such as rail yards and granaries, class D covers storage of explosives and high pressure gases, and class C includes manufacturing facilities that emit noise, smoke, odors, or pose a fire risk. We consider the noxious facilities in classes C separately in much of our analysis (only one instance of Class D manufacturing exists in our sample). Commercial use is indicated using only one category

and covers retail establishments, offices, and entertainment venues such as theaters.

Figure 2.4: Land use map sample.



2.3.3 Comprehensive Zoning Ordinance of 1923

We digitized the initial zoning ordinance for the same broad sample of Chicago as the land use survey, recording both volume zoning and use zoning. The volume districts in the zoning ordinance are essentially rough concentric rings radiating out from the central business district. Figure 2.5.A shows the digitization of these districts with each enumeration district assigned to the volume district most common within its borders. Our empirical work focuses on the two outermost rings, which were volume districts 1 and 2. Under zoning for volume district 1, buildings were capped at 5 to 6 stories and could cover only 50 percent of an interior lot. In volume district 2, apartment buildings could reach 12 to 13 stories and cover 60 percent of the lot. However, the effective difference in height and density limitations between these two districts was actually much greater due to restrictions on overall building volume. The volume district 1 maximum building height was effectively 33 feet, corresponding to roughly three stories, while in district 2 the maximum height was effectively 8 to 10 stories. The inner three volume districts allowed buildings with effective heights of 11, 16, and 22 stories, respectively, and were found only in the central business district and surrounding areas (see Figure 2.5.A). There were no density "minimums," only restrictions only the maximum volume, height, and lot coverage.

Figure 2.5: Volume and use zoning.

Panel A: Digitized Volume Zone Map *Panel B: Use Zoning Map Sample*

Use zoning delineated the city into four distinct districts: residential (single–family homes), apartment, commercial, and manufacturing. These use districts were hierarchical, with apartment districts allowing residential uses, commercial districts allowing both apartments and single–family homes, and manufacturing districts allowing any use.[14] The residential category was rarely used in the initial zoning ordinance; only three percent of the enumeration districts in our sample have any zoning of this type. Figure 2.5.B shows a section of a use zoning map from an area west of the downtown along the Chicago River. Zones for apartments, commercial activity, and manufacturing can all be seen.

2.4 Empirical Approach

Our empirical approach relies on the ability to observe the same land use data employed by the Chicago Zoning Commission when they drafted the ordinance. We pose two questions in our empirical work. First, how were minorities sorted across the city and within

[14]There were additional gradations within the commercial and manufacturing districts, with certain objectionable commercial uses barred if they were within 125 feet of a residential or apartment district, while certain manufacturing uses were barred if they were within 100 to 2000 feet of a residential, apartment, or commercial district. Some commercial uses within 125 feet of residential or apartment districts also saw restrictions on the hours during which trucking activities could occur.

neighborhoods with respect to existing land use and urban geography prior to the zoning ordinance? Second, accounting for geography and extant land use, what was the impact of various minority populations on zoning outcomes?

Crucial to the identification of the second question is that we sufficiently account for other causes of zoning that also influenced the demographic composition of enumeration districts. By conditioning on an extensive array of spatial, land use, and transportation variables, our empirical strategy attempts to block all "back–door" paths from our demographic variables to zoning outcomes (Pearl, 2009). In the language of Rosenbaum and Rubin (1983), we render the non–demographic causes of zoning "conditionally ignorable," and so the effect of demographic composition on zoning outcomes is identified. Recognizing the limits of our ability to block all alternate mechanisms via controls, we attempt to further verify our main results using a series of robustness checks in Section VI.

The models we estimate are all single index models, i.e., functions of a linear combination $x'\beta$ of our covariates. To permit nonlinearities in responses, we frequently allow covariates to enter through indicators as well as polynomials. Specifically, spatial variables such as distance to the central business district, distance to the nearest major street, distance to Lake Michigan, distance to the nearest river, distance to the nearest railroad, and distance to an ancillary railroad all enter as quartic polynomials, and we include indicators that equal one whenever an enumeration district is proximate to any of these features. We also include quartic polynomials for population density and the area of the enumeration districts. Indicators for overlapping a railroad or major street are included, as is a quartic polynomial for the distance to the nearest railroad.

To control for existing land use, we include variables measuring the density of commercial uses, warehouses, and each of the five different manufacturing use classes; these enter as both indicators and quadratic polynomials in the density of each type of use. To account for large industrial sites, we add an indicator equal to one if the enumeration district includes a contiguous area greater than 800,000 square feet (approximately four city blocks) populated by heavy industrial activity. We include separate indicators for enumeration districts that overlap the Union Stockyards and those that are within 1,000 feet of the Stockyards. To capture the industrial character of the area surrounding an enumeration district, we also include counts of different manufacturing uses in 500 and 1,000–foot rings around each enumeration district. To account for the existing distribution of building heights, we include the densities of four, five, six, seven, eight, nine, and ten story buildings. We also include the density of eleven through twenty–five story buildings; disaggregating this category has little impact on the analysis due to the concentration of these buildings in the central business district.

To address the possibility that recent immigrants and black migrants located in cheaper areas of the city that were also suitable for manufacturing activity, we include as a control a measure of land values transcribed by Gabriel Ahlfeldt and Daniel McMillen from the 1913 edition of Olcott's Blue Books.[15] Specifically, this variable is the average land value per front foot based on 125 foot tracts (see McMillen, 2012). As a further control for wealth, we use the head of household variable in the census to develop an income measure based on live–in hired help. For each enumeration district, we count the number of household heads as well as the number of individuals who report being a maid, cook, servant, or laborer in relation to the head of house.[16] We then compute the ratio of live–in hired help to heads of household and include this value in our regressions. We also include ward fixed effects to account for differential political influence exerted by alderman. There are approximately 51 enumeration districts per ward in our sample. Finally, to measure home neighborhood motivations for the zoning board members, we added an indicator for whether a zoning board member lived

[15] Land prices may have influenced zoning directly; for example, the zoning board may have considered areas with cheaper land to be more suited for large–scale industrial uses. Land prices may also proxy for unobservable neighborhood characteristics. Since both racial and ethnic composition and unobservable neighborhood characteristics can be expected to have had a causal effect on land prices, conditioning on land prices may induce a correlation between these variables even if they are unconditionally independent. This "collider–stratification" could bias the estimation of our coefficients of interest (Greenland 2003, Pearl 2009). However, despite the fact that land prices are strongly correlated with both our explanatory and outcome variables, their inclusion has a negligible effect on our coefficient estimates.

[16] We do not observe occupation in the Ancestry.com data, relation to head of house is our only opportunity to measure household employment status.

in the enumeration district.[17]

We measure zoning outcomes using both continuous and discrete variables as appropriate. For example, we assess the probability that an enumeration district contains any manufacturing zoning as well as the percentage of the enumeration district that is zoned for manufacturing uses. When the outcome is a binary indicator, we typically report results from a probit model in terms of average marginal effects. We consider only discrete outcomes for density zoning because there are relatively few enumeration districts straddling the relevant density zone borders. Each enumeration district is assigned to the volume district in which most of its area falls. When considering continuous outcomes, we typically report results from a Tobit model, which assumes the existence of an underlying variable that equals the index x' plus a normally distributed error term. The observable value of the latent variable is equal to zero if the latent variable is below zero; similarly, it is equal to one if the latent variable exceeds one. This model accounts for the fact that EDs receiving boundary values may differ substantially in their suitability for different types of zoning.[18]

Our baseline specification is thus

$$\% or indicator for zoning type_i = f\left(x_i'\beta + ward_i\right) + \epsilon_i$$

where the zoning type is manufacturing or commercial and x_i includes the extensive list

[17] Only one enumeration district with a board member received any industrial zoning. We explored a variety of political representation indicators in our analysis, including whether a ward's alderman served on the zoning board. We found small and insignificant results on manufacturing zoning for all variables relating to local representation on the board.

[18] In the Tobit model, β is the marginal effect of x on the underlying latent variable; the marginal effect over the uncensored range is obtained by multiplying this β by a shrinkage factor, which explains why it is generally larger than the estimates we obtain from the OLS specifications (McDonald and Moffit, 1980). An alternative estimation procedure involves fitting a beta distribution whose parameters are a function of our covariates. However, this is inappropriate since we observe many values at the boundary, and these values are discarded when estimating the parameters of the beta distribution because there is no support on the boundary. Papke and Wooldridge (1996) recommend the fractional logit estimation procedure in this context. The fractional logit estimator is a generalized linear model where the conditional expectation of the outcome variable is equal to the logit function evaluated at the index $x_i'\beta$. This ensures that the output from the model is always bounded between zero and one. As a robustness check, we also estimated all of the continuous dependent variable models reported here using the fractional logit specification. These results were qualitatively similar to those reported in the paper. For parsimony, we only report the OLS and Tobit results.

of spatial and land use controls described above as well as measures of the share of the enumeration district population composed of blacks, the share composed of first–generation immigrants, and the share composed of second–generation immigrants. We use robust standard errors throughout the analysis (White, 1980).[19] We also decompose the black share into southern– and northern–born blacks in much of the analysis.

2.5 Existing Patterns of Minority Residential Location

We begin by documenting the distribution of minority location across the city and within neighborhoods with respect to measures of urban density, proximity to commercial and manufacturing activity, and proximity to other demographic groups. We employ two approaches to measure pre–existing sorting associated with land use. First, we report the exposure to various uses experienced by the average member of each demographic group we study. Second, we regress a variety of land use variables on demographic composition along with basic spatial controls to understand the relationship between demographics and pre–existing land uses.

Table 2.2 reports the average exposure results. The first two columns of Panel A report the average number of four story and four to ten story buildings per acre experienced by the average member of each demographic group we study. Southern–born blacks had the highest exposure to both categories of tall structures, followed by northern blacks and then first–generation immigrants. However, first–generation immigrants experienced the highest population density (column 3). The ordering is similar for commercial enterprises per acre, noxious facilities per acre (defined as the number of Manufacturing class C uses), and general manufacturing facilities per acre (defined as Manufacturing classes B, C, and S uses) with both black groups and first–generation immigrants having the highest exposure (columns 4–6). Although industrial facility exposure was essentially equal across groups, southern

[19]Using the method of Conley (1999) to construct standard errors robust to spatial autocorrelation consistently resulted in smaller standard errors, which we do not report here.

Table 2.2: Average exposures.

Panel A Group	Number 4+ story buildings (1)	Number 4-10 story buildings (2)	Population density (3)	Commercial enterprises per acre (4)	Noxious facilities per acre (5)	Industrial facilities per acre (6)
Southern blacks	0.19	0.22	64.91	0.91	0.0072	0.02
Northern blacks	0.17	0.21	64.21	0.89	0.0060	0.02
First-gen. immigrants	0.12	0.15	70.09	1.01	0.0070	0.02
Second-gen. immigrants	0.08	0.11	58.01	0.72	0.0046	0.01
Third-gen. whites	0.10	0.14	55.00	0.64	0.0040	0.01
Sample Average	0.11	0.15	58.03	0.79	0.0071	0.02

Panel B Group	Share southern black (1)	Share northern black (2)	Share first gen. immigrant (3)	Share sec. gen. immigrant (4)	Share white 3rd gen. (5)	1913 avg. land prices (6)
Southern blacks	0.45	0.19	0.16	0.08	0.14	90.66
Northern blacks	0.42	0.19	0.17	0.09	0.15	96.69
First-gen. immigrants	0.01	0.01	0.60	0.19	0.20	93.11
Second-gen. immigrants	0.01	0.01	0.45	0.23	0.29	92.15
Third-gen. whites	0.02	0.01	0.38	0.23	0.36	125.67
Sample Average	0.04	0.02	0.46	0.21	0.27	103.37
Sample Std. Dev.	0.13	0.06	0.22	0.08	0.16	386.98

Notes: The numbers in panel A reflect the average value of the variable specified for the column experienced by the average member of the group specified for the row. For example, first two columns of Panel A report the average number of four story and four to ten story buildings per acre experienced by the average member of each demographic group we study. Panel B documents minority exposure to other demographic groups as well as the typical 1913 average land value experienced by the typical member of each group. The demographic data come from Ancestry.com and the land use counts were computed using the 1922 Land Use Survey created by the Chicago Zoning Commission. See Figure 1 for demographic group definitions.

blacks and first–generation immigrants were exposed to more noxious industrial uses than other groups (.007 uses per acre compared with .006 for northern blacks and .0046 for second–generation immigrants).

Minority exposure to other demographic groups is shown in Panel B. As we would expect, both northern and southern blacks live in enumeration districts with larger shares of other blacks. However, the sum of share northern and share southern black faced by the average southern black is only .64. We interpret this result as evidence that blacks were not completely segregated by race; we also note that many black individuals served as live–in maids in white neighborhoods and would have been enumerated in their employers' houses. Immigrants and native whites had very low exposure to blacks (average share .02 and .03, respectively). Finally, we observe that southern blacks lived on the cheapest land relative to other groups, with first–generation immigrants just behind them. The difference in land values faced by the average black and average third–generation white is a striking $35 ($90.66 versus $125.67 in 1913 dollars) and underscores the importance of controlling for land values in our regressions.

Table 2.3: Reverse exposure regressions.

Panel A (no controls)	Number 4+ story buildings (1)	Number 4-10 story buildings (2)	Population density (3)	Commercial enterprises per acre (4)	Noxious facilities per acre (5)	Industrial facilities per acre (6)
Southern black share	0.0167	-0.105	9.211	0.461	0.0553*	0.0750
	(0.125)	(0.162)	(18.55)	(0.393)	(0.0297)	(0.0464)
Northern black share	-0.718***	-0.849**	11.42	0.355	-0.115**	-0.130
	(0.255)	(0.372)	(37.47)	(0.804)	(0.0551)	(0.0886)
First-gen. immigrant share	-0.123***	-0.216***	38.85***	1.212***	0.0116***	0.0436***
	(0.0335)	(0.0466)	(5.800)	(0.106)	(0.00410)	(0.00865)
Second-gen. immigrant share	-1.239***	-1.655***	-75.03***	-1.965***	-0.0307***	-0.0575**
	(0.0971)	(0.136)	(16.22)	(0.304)	(0.0119)	(0.0272)
R-squared	0.128	0.104	0.112	0.243	0.030	0.058
Panel B (with controls)						
Southern black share	0.150	0.173	0.357	0.0187	0.0734**	0.0875*
	(0.121)	(0.133)	(16.17)	(0.366)	(0.0306)	(0.0471)
Northern black share	-0.307	-0.287	-13.74	0.886	-0.121**	-0.132*
	(0.225)	(0.298)	(32.05)	(0.733)	(0.0470)	(0.0769)
First-gen. immigrant share	-0.0592	-0.0446	24.39***	0.755***	0.00420	0.0106
	(0.0452)	(0.0606)	(7.188)	(0.141)	(0.00678)	(0.0135)
Second-gen. immigrant share	-0.303***	-0.188	-52.20***	-0.907***	0.0105	0.0104
	(0.0991)	(0.138)	(17.02)	(0.325)	(0.0158)	(0.0353)
R-squared	0.576	0.641	0.541	0.547	0.187	0.273

Notes: This table compares the sorting patterns of blacks and immigrants using a reverse regression analysis to identify the relationship between demographic groups and land uses while controlling for potentially confounding correlations with other demographic or spatial variables. Panel A includes no spatial or land use controls; the results can be thought of as the characteristics of areas in the cities where minority groups lived relative to third-generation whites. Panel B presents the results of the same specifications with the full set of geographic controls listed in Appendix Table 1; these results can be thought of as the urban characteristics faced by minorities relative to third-generation whites conditional on the particular neighborhood of the city in which they lived. See Figure 1 for demographic group definitions.

As a second approach, we compare the sorting patterns of blacks and immigrants using a reverse regression analysis to identify the relationship between demographic groups and land uses while controlling for potentially confounding correlations with other demographic or spatial variables. We regress land use variables on our slate of demographic variables and (in some cases) additional controls. Panel A of Table 2.3 includes no spatial or land use controls; the results can thus be thought of as the characteristics of areas in the cities where minority groups lived relative to third–generation whites (the omitted demographic group).[20] Panel B of Table 2.3 presents the results of the same specifications with the full set of spatial controls, including the area of the enumeration district, ward fixed effects,

[20]We include only our proxy for income, maids per head of household, as a control.

and distances to the central business district, major street, Lake Michigan, nearest river, and nearest railroad; these results can be thought of as the urban characteristics faced by minorities relative to third–generation whites conditional on the particular neighborhood of the city in which they lived.

The results from these regressions suggest relationships similar to those obtained from the average exposure exercise. Areas of the city with more second–generation immigrants and northern blacks had fewer tall structures compared with areas having more native whites. This finding is consistent with the pictorial evidence in Figure 2.1 showing that second–generation immigrants lived the furthest from the center city. Whether we look across the city (Panel A) or within neighborhoods (Panel B), first–generation immigrants lived in the densest, most commercial areas while southern blacks were exposed to more noxious and non–noxious manufacturing relative to third–generation whites (see columns 3 and 4 for first–generation immigrants and columns 5 and 6 for southern blacks). Furthermore, first–generation immigrants located in more industrial areas of the city (Panel A, columns 5 and 6).

These results underscore the need to control for existing sorting according to land use when asking how the spatial distribution of minorities shaped the zoning ordinance. We note, however, that the land use and demographic composition relationships identified in Panel B are in many instances at odds with the zoning findings we report in the next section, suggesting that our main results cannot driven solely by pre–existing relationships between land use and demography that later influenced the zoning ordinance.

2.6 The Impact of Minority Share on Zoning Outcomes

2.6.1 Density Zoning

We begin our analysis by exploring whether density zoning was used as a tool to concentrate blacks in higher density neighborhoods, a potential precursor to modern day arguments regarding exclusionary zoning.[21] Because the volume districts were essentially concentric rings radiating out from the central business district, the key tradeoff is between adjacent volume categories. We focus on the two outermost rings, which were volume districts 1 and 2 (see Figure 2.5.A). Under zoning for volume district 1, buildings were effectively capped at 3 stories In volume district 2, apartment buildings could reach as high as 8 to 10 stories. As a result, volume districts 1 and 2 effectively delineated the boundary between locations where 8 to 10 story tenements were allowed and locations where residential development was limited to structures of no more than 3 stories. This boundary represents the relevant margin for the proto–exclusionary zoning behavior we seek to analyze. We therefore focus our analysis on the border between volume districts 1 and 2.

Table 2.4: Density zoning results.

	Indicator for Receiving a Majority Zoning for Higher Density					
	(1)	(2)	(3)	(4)	(5)	(6)
Total black percent share	0.130*		0.0340		0.163**	
	(0.0757)		(0.0278)		(0.0682)	
Southern black share		0.211		-0.00348		0.187
		(0.227)		(0.0619)		(0.138)
Northern black share		-0.0456		0.0303		-0.00489
		(0.135)		(0.0446)		(0.103)
First-gen. immigrant share	0.00782	0.00789	-0.0729***	-0.0731***	-0.177**	-0.176**
	(0.0168)	(0.0168)	(0.0249)	(0.0249)	(0.0743)	(0.0741)
Second-gen. immigrant share	-0.0456***	-0.0454***	-0.00331	-0.00334	0.0626	0.0620
	(0.0175)	(0.0175)	(0.0179)	(0.0179)	(0.0473)	(0.0474)
1913 land values			2.520***	2.527***	6.091***	6.088***
			(0.488)	(0.490)	(1.347)	(1.351)
Diff. between black and first-gen. effect (p-value)	0.083		0.000		0.000	
Diff. between south. black and first-gen. effect (p-value)		0.368		0.273		0.012
Pseudo-R^2	0.037	0.037	0.675	0.676	0.529	0.529
Sample	Vol 1 & 2	Vol 1 & 2	Vol 1 & 2	Vol 1 & 2	Within 1000 ft of 1 & 2	
Controls	N	N	Y	Y	Y	Y
Observations	1,228	1,228	1,164	1,164	380	380

Notes: Columns (1)-(4) are restricted to areas that were zoned for volume districts 1 and 2 (the two lowest density areas). For columns (5)-(6), the sample is restricted to EDs within 1000ft of the border between volume districts 1 and 2 and excluding EDs containing any other type of volume zoning. The outcome indicator is equal to one if the enumeration district received a majority of volume district 2 zoning, the higher density type. The specifications in columns (3)-(6) include the full set of controls listed in Appendix Table 1. See Figure 1 for demographic group definitions.

To test for a potential exclusionary zoning motive in the location of these boundaries, in Table 2.4 we report the results from a probit analysis with the outcome variable equal to one if the enumeration district received a majority of zoning for volume district 2. To make the results readily comparable across groups, we report both coefficient estimates and standard errors in units of standard deviations for the relevant demographic variable (for instance, the coefficient on the variable "southern black" is reported in units of the standard deviation of

[21] A second potential vehicle through which the zoning ordinance could have been used to advance exclusionary motives would have been through the location of residential vs. apartment use zoning. However, in practice, residential zoning was restricted to outlying portions of the city in neighborhoods that were not proximate to significant numbers of black residents. Thus, there is little scope for an empirical analysis of tradeoffs along this margin.

southern black share). The standard deviations for each variable are reported in Table 2.1.

Columns 1–4 report the results with the sample consists of the entirety of volume districts 1 and 2. We begin with a simple specification omitting any controls (columns 1 and 2) and then add the full set of controls for geography, land use, political boundaries, and economic values (see Appendix Table 2.1 for list) in columns 3 and 4. In the no–controls specification, the presence of blacks appears positively correlated with higher density zoning (and second–generation immigrants negatively correlated). However, adding controls reduces the magnitude of the black effect and shows a precisely estimated negative first–generation immigrant effect on the likelihood of higher density zoning. The p–value of the difference between the effects of black share and first–generation immigrant share is .000 (column 3). These results are consistent with an exclusionary zoning strategy that, at the margin, sought to create low density neighborhoods for recent white immigrants while containing blacks in higher density areas where they had settled. We note that the main area of the "black belt" shown in Figure 2.2 contained none of the lowest density category.

We provide a further test of the exclusionary motive by examining black settlements that were located outside the main area of the "black belt" and nearer to areas that contained the lowest density category. In particular, we rerun the model limiting the sample to neighborhoods that were located along the boundary between volume districts 1 and 2 (within 1000 feet of both types and excluding neighborhoods that included any volume zoning other than districts 1 and 2). Our estimates suggest that either a one standard deviation increase in black share or a one standard deviation decrease in the first–generation immigrant share was associated with a 16 to 17 percentage point increase in the likelihood that an enumeration district received a majority of higher density zoning (column 5).[22]

We highlight that, in general, first–generation immigrants lived in more densely populated neighborhoods (see Tables 2 and 3) than did blacks prior to the zoning. This fact implies

[22]We note that some caution is warranted as these estimates leverage a much smaller number of black neighborhoods than was the case for the sample which included the entire coverage of volume districts one and two.

that these findings are unlikely to be driven by ex ante sorting and helps to explain why the inclusion of spatial controls makes such a difference for the estimated coefficients. In column 6, we divide the black population by origin, and our results suggest that the black effect is being driven by southern migrants (although these findings are not significant). We do, however, show that the difference in the black and first–generation immigrant effects (column 5) and southern black and first–generation immigrant effects (column 6) are statistically different at the one percent level, underscoring the differential treatment of the two groups.

In some ways, these findings are unexpected because our reading of the history indicates that the overarching concern of the zoning board relating to density was to keep skyscrapers in the downtown area. However, our results also suggest that a pre–cursor to modern–day exclusionary zoning may be found in the implementation of Chicago's initial zoning law. At the time, both European immigrants and black migrants faced housing shortages. At the margin, the Chicago Zoning Board appeared to adopt a strategy designed to keep blacks in place through high–density housing. The tendency towards lower–density zoning in European immigrant neighborhoods suggests an expectation that these immigrants would spread out across the city. Given the existence at the time of public animus towards both recent European immigrants and blacks, one possibility is that this differential treatment reflected the 1921 passage of federal immigration restrictions. With the border closing, the tide of European immigration was effectively stemmed, while the inflow of southern blacks was likely to continue unabated. Nonetheless, our findings suggest an early form of exclusionary zoning that was applied to blacks only and altered the trajectory of neighborhood density faced by minority groups.

2.6.2 Manufacturing Zoning

We next examine the relationship between the size of various minority groups and the likelihood of being zoned for manufacturing uses, again scaling coefficients by the standard deviation of the respective minority group. Turning first to the presence of any manufacturing zoning in the neighborhood, columns 1 through 3 of Table 2.5 report coefficient estimates from versions of equation (1) where the dependent variable is an indicator for the presence of any manufacturing zoning in the neighborhood. We begin with a simple probit model omitting all controls; this specification can be thought of as the standard environmental justice regression that does not control for sorting into areas suited for manufacturing. The results show a significant positive relationship between black and first–generation immigrant share and the likelihood of receiving at least some zoning for manufacturing uses. In column 2 we include the full vector of controls described in Section IV. Although the pseudo R–squared rises from .038 to .739 with the addition of controls, the black share effect increases in magnitude to .053. The first–generation immigrant effect is reduced by 40 percent but it still significant (.050).

Table 2.5: Manufacturing zoning results.

	Indicator for Any Industrial Zoning in ED Probit			Percent of ED Zoned Industrial Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)
Total black share	0.0428***	0.0536***		0.0626***	0.0420***	
	(0.0164)	(0.0122)		(0.0238)	(0.0158)	
Southern blacks share			0.0770***			0.112***
			(0.0147)			(0.0208)
Northern black share			-0.0232**			-0.0655***
			(0.0118)			(0.0174)
First-gen. immigrant share	0.0806***	0.0496***	0.0496***	0.135***	0.0669***	0.0682***
	(0.0150)	(0.0154)	(0.0153)	(0.0228)	(0.0198)	(0.0196)
Second-gen. immigrant share	-0.0210	0.0323**	0.0316**	-0.0448*	-0.0175	-0.0179
	(0.0161)	(0.0127)	(0.0127)	(0.0245)	(0.0162)	(0.0160)
1913 land values		0.0371**	0.0408**		0.000608	0.00557
		(0.0180)	(0.0177)		(0.00995)	(0.0108)
Controls	N	Y	Y	N	Y	Y
Pseudo-R^2	0.038	0.739	0.742	0.056	0.916	0.923
Observations	1,800	1,789	1,789	1,800	1,800	1,800

Notes: All specifications include the full set of controls listed in Appendix Table 1. See Figure 1 for demographic group definitions.

In column 3 we replicate column 2 with northern and southern blacks included separately. It is immediately clear from these results that the entire positive relationship between black share and the presence of manufacturing zoning is being driven by the southern black share. The coefficient estimates presented in Columns 1 and 2 indicate that enumeration districts with more first–generation immigrants were also more likely to be zoned for manufacturing uses. The magnitudes of these estimates are economically significant. The results in column 3 imply that a one standard deviation increase (roughly 13 percentage points) in southern black share is associated with an 8 percentage point increase in the likelihood of an enumeration district being zoned to include manufacturing uses. A standard deviation increase in the first–generation immigrant share (roughly 22 percentage points) is associated with a 5 percentage point increase in the likelihood of an enumeration district being zoned for manufacturing uses. These estimates are particularly large given that only 26 percent of enumeration districts in our sample received any manufacturing zoning. In Contrast, northern blacks were less likely to get manufacturing zoning in their neighborhoods. This finding is consistent with the anecdotal evidence regarding the status of northern blacks in the zoning process. Neighborhoods with larger populations of northern blacks were likely wealthier, more exclusive, and better represented by the Zoning Commission. In particular, contemporary reports suggest that Charles S. Duke, an African American on the Zoning Commission, actively worked to protecting northern black interests during the zoning process (Schwieterman and Caspall, 2006).

So far, we have argued that manufacturing use zoning was unambiguously "bad" in the sense that minority communities thus zoned would face disproportionate environmental hazards and decreased future home values. However, it is also possible that poor minority groups benefited economically from living in close proximity to their places of employment due to lower transportation costs. While we do not believe this is a driving force in our results, it is possible that within this context a positive value for the indicator may reflect advantageously located manufacturing zoning at the neighborhood fringe. One response to this concern is to focus instead on the share of a neighborhood that is zoned for manufacturing uses. The motivation here is that a positive relationship between minority share and the percentage of manufacturing zoning may be more consistent with the notion of encroachment of industry into black and immigrant neighborhoods and a finding that minorities were disadvantageously zoned.

Thus, we replicate our basic model using the continuous outcome measure, the percent of the enumeration district zoned for manufacturing. Tobit results are presented in columns

4 through 6 of Table 2.5. The dichotomy between the experience of northern and southern blacks is highlighted in these specifications. Overall, a one standard deviation increase in black share is associated with a roughly 4 percent increase in the area of an enumeration district being zoned for manufacturing uses. This effect is again driven by southern blacks, with a standard deviation increase in southern black share associated with an 11 percent increase in manufacturing zoning. Northern blacks were protected from manufacturing zoning along the intensive margin as well. In standard deviation terms, the southern black effect is nearly twice as large as the effect on first–generation immigrant share (.112 versus .068). Finally, we do not see any evidence that second–generation immigrant neighborhoods were disadvantageously zoned relative to third–generation white neighborhoods on either the extensive or intensive margin. Thus, our primary finding on manufacturing zoning is that southern black and first–generation immigrant neighborhoods were more likely to be zoned for manufacturing uses and tended to receive a larger amount of such zoning.[23]

Table 2.6: Robustness check on manufacturing zoning results.

	Indicator for Any Industrial Zoning in ED Probit			Percent of ED Zoned Industrial Tobit		
	No C or S (1)	No C or S and 500 feet away (2)	No C or S and 1000 feet away (3)	No C or S (4)	No C or S and 500 feet away (5)	No C or S and 1000 feet away (6)
Southern black share	0.0924***	0.0621***	0.0797***	0.145***	0.136***	0.158***
	(0.0161)	(0.0165)	(0.0280)	(0.0277)	(0.0288)	(0.0371)
Northern black share	-0.0319**	-0.0260**	-0.0490**	-0.0651***	-0.0443**	-0.0485**
	(0.0124)	(0.0127)	(0.0235)	(0.0213)	(0.0217)	(0.0242)
First-gen. immigrant share	0.0498***	0.0398**	0.0353	0.0996***	0.110***	0.127***
	(0.0178)	(0.0195)	(0.0223)	(0.0244)	(0.0280)	(0.0357)
Second-gen. immigrant share	0.0367**	0.0282*	0.0176	0.0212	0.0378	0.0488
	(0.0144)	(0.0164)	(0.0191)	(0.0208)	(0.0246)	(0.0299)
Observations	1,481	1,147	765	1,504	1,199	838

Notes: All specifications include the full set of controls listed in Appendix Table 1. See Figure 1 for demographic group definitions. Columns (1) and (4) include only enumeration districts with no Class C or S manufacturing. Columns (2) and (5) include only enumeration districts with no Class C or S manufacturing that are at least 500 feet away from such uses. Columns (3) and (6) include only enumeration districts with no Class C or S manufacturing that are at least 1,000 feet away from such uses.

So far our identification strategy has relied on controlling for an extensive set of spatial and pre–existing land use variables in addition to land prices, political influence, and a

[23]One potential area of interest is the fact that the first–generation immigrant group is itself composed of immigrants from many countries. In Appendix Table 2.1 we present the results from the indicator and continuous measures of industrial zoning with the first–generation immigrants further divided by sending country; these results are also presented in standard deviation terms. We observe that no group was as disadvantageously zoned for industrial uses as were southern blacks; furthermore, the coefficients on the share of the enumeration district population composed of the main ethnic groups (Polish, Russian, Italian, Irish, and German) are all quantitatively similar. Thus, it does not appear that any particular immigrant group was singled out for industrial zoning in the same way as southern blacks.

wealth proxy. We may nonetheless be concerned that our findings are driven by unobserved sorting of blacks and immigrants into industrial areas in a manner that is correlated with the initial zoning but not fully captured by our specification. To investigate the robustness of our approach, we rerun the specifications from Table 2.5 on samples of the city that would provide fewer opportunities for poor minority groups to sort into areas with high potential for manufacturing. We begin by restricting our sample to enumeration districts with no existing large–scale or noxious manufacturing uses (manufacturing classes C and S). We then further restrict the sample to enumeration districts without heavy or noxious uses that are also at least 500 feet away from such uses. Finally, we restrict the sample to enumeration districts at least 1000 feet away from any heavy or noxious uses. The results from probit and Tobit analyses on these restricted samples are presented in Table 2.6. Columns 1 and 4 present results from the least restricted samples while columns 3 and 6 present results from the most restricted samples. Results from each of the 3 different sample restrictions are quantitatively similar to the baseline results presented in Table 2.5.

2.6.3 Commercial Zoning

We next turn our attention to commercial zoning. While zoning for this use was undesirable for the wealthiest of neighborhoods that were exclusively residential, poor black and immigrant populations would likely have viewed close proximity to food stores, shops and entertainment venues as a benefit and would have viewed proximity to commercial uses as preferable to manufacturing uses.[24] Table 2.7 reports Tobit estimates of the relationship between demographics and the percentage of the enumeration district zoned for commercial uses.[25]

[24] An African American member of the Zoning Commission, Charles S. Duke, succeeded in removing two objectionable parts of the zoning ordinance covering the Black Belt, one of which would have extended a commercial district through Grand Boulevard where most of the "better colored homes" were situated (Schwieterman and Caspall, 2006, p. 29).

[25] Commercial zoning was much more prevalent than manufacturing zoning: 86 percent of enumeration districts received at least some commercial zoning, while only 26 percent received any manufacturing zoning. Thus, there is little reason to model commercial zoning outcomes using an indicator variable.

Table 2.7: Commercial zoning results.

	Percent of ED Zoned Commercial		
	Tobit		
	(1)	(2)	(3)
Total black percent share	0.0108	-0.0161*	
	(0.00856)	(0.00828)	
Southern black share			-0.0514***
			(0.0163)
Northern black share			0.0351**
			(0.0137)
First-gen. immigrant share	-0.00726	-0.0420***	-0.0412***
	(0.00741)	(0.00848)	(0.00848)
Second-gen. immigrant share	-0.0470***	-0.0168**	-0.0155**
	(0.00739)	(0.00713)	(0.00713)
1913 land values		-0.0104**	-0.0113**
		(0.00517)	(0.00520)
Controls			
Observations	1,800	1,800	1,800

Notes: All specifications include the full set of controls listed in Appendix Table 1. See Figure 1 for demographic group definitions.

We begin with the standard specification without controls in column 1 of Table 2.7 (continuing to list outcomes in terms of standard deviations). There is no effect of either black or first–generation immigrant share on commercial zoning while second–generation immigrant share is associated with less commercial zoning. However, adding controls addresses the pre–zoning sorting shown in Tables 2 and 3, and these results are shown in column 2 (black share entered separately) and column 3 (northern and southern black share entered separately). Column 3 shows that the small negative effect on total black share is driven by the presence of southern blacks with northern blacks receiving more commercial zoning. Similarly to the manufacturing results, we also find that first–generation immigrant neighborhoods also received less commercial zoning. We investigate the channels through which various groups received more manufacturing or commercial zoning in the next section.

2.6.4 Decomposing the Commercial vs. Manufacturing Zoning Tradeoff

To fully understand the mechanisms through which minority neighborhoods received more manufacturing and less commercial zoning, we split the sample by pre–existing levels of manufacturing and commercial activity and reproduce our baseline specifications in Table

2.8. Panel A presents results by quartile of pre–existing commercial use density, and Panel B by quartile of pre–existing manufacturing use density. For parsimony, we only present the coefficient estimates for the enumeration district's percent southern black and percent foreign born, again scaled so that the coefficients reflect the estimated effect of a one standard deviation increase in the given demographic group. The underlying regressions include the entire set of control and demographic variables that were incorporated in the baseline specification (listed in Appendix Table 2.1). To give a sense of scale and overall zoning patterns, we also present the average percentage of the neighborhoods in each quartile that were zoned for commercial or manufacturing uses. We also report by quartile the number of neighborhoods whose population is at least 10 percent southern black populations and greater than 40 percent first–generation immigrant.[26]

Focusing first on the commercial density decomposition, we note that there is a systematic relationship between pre–existing commercial density and the zoning of land for manufacturing and commercial uses. Moving from the first quartile to the fourth quartile in commercial density (from low levels of pre–existing commercial activity to high levels of pre–existing commercial activity), the average percentage of a neighborhood that received manufacturing zoning decreases monotonically from 16 to 4 percent. Furthermore, the average percentage of a neighborhood receiving commercial zoning increases monotonically from 9 to 36 percent. This decomposition reinforces McMillan and McDonald's (1999) finding that Chicago's initial zoning ordinance was significantly influenced by pre–existing land uses.

The regression results in Panel A also shed light on our finding that neighborhoods containing larger numbers of southern blacks or first–generation immigrants received larger shares of manufacturing zoning and smaller shares of commercial zoning, controlling for pre–existing land uses and geography. The largest concentration of neighborhoods comprised of at least 10 percent southern blacks occurs in the third quartile of the commercial density distribution. On average, these neighborhoods received a high level of commercial zoning

[26] We use a 10 percent cutoff for southern blacks and a 40 percent cutoff for foreign immigrants to characterize the presence of "enclaves" because of the difference in their relative size in the overall population.

Table 2.8: Decomposition results.

Panel A: Commercial Density	1st Quart.	2nd Quart.	3rd Quart.	4th Quart.
Pct. Zoned Manufacturing				
Avg. pct. zoned manufacturing	15.97%	12.16%	6.78%	4.05%
Percent southern black share	0.0467***	0.0131	0.0315*	0.0236
	(0.0158)	(0.0193)	(0.0187)	(0.0182)
Percent foreign born share	0.0563***	0.00398	0.0184	-0.00531
	(0.0187)	(0.0150)	(0.0146)	(0.00997)
Pct. Zoned Commercial				
Avg. pct. zoned commercial	9.25%	16.52%	25.23%	36.11%
Percent southern black share	0.00618	-0.000764	-0.0966***	-0.0406
	(0.0393)	(0.0269)	(0.0321)	(0.0260)
Percent foreign born share	0.00604	-0.0243	-0.0465**	-0.0712***
	(0.0110)	(0.0177)	(0.0197)	(0.0203)
# of observations	450	450	450	450
# of observations w/ s. black > 10%	22	40	54	45
# of observations w/ for. born > 40%	164	233	256	325

Panel B: Manufacturing Density	1st Quart.	2nd Quart.	3rd Quart.	4th Quart.
Pct. Zoned Manufacturing				
Avg. pct. zoned manufacturing	1.76%	12.15%	12.97%	15.00%
Percent southern black share	0.0188*	0.00256	0.0328*	0.0117
	(0.0109)	(0.0525)	(0.0188)	(0.0187)
Percent foreign born share	0.0134**	0.0613***	-0.00725	-0.0107
	(0.00645)	(0.0215)	(0.0155)	(0.0174)
Pct. Zoned Commercial				
Avg. pct. zoned commercial	13.88%	17.21%	23.37%	33.59%
Percent southern black share	0.0237	-0.161**	-0.0625**	-0.0283
	(0.0258)	(0.0628)	(0.0290)	(0.0272)
Percent foreign born share	-0.0116	-0.000510	-0.0172	-0.0561**
	(0.0124)	(0.0201)	(0.0177)	(0.0239)
# of observations	577	323	450	450
# of observations w/ s. black > 10%	38	11	61	51
# of observations w/ for. born > 40%	233	206	239	300

Notes: All specifications include the full set of controls listed in Appendix Table 1. See Figure 1 for demographic group definitions. All models are estimated using OLS.

and relatively low levels of manufacturing zoning. However, our regression results for these neighborhoods indicate that a one standard deviation increase in southern black share is associated with an almost 10 percentage point decrease in commercial zoning and a 3 percentage point increase in manufacturing zoning (relative to baseline averages of 25 percent and 7 percent, respectively). Thus, the presence of southern blacks appears to be associated with a significant shift away from potentially more desirable commercial zoning and towards manufacturing zoning in these neighborhoods.

A second dimension of the manufacturing effect is evident in the first quartile neighborhoods, which on average received high levels of manufacturing zoning. While these neighborhoods contain fewer southern blacks than those in any other quartile, when southern blacks are present, they are associated with a significant increase in the level of manufacturing zoning. A one standard deviation increase in southern black share is associated with a 4.6 percentage point increase, relative to a base of 16 percent. The first–generation immigrant results are generally similar to those for southern blacks with the exception that we do not see clear evidence of substitution between commercial and manufacturing in the third quartile of commercial density.

Panel B of Table 2.8 replicates the top panel with the sample decomposed based on pre–existing manufacturing density.[27] Very little manufacturing zoning was applied in these first quartile neighborhoods, all of which had no pre–existing manufacturing; on average, only 1.8 percent of these neighborhoods were zoned for manufacturing. The coefficient estimates from this quartile suggest that a large portion of the manufacturing zoning that did occur in these areas which had no extant manufacturing activity was concentrated in neighborhoods with large southern black and immigrant populations. The second quartile reveals a similar result for immigrants but not for southern blacks, although there were very few neighborhoods with a large number of southern blacks in this quartile.

Panel B also shows that higher levels of pre–existing manufacturing were generally associated with higher proportions of commercial zoning. The largest concentrations of southern blacks occurred in the third quartile of pre–existing manufacturing, while the largest concentrations of first–generation immigrants occurred in the fourth quartile. Both groups were associated with significantly lower levels of commercial zoning in these quartiles: a one standard deviation increase in southern black share in the third quartile led to 6.3 percentage points less commercial zoning, relative to an average of 23.4 percent, while a one standard deviation increase in first–generation immigrant share in the fourth quartile led to 5.6 percentage points less commercial zoning, relative to an average of 33.6 percent. We also note that for southern blacks, there is evidence that, in the third quartile, they are associated with substitution from commercial zoning to manufacturing zoning. This last result mimics

[27]Here, there are 577 enumeration districts with no pre–existing manufacturing uses. As a result, the first and second quartiles differ in their number of observations.

the finding from Panel A: the presence of southern blacks led to an overall shift out of commercial zoning and into manufacturing zoning in neighborhoods that could have received either type based on existing uses.

2.6.5 Impact of 1923 Zoning on 1940 Housing Density and Zoning Revisions

In Table 2.9 we explore whether inequitable treatment in the initial zoning ordinance had persistent effects. We begin with the density component of the ordinance, linking the volume zoning outcome in 1923 to housing and population density from the 1940 census. We are also interested in the impact of the use zoning ordinance on the location of industrial and commercial activity over time; however, the limited availability of land use data in the early twentieth century makes it difficult to undertake a similar analysis for this part of the ordinance.[28] Instead, we digitized the first major revision to the Chicago zoning ordinance, which occurred in 1942, to examine the persistence of use zoning. We show in a companion paper (Shertzer, Twinam, and Walsh, 2014) that the 1923 zoning ordinance had robust effects on the location of commercial and industrial activity in 2005. Assessing the persistence in zoning over the 1923 to 1942 period thus sheds light on the channels through which the initial zoning ordinance affected minority exposure to industry and commerce over the ensuing decades.

Table 2.9: Intermediate run zoning results.

	Housing Unit Density (1)	Population Density (2)	Percent Zoned Industrial (3)	Indicator for Industrial Zoning (4)	Percent Zoned Commercial (5)
Percent 1923 lowest density zoning	-1.484***	-3.638*	0.0521	-0.0205	0.0195
	(0.548)	(1.909)	(0.0941)	(0.116)	(0.0446)
Percent 1923 industrial zoning	-4.747**	-13.82*	0.952***	0.648***	0.0643**
	(2.285)	(8.146)	(0.0421)	(0.0554)	(0.0269)
Percent 1923 commercial zoning	0.167	1.093	0.225***	0.143***	0.668***
	(2.171)	(7.453)	(0.0551)	(0.0464)	(0.0289)
Model	OLS	OLS	Tobit	Probit	Tobit
Observations	347	347	1,800	1,687	1,800

Notes: The sample in columns (1) and (2) is restricted to EDs within 1000ft of the border between density districts 1 and 2. The housing density is defined as the number of housing units per acre in the 1940 census. Population density is defined as individuals per acre from the 1940 census. The zoning outcomes in columns (3)-(5) are from the revision to the Chicago zoning ordinance issued in 1942. The unit of observation for all specifications is the 1920 enumeration district. Zoning and density measures from the 1940 decade are assigned to these geographic areas using areal interpolation. All specifications include the full vector of controls listed in Appendix Table 1.

[28] For instance, the microdata for the census of manufacturers were not generally preserved in the same manner as the microdata for the census of the population in the early twentieth century.

For the density persistence analysis, we begin with the sample of 1920 census enumeration districts that were located 1000 feet from the border between the two most restrictive volume zoning categories from the 1923 ordinance and proceed in a similar manner to our exclusionary zoning analysis in part a. The population and housing unit density of these geographic units in 1940 is interpolated using the 1940 census tracts. Our specifications include the full set of controls for 1922 land use, building characteristics, population density, geography, and land values employed in the main analysis (see Appendix Table 2.I for the full list), plus the 1923 zoning shares. Column 1 shows that moving to the lowest density category from the second lowest (from volume category 2 to 1) is associated with 1.6 fewer housing units per acre in 1940. The average housing unit density in this sample is 10.9, so this effect represents a 15 percent decrease with respect to the mean. These results suggest that zoning had a causal effect on the subsequent development of the housing stock. Taken together with our results from part a., these findings suggest that black neighborhoods became more densely developed relative to immigrant neighborhoods within two decades of the zoning ordinance. The effect of lower density zoning on population density is negative and significant at the 10 percent level (column 2).

Turning to use zoning, we find strong evidence of persistence. Column 3 indicates that a standard deviation increase in 1923 industrial zoning share is associated with an 18.6 percent increase in industrial zoning share in 1942 (.196*.952=.186) off a base of 9 percent. The effect is similarly large if we use an indicator for any industrial zoning (column 4), with the presence of industrial zoning in 1923 associated with a 65 percentage point increase in the likelihood of industrial zoning in 1942. Finally, we find that commercial zoning is persistent to a similar degree (column 5). Taken together with our main findings, these results suggest that the inequitable treatment of minorities in the use zoning ordinance had meaningful impacts and persisted for decades.

2.7 Conclusion

This paper examines the introduction of zoning in Chicago and asks whether ostensibly race blind comprehensive zoning ordinances discriminated against minorities. We find evidence that neighborhoods with more black residents were more likely to be zoned for higher density buildings, suggesting that volume restrictions were used as an early form of exclusionary zoning. We also find robust and quantitatively important evidence that otherwise comparable neighborhoods with larger populations of blacks or immigrants were zoned disproportionately for manufacturing, suggesting environmental racism was present in the zoning process. Our results are robust to the inclusion of an extensive set of controls for geography, existing land use, land prices, and political factors; it is thus unlikely that sorting of minorities into neighborhoods suitable for industry can explain our results. These findings suggest that zoning reshaped the urban landscape faced by black and immigrant residents of the city of Chicago. Immigrants had selected into more densely populated neighborhoods in the early twentieth century, but one result of the zoning ordinance was to reduce the density of immigrant neighborhoods in the future via constraints on building height. Meanwhile, black neighborhoods were zoned for higher building density along the same margin. Zoning for higher density and mixed uses meant that minorities were excluded from the economic benefit of low density, purely residential zoning in the 1923 ordinance in terms of increased property values. Moreover, greater exposure to industrial uses may have adversely affected the health of blacks and immigrants relative to native whites. The findings of this paper indicate that zoning may have played a significant causal role in the adverse experience of minorities documented in the environmental justice and exclusionary zoning literature, and further research is needed to study the long–term impacts of land use regulation.

Chapter 3

Danger zone

3.1 Introduction

Crime is an important determinant of the quality of neighborhoods and cities. A substantial portion of central–city depopulation beginning in the 1970s can be attributed directly to crime, and rising crime is associated with neighborhood decline and increased isolation of minorities within cities (Cullen and Levitt 1999, Morenoff and Sampson 1997). The negative consequences of these developments, such as deteriorating public services and higher rates of poverty, are well documented (Bradbury, Downs and Small 1982, Massey and Denton 1993). A recent estimate found the total cost of a single armed robbery to be approximately $42,310 (McCollister, French and Fang 2010).[1] Using this estimate, the total cost of the 31,123 robberies in my sample is $1.3 billion – a considerable economic burden. Street crime patterns exhibit dramatic spatial heterogeneity, both between and within neighborhoods. Many have conjectured that land use patterns are an major determinant of street crime, and planners have embraced the notion that cities can use zoning regulations to shape land use patterns in a manner that will cultivate safe, vibrant neighborhoods.

[1]Estimates of the cost of crime vary widely; the estimated cost of an armed robbery ranges from $18,591 to $280,237 in 2008 dollars (Cohen, Rust, Steen and Tidd 2004, McCollister et al. 2010, Miller, Cohen and Wiersema 1996). The McCollister et al. (2010) estimate includes tangible and intangible costs to victims as well as costs incurred by the justice system.

Since the seminal work of Jane Jacobs, it has become conventional wisdom among both academic and professional urban planners that mixing commercial and residential land uses will lead to fewer street crimes by increasing pedestrian traffic and generating more supervision of street activities (Jacobs 1961). Glaeser (2011) has argued that high residential densities should operate against crime through the same channel. These ideas have been widely influential in practice; for example, Mayor Bloomberg presided over the rezoning of 37% of New York City, much of it for high-density, mixed-use developments encouraged by these theories (Silverman 2013). Many other major cities, such as Houston, Texas and Vancouver, British Columbia, have embraced the trend towards mixed–use and high–density development (Punter 2007, Sarnoff and Kaplan 2007); even smaller cities such as Sarasota, Florida have pursued rezoning plans to generate greater pedestrian traffic in high–crime areas through a greater availability and variety of commercial uses (Carter, Carter and Dannenberg 2003). Anderson, MacDonald, Bluthenthal and Ashwood (2013) refer to the argument that commercial and mixed–use zoning reduce crime as a "common–sense notion" and Geraldine Pettersson claims that "most of the present–day assumptions about the relationship between mixed uses and crime prevention appear to draw heavily on the arguments of Jane Jacobs and little else" (Coupland 1997).

In contrast, criminologists emphasize that mixed uses and high residential density generate more contact between potential offenders and potential victims. The "routine activities" theory of Cohen and Felson (1979) argues that direct–contact predatory crime requires the "convergence in space and time of *likely offenders*, *suitable targets* and the *absence of capable guardians*," which is arguably more likely to occur in higher–density, mixed–use areas. Stark (1987) argues that mixed uses and high density result in greater transience, anonymity, and "moral cynicism among residents," reducing neighborhood collective efficacy. This follows a long tradition in the sociology literature of linking high densities to pathological behavior (Sampson 1983, Wirth 1938). Additionally, specific commercial uses such as bars and liquor stores may serve as crime generators (Roncek and Bell 1981). The fact that crime is typically concentrated on a small number of street segments and intersections ("hot spots") lends further credence to the notion that place characteristics can be criminogenic (Weisburd, Groff and Yang 2012).[2]

The empirical evidence for these theories is limited, and existing studies suffer from a variety of measurement and identification problems. Since crime is an enormously costly burden on cities, and local governments exert substantial influence over the built environment through zoning, quantifying the criminogenic externalities of commercial and residential land use is of first–order importance. To this end, I study the effect of commercial and high–density residential use on street crime. I develop a unique high–resolution dataset on land use types in the City of Chicago using a comprehensive 2005 land use survey supplemented with exact locations and descriptions of every licensed restaurant, (late–hour) bar, and liquor store in the city. I combine this with detailed, spatially–referenced crime data covering all reported crime incidents over the period 2008–2013. My sample consists of approximately 20,000 street segments. This fine spatial scale implies that the analysis maps directly to the theory, allowing me to avoid the ecological inference problems which made the results of previous studies difficult to interpret. This approach also allows for the measurement of the spatial scale of land use effects, which has been largely ignored by the previous literature despite its important implications for the extent to which negative land use externalities can be mitigated through alternative policing strategies. I am also able to determine the extent to which the effect of commercial activity on crime is driven by particular uses, which has not been previously documented.

To address unobserved neighborhood characteristics and reverse causality, I employ an instrumental variables approach, using the city's 1923 zoning code as an instrument for modern land use. I show that historical zoning is a strong predictor of modern land use,

[2] Sherman, Gartin and Buerger (1989) find that 3% of addresses/intersections in Minneapolis are resposible for 50% of calls to the police. Braga, Papachristos and Hureau (2010) find a similar result for gun crime in Boston and show that these hot spots tend to persist over long time horizons. This pattern has been documented in Seattle and Tel Aviv–Jaffa as well, suggesting that this is a general feature of urban areas (Weisburd and Amram 2014, Weisburd, Bushway, Lum and Yang 2004).

and I validate the assumption of exogeneity by showing that unobservable neighborhood characteristics affecting crime and zoning in the 1920s were not persistent. To identify the impact of specific commercial uses such as restaurants, (late–hour) bars, and liquor stores, I apply a spatial matching approach, examining how the level of crime differs within pairs of street segments that differ in their land use composition but are so proximate spatially that they arguably share the same unobservable neighborhood characteristics. Previous empirical studies in this area relied on a very limited set of control variables to account for neighborhood characteristics. This is the first study to use these more rigorous approaches to identify the causal effects of land use.

My results indicate that commercial uses lead to substantially more street robberies and assaults/batteries in their immediate vicinity. However, this result hinges critically on density: Commercial activity actually reduces street crime in denser areas. The spillover effect of commercial uses into neighboring areas is negligible for robberies and relatively small for assaults/batteries. My findings indicate that the effect of commercial activity on assaults/batteries is driven almost entirely by liquor stores and bars, and that these uses contribute substantially to robberies as well. Per capita crime rates generally decline with residential density, a striking finding given that larger cities are known to have higher crime rates.

The experimental literature on hot spots policing provides some insight into how the externalities of commercial land use might be curtailed. Randomized controlled trials have demonstrated that concentrating policing in a localized area of high crime can substantially reduce violent crime in that area without displacement to nearby areas (Braga and Weisburd 2010). The limited spillover effect of commercial uses indicated by my results suggests that hot spots policing could be an effective response. Zoning could potentially be used to limit the number and diffusion of particularly criminogenic uses, facilitating the efficient use of police resources. My findings on the role of population density suggest that zoning which favors higher residential density could improve neighborhood safety, and that zoning which allows for mixed use structures may be preferable to more restrictive rules that aim for strictly residential or commercial use. More broadly, my finding that land use is a major determinant of crime patterns further establishes the importance of understanding this relationship.

3.2 Previous literature

Economists have largely ignored intra–metropolitan variation in crime, instead focusing on temporal and inter–metropolitan variation (O'Flaherty and Sethi 2014).[3] However, there is an extensive empirical literature in criminology and sociology on the relationship between crime and land use. This literature is largely descriptive, giving limited attention to the causal inference challenges present in this context. I review this literature here and discuss how my work improves upon the existing approaches.

Bernasco and Block (2009) study the location selection behavior of robbers in Chicago at the census tract level. Their results indicate that robbers frequently choose to offend in the census tract in which they reside or one which has a racial composition similar to that of their tract of residence; this is consistent with the interview–based evidence presented in Wright and Decker (1997). They find that individuals rarely travel far to offend and that census tracts with greater retail employment are more likely to be chosen. Browning, Byron, Calder, Krivo, Kwan, Lee and Peterson (2010) study the relationship between crime and commercial and residential density in a sample of census tracts from Columbus, Ohio. They find that, at low levels, an increase in a variable measuring commercial/residential density is associated with more crimes; at high levels, this relationship becomes negative.

Stucky and Ottensmann (2009) examine the relationship between violent crimes and land use patterns in Indianapolis. They find that robberies are much more common in commercial areas, even when the comparison is between commercial areas with above–average measured

[3]There are some exceptions. Cui and Walsh (2014) show that residential foreclosures resulting in long–term vacancies increase violent crime nearby. O'Flaherty and Sethi (2010) develop a sorting model to explain the concentration of street vice (such as prostitution and drug selling) in poor central city neighborhoods.

socioeconomic status and non–commercial areas with below–average socioeconomic status; however, they find the reverse pattern for homicides. Anderson et al. (2013) use zoning as a proxy for land use and study the relationship between crime, land use, and other built environment characteristics such as physical disorder, territoriality, and the condition of buildings, sidewalks, and streets. They measure the number of crimes within 100 and 250 meters of each of 205 blocks in Los Angeles County. They match blocks so that they have a comparable demographic composition. They find that residential zoning is associated with less crime than mixed–use zoning, and that commercial zoning is associated with substantially more crime than mixed–use zoning.

Sampson (1983) argues that the defensible–space and routine–activities theories support the idea that high residential densities will lead to more violent crime. He tests this hypothesis using National Crime Survey victimization data combined with roughly tract–level data on the residential density experienced by the respondents. He finds the expected positive relationship. White (1990) studies neighborhood permeability and burglary; a secondary finding is that residential density is negatively associated with burglary rates.

Some studies have examined the extent to which specific commercial uses are correlated with crime. Using data from Cleveland, Roncek and Maier (1991) document that city blocks containing bars see substantially more violent and property crime. Bernasco and Block (2011) study the spatial pattern of street robberies in Chicago. Their measure of commercial land use is derived from retail business counts collected by the marketing firm Claritas. They focus on a subset of these businesses selected so that the proportion of cash transactions would be high; this subset includes small bars, fast–food restaurants, liquor stores, laundromats, as well as other businesses. They find that every in–block commercial use they measure has a statistically significant positive relationship with the number of robberies, as does almost every adjacent–block commercial use. Of particular relevance to my analysis, they find that bars, fast–food restaurants, and liquor stores are associated with more robberies. Teh (2008) uses an event–study methodology to show that the introduction of liquor stores into Los Angeles neighborhoods with low socioeconomic status is associated with more violent and property crime.

The existing literature has been largely descriptive, with very limited attention to identifying the causal effects of land use patterns. In addition to using a much more complete set of control variables than existing studies, my study is the first to employ instrumental variable and spatial matching approaches to identification. Land use patterns and crime are confounded by unobservable neighborhood characteristics such as collective efficacy (Fischel 2001, Sampson, Raudenbush and Earls 1997). Crime also influences land use patterns, leading to a reverse causality problem (Rosenthal and Ross 2010). The novel historical zoning instrument employed here is both highly predictive of land use and demonstrably unrelated to unobservable neighborhood characteristics. The spatial matching approach yields an additional verification of my IV results. This multi–pronged approach to identifying causality arguably lends considerable credence to my findings and improves substantially on the existing literature.

I build on the existing literature in a number of other ways, using higher quality data as well as improved measurement strategies. I emphasize the role of population density, which has been marginalized in previous work, and I study how the interaction of commercial land use and population density affects crime outcomes. The unique detail of my crime data allows me to separate street crimes from crimes occurring indoors, which has not been possible in previous work. Since robberies of commercial establishments can only occur where such establishments exist, separating street robberies from business robberies eliminates an clear source of bias. I also consider crimes disaggregated by type, which is advantageous if different crimes have different relationships to land use, as my results indicate.

My study aggregates crimes to very small units of observation that effectively capture the land use immediately surrounding the crimes while separately accounting for ambient, "down the street" land uses. This avoids the numerous problems associated with aggregating crime and land use measurements to larger geographic areas such as census tracts, the standard approach in the previous literature. Higher level aggregation leads to an ecological inference problem; one cannot use the results to determine if crimes are concentrated close to commercial uses. Higher level aggregation also eliminates the possibility of determining the spatial range of land use effects and exacerbates the problem of confounding by unmeasured neighborhood characteristics.[4] I argue that my study is the first to effectively capture the spatial range of land use effects on crime; the few studies that measured crime and land use at the block face level failed to account for nearby land uses. When estimating the impact of specific uses, such as bars and liquor stores, I account for the general commercial character of the area. This allows me to precisely attribute differences in crime to the specific uses I consider, which was not possible in previous studies. It also allows me to estimate the "residual" effect of commercial activity after accounting for particularly criminogenic uses.

3.3 Data

This section describes the seven components of the dataset compiled for this paper. Land use data is drawn from two sources: A 2005 comprehensive survey of land use in Chicago and a registry of business licenses. Modern demographic data is derived from the 2010 Decennial Census as well as the American Community Survey. Crime data is derived from incident report records provided by the Chicago Police Department. Historical zoning data was geocoded from the original 1923 zoning ordinance and associated maps. Historical demographic data comes from the 1920 Decennial Census and the 1938 Local Community Fact Book. Historical homicide data is taken from the Chicago Historical Homicide Project. Historical land use data was geocoded from a comprehensive 1922 land use survey.

[4]Aggregating to the census block level, as some studies have done, is problematic as well; crimes that occurred on the residential side of a block could be associated with commercial uses on the other side of the block, despite the fact that these commercial uses are not proximate to the crime. Measuring land use at the block level also ignores the fact that crimes will be directly influenced by land use on proximate block faces. My approach to defining observations avoids these problems.

3.3.1 Land use

My primary land use data comes from a 2005 comprehensive survey conducted by the Chicago Metropolitan Agency for Planning (CMAP). From the CMAP classification I derive the following mutually exclusive and exhaustive land use categories: Single–family residential, multi–family residential, commercial (including residential with ground–level retail), industrial, institutional, open space, transportation, infrastructure, vacant, and under construction. Virtually all of the land in the city is coded as residential, commercial, industrial, institutional, or open space. The variables included in the analysis are discussed in section 3.4.2.

There are a number of reasons to believe that specific commercial uses may have an outsized effect on crime. I obtained data on specific uses from the registry of business licenses maintained by the Chicago Department of Business Affairs and Consumer Protection over the period 2008–2013. This registry includes coordinates which were used to geocode the establishments. I use data on the following license types: "Tavern," "Retail Food Establishment," "Late Hour," "Consumption on Premises - Incidental Activity," "Package Goods," and "Tobacco Retail Over Counter." I use the particular set of licenses held by an establishment to determine whether it is a restaurant, bar, late–hour bar, or liquor store.

3.3.2 Demographics

Demographic data is drawn from the 2010 Decennial Census and the 2006–2010 American Community Survey. The 2010 Census provides total population counts, counts by race and Hispanic/Latino origin, age composition, and counts of housing units and tenure status at the block level.[5] The 2006–2010 American Community Survey provides data on median household income, counts of individuals on public assistance, and poverty status. The block– and block–group–level data was attached to my sampling units via areal interpolation. Census data and associated GIS maps were taken from NHGIS.

[5] Census blocks roughly correspond to standard city blocks throughout much of Chicago.

3.3.3 Crime

Information on crimes is drawn from a publicly–accessible database of crime incident report data provided by the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting system. It includes every instance of robbery, battery, and assault over the period 2008–2013 for which an incident report was filed. Robbery is defined as the intentional taking of property from a person "by the use of force or by threatening the imminent use of force." A person commits battery if they knowingly cause "bodily harm to an individual" or make "physical contact of an insulting or provoking nature with an individual." A person commits an assault when they knowingly engage in "conduct which places another in reasonable apprehension of receiving a battery."

The publicly–available data includes coordinates corresponding to the most proximate address, which were used to geocode the crimes.[6] Crucial for my study is the fact that each incident report includes a brief description of the location of the crime, such as sidewalk, apartment, or small retail store. This location description allows me to isolate street robberies, assaults, and batteries from those occurring inside businesses.

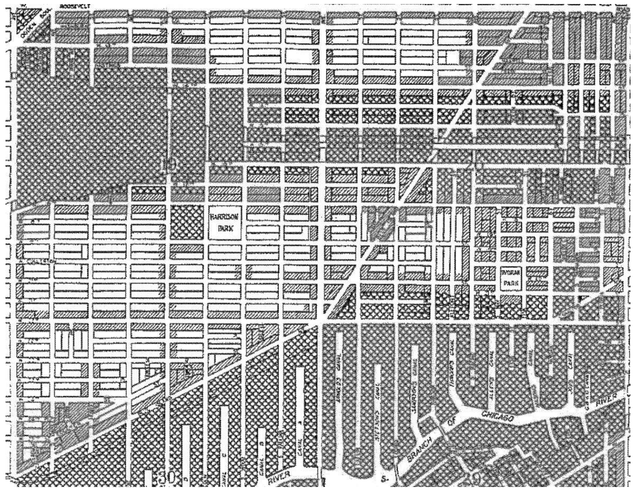
3.3.4 Historical zoning

To deal with potential confounding between land use and crime, I adopt an instrumental variable approach, using Chicago's original 1923 zoning code as an instrument for modern land use. This was the city's first comprehensive zoning ordinance. The ordinance established districts regulating both land use types ("use districts") and building density ("volume districts"). Four use districts were created: Residential (single–family housing), apartment, commercial, and manufacturing. These use districts were hierarchical, with apartment districts allowing residential uses, commercial districts allowing both apartments and single–family homes, and manufacturing districts allowing any use. Figure 3.1a provides a sample of the 1923 use zoning map.

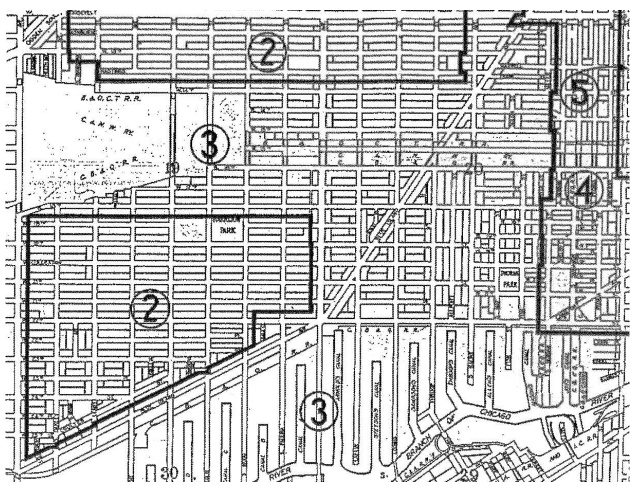
[6] There is no evidence that crimes were coarsely geocoded to, e.g., the nearest street intersection.

Figure 3.1: 1923 use and density zoning maps

(a)



(b)



Volume districts imposed restrictions on maximum lot coverage, aggregate volume, and height. Five volume districts were established, with district 1 restricted to the lowest density while district 5 permitted skyscrapers. Figure 3.1b provides a sample of the 1923 volume zoning map. Shertzer et al. (2014b) demonstrate that this zoning ordinance had a substantial causal effect on the spatial evolution of land use patterns in Chicago. This makes the zoning code a powerful instrument, as I document in section 3.4.3. The specific variables I derive from the zoning ordinance are discussed in section 3.4.3.

3.3.5 Historical land use

In section 3.4.3, I use historical land use data as part of a test for persistent unobservable neighborhood characteristics which may influence crime. I geocoded this data from a comprehensive 1922 land use survey conducted by the Chicago Zoning Commission to inform the process of drafting the 1923 zoning ordinance. This data contains the location of every commercial and manufacturing use in the city, with the latter subdivided into five subcategories, as well as the location and number of stories for every building with four or more stories.

3.3.6 Historical demographics

During the late 1920's, a group of sociologists at the University of Chicago divided the city into 75 mutually exclusive and exhaustive "community areas." These were considered "natural areas," the divisions reflecting distinct and identifiable clusters of related neighborhoods (Bulmer 1986). I use fixed effects based on these community areas to partially mitigate biases due to unmeasured neighborhood characteristics.

The Chicago Recreation Committee prepared an extensive handbook on community area characteristics in 1930 and 1934 for use by civic and social agencies; the 1938 Local Community Fact Book that resulted contains data on the share of households receiving public assistance, which I utilize in section 3.4.3 to argue for the validity of my instrumental variables strategy (Wirth and Furez 1938). Historical data on tract–level population and racial composition comes from the 1920 Decennial Census. The data and associated GIS maps were taken from NHGIS.

3.3.7 Historical crime

In section 3.4.3, I compare historical and modern patterns of homicide to argue for the validity of my instrumental variables strategy. Historical homicide data is taken from the

Chicago Historical Homicide Project, which digitized a continuous record of approximately 11,000 homicide cases maintained by the Chicago Police Department over the period 1870–1930 (Bienen and Rottinghaus 2002). Many of these records contained an address for the location of the crime. 4,528 of these were geocoded to a specific street address, while another 742 were matched to the nearest street intersection. Of these 5,270 homicides, 4,290 are dated between 1910 and 1930.

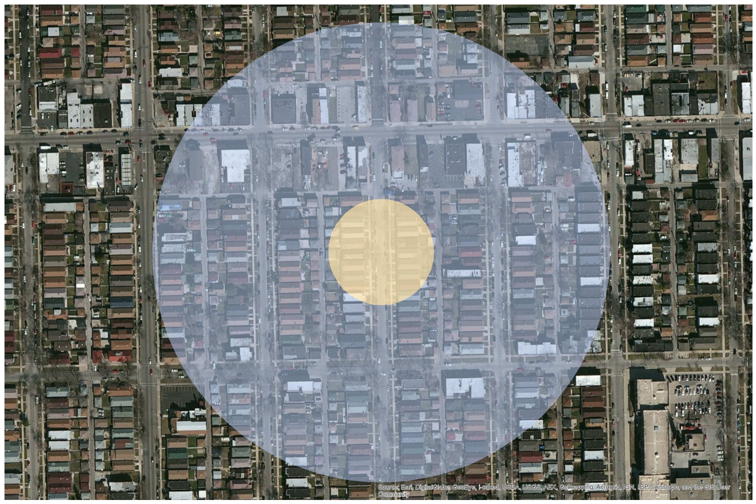
3.4 Methodology

In section 3.4.1, I define and motivate my unit of observation. In section 3.4.2, I describe the basic empirical approach. In section 3.4.3, I outline my instrumental variable strategy and provide evidence for the relevance and exogeneity of the instruments. In section 3.4.4, I present a solution to the problem of identifying the effects of specific commercial uses based on matching proximate observations.

3.4.1 Unit of observation

The goal of the empirical analysis is to determine the effect of proximate and nearby commercial uses on crime, as well as the influence of population density and the interaction of these effects. Given a small street segment, I want to determine how commercial uses on the street segment influence crime, and contrast this effect with that of more distant commercial uses. Theory suggests that commercial uses may affect crime in their immediate vicinity by increasing pedestrian traffic and contributing to social norm enforcement via monitoring by business proprietors. Commercial uses may have an effect over a longer range by generating street traffic that spills over into neighboring residential areas. The ideal unit of observation should capture crimes and their immediate surrounding land uses while also measuring proximity to neighboring land use types. For example, crimes that occurred in front of a commercial establishment should be distinguishable from crimes that occurred in front of a home but down the street from a commercial use, and these latter crimes should be distinguishable from crimes that occurred in isolated residential areas.

Figure 3.2: Sample unit of observation with annulus



To accomplish this, I aggregate crimes within small (300–ft–wide) street–centered circles and measure the land use within these circles. The circles are small enough so that the land use captured is only that which immediately surrounds the location of the crimes.[7] To analyze the spatial range of effects, I also measure land use in an annulus extending 500 feet from the boundary of each circle. This captures the effect of "down the street" land uses. An example is given in figure 3.2.

These (non–overlapping) circles are centered on points selected along the street grid. Ideally, my sample would cover the entire street area in the portion of the city for which I have data. However, this is not feasible, since it would be impossible to avoid generating circles that overlap. The algorithm I use approximates this ideal:

1. Start with all street intersections and midpoints.

2. Drop midpoints within 300 feet of an intersection.

3. Drop intersections within 300 feet of each other.

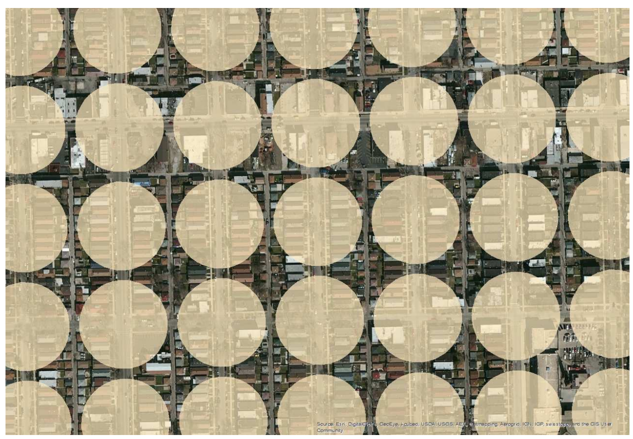
[7]This method also ensures that the land use on the sides of the street opposite the location of the crime are effectively captured, which is not the case when census blocks are used as the unit of analysis.

4. Randomly sample points on portions of the street grid that are more than 300 feet away from any remaining points.

The first three steps of this algorithm yield a dense, regular array of sample points in the majority of the city, due to the ubiquitous rectangular grid street system. An example is given in figure 3.3a. In the portions of the city with an irregular street grid, the sample points are less densely packed. An example is given in figure 3.3b.

Figure 3.3: Sampling in regular and irregular portions of the street grid

(a)



(b)

My circle–level data consists of crime counts as well as land use (including counts of business types) and housing data. I also measure ambient land use and businesses in the 500–foot annulus. Demographic data is attached to the combined circle–annulus area via areal interpolation.

My data covers the portion of Chicago south of Irving Park Road and north of 87th Street. This is approximately the middle two–thirds of the city and it includes the central business district, the historic Black Belt, and many of the largely black or Hispanic enclaves that have developed since the early twentieth century. Since the core of the central business district and the waterfront are not representative of the city as a whole, I exclude circles whose annuli overlap the central business district or lie within 500 feet of Lake Michigan.

3.4.2 Estimation: Baseline specification

The main outcomes of interest are counts of robberies and assaults/batteries.[8] A Poisson regression is the standard approach for analyzing data with nonnegative outcomes (Cameron and Trivedi 2013). This approach assumes that the pdf of the data generating process is

$$f\left(y_i \mid x_i\right) = \frac{e^{-e^{x_i'\beta}} e^{x_i'\beta y_i}}{y_i!}, \qquad i \in \{1, 2, \ldots, n\} \tag{3.4.1}$$

This implies that

$$\mathrm{E}\left[\, y_i \mid x_i \,\right] = e^{x_i'\beta} \tag{3.4.2}$$

If this functional form is correctly specified, a consistent, efficient, and asymptotically normal estimator of β can be obtained via maximum likelihood under standard regularity conditions (Cameron and Trivedi 2013).[9]

The primary explanatory variables of interest are the percentage of the circle and annulus occupied by commercial uses (including apartment buildings with ground–level retail) and the population density of the combined circle–annulus area. I allow population density to enter as a quadratic polynomial and I include an interaction between population density and the percentage of the circle occupied by commercial uses. Annulus commercial use and

[8] I aggregate assaults and batteries due to the hierarchical nature of incident reporting: Batteries are a class A misdemeanor in Illinois, so an incident involving an assault and a battery will be classified as a battery, since assaults are a (lower) class C misdemeanor.

[9] If (3.4.2) is correctly specified and standard regularity conditions hold, then the quasi–maximum likelihood estimator of β is consistent and asymptotically normal even if (3.4.1) is misspecified, as is the case when overdispersion is present (Gourieroux, Monfort and Trognon 1984a,b, White 1982).

population density are standardized. Other land use variables include the percentage of the circle and annulus occupied by single–family residences and industrial uses; the share devoted to multi–family residences is left as the omitted category. I also include distances to the nearest commercial and industrial use. These are the primary land use variables, and I instrument for all of them in the second part of the empirical analysis. I also account for a variety of auxiliary land uses, such as the percentage of the circle and annulus occupied by institutional and large–scale transportation uses as well as the percentage that is vacant or open space.

I include an indicator for whether the circle contains a street intersection, following the evidence presented by Wright and Decker (1997) that armed robbers prefer to commit offenses near intersections. White (1990) suggests that neighborhood permeability, defined as access to major traffic arteries, may have a positive impact on crime, and he provides some evidence for this hypothesis. To account for this possibility, I include a measure of ambient street density, an indicator for location on a major street, a quadratic polynomial in the distance to a major street, and the percentage of the circle and annulus occupied by a major transportation corridor. The concentration of crime around bus stops is well documented (see, e.g., Loukaitou-Sideris (1999)), and bus stops are frequently located along streets occupied by commercial uses, so I include counts of bus stops in each circle.

Other control variables include the percentage of housing units which are vacant, the percentage which are owner–occupied, the percentage of the population that is black, Hispanic, or under 18, the percentage of households with members over the age of 65, and the average household size. The share of the population that is black or Hispanic enters quadratically, and I also include an interaction between these shares as well as four indicator variables for highly segregated neighborhoods (those with shares black or Hispanic above 90% or below 10%). The percentage of households on public assistance is included, as is the share of households falling into each of seven bins defined by household income relative to the poverty level. I include quadratic polynomials in the distance to the central business district, Lake Michigan, the nearest river, nearest railroad, nearest park, and the nearest CTA station. I also include community area fixed effects to mitigate the bias due to unmeasured neighborhood characteristics.

For ease of interpretation, reported estimates are average marginal effects of the variables of interest. For the interaction between commercial uses and population density, I report the average cross–partial derivative. Bootstrap standard errors for the baseline specification.[10]

A Poisson regression is preferable to a standard linear regression for two reasons. First, the exponential conditional mean assumption (3.4.2) ensures that predicted values of y will be nonnegative. Second, the Poisson model substantially outperforms the linear model in out–of–sample prediction.[11] A negative binomial model is an alternative approach suited to count data, however it is more complex to estimate and does not offer a clear advantage over a simpler Poisson model (Blackburn 2014).[12]

3.4.3 Identification: Instrumental variables

To address the potential endogeneity of land use patterns, I adopt an instrumental variables strategy, using Chicago's 1923 zoning code to instrument for modern land use. There are a number of reasons why one might suspect that unobservable confounders or reverse causality between crime and land use are biasing the results obtained using the baseline approach. There is substantial evidence that crime rates are related to (difficult–to–measure) neighborhood social cohesion (Martin 2002, Morenoff, Sampson and Raudenbush 2001, Sampson et al. 1997). Homeowners have substantial incentives to exert control over changes in nearby land use patterns which may affect their property values (Fischel 2001). The extent to which they can do so depends on neighborhood social cohesion, since influ-

[10] I also estimated these models using the Conley (1999) approach to adjust for spatial autocorrelation; the standard errors were similar.

[11] In a 2–fold cross–validation test using counts of street robberies as the outcome, the average out–of–sample mean squared prediction error of the baseline Poisson model was 72% of that of the linear model.

[12] In a 2–fold cross–validation test using counts of street robberies as the outcome, the average out–of–sample mean squared prediction error of the negative binomial model was 103.5% of that of the baseline Poisson model.

encing the political process of zoning requires the concerted effort of many residents, which may be undermined by free–riding. Thus, neighborhood social cohesion may confound the relationship between land use patterns and crime.

Furthermore, reverse causality is a concern because high levels of crime or rising crime rates may alter the incentives determining land use patterns. For example, crime may discourage the construction of new high–density residences, or it could lower property values, encouraging the encroachment of industrial or commercial uses into previously residential areas. It could also have the opposite effect, diminishing the incentives for new business formation. Rosenthal and Ross (2010) document this kind of sorting behavior by entrepreneurs.

Two other factors merit consideration. To the extent that policing behavior is correlated with land use due to a common excluded cause, this IV strategy will isolate the effect of land use patterns on crime. If land use patterns directly affect policing behavior, as they likely do, then this approach identifies a "net" treatment effect, where police behavior is one causal channel through which land use affects crime. Land use also influences residential sorting behavior; individuals with a higher propensity to commit crime may sort towards areas with certain land use characteristics. If this is the case, sorting is another channel through which land use affects street crime. Since the goal of the analysis is to understand how changes in land use policy will influence street crime, capturing the indirect impact of land use through induced changes in policing behavior and individual sorting is essential.

Instrument set

I include the percentage of each circle zoned for commercial and manufacturing use in 1923 as well as the percentage falling into volume districts 1, 2, and 3, with the omitted density category comprised of districts 4 and 5. The same variables are computed for the annulus around each circle. Quadratic terms and interactions between use and density variables are included as well. A quadratic in the distance to the nearest commercial and manufacturing zoning is included, and each distance is interacted with its circle's density zoning variables.

Each circle use variable is interacted with each annulus use variable.

Estimation

I estimate the model

$$y_i = e^{x_i'\beta} + u_i$$

using generalized method of moments (GMM) (Hansen 1982). The moment conditions are

$$\mathrm{E}\left[z_i \left(y_i - e^{x_i'\beta} \right) \right] = 0 \tag{3.4.3}$$

where z_i includes the instruments discussed in section 3.4.3 as well as the covariates described in section 3.4.2, excluding the potentially–endogenous primary land use variables. In particular, the circle and annulus shares of single–family residential, commercial, and industrial uses are excluded, as is population density. The distances to the nearest commercial and industrial uses are omitted from z_i as well. To obtain standard errors for the average marginal effects of interest, I use an m out of n without replacement bootstrap with 50 iterations and $\frac{m}{n} \approx \frac{1}{2}$. The m out of n without replacement bootstrap is known to be consistent under minimal assumptions (Bickel, Götze and van Zwet 1997, Politis and Romano 1994).

There are more moment conditions than parameters to estimate, so Hansen's J statistic could be used to test the validity of the moment conditions (Hansen 1982). However, the effectiveness of this test is questionable; the finite–sample size and power appear to be complex, nonlinear functions of the sample size, number of overidentifying restrictions, and instrument strength (Hansen, Heaton and Yaron 1996). The J statistic does provide further evidence that the exponential mean specification is superior to a simple linear model.[13]

[13] GMM estimation of the robbery model on the full sample using the exponential mean specification yields a Hansen J statistic of 87.36; the same estimation using a linear model yields a J statistic of 161.98.

Relevance

Table 3.1 presents the F statistic and R^2 from a linear regression of each endogenous variable on the set of instruments outlined in section 3.4.3. It is clear that historical zoning is a strong predictor of modern land use, and in fact it explains much of the variation in present–day exposure to different use types.

Table 3.1: IV first stage: Predicted land use using historical zoning

	Circle		Annulus	
Modern land use	F–statistic	R^2	F–statistic	R^2
% single–family housing	325.492	0.392	514.797	0.505
% commercial	330.164	0.395	229.447	0.312
% industrial	183.142	0.266	303.275	0.375
Distance to commercial use	275.091	0.353		
Distance to industrial use	1391.748	0.734		

	Circle–annulus	
	F–statistic	R^2
Population	114.836	0.185
Population^2	49.289	0.089
Population × % commercial	35.611	0.066
	1% critical value for the F–test: 1.60	

Results from linear regressions of land use variables on the historical zoning instruments outlined in section 3.4.3. Regression F–statistics and R^2 are reported. Results for circle land uses are reported in the first two columns of the upper panel, while results for annulus land uses are reported in the second two columns. The bottom panel reports results for variables measured at the combined circle–annulus level. Models are estimated on the full sample of 18,712 observations.

However, in the case of multiple endogenous variables, the standard approach to measuring instrument strength is not sufficient. If there is insufficient variation in the instruments which can be uniquely attributed to each endogenous variable, then predicted values will be highly correlated and inferences will suffer. Currently, there is no test for weak instruments in nonlinear models with multiple endogenous variables. Angrist and Pischke (2009) reports a method for constructing correct first–stage F statistics in linear models with multiple endogenous variables; similarly, Shea (1997) describes a method for adjusting the first–stage

R^2 in this context. I report these results in table 3.2.

Table 3.2: IV first stage: Angrist–Pischke F–statistics and Shea R^2

	Circle		Annulus	
Modern land use	A–P F–stat	Shea R^2	A–P F–stat	Shea R^2
% single–family housing	12.513	0.032	11.859	0.033
% commercial	34.767	0.066	19.463	0.052
% industrial	10.190	0.036		0.044
Distance to commercial use	53.869	0.087		
Distance to industrial use	454.282	0.345		

	Circle–annulus	
	A–P F–stat	Shea R^2
Population	5.463	0.018
Population^2	11.582	0.025
Population $\times$ % commercial	11.257	0.019

Angrist–Pischke F–statistics and Shea R^2 for each endogenous land use variable (Angrist and Pischke 2009, Shea 1997). Results for circle land uses are reported in the first two columns of the upper panel, while results for annulus land uses are reported in the second two columns. The bottom panel reports results for variables measured at the combined circle–annulus level. Models are estimated on the full sample of 18,712 observations.

While the F–statistics and R^2 are substantially smaller than the unadjusted values, it is clear that near–perfect multicollinearity is not an issue.[14] As will be seen in section 3.5.2, the standard errors increase when I move from the baseline approach to GMM, however they are not so large as to make inference impossible.

Exogeneity

The validity of the exclusion restriction implied by (3.4.3) hinges on the assumption that unobservable neighborhood characteristics which may have influenced crime and zoning in 1923 have not persisted to the present. In this section, I argue that large–scale demographic

[14]These partial R^2 are comparable to the first–stage R^2 in some other well–known studies using instrumental variables.The first–stage R^2 in the Levitt (1997) study on policing and crime ranges from 0.06 to 0.11 (see his table 2). The first–stage R^2 in the Angrist and Evans (1998) study on fertility and female labor supply ranges from 0.004 to 0.084 (see their table 6).

changes preclude this possibility, and I use historical data on crime, land use, and demographics to rigorously test for the persistence of unobservable confounders.

Substantial neighborhood transformation has taken place throughout Chicago over the past 90 years. The closure of the border following the 1921 Emergency Quota Act and the Immigration Act of 1924 led to the assimilation of the city's theretofore marginalized immigrant population. Deindustrialization and suburbanization following World War II caused a dramatic shift in the demographics of the city; Chicago lost nearly 22% of its population between 1960 and 1990 (Hunt and DeVries 2013). Bursik and Webb (1982) document that demographic changes in Chicago over the period 1940–1970 were strongly related to changes in delinquency, which is highly correlated with the crimes I consider. Many of the most segregated and violent enclaves today are located in outlying areas of the city that were largely inhabited by relatively high–status second–generation immigrants of western European descent in 1920 (Shertzer, Twinam and Walsh 2014a).

The unique range of data available for Chicago allows me to present some quantitative evidence of neighborhood change. As discussed in section 3.3.7, counts of homicides over the period 1870–1930 (largely concentrated between 1910 and 1930) are available for the 49 Chicago community areas that overlap my sample area. Homicide is a strong proxy for unmeasured neighborhood characteristics which may influence crime. If the factors that led to high crime in the early twentieth century are persistent, one would expect to find that historically high–crime areas continue to see a relatively high level of crime today. However, the correlation between historical and modern homicide counts is only -0.0465.

Historical data on the percentage of families on public relief in 1934 is also available by community area. There is strong evidence suggesting that economic conditions influence crime by affecting individuals' incentives to offend (Becker 1968, Cantor and Land 1985, Fishback, Johnson and Kantor 2010). Historical public relief shares can be compared to modern public assistance shares derived from American Community Survey data. The correlation between historical and modern shares of households receiving public assistance is

-0.0071. These simple correlations suggest that the character of community areas in Chicago has changed dramatically.

The qualitative and quantitative evidence presented thus far suggests that unobservable neighborhood characteristics which may have influenced both zoning and crime in 1923 are unlikely to have persisted over the 90 years to the present. To further validate this supposition, I develop a rigorous test of the exclusion restriction utilizing the unique range of historical data available for Chicago.

Essentially, I argue that modern crime in my sample circles should only be related to historical crime to the extent that historical causes of crime have persisted to the present. Such causes include (measurable) land use patterns, zoning, and demographics as well as other (unmeasured) neighborhood characteristics. Thus, if historical crime is independent of modern crime, conditional on land use, zoning, and demographics, that strongly suggests that unobservable neighborhood characteristics that influenced crime in the past have not persisted to the present. This can be formalized most transparently using the language of causal graphical models; I relegate this discussion to a technical appendix.

Following this argument, I test for a relationship between historical and modern crime by estimating a Poisson regression of modern street homicide counts in my sample circles on historical homicide counts. I include only those historical homicides that can be geocoded to an exact street address. I also condition on the full set of zoning variables I use as instruments as well as historical land use data (attached to the circle as well as the associated annulus) and the 1920 population and share of the population that is black; this data is described in more detail in sections 3.3.4, 3.3.5, and 3.3.6. As a robustness check, I estimate the same model with street robberies as the outcome, since these are much more common and should allow for better inference. The results are given in table 3.3.

Columns (2) and (4) present results which include the necessary historical control variables. With either modern homicide or robbery counts as the outcome variable, the influence of historical homicides is very small and not statistically different from zero. This is strong

Table 3.3: Relationship Between Modern and Historical Crime

	# of modern homicides		# of modern robberies	
	(1)	(2)	(3)	(4)
# of historical homicides	-0.0001 (0.00341)	-0.0002 (0.00436)	0.1529*** (0.03871)	0.0252 (0.04032)
Controls	N	Y	N	Y
Observations	18,563	18,563	18,563	18,563

Results from Poisson regressions of modern street homicide or robbery counts on historical homicide counts. Regressions in columns (2) and (4) include 1922 land use, 1920 population and racial composition, and 1923 zoning; see sections 3.3.5, 3.3.6, and 3.4.3 for details. Results are average marginal effects. Sample excludes some circles for which historical land use data is not available due to damaged land use maps. Bootstrap standard errors are reported in parentheses.

evidence in favor of the exogeneity assumptions underlying my instrumental variable strategy. Excluding the historical control variables actually yields a positive and significant relationship when the outcome is robbery counts; this suggests that the test has sufficient power to detect an exogeneity violation if one existed.

3.4.4 Identification: Spatial matching

In section 3.5.3, I test for the influence of specific commercial land uses (such as bars) on crime. Unfortunately, the instrumental variable strategy described above is not applicable here, since historical zoning can only predict general land use patterns and not specific commercial uses. I adopt an alternative approach, matching sample circles whose boundaries lie not more than 200 feet apart. I then analyze differences in outcomes between these matched observations as a function of differences in covariates. Assuming that unobservable neighborhood characteristics vary smoothly across space, they should be largely constant between matched observations, so that the effects of differences in land use can be identified.

I estimate models of the form

$$y_i - y_j = (x_i - x_j)' \beta + \epsilon_{ij}$$

using ordinary least squares. Observations are paired so that the centroid of circle i is within 500 feet of the centroid of circle j. The argument is that confounding factors will be differenced out; this should be the case if unobservable neighborhood characteristics which may influence crime vary smoothly over space. To gauge the effectiveness of this identification strategy, I use it to replicate the instrumental variables analysis. Estimation using OLS is arguably appropriate here since the estimated residuals are approximately normal.[15]

3.5 Results

I first present descriptive statistics and discuss the spatial pattern of crime in Chicago. I then present results from baseline Poisson regressions without instruments in section 3.4.2. In section 3.5.2, I reestimate these models using GMM with historical zoning instruments. In section 3.5.3, I use the spatial matching approach to study the role of specific commercial land uses.

Table 3.4: Descriptive Statistics: Outcomes

Robberies	1.7 (3.03)
Robberies (per 1000 residents)	2.6 (6.92)
Assaults/batteries	5 (7.85)
Assaults/batteries (per 1000 residents)	7.3 (14.12)
Observations	18,712

[15]The residuals display heavy tails due to the right–skewed distribution of crime. However, they are approximately normally distributed over most of the range of the differenced outcome variables.

Table 3.4 provides means and standard deviations of crime counts in my sample. Street crime in my data is highly concentrated spatially. The median number of street robberies is one and the median number of assaults/batteries is two. 42% of observations see no robberies at all over the period 2008–2013; similarly, 23% see no batteries or assaults. Sample points with four or more robberies, the top 13%, account for 56% of the 31,123 robberies I observe. This is typical of urban crime and has been well documented in other cities such as Boston, Minneapolis, Seattle, and Tel–Aviv (Braga et al. 2010, Sherman et al. 1989, Weisburd and Amram 2014, Weisburd et al. 2012).

Table 3.5: Descriptive Statistics: Land Use

% commercial	0.12 (0.27)
% ambient commercial	0.12 (0.14)
Population density	835.4 (426.43)
Observations	18,712

Local and ambient commercial uses as well as population density are the primary predictors of interest in the baseline and instrumental variables analyses. Table 3.5 provides basic descriptive statistics for these variables. 21% of my sample points contain some commercial use, 10% contain some industrial use, and 49% are strictly residential. The average population in the combined circle–annulus area is 835, with an interquartile range of $[529, 1091]$. The distribution of population is very similar for observations with and without any commercial uses.

In the matching analysis, I focus on specific commercial uses. In particular, I examine the effects of restaurants, bars, late–hour bars (those bars permitted to continue serving alcohol past 2 a.m.), and liquor stores. There are 8,414 matched pairs of circles in my sample. 9.5% of these pairs contain at least one restaurant, 2.5% contain at least one bar, and 2.4% contain at least one liquor store. Late–hour bars are considerably less common; only 40 pairs

(0.05%) contain at least one.

3.5.1 Baseline results

Column (1) of table 3.6 reports the baseline Poisson results for street robbery counts. Interpreting the magnitudes of the marginal effects of commercial and ambient commercial use requires some attention to the typical variation in these explanatory variables observed in the data. Since the circles are small and capture areas within opposing block faces, they are typically homogeneous, with half of the circles in my sample devoted exclusively to residential use. Circles that contain any commercial use are frequently dominated by such use. It is most natural to evaluate the impact of commercial use by considering the difference in crime between a fully commercial and fully residential circle. The variation in ambient commercial use is considerably less stark and is more effectively summarized by its standard deviation, so I standardize the variable; the reported marginal effects reflect the impact of a one standard deviation change.[16]

Fully commercial circles are associated with 0.5 more street robberies than circles devoted exclusively to multi–family residential use. Since the median number of street robberies is one, this is a substantial difference. A one standard deviation increase in ambient commercial use is associated with 0.16 additional street robberies. A one standard deviation increase in population density is associated with 0.47 additional street robberies.

The strong positive relationship between commercial uses and crime in their immediate vicinity is consistent with the existing empirical literature. The relatively small variation in street robberies associated with differences in ambient commercial use is surprising given the small spatial scale and has not been documented previously. The relatively low spillover of crime from commercial areas to nearby residential areas has important policy implications, which I discuss in section 3.6.

[16]The standard deviation of ambient commercial use is 0.14, close to its mean of 0.12, so scaling the average marginal effect by the standard deviation yields an effect similar to that of moving from a fully residential annulus to one with the average level of ambient commercial use.

Table 3.6: Baseline results: Robberies and assaults/batteries

Land use	# of robberies	# of assaults/batteries
% commercial	0.508***	1.206***
	(0.0907)	(0.214)
Ambient % commercial	0.156***	0.577***
	(0.0266)	(0.0737)
Population density	0.471***	1.755***
	(0.0393)	(0.107)
Population density	-0.105**	0.625***
× % commercial	(0.048)	(0.1376)
Model	Poisson	Poisson
Observations	18,712	18,712
*** p<0.01, ** p<0.05, * p<0.1		

Results from baseline Poisson regressions of street robbery and assault/battery counts on the full set of land use, demographic, and geographic covariates; see section 3.4.2 for details. Results are average marginal effects. Ambient % commercial and population density are standardized. For the interaction term, I report the estimate of $\mathrm{E}\left[\frac{\partial^2 y}{\partial\, pop\, density\, \partial\, \%\, commercial}\right]$, where y is the outcome of interest. Bootstrap standard errors are reported.

Column (2) of table 3.6 reports the baseline Poisson results for street assault/battery counts. Fully commercial circles are associated with 1.2 more assaults/batteries than circles devoted exclusively to multi–family residential use; since the median number is two, land use appears to explain a considerable proportion of street crime. A one standard deviation increase in ambient commercial use is associated with 0.58 more street assaults/batteries, while a one standard deviation increase in population density is associated with 1.3 more street assaults/batteries. In relation to land use, assaults/batteries behave much like robbery.

The population density results reported in table 3.6 consistently show that denser areas see a larger number of street robberies and assaults/batteries. However, in per capita terms, crime rates are shrinking with population density.[17] Moving from an average density area to

[17] I do not normalize outcomes by population, as population density is accounted for in the model. Additionally, the per capita crime rate may not reflect the probability of victimization per unit of exposure time; Balkin and McDonald (1981) show that, when potential victims respond rationally to the possibility of victimization, this real crime rate may be inversely related to the per capita crime rate. Highly commercial areas may also see substantial pedestrian traffic from non–residents, so it is not clear how one would interpret

a one standard deviation denser area leads to a decline in the predicted number of robberies per 1000 residents from 2.1 to 1.8. The same transition leads to a decline from 6.4 to 5.7 assaults/batteries per 1000 residents. This is counterintuitive given the fact that larger cities have higher overall crime rates (Glaeser and Sacerdote 1999, Haynes 1973).

The interaction between commercial use and residential density is of independent interest, as it conveys the impact of mixing residential and commercial uses.[18] If the interaction is negative, one could argue that commercial uses accompanied by residences see less crime than standalone commercial uses. Returning to table 3.6, it is clear that no consistent pattern across crimes emerges. The interaction is negative and statistically significant for robbery counts but positive and significant for assaults/batteries.

In summary, the baseline results indicate a strong positive relationship between commercial uses and street robberies and assaults/batteries in their immediate vicinity. Nearby commercial uses are associated with more crime in neighboring areas, but this relationship is substantially weaker. Population density has a positive relationship with street crime counts, but the magnitude is small enough that per capita crime rates fall with population density. No consistent relationship between street crime and the interaction of commercial uses and residential density emerges.

3.5.2 IV results

In this section, I reestimate the models from section 3.5.1 using GMM with historical zoning instruments for the endogenous land use variables. Column (1) of table 3.7 reports the IV results for street robbery counts. Fully commercial circles are associated with 0.8 more street robberies than circles devoted exclusively to multi–family residential use. This estimate is

results from a model with normalized outcomes.

[18]I report the estimated interaction between % commercial and population density; specifically, the average cross–partial derivative

$$\mathrm{E}\left[\frac{\partial^2 y}{\partial\, pop\, density\, \partial\, \%\, commercial}\right]$$

where y is the outcome of interest.

substantially larger in magnitude that the baseline estimate. The IV estimate of the effect of ambient commercial use is half the size of the corresponding baseline estimate and not statistically significant. The IV estimate of the marginal effect of population density is slightly smaller than its baseline counterpart, strengthening the negative per capita relationship.

Table 3.7: IV results: Robberies and assaults/batteries

Land use	# of robberies	# of assaults/batteries
	(1)	(2)
% commercial	0.836***	1.67**
	(0.2537)	(0.7786)
Ambient % commercial	0.064	0.642**
	(0.142)	(0.2994)
Population density	0.368	2.82***
	(0.2246)	(0.7282)
Population density × % commercial	-0.725*	-0.567
	(0.4357)	(0.8462)
Model	Poisson IV	Poisson IV
Observations	18,712	18,712
*** p<0.01, ** p<0.05, * p<0.1		

Results from GMM estimation of Poisson regressions of street robbery and assault/battery counts on the full set of land use, demographic, and geographic covariates; see sections 3.4.2 and 3.4.3 for details. Historical zoning variables are used as instruments for modern land use; see section 3.4.3 for details. Results are average marginal effects. Ambient % commercial and population density are standardized. For the interaction term, I report the estimate of $\mathrm{E}\left[\frac{\partial^2 y}{\partial\, pop\, density\, \partial\, \%\, commercial}\right]$, where y is the outcome of interest. Bootstrap standard errors are reported.

Column (2) of table 3.7 reports the IV results for street assault/battery counts. As was the case for street robberies, the IV estimate of the marginal effect of commercial uses on crime in their immediate vicinity is larger than the baseline estimate. However, as I show in section 3.5.3, this effect is driven almost entirely by a small subset of commercial uses. The marginal effects of ambient commercial use and population density are also substantially larger than their associated baseline estimates. The negative per capita effect on population density is actually reversed here; moving from an average density area to a one standard deviation denser area leads to a increase in the predicted number of assaults/batteries per

1000 residents from 6.7 to 10.

Table 3.7 reports IV estimates of the interaction term between commercial use and population density. Unlike the mixed results obtained from the baseline regressions in table 3.6, the IV estimates of the interaction are consistently negative. For robbery counts, the interaction is substantially larger in magnitude; the transition from fully residential to fully commercial circles leads to 1.3 more robberies at average density, but 0.5 fewer robberies at density one standard deviation above mean. The same pattern holds for assaults/batteries: The residential to commercial transition leads to 3.2 more assaults/batteries at average density, but 1.6 fewer at density one standard deviation above mean.

Table 3.8: Predicted robbery and assault/battery counts from IV model

Land use	Predicted # of robberies	Predicted # of assaults/batteries
	(1)	(2)
Avg. density, no commercial	1.6 (0.11)	5.9 (0.47)
Avg. density, fully commercial	2.9 (0.38)	9.2 (1.21)
+1 s.d. density, no commercial	2.2 (0.33)	9.6 (1.53)
+1 s.d. density, fully commercial	1.7 (0.36)	8.1 (1.57)
Model	Poisson IV	Poisson IV

Predicted number of robberies and assaults/batteries from the GMM Poisson model of street crime as a function of the full set of land use, demographic, and geographic covariates; see sections 3.4.2 and 3.4.3 for details. Historical zoning variables are used as instruments for modern land use; see section 3.4.3 for details.

Table 3.8 reports predicted robbery and assault/battery counts for observations with different land use configurations. An apparent inverse–U shaped relationship emerges, with lower–density residential areas and higher–density commercial areas seeing lower numbers of robberies than higher–density residential areas and lower–density commercial areas. The robbery count in lower–density residential areas translates to a rate of 1.9 robberies per 1000 residents, while the similar count in higher–density mixed–use areas translates to 1.3 per

1000 residents. The assault/battery count in lower–density residential areas translates to a rate of 7.1 per 1000 residents, while the count in higher–density mixed–use areas translates to 6.4 per 1000 residents. For both classes of crime, higher–density mixed–use areas have lower crime rates than average residential areas.

In summary, the IV results show a strong positive effect of commercial uses on street robberies and assaults/batteries in their immediate vicinity, with relatively weak spillover effects into very proximate neighboring areas. However, this effect hinges critically on population density, so that dense mixed–use areas actually see lower rates of robbery and assault/battery than average residential areas.

3.5.3 Spatial matching results

The IV results establish that commercial areas have a strong effect on robberies and assaults/batteries in their immediate vicinity. In this section, I replicate those results using the spatial matching approach described in section 3.4.4. I also use this approach to measure the effects of specific commercial uses such as restaurants, bars, late–hour bars, and liquor stores. This allows me to determine the extent to which the commercial effect is driven by specific uses and how this extent differs across types of crime. Observable neighborhood characteristics vary smoothly over space, so the spatial matching approach employed here yields measurably similar pairs.[19]

Columns (1) and (3) of table 3.9 replicate the basic results from the IV estimation. Commercial uses have a large positive effect on crime in their immediate vicinity, but little spillover effect.[20] Columns (2) and (4) add differences in counts of restaurants, bars, late–

[19] The average difference in population between matched observations is 1 person, and the standard deviation of the difference is 181 persons. This is small relative to the average population (844 persons) and the standard deviation of population (419 persons). The average difference in the percentage of residents that are black (Hispanic) between matched observations is 0.03 (0.01) percentage points and the standard deviation of the difference is 4.97 (4.35) percentage points. The average difference in the percentage of households that are owner occupied is 0.06 percentage points, with a standard deviation of 6.39 percentage points.

[20] The interaction term was excluded to simplify the derivation of marginal effects with differenced outcomes/covariates. When included in either model, the interaction effect is comparable in magnitude and statistical significance to the IV results, lending further credence to those findings.

hour bars (those permitted to stay open past 2 a.m.) and liquor stores across both circles and annuluses. After accounting for these particular uses, roughly two–thirds of the effect of general commercial character on robberies remains; however, this effect disappears for assaults/batteries, indicating that the commercial effect is driven almost entirely by the specific uses accounted for in the model.

Table 3.9: Matching results: Robberies and assaults/batteries

Land use	Outcome			
	# of robberies		# of assaults/batteries	
	(1)	(2)	(3)	(4)
% commercial	1.402***	0.809***	3.567***	0.220
	(0.230)	(0.216)	(0.531)	(0.535)
Ambient % commercial	0.0354	-0.00133	0.563**	0.288
	(0.115)	(0.0978)	(0.224)	(0.248)
Population density	0.215	0.195*	1.376***	1.244***
	(0.158)	(0.114)	(0.337)	(0.375)
# of restaurants		0.155***		0.942***
		(0.0405)		(0.107)
# of bars		0.156		2.375***
		(0.203)		(0.481)
# of late–hour bars		1.430*		11.52***
		(0.821)		(3.268)
# of liquor stores		1.252***		4.886***
		(0.305)		(0.776)
# of nearby restaurants		0.00681		0.0837***
		(0.0160)		(0.0291)
# of nearby bars		0.0786		0.150
		(0.0611)		(0.134)
# of nearby late–hour bars		-0.0803		-0.708
		(0.232)		(0.581)
# of nearby liquor stores		0.00791		0.0414
		(0.0835)		(0.168)
Model	OLS	OLS	OLS	OLS
Observations	8,414	8,414	8,414	8,414
*** p<0.01, ** p<0.05, * p<0.1				

Results from linear regressions of differences in street robbery and assault/battery counts across matched pairs of observations on the full set of land use, demographic, and geographic covariates differenced across observations; see sections 3.4.2 and 3.4.4 for details. Columns (2) and (4) include 8 additional variables measuring differences in counts of restaurants, bars, late–hour bars, and liquor stores across both circles and annuluses. Bootstrap standard errors are reported.

An additional restaurant or bar in a circle leads to an additional 0.16 robberies; only for restaurants is this coefficient statistically significant. Late–hour bars have a considerably larger effect; an additional late–hour bar leads to 1.4 more robberies. Liquor stores have a comparable effect, with each additional store leading to 1.3 more robberies. As the median number of robberies is one, this is a substantial effect. The results are similar for assaults/batteries: Late–hour bars and liquor stores have an outsized effect, while bars and restaurants have a positive but substantially smaller impact. The fact that liquor stores have a substantially larger impact than the typical bar is surprising given the focus of the criminology literature on the role of bars. The large difference in impact between bars that close at 2 a.m. and those that stay open late is also striking and has important policy implications.

The annulus use results in columns (2) and (4) indicate that restaurants, (late–hour) bars, and liquor stores all have a negligible effect on robberies in nearby areas; their effects are restricted to the immediate area surrounding the establishment. The same pattern holds for assaults/batteries. This is consistent with my earlier results suggesting that the effect of commercial uses decays rapidly over space.

3.6 Discussion and policy implications

My findings indicate that land use is a major determinant of street crime patterns. The results indicate that commercial uses lead to substantially more street crime in their immediate vicinity. However, this hinges critically on density: Commercial activity actually reduces street crime in denser areas. This finding is a partial vindication of the hypothesized relationship between land use and crime suggested by Jacobs (1961). Contra Jacobs, commercial uses attract a substantial amount of crime, and this relationship is particularly strong for commercial activities that generate pedestrian traffic at late hours (like bars and liquor stores). However, this effect is ameliorated by higher residential densities, indicating that more mixed use areas attract less crime than exclusively commercial areas. The results suggest that a critical mass of pedestrian traffic may be necessary to create safe neighborhoods. Overall, per capita crime rates are actually declining with residential density, a striking finding given that larger cities are known to have higher crime rates. My results suggest that land use regulations which favor higher residential density could improve neighborhood safety, and that zoning which allows for mixed use structures may be preferable to more restrictive rules that aim for exclusively residential or commercial use.

The spillover effect of commercial uses into neighboring areas is negligible for robberies and relatively small for assaults/batteries. My findings indicate that the effect of commercial activity on assaults/batteries is driven almost entirely by liquor stores and (largely late–hour) bars. The prominence of liquor stores over typical bars as crime generators/attractors is striking given the common perception of bars as hot spots of crime. Since proximity to commercial activity is desirable for a variety of reasons, it is worth considering methods of mitigating its criminogenic externalities.

There is some evidence that the establishment of business improvement districts, where businesses pool resources to provide for additional local security, leads to substantial reductions in crime (Brooks 2008, Cook and MacDonald 2011). There is considerable evidence that the particular strategies employed by the police are an important determinant of their success in combating crime (Braga and Weisburd 2010). As discussed previously, crime is highly concentrated spatially, and this concentration is generally stable over time (Weisburd et al. 2012). Numerous strategies have been devised which focus police attention on these crime hot spots, including directed patrol and problem–oriented policing. There is a large experimental literature evaluating the impacts of these interventions, measuring their effects on crime and community relations as well as the extent to which they result in displacement, i.e., the shifting of crime to nearby areas not targeted by the intervention (Braga 2005). This literature convincingly demonstrates that intensive and problem–oriented policing applied to crime hot spots can result in sizable reductions in violent street crime without displacing crime to nearby areas or straining the relationship between police and the community (Braga and Bond 2008, Braga, Weisburd, Waring, Mazerolle, Spelman and Gajewski 1999, Sherman and Rogan 1995).

The findings of this literature suggest that crime concentrations resulting from specific uses (like liquor stores and late–hour bars) could be partially mitigated by strategic applications of police resources. However, such resources are costly. Zoning is a powerful and flexible tool for controlling land use patterns. It could potentially be employed to constrain the number and diffusion of such uses, limiting the strain they impose on police resources. The value of hot spots policing is particular high given the limited spillover of crime into areas neighboring commercial uses.

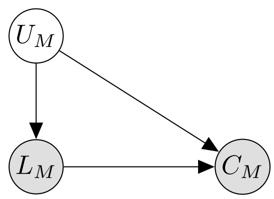
Future work should focus on the sources of heterogeneity in the extent to which commercial uses drive local crime; why do some commercial areas become crime hot spots while others do not? The distinction between attracting and generating crime is important as well. If commercial uses merely attract a finite local supply of potential offenders, an increase in the amount of commercial activity in an area may affect the spatial distribution of crime but leave the total amount of crime unchanged. Closely related to this is the question of how the extent to which commercial uses are concentrated or diffuse influences the overall crime rate. Jacobs (1961) argues that diffusing commercial uses results in less crime, while criminology research on offender behavior would suggest the opposite (Bernasco and Block 2009, Wright and Decker 1997). I am exploring this question in ongoing research.

3.7 Technical appendix

In this section, I discuss the theoretical justification for the exogeneity test proposed in section 3.4.3. Intuitively, I argue that historical crime should only be related to modern crime to the extent that historical causes of crime have persisted to the present. Such causes include (measurable) land use patterns, zoning, and demographics as well as other (unmeasured) neighborhood characteristics. Thus, if historical crime is independent of modern crime, conditional on historical land use, zoning, and demographics, that strongly suggests that unobservable neighborhood characteristics which influenced crime in the past have not persisted to the present. A causal graphical model provides a convenient and compact way to formalize and visualize this argument.

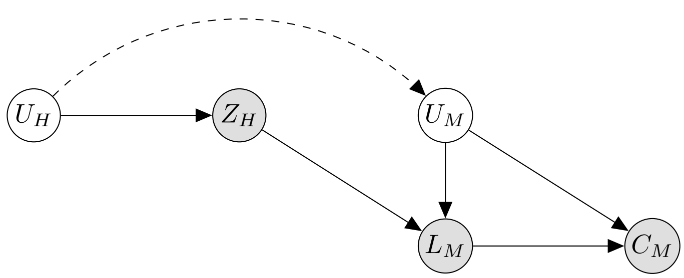
Figure 3.4 illustrates the basic identification problem: The effect of modern land use L_M on modern crime C_M is confounded by unobservable neighborhood characteristics U_M. This is a causal graphical model, which encodes conditional (in)dependences implied by a full nonparametric structural equation model (Pearl 2009). Arrows can be read as directional causal statements, so that L_M has a causal effect on C_M, and U_M has a causal effect on both. Grey nodes denote observable variables while white nodes denote unobservable variables; for readability, and without loss of generality, I present certain categories of related variables as a single node.

Figure 3.4: Confounding



A solution to the implied identification problem is the introduction of historical zoning Z_H as an instrumental variable (figure 3.5). If Z_H is unconditionally independent of U_M, i.e., if the dashed link between historical unobservable neighborhood characteristics U_H and its modern counterpart U_M is absent (so that unobservable neighborhood characteristics are not persistent), then the effect of land use on crime can be identified.[21] However, if this link is present, the instrument is contaminated and it is likely that the exclusion restriction does not hold.

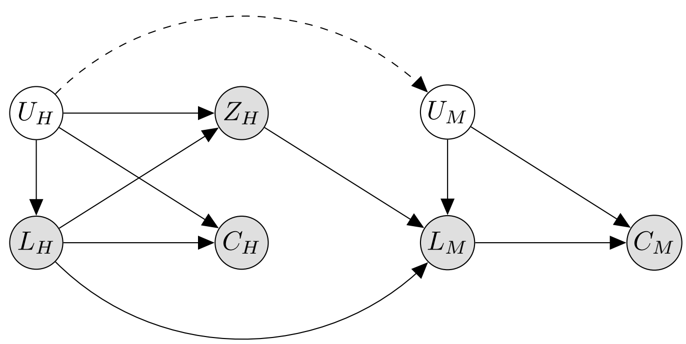
Figure 3.5: Contaminated instrument



[21] Under additional restrictions on functional forms; see section 3.4.3 for the formal statement of the identifying assumptions.

The unique data available for Chicago allows for a strong test of this independence assumption. Figure 3.6 shows how the zoning variable Z_H is embedded in an analogous historical version of figure 3.4. The availability of geocoded historical land use, demographic, and homicide data (described in sections 3.3.5, 3.3.6, and 3.3.7) means that the historical analogs of L_M and C_M, L_H and C_H, can be observed. From the graph, one can deduce that C_H is independent of C_M conditional on Z_H and L_H if U_H is unconditionally independent of U_M, i.e., if the dashed link is absent.[22] In the language of Pearl (2009), Z_H and L_H d–separate C_H and C_M when U_H is unconditionally independent of U_M. This suggests that testing for a relationship between historical and modern crime will provide a test of the IV exclusion restriction. C_H may be uncorrelated with C_M conditional on Z_H and L_H even if the dashed link in figure 3.5 is present, however this would require a level of fine–tuning that seems unlikely to occur in practice.

Figure 3.6: Nested IV



[22]The graph must be correctly specified for this statement to hold, however my argument is robust to a variety of changes in graph structure. Including an arrow between Z_H and C_H (regardless of orientation) or reversing the orientation of the arrow between L_H and C_H or L_H and U_H does not affect my argument.

Bibliography

Ancestry.com (2010). 1920 United States Federal Census.

Anderson, James M., John M. MacDonald, Ricky Bluthenthal, and J. Scott Ashwood (2013). Reducing crime by shaping the built environment with zoning: An empirical study of Los Angeles. *University of Pennsylvania Law Review*, **161**, pp. 699–861.

Angrist, Joshua D. and William N. Evans (1998). Children and their parents labor supply: Evidence from exogenous variation in family size. *American Economic Review*, **88**(3), pp. 450–477.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business & Economic Statistics*, **26**(3), pp. 261–283.

Baden, Brett M and Don L Coursey (2002). The locality of waste sites within the city of Chicago: a demographic, social, and economic analysis. *Resource and Energy Economics*, **24**(1), pp. 53–93.

Balkin, Steven and John F. McDonald (1981). The market for street crime: An economic analysis of victim–offender interaction. *Journal of Urban Economics*, **10**(3), pp. 390–405.

Becker, Gary S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, **76**(2), pp. 169–217.

Been, Vicki and Francis Gupta (1997). Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims. *Ecological Law Quarterly*, **24**(1), pp. 1–56.

Beresteanu, A. (2005). Nonparametric analysis of cost complementarities in the telecommunications industry. *RAND Journal of Economics*, **36**(4), pp. 870–889.

——— (2007). Nonparametric estimation of regression functions under restrictions on partial derivatives. *Unpublished manuscript*.

Bernasco, Wim and Richard Block (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, **47**(1), pp. 93–130.

——— (2011). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency*, **48**(1), pp. 33–57.

Berry, Christopher (2001). Land use regulation and residential segregation: Does zoning matter? *American Law and Economics Review*, **3**(2), pp. 251–274.

Bhattacharya, J., A.M. Shaikh, and E. Vytlacil (2008). Treatment effect bounds under monotonicity assumptions: An application to Swan–Ganz catheterization. *American Economic Review*, **98**(2), pp. 351–356.

Bhattacharya, Jay, Azeem M. Shaikh, and Edward Vytlacil (2012). Treatment effect bounds: An application to Swan–Ganz catheterization. *Journal of Econometrics*, **168**(2), pp. 223–243.

Bickel, P.J., F. Götze, and W.R. van Zwet (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, **7**, pp. 1–31.

Bienen, Leigh B. and Brandon Rottinghaus (2002). Learning from the past, living in the present: Understanding homicide in Chicago, 1870–1930. *Journal of Criminal Law and Criminology*, **92**(3), p. 437.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, **96**(4), pp. 988–1012.

——— (2008). Distributional impacts of the self–sufficiency project. *Journal of Public Economics*, **92**(3), pp. 748–765.

——— (2014). Can variation in subgroups' average treatment effects explain treatment effect heterogeneity? Evidence from a social experiment. *NBER Working Paper 20142*.

Blackburn, McKinley L. (2014). The relative performance of Poisson and negative binomial regression estimators. *Oxford Bulletin of Economics and Statistics, forthcoming*.

Bracey Jr., John H. and August MeierPapers of the NAACP, Part 12: Selected Branch Files, Series C: The Midwest. Reel 1..

Bradbury, Katharine L., Anthony Downs, and Kenneth A. Small (1982). *Urban decline and the future of American cities*. Brookings Institution.

Braga, Anthony A. (2005). Hot spots policing and crime prevention: A systematic review of randomized controlled trials. *Journal of Experimental Criminology*, **1**(3), pp. 317–342.

Braga, Anthony A. and Brenda J. Bond (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology*, **46**(3), pp. 577–607.

Braga, Anthony A. and David L. Weisburd (2010). *Policing problem places*. Oxford University Press.

Braga, Anthony A., Andrew V. Papachristos, and David M. Hureau (2010). The concentration and stability of gun violence at micro places in Boston, 1980–2008. *Journal of Quantitative Criminology*, **26**(1), pp. 33–53.

Braga, Anthony A., David L. Weisburd, Elin J. Waring, Lorraine G. Mazerolle, William Spelman, and Francis Gajewski (1999). Problem–oriented policing in violent crime places: A randomized controlled experiment. *Criminology*, **37**(3), pp. 541–580.

Brooks, Leah (2008). Volunteering to be taxed: Business improvement districts and the extra–governmental provision of public safety. *Journal of Public Economics*, **92**(1), pp. 388–406.

Brooks, Richard RW (2011). Covenants without courts: Enforcing residential segregation with legally unenforceable agreements. *American Economic Review*, **101**(3), pp. 360–365.

——— (2013). *Saving the Neighborhood*. Harvard University Press.

Browning, Christopher R., Reginald A. Byron, Catherine A. Calder, Lauren J. Krivo, Mei-Po Kwan, Jae-Yong Lee, and Ruth D. Peterson (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency*, **47**(3), pp. 329–357.

Bulmer, Martin (1986). *The Chicago school of sociology: Institutionalization, diversity, and the rise of sociological research*. University of Chicago Press.

Bursik, Robert J. and Jim Webb (1982). Community change and patterns of delinquency. *American Journal of Sociology*, **88**(1), pp. 24–42.

Cameron, A. Colin and Pravin K. Trivedi (2013). *Regression analysis of count data*. Cambridge University Press.

Cantor, David and Kenneth C. Land (1985). Unemployment and crime rates in the post–World War II United States: A theoretical and empirical analysis. *American Sociological Review*, **50**(3), pp. 317–332.

Carter, Sherry P., Stanley L. Carter, and Andrew L. Dannenberg (2003). Zoning out crime and improving community health in Sarasota, Florida: ”Crime prevention through environmental design”. *American Journal of Public Health*, **93**(9), pp. 1442–1445.

Cattaneo, Matias D. (2010). Multi–valued treatment effects. In Steven N. Durlauf and Lawrence E. Blume, eds., *New Palgrave Dictionary of Economics*, Palgrave Macmillan.

Chernozhukov, V., S. Lee, and A.M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica*, **81**(2), pp. 667–737.

Cohen, Lawrence E. and Marcus Felson (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, **44**(4), pp. 588–608.

Cohen, Mark A., Roland T. Rust, Sara Steen, and Simon T. Tidd (2004). Willingness–to–pay for crime control programs. *Criminology*, **42**(1), pp. 89–110.

Commission, Chicago Zoning (1922). Zoning Chicago.

Conley, Timothy G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, **92**(1), pp. 1–45.

Cook, Philip J. and John MacDonald (2011). Public safety through private action: An economic assessment of BIDs. *Economic Journal*, **121**(552), pp. 445–462.

Coupland, Andy (1997). *Reclaiming the city: Mixed use development.* Taylor & Francis.

Cox, D.R. (1958). *Planning of experiments.* Wiley.

Cui, Lin and Randall Walsh (2014). Foreclosure, vacancy and crime. *SSRN Working Paper #1773706.*

Cullen, Julie Berry and Steven D. Levitt (1999). Crime, urban flight, and the consequences for cities. *Review of Economics and Statistics*, **81**(2), pp. 159–169.

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor (2008). Is the melting pot still hot? Explaining the resurgence of immigrant segregation. *Review of Economics and Statistics*, **90**(3), pp. 478–497.

Daniels, Glynis and Samantha Friedman (1999). Spatial Inequality and the Distribution of Industrial Toxic Releases: Evidence from the 1990 TRI. *Social Science Quarterly*, **80**(2), pp. 244–262.

Depro, Brooks, Christopher Timmins, and Maggie ONeil (2014). White Flight and Coming to the Nuisance: Can Residential Mobility Explain Environmental Injustice? *Unpublished Manuscript.*

Djebbari, H. and J. Smith (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, **145**(1), pp. 64–80.

Eeckhout, J. and P. Kircher (2011). Identifying sorting – In theory. *Review of Economic Studies*, **78**(3), pp. 872–906.

Fan, Y. and S.S. Park (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, **26**(3), p. 931.

Feller, A. and C.C. Holmes (2009). Beyond toplines: Heterogeneous treatment effects in randomized experiments. *Unpublished manuscript.*

Firpo, S.P. and G. Ridder (2010). Bounds on functionals of the distribution treatment effects. *Unpublished manuscript.*

Fischel, William A. (2001). *The homevoter hypothesis.* Harvard University Press.

Fishback, Price V., Ryan S. Johnson, and Shawn Kantor (2010). Striking at the roots of crime: The impact of welfare spending on crime during the Great Depression. *Journal of Law and Economics*, **53**(4), pp. 715–740.

Flint, Barbara J (1977). Zoning and residential segregation: A social and physical history, 1910–40. Ph.D. dissertation, University of Chicago.

Frank, M.J., R.B. Nelsen, and B. Schweizer (1987). Best–possible bounds for the distribution of a sum – A problem of Kolmogorov. *Probability Theory and Related Fields*, **74**(2), pp. 199–211.

Gelman, A. (2013). The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective. *Unpublished manuscript*.

Giustinelli, P. (2011). Non-parametric bounds on quantiles under monotonicity assumptions: With an application to the Italian education returns. *Journal of Applied Econometrics*, **26**(5), pp. 783–824.

Glaeser, Edward (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier and happier*. Penguin Press.

Glaeser, Edward L. and Bruce Sacerdote (1999). Why is there more crime in cities? *Journal of Political Economy*, **107**(6), pp. S225–S258.

Gourieroux, Christian S., Alain Monfort, and Alain Trognon (1984a). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, **52**(3), pp. 701–20.

——— (1984b). Pseudo maximum likelihood methods: Theory. *Econometrica*, **52**(3), pp. 681–700.

Graham, Bryan S., Guido W. Imbens, and Geert Ridder (2014). Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics*, **5**(1), pp. 29–66.

Greenland, Sander (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, **14**(3), pp. 300–306.

Grossman, James R (1989). *Land of hope: Chicago, black southerners, and the great migration*. University of Chicago Press.

Gundersen, Craig, Brent Kreider, and John V. Pepper (2012). The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. *Journal of Econometrics*, **166**(1), pp. 79–91.

Hall, Peter (2014). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design Since 1880*. John Wiley & Sons.

Hansen, Lars Peter (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), pp. 1029–54.

Hansen, Lars Peter, John Heaton, and Amir Yaron (1996). Finite–sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics*, **14**(3), pp. 262–280.

Haynes, Robin M. (1973). Crime rates and city size in America. *Area*, **5**(3), pp. 162–165.

Heckman, J.J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, **64**(4), pp. 487–535.

Hunt, D. Bradford and Jon B. DeVries (2013). *Planning Chicago*. American Planning Association Planners Press.

Ihlanfeldt, Keith R (2004). Exclusionary land–use regulations within suburban communities: A review of the evidence and policy prescriptions. *Urban Studies*, **41**(2), pp. 261–283.

Jacobs, Jane (1961). *The death and life of great American cities*. Random House.

Kennedy, Louise Venable (1930). *The Negro Peasant Turns Cityward: Effects of Recent Migrations to Northern Centers*. New York: Columbia University Press.

Kreider, Brent and Steven C. Hill (2009). Partially identifying treatment effects with an application to covering the uninsured. *Journal of Human Resources*, **44**(2), pp. 409–449.

Kreider, Brent and John V. Pepper (2007). Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, **102**(478), pp. 432–441.

Kreinovich, V. and S. Ferson (2006). Computing best–possible bounds for the distribution of a sum of several variables is NP–hard. *International Journal of Approximate Reasoning*, **41**(3), pp. 331–342.

Lalive, Rafael, Jan Van Ours, and Josef Zweimüller (2006). How changes in financial incentives affect the duration of unemployment. *Review of Economic Studies*, **73**(4), pp. 1009–1038.

Lazzati, Natalia (2014). Treatment response with social interactions: Partial identification via monotone comparative statics. *Unpublished manuscript*.

Levitt, Steven D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review*, **87**(3), pp. 270–290.

Loukaitou-Sideris, Anastasia (1999). Hot spots of bus stop crime: The importance of environmental attributes. *Journal of the American Planning Association*, **65**(4), pp. 395–411.

Maantay, Juliana (2001). Zoning, equity, and public health.. *American Journal of Public Health*, **91**(7), p. 1033.

Makarov, G.D. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, **26**(4), pp. 803–806.

Manski, C.F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, **24**(3), pp. 343–360.

——— (1990). Nonparametric bounds on treatment effects. *American Economic Review*, **80**(2), pp. 319–323.

——— (1997). Monotone treatment response. *Econometrica*, **65**(6), pp. 1311–1334.

——— (2003). *Partial identification of probability distributions*. Springer.

——— (2013). Identification of treatment response with social interactions. *Econometrics Journal*, **16**(1), pp. S1–S23.

Manski, Charles F. and Daniel S. Nagin (1998). Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological methodology*, **28**(1), pp. 99–137.

Manski, C.F. and J.V. Pepper (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, **68**(4), pp. 997–1010.

——— (2009). More on monotone instrumental variables. *Econometrics Journal*, **12**, pp. S200–S216.

Martin, David (2002). Spatial patterns in residential burglary Assessing the effect of neighborhood social capital. *Journal of Contemporary Criminal Justice*, **18**(2), pp. 132–146.

Massey, Douglas S. and Nancy A. Denton (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

McCollister, Kathryn E., Michael T. French, and Hai Fang (2010). The cost of crime to society: New crime–specific estimates for policy and program evaluation. *Drug and Alcohol Dependence*, **108**(1), pp. 98–109.

McDonald, John F and Robert A Moffitt (1980). The Uses of Tobit Analysis. *Review of Economics and Statistics*, **62**(2), pp. 318–321.

McMillen, Daniel P. (2012). Conditionally Parametric Quantile Regressions for Spatial Data: An Analysis of Land Values in Early Nineteenth Century Chicago.

McMillen, D.P. and J.F. McDonald (1991). A Markov chain model of zoning change. *Journal of Urban Economics*, **30**(2), pp. 257–270.

McMillen, Daniel P and John F McDonald (1999). Land use before zoning: The case of 1920's Chicago. *Regional Science and Urban Economics*, **29**(4), pp. 473–489.

——— (2002). Land values in a newly zoned city. *Review of Economics and Statistics*, **84**(1), pp. 62–72.

Miller, Ted R., Mark A. Cohen, and Brian Wiersema (1996). Victim costs and consequences: A new look. *NIJ Research Report NCJ 155282*.

Mills, Edwin S (1979). *Economic analysis of urban land use controls*. Current Issues in Urban Economics.

Molinari, F. and A.M. Rosen (2008). The identification power of equilibrium in games: The supermodular case. *Journal of Business & Economic Statistics*, **26**(3), pp. 297–302.

Morenoff, Jeffrey D. and Robert J. Sampson (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces*, **76**(1), pp. 31–64.

Morenoff, Jeffrey D., Robert J. Sampson, and Stephen W. Raudenbush (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, **39**(3), pp. 517–558.

Neumark, David and William Wascher (2011). Does a higher minimum wage enhance the effectiveness of the Earned Income Tax Credit? *Industrial and Labor Relations Review*, **64**(4), pp. 712–746.

of Christ Commission for Racial Justice, United Church and Inc. Public Data Access (1987). Toxic Waste and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites.

O'Flaherty, Brendan and Rajiv Sethi (2010). The racial geography of street vice. *Journal of Urban Economics*, **67**(3), pp. 270–286.

——— (2014). Urban crime. In *Handbook of Urban and Regional Economics*, Elsevier.

on Race Relations, Chicago Commission (1922). The Negro in Chicago: A Study of Race Relations and a Race Riot.

Papke, Leslie E and Jeffrey M Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics*, **11**(6), pp. 619–32.

Pearl, Judea (2009). *Causality: Models, reasoning and inference*. Cambridge University Press.

Pepper, John V. (2000). The intergenerational transmission of welfare receipt: A nonparametric bounds analysis. *Review of Economics and Statistics*, **82**(3), pp. 472–488.

Pogodzinski, J Michael (1991). The effects of fiscal and exclusionary zoning on household location: A critical review. *Journal of Housing Research*, **2**(2), pp. 145–160.

Politis, Dimitris N. and Joseph P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, **22**(4), pp. 2031–2050.

Punter, John (2007). *The Vancouver achievement: Urban planning and design*. UBC Press.

Rex, F., Frederick (1937). Centennial List of Mayors, City Clerks, City Attorneys, City treasurers, and Aldermen Elected by the People of the City of Chicago.

Roncek, Dennis W. and Ralph Bell (1981). Bars, blocks, and crimes. *Journal of Environmental Systems*, **11**(1), pp. 35–47.

Roncek, Dennis W. and Pamela A. Maier (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots". *Criminology*, **29**(4), pp. 725–753.

Rosenbaum, Paul R and Donald B Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), pp. 41–55.

Rosenthal, Stuart S. and Amanda Ross (2010). Violent crime, entrepreneurship, and cities. *Journal of Urban Economics*, **67**(1), pp. 135–149.

Rothwell, Jonathan T (2011). Racial enclaves and density zoning: the institutionalized segregation of racial minorities in the United States. *American Law and Economics Review*, **13**(1), pp. 290–358.

Rothwell, Jonathan and Douglas S Massey (2009). The Effect of Density Zoning on Racial Segregation in US Urban Areas. *Urban Affairs Review*, **44**(6), pp. 779–806.

Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, **6**(1), pp. 34–58.

Sampson, Robert J. (1983). Structural density and criminal victimization. *Criminology*, **21**(2), pp. 276–293.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, **277**(5328), pp. 918–924.

Sarnoff, Nancy and David Kaplan (2007). High–density, mixed–use trend takes root in Houston. *Houston Chronicle*.

Schwieterman, Joseph P. and Dana M. Caspall (2006). *The politics of place: A history of zoning in Chicago*. Lake Claremont Press.

Sharkey, Patrick (2013). *Stuck in place: Urban neighborhoods and the end of progress toward racial equality*. University of Chicago Press.

Shea, John (1997). Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics*, **79**(2), pp. 348–352.

Sherman, Lawrence W. and Dennis P. Rogan (1995). Effects of gun seizures on gun violence: "Hot spots" patrol in Kansas City. *Justice Quarterly*, **12**(4), pp. 673–693.

Sherman, Lawrence W., Patrick R. Gartin, and Michael E. Buerger (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, **27**(1), pp. 27–56.

Shertzer, Allison, Tate Twinam, and Randall P. Walsh (2014a). Race, ethnicity, and discriminatory zoning. *NBER Working Paper #20108*.

——— (2014b). Zoning and urban persistence. *Unpublished manuscript*.

Shlay, AB and PH Rossi (1981). Keeping Up the Neighborhood-Estimating Net Effects of Zoning. *American sociological review*, **46**(6), pp. 703–719.

Silverman, Gary (2013). New York: A concrete legacy. *Financial Times*.

Stark, Rodney (1987). Deviant places: A theory of the ecology of crime. *Criminology*, **25**(4), pp. 893–910.

Stucky, Thomas D. and John R. Ottensmann (2009). Land use and violent crime. *Criminology*, **47**(4), pp. 1223–1264.

Teh, Bing-Ru (2008). Do liquor stores increase crime and urban decay? Evidence from Los Angeles. *Working Paper University of California, Berkeley*.

Tsunao, Okumura and Emiko Usui (2014). Concave–monotone treatment response and monotone treatment selection: With an application to the returns to schooling. *Quantitative Economics*, **5**(1), pp. 175–194.

Villareal, Carlos, Brian Bettenhausen, Eric Hanss, and Jonathan Hirsch (2014). Historical Health Conditions in Major US Cities: The HUE Dataset.

Weisburd, David and Shai Amram (2014). The law of concentrations of crime at place: The case of Tel Aviv–Jaffa. *Police Practice and Research*, **15**(2), pp. 101–114.

Weisburd, David, Shawn Bushway, Cynthia Lum, and Sue-Ming Yang (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology*, **42**(2), pp. 283–322.

Weisburd, David L., Elizabeth R. Groff, and Sue-Ming Yang (2012). *The criminology of place: Street segments and our understanding of the crime problem*. Oxford University Press.

White, Halbert (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**(1), pp. 1–25.

White, Garland F. (1990). Neighborhood permeability and burglary rates. *Justice Quarterly*, **7**(1), pp. 57–67.

Williamson, R.C. and T. Downs (1990). Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, **4**(2), pp. 89–158.

Wilson, Sacoby, Malo Hutson, and Mahasin Mujahid (2008). How planning and zoning contribute to inequitable development, neighborhood health, and environmental injustice. *Environmental Justice*, **1**(4), pp. 211–216.

Wirth, Louis (1938). Urbanism as a way of life. *American Journal of Sociology*, **44**(1), pp. 1–24.

Wirth, Louis and Margaret Furez (1938). *Local community fact book*. Chicago Recreation Commission.

Wright, Richard T. and Scott H. Decker (1997). *Armed robbers in action: Stickups and street culture*. Northeastern University Press.